



**ФАМЭБ'2013  
XII**

*Р о с с и й с к а я   а к а д е м и я   н а у к*  
*М и н и с т е р с т в о   о б р а з о в а н и я   и   н а у к и   Р Ф*  
*Институт математики и фундаментальной информатики*  
*Торгово-экономический институт*  
*Сибирский федеральный университет*  
*Научно-исследовательский институт проблем пожарной безопасности*  
*Институт вычислительного моделирования СО РАН*

*Т   Р   У   Д   Ы*  
*ДВЕНАДЦАТОЙ МЕЖДУНАРОДНОЙ КОНФЕРЕНЦИИ*  
*ПО ФИНАНСОВО–АКТУАРНОЙ МАТЕМАТИКЕ*  
*И ЭВЕНТОЛОГИИ БЕЗОПАСНОСТИ*

*К р а с н о я р с к*  
*2013*



УДК 519.248: [004.8+33+301+159.9]

Т 78

Труды XII международной ФАМЭБ'2013 конференции. Под ред. Олега Воробьёва. — Красноярск: НИИППБ, СФУ, 2013. — 372 с.

ISBN 978-5-9903358-3-7

Конференция поддержана

Российским фондом фундаментальных исследований:

гранты №№ 02-01-10015-г; 03-01-10020-г; 04-01-10051-г;

07-01-06011-г; 08-06-06016-г; 09-06-06028-г;

Торгово-экономическим институтом СФУ:

гранты №№ ГВ-09-11; ГВ-09-04;

и Научно-исследовательским институтом проблем пожарной безопасности

Редакционная коллегия:

Амельчугов С.П., д. техн. наук, профессор

Баранова И.В., канд. физ.-мат. наук

Быкова В.В., д-р физ.-мат. наук, профессор

Воробьёв О.Ю., д-р физ.-мат. наук, профессор (редактор)

Голденко Е.Е., канд. физ.-мат. наук

Клочков С.В., канд. физ.-мат. наук

Лукьянова Н.А.

Мажаров В.Ф., д-р мед. наук

Новосёлов А.А., канд. физ.-мат. наук

Семёнова Д.В., канд. физ.-мат. наук (помощник редактора)

© Институт математики и фундаментальной информатики, 2013

© Торгово-экономический институт, 2013

© Сибирский федеральный университет, 2013

© НИИ проблем пожарной безопасности, 2013

© Институт вычислительного моделирования СО РАН, 2013

ISBN 978-5-9903358-3-7

*Russian Academy of Sciences*  
*Ministry of Education and Science of RF*  
*Institute of Mathematics and Computer Science*  
*Trade Economy Institute*  
*Siberian Federal University*  
*Research Institute for Fire Safety Problems*  
*Institute of Computational Modeling of SB RAS*

*P R O C E E D I N G S*  
*OF XII INTERNATIONAL CONFERENCE*  
*ON FINANCIAL AND ACTUARIAL MATHEMATICS*  
*AND EVENTOLOGY OF SAFETY*

*K r a s n o y a r s k*  
*2013*



УДК 519.248: [004.8+33+301+159.9]

T78

Proceedings of the XII International FAMES'2013 Conference. Oleg Vorobyev, ed. — Krasnoyarsk: RIFS, SFU, 2013. — 372 p.

ISBN 978-5-9903358-3-7

The conference is supported by  
the Russian Fund for Basic Research:  
grants №№ 02-01-10015-з; 03-01-10020-з; 04-01-10051-з;  
07-01-06011-з; 08-06-06016-з; 09-06-06028-з;  
the Krasnoyarsk State Trade Economic Institute:  
grants №№ FB-09-11; FB-09-04;  
and Research Institute for Fire Safety Problems

Editorial board:

Amelchugov S.P., professor

Baranova I.V., Ph.D.

Bykova V.V., professor

Goldenok E.E., Ph.D.

Klochkov S.V., Ph.D.

Lukyanova N.A.

Mazharov V.F., Ph.D.

Novosyolov A.A., Ph.D.

Semenova D.V. Ph.D. (Assistant Editor)

Vorobyev O.Yu., professor (Editor)

© Institute of Mathematics and Computer Science, 2013

© Trade Economy Institute, 2013

© Siberian Federal University, 2013

© Research Institute for Fire Safety Problems, 2013

© Institute of Computational Modeling of RAS, 2013

ISBN 978-5-9903358-3-7

ДВЕНАДЦАТАЯ МЕЖДУНАРОДНАЯ КОНФЕРЕНЦИЯ  
ПО ФИНАНСОВО-АКТУАРНОЙ МАТЕМАТИКЕ  
И ЭВЕНТОЛОГИИ БЕЗОПАСНОСТИ  
19 ~ 20 апреля 2013 года, Красноярск

Т е м ы

- Финансово-актуарная математика
- Конвергенция нано-био-инфо-когно наук и технологий для совершенствования человеческой деятельности
- Эвентологическая математика в гуманитарных, социально-экономических и естественнонаучных областях
- Эвентология безопасности и риска
- Теория вероятностей и математическая статистика
- Эвентология и математическая эвентология
- Философская эвентология и философия вероятности
- Практическая эвентология
- Эвентологическая экономика
- Эвентологическая психология
- Эвентологические проблемы искусственного интеллекта
- Принятие решений в условиях риска и неопределённости
- Измерение и модели риска
- Теория нечётких событий, обобщенная теория неопределённости
- Математические начала хаоса в экономике
- Системный анализ и управление случайными событиями
- Теория механизмов создания распределений событий

С о п р е д с е д а т е л и   к о н ф е р е н ц и и

Воробьёв О.Ю. (Красноярск)

Комаров С.Ю. (Красноярск)

П р о г р а м м н ы й   к о м и т е т

Абдушукуров А.А. (Узбекистан)  
Алахверди Н. (Турция)  
Александров Ю.Л. (Красноярск)  
Алиев Т. (Азербайджан)  
Ализаде Т. (Азербайджан)  
Амельчугов С.П. (Красноярск)  
Ангелов П. (Великобритания)  
Андреев Ю.А. (Красноярск)  
Витлинский В.В. (Украина)  
Воробьёв А.О. (Москва)  
Голдблатт Дж.Дж. (Великобритания)  
Гианнопулос К. (ОАЭ)  
Горбань А.Н. (Великобритания)  
Городницкая И. (США)  
Грановская Р.М. (С.-Петербург)  
Григорьев Ю.Д. (С.-Петербург)  
Гуц А.К. (Омск)  
Добронец Б.С. (Красноярск)  
Доррер Г.А. (Красноярск)

Зандер Е.В. (Красноярск)  
Киселёва Т. (Германия)  
Кокс Л.Э. (США)  
Куимов В.В. (Красноярск)  
Кытманов А.М. (Красноярск)  
Лепский В.Е. (Москва)  
Лефевр В.А. (Ирвин, США)  
Лукин В.Н. (Канада)  
Луценко Е.В. (Краснодар)  
Мажаров В.Ф. (Красноярск)  
Немировский В.Г. (Красноярск)  
Новосёлов А.А. (Красноярск)  
Орсингер Э. (Италия)  
Охорзин В.А. (Красноярск)  
Перфильева И. (Чехия)  
Поддубный В.В. (Томск)  
Подопригора В.Г. (Красноярск)  
Рубан А.И. (Красноярск)  
Садовский М.Г. (Красноярск)

Семёнкин Е.С. (Красноярск)  
Семёнов А.Т. (Новосибирск)  
Сороко Э.М. (Белоруссия)  
Стоянов Й. (Великобритания)  
Урясев С. (США)  
Финкель Р. (Великобритания)  
Хасан Б.И. (Красноярск)  
Хлебопрос Р.Г. (Красноярск)  
Холтон Г. (США)  
Цих А.К. (Красноярск)  
Цициашвили Г. Ш. (Владивосток)  
Черемисин А. А. (Красноярск)  
Шайдуров В.В. (Красноярск)  
Шмидт Э. (Польша)  
Шокин Ю.И. (Новосибирск)  
де Шпинадель В. (Аргентина)  
Янковская А.Е. (Томск)

О р г а н и з а ц и о н н ы й   к о м и т е т

Амельчугова С.В. (Красноярск)  
Андрейчикова Т.В. (Красноярск)  
Баранова И.В. (Красноярск)  
Голденюк Е.Е. (Красноярск)  
Клочков С.В. (Красноярск)  
Лукьянова Н.А. (Красноярск)  
Семенова Д.В. (Красноярск)

К о н т а к т ы

- Оргкомитет XII международной ФАМЭБ'2013 конференции: Торгово-экономический институт СФУ, ул. Лиды Прушинской 2, к.7-40, Красноярск, 660075, тел. +73912219842, Голденюк Елена Евгеньевна, e-mail: ellen\_gold@rambler.ru; Сибирский федеральный университет, пр. Свободный 79, к.34-12, Красноярск, 660041, тел. +73912443149, Семёнова Дарья Владиславовна, e-mail: famemconf@gmail.com, url: <http://fam.conf.sfu-kras.ru>.
- Сопредседатель конференции: Воробьёв Олег Юрьевич, т. +73912494795; e-mail: oleg.yu.vorobyev@gmail.com, url: [eventology-theory.ru](http://eventology-theory.ru)
- Секретарь оргкомитета: Семёнова Дарья Владиславовна, т. +79029468205; e-mail: famemconf@gmail.com

*THE XII INTERNATIONAL CONFERENCE  
ON FINANCIAL AND ACTUARIAL MATHEMATICS  
AND EVENTOLOGY OF SAFETY  
19 ~ 20 April, 2013, Krasnoyarsk*

**Topics**

- Financial and actuarial mathematics
- Converging nano-bio-info-cogno sciences and technologies for improving human performance
- Eventological mathematics in the humanities, socioeconomic and natural sciences
- Eventology of safety and risk
- Probability theory and statistics
- Eventology and mathematical eventology
- Philosophical eventology and philosophy of probability
- Practical eventology
- Eventological economics
- Eventological psychology
- Eventological problems of artificial intelligence
- Decision-making under risk and uncertainty
- Risk measurement and risk models
- Theory of fuzzy events and generalized theory of uncertainty
- Mathematical onset to chaos in economy
- System analysis and events management
- Events distributions mechanism design theory

**Conference CoChairs**

Vorobyev Oleg (Russia)

Komarov Sergey (Russia)

**Program Committee**

Abdushukurov Abdurahim (Uzbekistan)  
Allahverdi Novruz (Turkey)  
Alexandrov Yuri (Russia)  
Aliev Telman (Azerbaijan)  
Alizada Tahir (Azerbaijan)  
Amelchugov Sergey (Russia)  
Angelov Plamen (UK)  
Andreyev Jury (Russia)  
Cheremisin Alexander (Russia)  
Cox Louis (USA)  
Dobronec Boris (Russia)  
Dorrer Georgy (Russia)  
Finkel Rebecca (UK)  
Giannopoulos Kostas (UAE)  
Goldblatt Joe (UK)  
Gorban' Alexander (UK)  
Gorodnitsky Irina (USA)  
Granovskaya Rada (Russia)  
Grigor'ev Yuri (Russia)

Guts Alexander (Russia)  
Holton Glyn (USA)  
Khasan Boris (Russia)  
Khlebopros Rem (Russia)  
Kiseliova Tatiana (Germany)  
Kuimov Basil (Russia)  
Kytmanov Alexander (Russia)  
Lefebvre Vladimir (USA)  
Lepsky Vladimir (Russia)  
Lukin Vladimir (Canada)  
Lutsenko Euheny (Russia)  
Mazharov Vladimir (Russia)  
Nemirovsky Valentin (Russia)  
Novosyolov Arcady (Russia)  
Okhorzin Vladimir (Russia)  
Orsingher Enzo (Italy)  
Perfil'eva Irina (Czech Republik)  
Poddubny V.V. (Toms)  
Podoprigora Vladimir (Russia)

Ruban Anatoly (Russia)  
Sadovsky Michael (Russia)  
Semenkin Euheny (Russia)  
Semenov Alexander (Russia)  
Shaidurov Vladimir (Russia)  
Shokin Yuri (Russia)  
Soroko Eduard (Belarus)  
de Spinadel Vera (Argentina)  
Stoyanov Jordan (UK)  
Szmidi Eulalia (Poland)  
Tsykh August (Russia)  
Tsytsyhashvili Gurami (Russia)  
Uryasev Stanislav (USA)  
Vitlinsky Valdemar (Ukraine)  
Vorobyev Alexey (Russia)  
Yankovskaya Anna (Russia)  
Zander Euheniya (Russia)

**Organizing Committee**

Amelchugova Svetlana (Russia)  
Andreychikova Tatyana (Russia)  
Baranova Irina (Russia)  
Goldenok Ellen (Russia)  
Klochkov Svyatoslav (Russia)  
Lukyanova Nataly (Russia)  
Semenova Daria (Russia)

**Contacts**

- Organizing Committee of the XII international FAMES'2013 Conference: Trade Economy Institute of SFU, ul. Lidy Prushinskoi 2, r.7-40, Krasnoyarsk, 660075, Ph. +73912219842, Goldenok Ellen, e-mail: ellen\_gold@rambler.ru; Siberian Federal University, pr.Svobodny 79, r.34-12, Krasnoyarsk, 660041, Ph. +73912443149, Semenova Daria, e-mail: famemconf@gmail.com, url: <http://fam.conf.sfu-kras.ru>.
- Conference CoChair: Vorobyev Oleg, tel. +73912494795; e-mail: oleg.yu.vorobyev@gmail.com, url: [eventology-theory.com](http://eventology-theory.com)
- Organizing Secretary: Semenova Daria, tel. +79029468205; e-mail: famemconf@gmail.com



## С о д е р ж а н и е

ABDUSHUKUROV A.A., AND MURADOV R.S. ( <i>Tashkent, Uzbekistan</i> ) .....	15
<i>Dependent random censoring from the right and the survival function copula estimates</i>	
ABDUSHUKUROV A.A., AND NURMUHAMEDOVA N.S. ( <i>Tashkent, Uzbekistan</i> ) .....	20
<i>On asymptotic minimax efficiency of maximum likelihood estimates by incomplete data</i>	
DEMOTCHKINE VAL ( <i>London, United Kingdom</i> ) .....	22
<i>Hedge Fund Strategies, Financial Instability, and Stock Prices Stabilization</i>	
MASAEV SERGEY ( <i>Krasnoyarsk</i> ) .....	29
<i>Company management system estimation on the basis of integrated indicators</i>	
VOROBYEV OLEG ( <i>Krasnoyarsk</i> ) .....	33
<i>In search of a primary source: remaking the paper (1975) where at the first time a definition of the lattice (Vorob'ev) expectation of a random set was given</i>	
VOROBYEV OLEG ( <i>Krasnoyarsk</i> ) .....	40
<i>Applicable eventology of safety: inconclusive totals</i>	
АБАСОВ Р.З. и З.А. САФАРОВА ( <i>Баку, Азербайджан</i> ) .....	50
<i>Не пора ли думать об универсальных методах решения для олимпиадных задач</i>	
АБДУШУКУРОВ А.А. и Л.Р.КАКАДЖАНОВА ( <i>Ташкент, Узбекистан</i> ) .....	55
<i>О некоторых свойствах одного класса эмпирических процессов</i>	
АБДУШУКУРОВ А.А. и К.С. САГИДУЛЛАЕВ ( <i>Ташкент, Узбекистан</i> ) .....	57
<i>Оценивание производных функции распределения в модели пропорциональных интенсивностей</i>	
АЗИМОВ Ж.Б. ( <i>Ташкент, Узбекистан</i> ) .....	60
<i>Об асимптотических поведениях ветвящихся процессов с неоднородной иммиграцией, зависящей от состояния</i>	
БАРАНОВА И.В. ( <i>Красноярск</i> ) .....	62
<i>Регрессионный анализ двудольных множеств событий</i>	
БАРАНОВА И.В. и Е.Н. ТЕМЕРОВА ( <i>Красноярск</i> ) .....	73
<i>Применение энтропии для исследования распределений двудольных множеств событий</i>	

БЕЛОСОХОВА А.В. и В.Е. РАСПОПОВ (Красноярск) .....	77
<i>Численная идентификация правой части специального вида системы двух уравнений параболического типа</i>	
БЕСКОРСКИЙ Н.С. и Б.В. ОЛЕЙНИКОВ (Красноярск) .....	80
<i>Решение обратной задачи многокритериального выбора в методологии метода анализа иерархий</i>	
БОРОВИКОВ И.А. и М.Г. САДОВСКИЙ (Калифорния, Красноярск) .....	83
<i>Приложение N-граммного анализа в финансовой математике</i>	
ВЕРЕТНОВА К.Ю. (Красноярск) .....	92
<i>Задача построения множества наиболее независимых событий, которые порождают заданное террасное разбиение пространства всеобщих элементарных исходов</i>	
ВОРОБЬЕВ О.Ю. (Красноярск) .....	97
<i>В поисках первоисточника: римейк статьи (1975), где впервые дается определение решетчатого (воробьевского) ожидания случайного множества</i>	
ВОРОБЬЕВ О.Ю. (Красноярск) .....	104
<i>Применимая эвентология безопасности: неокончательные итоги</i>	
ВОРОБЬЕВ О.Ю. (Красноярск) .....	115
<i>Голубь и разум: эвентологическая трансдукция голубинового эксперимента Геррнштейна и проблема вероятностного выбора</i>	
ВОРОБЬЕВ О.Ю. (Красноярск) .....	133
<i>Эвентологические особенности выбора точки из Фреше-многогранника в симплексе распределений</i>	
ВОРОБЬЕВ О.Ю. (Красноярск) .....	142
<i>Измерение многомерных событийных зависимостей в эвентологическом портфельном анализе финансово-актуарных, гуманитарных и социально-экономических рынков</i>	
ВОРОБЬЕВ О.Ю. (Красноярск) .....	150
<i>Гиббсовская характеристика множеств событий</i>	
ВОРОБЬЕВ О.Ю. (Красноярск) .....	158
<i>Аппроксимация множества событий в средневероятном</i>	
ВОРОБЬЕВ О.Ю. и А.С. НИФОНТОВ (Красноярск) .....	162
<i>Портфельный анализ событий: расширенная версия</i>	

Глик Л.А и М.Е. СЕМЕНОВ (Томск) .....	165
<i>Статистический анализ внутридневных колебаний спреда на российском фондовом рынке</i>	
ГНЕЗДИЛОВА М.А и М.Г. САДОВСКИЙ (Красноярск) .....	169
<i>Приложение N-граммного анализа в финансовой математике</i>	
ГОРДЕЕВА К.В. (Красноярск) .....	175
<i>Изучение потребительского спроса с помощью методов ассоциативных правил</i>	
ГОРОХОВ А.А. и В.Е. РАСПОПОВ (Красноярск) .....	180
<i>Численная идентификация правой части специального вида системы двух уравнений параболического типа</i>	
ГРИГОРЬЕВ Ю.Д. (Санкт-Петербург) .....	183
<i>Размышления по поводу судьбы нашей вузовской науки</i>	
ГРИГОРЬЕВ Ю.Д. (Санкт-Петербург) .....	195
<i>Определители Ганкеля и полиномиальная регрессия</i>	
ДОБРОНЕЦ Б.С. и О.А. ПОПОВА (Красноярск) .....	199
<i>Численная оптимизация со случайными данными</i>	
ДОРРЕР М.Г., ЗЫРЯНОВ А.А. и Е.А. ЛАНЦЕВ (Красноярск) .....	203
<i>Идентификация и прогноз функционирования модели бизнес-процессов дистанционного обучения</i>	
ЕЛИНА М.В, ЛОГИНОВА Е.Б и М.Г. САДОВСКИЙ (Красноярск) .....	208
<i>Об определении отношения избирателей к кандидату (на примере досрочных выборов Главы г. Красноярска)</i>	
ЕСИН Р.В. (Красноярск) .....	210
<i>Решение обратной задачи теории риска в рамках модели CVaR</i>	
ЗЫРЯНОВ А.А. (Красноярск) .....	215
<i>Расчет числовых характеристик однородной GERT-сети на основе эквивалентных преобразований</i>	
ЖАТКИНА К.В. и И.В. БАРАНОВА (Красноярск) .....	220
<i>Решение задачи скоринга на основе статистик разнотипных данных</i>	
ИМОМОВ А.А. и Ш.Г. МУСУРМОНОВА (Карши, Узбекистан) .....	224
<i>О сходимости к экспоненциальному закону и дифференциальный аналог основной леммы теории критических процессов Гальтона-Ватсона</i>	



КИРЕЕВ В.А. (Красноярск) .....	227
<i>О численном решении изотермической модели фильтрации несжимаемой вязкой жидкости в пористой среде</i>	
КИРИЛЮК А.А. (Красноярск) .....	231
<i>Разработка и применение электронного образовательного ресурса «Алгоритмы быстрых вычислений»</i>	
КЛОКОВА К.А и М.Г. САДОВСКИЙ (Красноярск) .....	237
<i>О распределении расстояний до ближайшего соседа в Марковских цепях</i>	
КОМОРОВСКИЙ В.С. и Г.А. ДОРРЕР (Красноярск) .....	244
<i>Когнитивное моделирование в управлении рисками от лесных пожаров</i>	
КУЧЕРОВ М.Н. (Красноярск) .....	249
<i>Применение методов факторного анализа для исследования социально-экономического положения населения</i>	
КУСТИЦКАЯ Т.А. (Красноярск) .....	255
<i>Нахождение эллиптического конуса приемлемых рисков для некоторых норм Гельдера</i>	
ЛАНЦЕВ Е.А. (Красноярск) .....	261
<i>Применение process mining для разработки агентных имитационных моделей на примере процесса дистанционного обучения</i>	
ЛУКЬЯНОВА Н.А. (Красноярск) .....	264
<i>Энтропия множества событий и его подмножества</i>	
МАКСИМОВА К.И. и И.В. БАРАНОВА (Красноярск) .....	272
<i>Нахождение ассоциативных правил в статистике покупок продуктов</i>	
МЕДНИЦКИЙ А.Н. и А.Ю. ТРУСОВА (Самара) .....	277
<i>Актuarные расчеты по добровольному медицинскому страхованию</i>	
МЕДНИЦКИЙ А.Н. и А.Ю. ТРУСОВА (Самара) .....	280
<i>Расчет франшизы для бета, гамма и экспоненциального распределения</i>	
НАРТОВ Я.В. (Красноярск) .....	284
<i>Событийная визуализация распределения Пуассона и его средневероятного события</i>	
НИФОНТОВ А.С. (Красноярск) .....	287
<i>Алгоритмы Монте-Карло генерирования распределений множеств событий и их применение в портфельном анализе валютных рынков и рынков драгоценных металлов</i>	

НОВОСЕЛОВ А.А. (Красноярск) .....	293
<i>Дележ риска между факторами и активами</i>	
НОВОСЕЛОВ А.А. (Красноярск) .....	301
<i>Некоторые задачи управления пожарным риском</i>	
ОЛЕЙНИКОВ Б.В. и А.И. ШАЛАБАЙ (Красноярск) .....	305
<i>Алгоритм тематической классификации научных текстов</i>	
ПАЗЕНКО О.В и М.Г. САДОВСКИЙ (Красноярск) .....	308
<i>Восприятие рекламы в малых населённых пунктах на примере сотового оператора Билайн</i>	
ПОДДУВНЫЙ В.В., ПЕХТЕРЕВ А.С. (Томск) .....	311
<i>Копулы сглаженных эмпирических распределений при наличии связей (совпадений) и их применение в имитационном моделировании</i>	
ПОПОВ А.В. и В.С. ШАБАНОВ (Красноярск) .....	321
<i>Информационно-поисковые системы для электронных библиотек</i>	
ПОПОВА Е.А. (Красноярск) .....	324
<i>К истории закона больших чисел</i>	
ПОПОВА О.А. (Красноярск) .....	328
<i>О подходах к постановкам оптимизационных задач с неопределёнными данными</i>	
СЕВАСТЬЯНОВА Н.А. (Красноярск) .....	333
<i>Об обобщении результатов Никольского С.М. для погрешности усложнённых квадратурных формул на случай класса функций, имеющих дробные производные</i>	
СЕМЕНОВА Д.В. и Ю.С. КОЧАНОВА (Красноярск) .....	335
<i>Эвентологические распределения случайных множеств событий на основе ассоциативных функций</i>	
СЕМЕНОВА Д.В. и Ю.С. КОЧАНОВА (Красноярск) .....	341
<i>Примеры семейств ассоциативных случайных множеств</i>	
ФЕДОРОВ Е.В. (Красноярск) .....	348
<i>Восстановление модели бизнес-процессов дистанционного обучения на основе данных информационной системы. Постановка задачи</i>	
ЦИЦИАШВИЛИ Г.Ш. (Владивосток) .....	349
<i>Распознавание образов с помощью многомерных интервалов</i>	

ШИРЯЕВА Л.К. (Самара) .....	351
<i>О распределении статистики отношения правдоподобия, применяемой для тестирования нормальной выборки на один выброс</i>	
ШИХЛИНСКАЯ Р.Ю. (Баку, Азербайджан) .....	358
<i>Пример обработки нечетких правил вывода в экспертных системах</i>	
ЮМАГУЛОВ Д.Т. (Уфа, Россия) .....	359
<i>Функциональные моментные стратегии управления портфелем ценных бумаг</i>	
<i>С п и с о к а в т о р о в</i> .....	361
<i>L i s t o f a u t h o r s</i> .....	361
<i>A b s t r a c t s</i> .....	362



## Желаемое

*Иди спокойно среди шума и суеты, и помни, покой — в тишине. Будь в хороших отношениях со всеми насколько возможно без уступок. Говори свою правду спокойно и ясно; и слушай других, даже тупых и несведущих — им тоже есть что сказать. Избегай крикливых и агрессивных, они — досада твоему духу.*

*Сравнивая себя с другими, можешь стать тщеславным или ожесточенным; ибо всегда найдутся люди выше или ниже тебя. Радуйся как своим достижениям, так и своим планам. Проявляй интерес к своему жизненному пути, даже скромному; он — твоё единственное подлинное достояние в изменчивых превратностях судьбы.*

*Будь осторожен в своих делах, ибо мир полон обмана. И не позволяй ему ослепить тебя, чтобы скрыть существование добродетели; многие люди стремятся к высоким идеалам, и всюду жизнь полна героизма. Будь самим собой. В особенности не симулируй привязанность. Не будь циничным к любви; поскольку перед лицом всех разочарований и бесплодности только она так же вечна, как трава. Принимай с благодарностью уроки лет, изящно расставаясь с приметами молодости.*

*Укрепляй силу духа — твой щит от ударов судьбы. И не беспокой себя мрачными мыслями. Много страхов рождаются от усталости и одиночества. Не забывая о полезной дисциплине, оставайся нежным к себе. Ты — дитя вселенной, не меньше, чем деревья и звезды; и ты имеешь право быть здесь. Осознаешь ли ты это или нет, но вне всякого сомнения Вселенная развивается так, как должно.*

*Поэтому, будь в мире с Богом, каким бы ты его себе не представлял. И чем бы ты не занимался, и к чему бы не стремился среди шумной неразберихи жизни, сохраняй покой в своей душе. Со всем его обманом, тяжелой работой и разбитыми мечтами; это — все же прекрасный мир. Радуйся жизни, стремись быть счастливым.*

Макс Эрманн, 1927.

(перевод: Олег Воробьев, 2008)

## Desiderata

*Go placidly amid the noise and the haste, and remember what peace there may be in silence. As far as possible without surrender be on good terms with all persons. Speak your truth quietly and clearly; and listen to others, even to the dull and the ignorant, they too have their story. Avoid loud and aggressive persons, they are vexations to the spirit.*

*If you compare yourself with others, you may become vain or bitter; for always there will be greater and lesser persons than yourself. Enjoy your achievements as well as your plans. Keep interested in your own career, however humble; it is a real possession in the changing fortunes of time.*

*Exercise caution in your business affairs, for the world is full of trickery. But let not this blind you to what virtue there is; many persons strive for high ideals, and everywhere life is full of heroism. Be yourself. Especially do not feign affection. Neither be cynical about love; for in the face of all aridity and disenchantment it is as perennial as the grass. Take kindly the counsel of the years, gracefully surrendering the things of youth.*

*Nurture strength of spirit to shield you in sudden misfortune. But do not distress yourself with dark imaginings. Many fears are born of fatigue and loneliness. Beyond a wholesome discipline, be gentle with yourself. You are a child of the universe, no less than the trees and the stars; you have a right to be here. And whether or not it is clear to you, no doubt the universe is unfolding as it should.*

*Therefore, be at peace with God, whatever you conceive Him to be. And whatever your labors and aspirations in the noisy confusion of life, keep peace in your soul. With all its sham, drudgery and broken dreams; it is still a beautiful world. Be cheerful. Strive to be happy.*

# Dependent random censoring from the right and the survival function copula estimates

Abdushukurov Abdurahim Akhmedovich

National University of Uzbekistan  
Uzbekistan, Tashkent  
a\_abdushukurov@rambler.ru

Muradov Rustamjon Sobitkhonovich

National University of Uzbekistan  
Uzbekistan, Tashkent  
r\_muradov1985@rambler.ru,

**Abstract.** In this article a new estimators for survival functions with using Archimedean copulas under onedimensional and multidimensional random censoring from the right are constructed. The property of uniform consistency of estimators is proved.

**Keywords.** Random censorship, survival function, Archimedean copulas, relative-risk power estimator.

## 1 Introduction

In survival analysis our interest focuses on a nonnegative random variables (r.v.-s) denoting death times of biological organisms or failure times of mechanical systems. A difficulty in the analysis of survival data is the possibility that the survival times can be subjected to random censoring by other nonnegative r.v.-s and therefore we observe incomplete data. There are various types of censoring mechanisms. In this article we consider only right censoring model and problem of estimation of survival function when the survival times and censoring times are dependent and propose new estimates of survival function assuming that the dependence structure is described by a known copula function.

## 2 The copula dependence and estimator: onedimensional case

On the probability space  $(\Omega, \mathcal{A}, P)$  we consider  $\{(X_k, Y_k), k \geq 1\}$  — a sequence of independent and identically distributed pairs of nonnegative r.v.-s with common joint distribution function (d.f.)

$$H(x, y) = P(X_1 \leq x, Y_1 \leq y), (x, y) \in \overline{R}^{+2} = [0, \infty]^2.$$

We suppose that the marginal d.f.

$$F(x) = P(X_1 \leq x) = H(x, +\infty)$$

and

$$G(y) = P(Y_1 \leq y) = H(+\infty, y), x, y \in \overline{R}^+,$$

are continuous and  $F(0) = G(0) = 0$ . Assume that the sequence  $\{X_k, k \geq 1\}$  is right censored by the sequence

$\{Y_k, k \geq 1\}$  and at  $n$ -th stage of the experiment the observation available the sample

$$\mathbb{V}^{(n)} = \{(Z_k, \delta_k), 1 \leq k \leq n\},$$

where  $Z_k = \min(X_k, Y_k)$ ,  $\delta_k = I(Z_k = X_k)$  and  $I(A)$  is the indicator of the event  $A$ . Should be noted that it does not require independence of sequences  $\{X_k\}$  and  $\{Y_k\}$  (see [1, 2]). The problem is consist in estimating of the survival function

$$S^X(x) = P(X_1 > x) = 1 - F(x), x \in \overline{R}^+,$$

from the sample  $\mathbb{V}^{(n)}$ . Let

$$\overline{H}(x, y) = P(X_1 > x, Y_1 > y), (x, y) \in \overline{R}^{+2}$$

— a joint survival function of the pairs  $(X_k, Y_k)$ . According to Theorem of Sclar  $H$  and  $\overline{H}$  can be submitted through the appropriate copula functions (see [3, 4]):

$$H(x, y) = C(F(x); G(y)), (x, y) \in \overline{R}^{+2}, \quad (1)$$

$$\overline{H}(x, y) = C^*(S^X(x); S^Y(y)), (x, y) \in \overline{R}^{+2},$$

where copulas  $C$  and  $C^*$  are related as

$$C^*(u, v) = u + v - 1 + C(1 - u; 1 - v), \quad (2)$$

$$(u, v) \in [0, 1]^2.$$

In the sequel in order to construct estimates for the survival function  $S^X$ , assume that  $C^*$  is Archimedean copula, i.e.

$$C^*(u, v) = \varphi^{[-1]}[\varphi(u) + \varphi(v)], (u, v) \in [0, 1]^2,$$

where

$$\varphi: [0, 1] \rightarrow \overline{R}^+$$

is some generator function with the pseudo inverse  $\varphi^{[-1]}$ . Thus, by (1) and (2)

$$\overline{H}(x, y) = \varphi^{[-1]}[\varphi(S^X(x)) + \varphi(S^Y(y))], \quad (3)$$

$$(x, y) \in \overline{R}^{+2},$$

$$S^Z(x) = \varphi^{[-1]}[\varphi(S^X(x)) + \varphi(S^Y(x))], x \in \overline{R}^+.$$



We introduce a usual  $\lambda^X$ ,  $\lambda^Z$  and "crude"  $\lambda$ - hazard functions

$$\begin{aligned}\lambda^X(x) &= \lim_{\Delta \downarrow 0} \frac{1}{\Delta} P(x < X_1 \leq x + \Delta / X_1 > x), \\ \lambda^Z(x) &= \lim_{\Delta \downarrow 0} \frac{1}{\Delta} P(x < Z_1 \leq x + \Delta / X_1 > x, Y_1 > x), \\ \lambda(x) &= \lim_{\Delta \downarrow 0} \frac{1}{\Delta} P(x < X_1 \leq x + \Delta / X_1 > x, Y_1 > x).\end{aligned}\quad (4)$$

In order to construct a copula estimates for  $S^X$  consider the following easily verifiable equality:

$$\begin{aligned}\lambda^X(x) S^X(x) \varphi'(S^X(x)) &= \\ &= \lambda(x) S^Z(x) \varphi'(S^Z(x)).\end{aligned}\quad (5)$$

Integrating (5) over the interval  $[0, x]$  and denoting by  $\Lambda(x) = \int_0^x \lambda(t) dt$  and  $\Lambda^X(x) = \int_0^x \lambda^X(t) dt$  corresponding cumulative hazard functions we obtain the integral equation

$$\begin{aligned}\int_0^x S^X(t) \varphi'(S^X(t)) d\Lambda^X(t) &= \\ &= \int_0^x S^Z(t) \varphi'(S^Z(t)) d\Lambda(t), \quad x \in \overline{R}^+.\end{aligned}\quad (6)$$

Integral on the left side of (6) is equal to  $-\varphi(S^X(x))$  and then (6) takes the form

$$\varphi(S^X(x)) = - \int_0^x S^Z(t) \varphi'(S^Z(t)) d\Lambda(t), \quad (7)$$

$x \in \overline{R}^+$ .

Hence we find the expression for the survival function  $S^X$ :

$$S^X(x) = \varphi^{[-1]} \left[ - \int_0^x S^Z(t) \varphi'(S^Z(t)) d\Lambda(t) \right], \quad (8)$$

$x \in \overline{R}^+$ .

Note that the survival function  $S^Z$  permit usual empirical estimation by the values  $Z_k$  observed in the sample  $\mathbb{V}^{(n)}$ :

$$S_n^Z(x) = \frac{1}{n} \sum_{k=1}^n I(Z_k > x), \quad x \in \overline{R}^+. \quad (9)$$

Substituting (9) to the right of representation (8), we obtain a preliminary estimate of  $S^X$  as

$$\tilde{S}_n^X(x) = \varphi^{[-1]} [\mu_n(x)], \quad (10)$$

where

$$\mu_n(x) = - \int_0^x I(S_n^Z(t-) > 0) S_n^Z(t-) \varphi'(S_n^Z(t)) d\Lambda_n(t),$$

$$\Lambda_n(t) = \frac{1}{n} \sum_{k=1}^n \frac{I(Z_k \leq t, \delta_k = 1)}{S_n^Z(Z_k) - \frac{1}{n}}, \quad (11)$$

— the corresponding estimate for

$$\Lambda(t) = \int_0^t \frac{dP(Z_1 \leq s, \delta_1 = 1)}{P(Z_1 > s)}.$$

Estimate (10) plays a supporting role in the construction of the main estimates for  $S^X$  in the future. Since the estimates will mainly be studied using the martingale method, the more convenient for us to present them through the counting processes. Let

$$N_k(t) = I(Z_k \leq t, \delta_k = 1).$$

Define the counting processes

$$\overline{N}_n(t) = \sum_{k=1}^n N_k(t)$$

and

$$\mathbb{J}_n(t) = n S_n^Z(t-) = \sum_{k=1}^n I(Z_k \geq t).$$

Then the estimates (10) and (11) can be represented as

$$\begin{aligned}\tilde{S}_n^X(x) &= \\ &= \varphi^{[-1]} \left[ - \frac{1}{n} \int_0^x I(\mathbb{J}_n(t) > 0) \varphi' \left( \frac{\mathbb{J}_n(t)}{n} \right) d\overline{N}_n(t) \right],\end{aligned}\quad (12)$$

$$\Lambda_n(t) = \int_0^t \frac{I(\mathbb{J}_n(s) > 0)}{\mathbb{J}_n(s)} d\overline{N}_n(s)$$

Given the analog left side of (6), i.e.

$$\begin{aligned}\varphi(S^Z(x)) &= \\ &= - \int_0^x S^Z(t) \varphi'(S^Z(t)) d\Lambda^Z(t),\end{aligned}\quad (13)$$

where

$$\Lambda^Z(t) = \int_0^t \lambda^Z(s) ds,$$

together with (9) also obtain other estimate for  $S^Z$  as

$$\begin{aligned}\tilde{S}_n^Z(x) &= \\ &= \varphi^{[-1]} \left[ - \frac{1}{n} \int_0^x I(\mathbb{J}_n(t) > 0) \varphi' \left( \frac{\mathbb{J}_n(t)}{n} \right) d\overline{N}_n^Z(t) \right],\end{aligned}\quad (14)$$

where

$$\Lambda_n^Z(t) = \int_0^t \frac{I(\mathbb{J}_n(s) > 0)}{\mathbb{J}_n(s)} d\overline{N}_n^Z(s)$$

is estimate for  $\Lambda^Z(t)$  and

$$\begin{aligned}\bar{N}_n^Z(t) &= n(1 - S_n^Z(t)) = n - \mathbb{J}_n(t+) = \\ &= \sum_{k=1}^n N_k^Z(t) = \sum_{k=1}^n I(Z_k \leq t)\end{aligned}$$

— the counting process. For  $S^X$  have the following obvious identity obtained from the representations (7) and (13):

$$S^X(x) = \varphi^{[-1]}[\varphi(S^Z(x)) \cdot \chi(x)], \quad (15)$$

where

$$\chi(x) = \frac{\left(-\int_0^x S^Z(t) \varphi'(S^Z(t)) d\Lambda(t)\right)}{\left(-\int_0^x S^Z(t) \varphi'(S^Z(t)) d\Lambda^Z(t)\right)}.$$

Now substituting the empirical estimate of (9) under the first factor on the right of representation (15) and the corresponding estimates (12) and (14) instead of integrals we obtain the desired estimate  $S^X$  in the form

$$S_n^X(x) = \varphi^{[-1]}[\varphi(S_n^Z(x)) \cdot \chi_n(x)] \quad (16)$$

where

$$\begin{aligned}\chi_n(x) &= \frac{\left(-\int_0^x I(\mathbb{J}_n(t) > 0) \varphi'\left(\frac{\mathbb{J}_n(t)}{n}\right) d\bar{N}_n(t)\right)}{\left(-\int_0^x I(\mathbb{J}_n(t) > 0) \varphi'\left(\frac{\mathbb{J}_n(t)}{n}\right) d\bar{N}_n^Z(t)\right)}, \\ \varphi(S_n^Z(x)) &= -\int_0^x I(\mathbb{J}_n(s) > 0) \cdot \left[\varphi\left(\frac{\mathbb{J}_n(s)}{n}\right) - \varphi\left(\frac{\mathbb{J}_n(s)}{n} - \frac{1}{n}\right)\right] d\bar{N}_n^Z(s) \quad (17)\end{aligned}$$

— estimator of  $\varphi(S^Z(x))$ . Investigate the properties of new estimate (16) for the survival function  $S^X$ .

Denote

$$Z^{(n)} = \sup\{x \geq 0 : \mathbb{J}_n(x) > 0\},$$

$$T_Z = \sup\{x \geq 0 : S^Z(x) > 0\}, \quad \Psi(x) = -x\varphi'(x).$$

Introduce the regularity conditions with respect to  $S^X$ ,  $S^Z$  and the copula generator  $\varphi$ . By  $\Lambda^*$  in conditions below denote both of  $\Lambda$  and  $\Lambda^Z$ :

(C1) The function  $\varphi(\cdot)$  is strictly decreasing on  $(0,1]$  and is sufficiently smooth in the sense that the first two derivatives of the functions  $\varphi(x)$  and  $\Psi(x)$  are bounded for  $x \in [\varepsilon, 1]$ , where  $\varepsilon > 0$  is arbitrary. Moreover, the first derivative  $\varphi'$  is bounded away from zero on  $[0, 1]$ ;

(C2)  $0 < \int_0^{T_Z} [\Psi(S^Z(x))]^m d\Lambda^*(x) < \infty$  for  $m = 0, 1, 2$ ;

(C3)  $\int_0^{T_Z} |\Psi'(S^Z(x))| d\Lambda^*(x) < \infty$ ;

We prove the following statement about the uniform consistency of all input estimators using some ideas from [5, 4].

**Theorem 1.** Let conditions (C1)-(C3) are hold. Then for  $n \rightarrow \infty$

$$\sup_{0 \leq x \leq T_Z} |S_n^X(x) - S^X(x)| \xrightarrow{P} 0.$$

**Proof.** See in [7].

**Remark 1.** Consider independent censoring model (i.e.  $\{X_k\}$  and  $\{Y_k\}$  are mutually independent). In this case in (2)

$$C(u; v) = uv = C^*(u; v), u, v \in [0, 1]$$

and hence  $\varphi(u) = -\log u$ ,  $u \in (0, 1]$  and

$$\varphi^{[-1]}(t) = \varphi^{-1}(t) = \exp(-t), t \geq 0,$$

so that

$$S^Z(x) = S^X(x)S^Y(x), x \in \bar{R}^+. \quad (18)$$

Now it is easy to verify that from (12) and (16) respectively we obtain the exponential-hazard estimator

$$\tilde{S}_n^X(x) = \exp\left\{-\int_0^x \frac{I(\mathbb{J}_n(t) > 0)}{\mathbb{J}_n(t)} d\bar{N}_n(t)\right\}, \quad (19)$$

and relative-risk power estimator of Abdushukurov (1998) (see [1]):

$$S_n^X(x) = [S_n^Z(x)]^{R_n(x)}, \quad R_n(x) = \frac{\Lambda_n(x)}{\Lambda_n^Z(x)}. \quad (20)$$

Moreover the Zeng-Klein's (1994) copula-graphic estimator is (see [5, 4]):

$$\begin{aligned}\hat{S}_n^X(x) &= \varphi^{[-1]}\left[\int_0^x I(\mathbb{J}_n(t) > 0) \left(\varphi\left(\frac{\mathbb{J}_n(t)-1}{n}\right) - \right.\right. \\ &\quad \left.\left. - \varphi\left(\frac{\mathbb{J}_n(t)}{n}\right)\right) d\bar{N}_n(t)\right],\end{aligned} \quad (21)$$

which in independence model is reduced to well - known Kaplan - Meier product - limit estimator

$$\hat{S}_n^X(x) = \prod_{t \leq x} \left\{1 - \frac{d\bar{N}_n(t)}{\mathbb{J}_n(t)}\right\}. \quad (22)$$

Let  $\tilde{S}_n^Y$ ,  $S_n^Y$  and  $\hat{S}_n^Y$  are respectively estimators of  $S^Y$  of exponential-hazard, relative-risk power and product-limit structures obtained from formulas (20), (21) and (22) by using events  $\delta_k = 0$  instead of  $\delta_k = 1$ . Then we have:

(a)  $\tilde{S}_n^X(x)\tilde{S}_n^Y(x) = \exp\{-\Lambda_n^Z(x)\} \neq S_n^Z(x)$   
and for  $x \geq Z_{(n)} = \max\{Z_k, 1 \leq k \leq n\}$ ,  
 $\max\{\tilde{S}_n^X(x), \tilde{S}_n^Y(x)\} < 1$ ;

(b)  $S_n^X(x)S_n^Y(x) = S_n^Z(x)$ , for all  $x \in \overline{R}^+$  and  $S_n^X(x) = S_n^Y(x) = 0$  for  $x \geq Z_{(n)}$ ;

(c)  $\hat{S}_n^X(x)\hat{S}_n^Y(x) \neq S_n^Z(x)$  and for  $x \geq Z_{(n)}$  the estimators  $\hat{S}_n^X$  and  $\hat{S}_n^Y$  are undefined. Moreover the estimators  $\hat{S}_n^X$  and  $\hat{S}_n^Y$  require also the condition  $P(X_k = Y_k) = 0, k = 1, 2, \dots$ , which in many practical situations is not hold. Thus only the relative-risk power estimators have identifiability properties with independence censoring model satisfying empirical analogue of equality (18). Consequently a new estimator (16) is more suitable estimator for  $S^X$  than the estimators (12) and (20).

### 3 Multidimensional dependent censoring model

For convenience of a statement of article we shall consider a case of two-dimensional random vectors. On the probability space  $(\Omega, \mathcal{A}, P)$  we consider

$$\mathbf{X} = \{(X_{1i}, X_{2i}), i \geq 1\}$$

and

$$\mathbf{Y} = \{(Y_{1i}, Y_{2i}), i \geq 1\}$$

— two sequences of independent and identically distributed random vectors with general survival functions

$$\overline{F}(x, y) = P(X_{11} > x, X_{21} > y)$$

and

$$\overline{G}(x, y) = P(Y_{11} > x, Y_{21} > y),$$

$(x, y) \in \overline{R}^{+2}$ . Note that these sequences may be dependent. Let's enter into consideration marginal survival functions

$$\begin{aligned} S_1^X(x) &= P(X_{11} > x), \\ S_2^X(y) &= P(X_{21} > y), \\ S_1^Y(z) &= P(Y_{11} > z), \end{aligned} \quad (23)$$

$$S_2^Y(\nu) = P(Y_{21} > \nu), x, y, z, \nu \geq 0,$$

and we suppose that they are continuous. Assume that the sequence  $\mathbf{X}$  is right censored by the sequence  $\mathbf{Y}$  and at the  $n$ -th stage of the experiment the observation available the sample  $\mathbf{V}^{(n)} = \{(Z_i, \delta_i), i = 1, \dots, n\}$ , where  $Z_i = (Z_{1i}, Z_{2i})$ ,  $Z_{ki} = (X_{ki}, Y_{ki})$ ,  $\delta_i = (\delta_{1i}, \delta_{2i})$ ,  $\delta_{ki} = I(Z_{ki} = X_{ki}), k = 1, 2; i = 1, \dots, n$ . The problem is consist in estimating of the survival function  $\overline{F}$  from the sample  $\mathbf{V}^{(n)}$  with nuisance function  $\overline{G}$ . Let's notice that the r.v.-s of interest  $X_{ki}$  are observed only at  $\delta_{ki} = 1$ . Let's enter the joint survival function of a vector  $(X_{11}, X_{21}, Y_{11}, Y_{21})$ :

$$\begin{aligned} K(x, y, z, \nu) &= P(X_{11} > x, X_{21} > y, Y_{11} > z, Y_{21} > \nu), \\ (x, y, z, \nu) &\in \overline{R}^{+4} = [0, \infty]^4. \end{aligned}$$

Then we have the following representations for survival functions:

$$\overline{F}(x, y) = K(x, y, 0, 0), \overline{G}(x, y) = K(0, 0, x, y),$$

$$\overline{H}(x, y) = P(Z_{11} > x, Z_{21} > y) = K(x, y, x, y). \quad (24)$$

According to Sklar's theorem [3] there exist a copula survival function  $C^*(\overline{u}), \overline{u} = (u_1, u_2, u_3, u_4) \in [0, 1]^4$  such that we have the following representation:  $(x, y, z, \nu) \in \overline{R}^{+4}$ :

$$K(x, y, z, \nu) = C^*(S_1^X(x), S_2^X(y), S_1^Y(z), S_2^Y(\nu)). \quad (25)$$

Further we assume, that  $C^*$  is Archimedean copula:

$$C^*(u_1, u_2, u_3, u_4) = \varphi^{[-1]}[\varphi(u_1) + \varphi(u_2) + \varphi(u_3) + \varphi(u_4)],$$

$$\overline{u} = (u_1, u_2, u_3, u_4) \in [0, 1]^4, \quad (26)$$

where  $\varphi$  is generator function of copula  $C^*$ . Then according to (24)-(26) we have

$$\begin{aligned} \overline{H}(x, y) &= \varphi^{[-1]}[\varphi(S_1^X(x)) + \varphi(S_2^X(y)) + \\ &+ \varphi(S_1^Y(x)) + \varphi(S_2^Y(y))], (x, y) \in \overline{R}^{+2}. \end{aligned} \quad (27)$$

Since

$$S_1^Z(x) = P(X_{11} > x, Y_{11} > x),$$

$$S_2^Z(y) = P(Y_{11} > y, Y_{21} > y), x \geq 0, y \geq 0,$$

then from the representation (27) we'll get following important formulas:

$$\begin{aligned} S_1^Z(x) &= \varphi^{[-1]}[\varphi(S_1^X(x)) + \varphi(S_1^Y(x))], \\ S_2^Z(y) &= \end{aligned} \quad (28)$$

$$= \varphi^{[-1]}[\varphi(S_2^X(y)) + \varphi(S_2^Y(y))], x \geq 0, y \geq 0.$$

Now from (28) also follows, that

$$\varphi(\overline{H}(x, y)) = \quad (29)$$

$$= \varphi(\overline{F}(x, y)) + \varphi(\overline{G}(x, y)), (x, y) \in \overline{R}^{+2}.$$

It is necessary to note, that formulas (28) allows to estimate two dimensional survival functions through one dimensional survival functions  $S_1^X$  and  $S_2^X$  accordingly by samples  $\mathbf{V}_1^{(n)} = \{(Z_{1i}, \delta_{1i}), i = 1, \dots, n\}$  and  $\mathbf{V}_2^{(n)} = \{(Z_{2i}, \delta_{2i}), i = 1, \dots, n\}$ ,  $\mathbf{V}_1^{(n)} + \mathbf{V}_2^{(n)} = \mathbf{V}^{(n)}$ . Then by the sample  $\mathbf{V}^{(n)}$  we estimate  $F$  using the formulas (27)-(29). First of all, functions  $\overline{H}$ ,  $S_1^Z$  and  $S_2^Z$  we shall estimate with the help its empirical analogies:

$$\overline{H}_n(x, y) = \frac{1}{n} \sum_{i=1}^n I(Z_{1i} > x, Z_{2i} > y),$$

$$S_{1n}^Z(x) = \frac{1}{n} \sum_{i=1}^n I(Z_{1i} > x),$$

$$S_{2n}^Z(x) = \frac{1}{n} \sum_{i=1}^n I(Z_{2i} > y), x \geq 0, y \geq 0. \quad (30)$$

Let's enter on the samples  $\mathbf{V}_1^{(n)}$  and  $\mathbf{V}_2^{(n)}$  also corresponding counting processes

$$\bar{N}_{nm}(t) = \sum_{k=1}^n I(Z_{mk} \leq t, \delta_{mk} = 1),$$

$$\bar{N}_{nm}^Z(t) = \sum_{k=1}^n I(Z_{mk} \leq t),$$

$$\mathbf{J}_m(t) = nS_{mn}^Z(t-) = \sum_{k=1}^n I(Z_{mk} \geq t), m = 1, 2.$$

Then respectively estimates of  $S_m^X, m = 1, 2$  we form as

$$S_{mn}^X(x) = \varphi^{[-1]}[\varphi(S_{mn}^Z(x)) \frac{\varphi(\tilde{S}_{mn}^X(x))}{\varphi(\tilde{S}_{mn}^Z(x))}], m = 1, 2, \quad (31)$$

where

$$\varphi(S_{mn}^Z(x)) = - \int_0^x I(\mathbf{J}_{mn}(t) > 0) [\varphi(\frac{\mathbf{J}_{mn}(t)}{n}) -$$

$$- \varphi(\frac{\mathbf{J}_{mn}(t)}{n} - \frac{1}{n})] d\bar{\mathbf{N}}_{mn}^Z(t),$$

$$\varphi(\tilde{S}_{mn}^Z(x)) = - \int_0^x I(\mathbf{J}_{mn}(t) > 0) \varphi'(\frac{\mathbf{J}_{mn}(t)}{n}) d\bar{\mathbf{N}}_{mn}^Z(t),$$

$$\varphi(\tilde{S}_{mn}^X(x)) = - \int_0^x I(\mathbf{J}_{mn}(t) > 0) \varphi'(\frac{\mathbf{J}_{mn}(t)}{n}) d\bar{\mathbf{N}}_{mn}(t).$$

Now joint survival function  $\bar{F}(x, y)$  we estimate as:

$$\bar{F}_n(x, y) = \varphi^{[-1]}[\varphi(S_{1n}^X(x)) + \varphi(S_{2n}^X(y))]. \quad (32)$$

Let  $\Delta = [0, T^{(1)}] \times [0, T^{(2)}]$ , where  $T^{(m)} = \sup\{x \geq 0 : S_m^Z(x) > 0\}, m = 1, 2$ . We have multidimensional analogue of theorem 1.

**Theorem 2.** Under regularity conditions (C1)-(C3) on functions  $\varphi, S_m^X, S_m^Z, m = 1, 2$  we have

$$\sup_{(x,y) \in \Delta} |\bar{F}_n(x, y) - \bar{F}(x, y)| \xrightarrow{P} 0. \quad (33)$$

Since the proof of theorem 2 is analogue to theorem 1, therefore it is omitted (see [6] for details).

**Remark 2.** It is necessary to note that for estimators (16) and (32) we prove also the properties of convergence to appropriate Gaussian processes.

## References

- [1] Abdushukurov A.A. *Estimators of Unknown Distributions from Incomplete Observations and its Properties*. LAMBERT Academic Publishing, 2011 (in Russian).
- [2] Muradov R.S. Abdushukurov A.A. Copula — graphical power estimators of the distribution function for the dependent model of random censoring on the right. *Acta of National University of Uzbekistan*, №1/2:201–204, 2010 (in Russian).
- [3] Nelsen R.B. *An introduction to copulas*. Springer, New York, 1999 (Second edition in 2006.).
- [4] Wells M.T. Rivest L.-P. A martingale approach to the copula-graphic estimator for the survival function under dependent censoring. *J. Multivar. Anal.*, 79:138–155, 2001.
- [5] Guha S. Li Y., Tiwari R.C. Mixture cure survival models with dependent censoring. *J. Royal Statist. Soc. B.*, 69, Part 3:285–306, 2007.
- [6] Muradov R.S. Abdushukurov A.A., Dushatov N.T. Estimating of functionals of a multidimensional distribution by censored observations with using copula functions. *In: Statistical Methods of Estimation and Hypotheses Testing. Perm State University*, Issue 23:36–47, 2011.
- [7] Abdushukurov A.A. Muradov R.S. *Estimation of multivariate distributions and its mixtures by incomplete data*. LAMBERT Academic Publishing, 2011 (in Russian).

# On asymptotic minimax efficiency of maximum likelihood estimates by incomplete data

Abdushukurov Abdurahim Akhmedovich

National University of Uzbekistan  
Tashkent  
a\_abdushukurov@rambler.ru

Nurmuhamedova Nargiza Saydillayevna

National University of Uzbekistan  
Tashkent  
rasulova\_nargiza@mail.ru

**Abstract.** We demonstrate asymptotic minimax efficiency of maximum likelihood estimate of unknown parameter in competing risk model under random censoring from both sides.

**Keywords.** Maximum likelihood estimate, competing risk model.

Fisher [1] introduced the concept of effective asymptotically normal estimates with asymptotically minimum variance. This approach involved the asymptotically best estimate, satisfying following two conditions:

1. if  $\theta_n$  is maximum likelihood estimate (MLE) for  $\theta$ , then in usual regularity conditions

$$\mathcal{L}(\sqrt{n}(\theta_n - \theta)) \xrightarrow{n \rightarrow \infty} N(0, I^{-1}(\theta));$$

2. if  $T_n$  is sequence of asymptotically normal estimates, then

$$\lim_{n \rightarrow \infty} M_\theta(\sqrt{n}(T_n - \theta))^2 \geq I^{-1}(\theta), \quad \theta \in \Theta \subset R^1.$$

However, as noted in [1, §1.9], there are super-efficient estimates with regards to the quadratic loss function and therefore this definitions for asymptotically best estimation is inconsistent. Hajek [2] (see also [1]) established the asymptotic minimax bound for risks of any statistical estimates, if we know the asymptotic properties of Bayesian estimators. He proved this result by using the property of local asymptotic normality (LAN) for likelihood ratio statistics (LRS). In [3] in competing risk model (CRM) under random censorship from both sides under regularity condition the proved of LAN property for the LRS. Here we prove the property of asymptotic minimax efficiency of MLE. Particular, we'll formulate the Hajek's theorem for minimax lower bound on properties of various estimates for a wide class of loss functions under certain regularity conditions which gives a necessary and sufficient condition for the achieved of this bound. We use this theorem to determine the asymptotic minimax efficiency of MLE.

We consider  $(L \vee (X \wedge Y); D)$  - model from [4, §1.3], which described as follows. We consider the  $(X; A)$  - model in which we observe the random aggregate

$(X, A^{(1)}, \dots, A^{(k)})$ , where  $X$  is random variable (r.v.) of interest with continuous distribution function (d.f.)  $H(x; \theta)$ , depending on unknown parameter  $\theta$ ,  $\theta \in \Theta \subseteq R^1$  and  $\{A^{(i)}, i = \overline{1, k}\}$  are pairwise disjoint events, with

$$A^{(1)} + \dots + A^{(k)} = \Omega.$$

We interested in joint properties of pairs  $(X; A^{(i)})$ ,  $i = \overline{1, k}$ . Consequently  $(X; A)$  - model is usual CRM in survival analysis. Then  $(L \vee (X \wedge Y); D)$ -model also a CRM model in which aggregate  $(X, A^{(1)}, \dots, A^{(k)})$  is censored from the left and right by r.v.-s  $L$  and  $Y$  with continuous d.f.-s  $L(y)$  and  $K(y)$ ,  $y \in R^1$  respectively. We observe  $(Z; D^{(-1)}, D^{(0)}, D^{(1)}, \dots, D^{(k)})$ , where  $Z = L \vee (X \wedge Y)$ , events  $D^{(-1)} = \{\omega : X(\omega) \wedge Y(\omega) < L(\omega)\}$ ,  $D^{(0)} = \{\omega : L(\omega) \leq Y(\omega) < X(\omega)\}$ ,  $a \wedge b = \min(a, b)$ ,  $a \vee b = \max(a, b)$  and

$$D^{(i)} = A^{(i)} \cap \{\omega : L(\omega) \leq X(\omega) \leq Y(\omega)\}, \quad i = \overline{1, k}.$$

Note that the events  $\{D^{(-1)}, D^{(0)}, D^{(1)}, \dots, D^{(k)}\}$  also have properties of events  $A^{(1)}, \dots, A^{(k)}$ . In this  $(L \vee (X \wedge Y); D)$  - model the r.v.-s  $Y$ ,  $L$  and d.f.-s  $K$ ,  $L$  are considered as nuisance. Let  $\{X_j, L_j, Y_j; D^{(-1)}, D^{(0)}, D^{(1)}, \dots, D^{(k)}\}_{j=1}^\infty$  be a sequence of independent replicas of the aggregate  $(X, L, Y; D^{(-1)}, D^{(0)}, D^{(1)}, \dots, D^{(k)})$ . On  $n$  th stage of the experiment, we observe the sample of size  $n$ :

$$\tilde{Z}^{(n)} = (\tilde{Z}_1, \tilde{Z}_2, \dots, \tilde{Z}_n), \quad 3$$

where  $\tilde{Z}_j = (Z_j; \Delta_j^{(-1)}, \Delta_j^{(0)}, \Delta_j^{(1)}, \dots, \Delta_j^{(k)})$ ,  $Z_j = L_j \vee (X_j \wedge Y_j)$  and  $\Delta_j^{(i)} = I(D_j^{(i)})$ ,  $i = -1, 0, 1, \dots, k$ . Note that in (3), pairs  $(X_j; A_j^{(i)})$  are observable only in the case of  $\Delta_j^{(i)} = 1$ ,  $i = \overline{1, k}$ . The r.v.-s  $Z_j$  have a d.f.

$$N(x; \theta) = L(x)(1 - (1 - K(x))(1 - H(x; \theta))).$$

We introduce subdistributions

$$T^{(i)}(x; \theta) = P_\theta(Z_j < x; D_j^{(i)}), \quad i = -1, 0, 1, \dots, k,$$

where

$$T^{(-1)}(x; \theta) + T^{(0)}(x; \theta) + T^{(1)}(x; \theta) + \dots + T^{(k)}(x; \theta) = N(x; \theta).$$

Let  $\{\mathcal{Y}^{(n)}, \mathcal{U}^{(n)}, \tilde{Q}_\theta^{(n)}\}$  denote the sequence of statistical experiments generated by observations (1). Denoting  $\tilde{Z}$  the set of values of r.v.  $Z$ , we have

$$\mathcal{Y}^{(n)} = \{\tilde{Z} \times \{0, 1\}^{(k+2)}\}^{(n)} =$$

$$= \overbrace{\{\tilde{Z} \times \{0, 1\}^{(k+2)}\} \times \cdots \times \{\tilde{Z} \times \{0, 1\}^{(k+2)}\}}^n,$$

$$\{0, 1\}^{(k+2)} = \overbrace{\{0, 1\} \times \cdots \times \{0, 1\}}^{k+2},$$

$\mathcal{U}^{(n)}$ -  $\sigma$  algebra of Borel sets in  $\mathcal{Y}^{(n)}$ ,  $\tilde{Q}_\theta^{(n)}$ - distribution on  $(\mathcal{Y}^{(n)}, \mathcal{U}^{(n)})$ , being " $n$ -fold product of one-dimensional" distributions  $\tilde{Q}_\theta(x, y^{(-1)}, y^{(0)}, y^{(1)}, \dots, y^{(k)}) = P_\theta(Z_j < x, \Delta_j^{(-1)} = y^{(-1)}, \Delta_j^{(0)} = y^{(0)}, \Delta_j^{(1)} = y^{(1)}, \dots, \Delta_j^{(k)} = y^{(k)})$ . Then the distribution  $\tilde{Q}_\theta^{(n)}$  is absolutely continuous with respect to measure  $\nu^{(n)}$  and its density is given by

$$\frac{d\tilde{Q}_\theta^{(n)}(\tilde{z}^{(n)})}{d\nu^{(n)}(\tilde{z}^{(n)})} = p_n(\tilde{Z}^{(n)}; \theta) =$$

$$= \prod_{m=1}^n \left\{ \prod_{i=1}^k [h^{(i)}(Z_m; \theta)]^{y_m^{(i)}} \right\}.$$

$$\cdot [\gamma(Z_m; \theta)]^{y_m^{(-1)}} [(1 - H(Z_m; \theta))]^{y_m^{(0)}}, \tilde{Z}^{(n)} \in Y^{(n)}, \quad (4)$$

where  $k(x) = K'(x)$ ,  $l(x) = L'(x)$ ,  $\gamma(x; \theta) = 1 - (1 - K(x))(1 - H(x; \theta))$ ,  $d\nu^{(n)}(\tilde{z}^{(n)}) = d\nu(\tilde{z}_1) \times \cdots \times d\nu(\tilde{z}_n)$ ,  $d\nu(\tilde{z}_m) = \varepsilon_{y_m^{(i)}} \times dx_m$ ,  $i = \overline{-1, 0, k}$ ,  $m = \overline{1, n}$  and  $\varepsilon_{y_m^{(i)}}$ -counting measures concentrated at a point  $y_m^{(i)} \in \{0, 1\}$ .

We introduce a regularity condition:

(C1) Supports  $N_{f^{(i)}} = \{x : f^{(i)}(x; \theta) > 0\}$ ,  $i = \overline{1, k}$  are independent on parameter  $\theta$  and  $\bigcap_{i=1}^k N_{f^{(i)}} \neq \emptyset$ ;

(C2) For any  $\theta_1, \theta_2 \in \Theta$ ,  $\theta_1 \neq \theta_2$  and  $x \in N_{f^{(i)}}$ ,  $f^{(i)}(x; \theta_1) \neq f^{(i)}(x; \theta_2)$ ,  $i = \overline{1, k}$ ;

(C3) For all  $x$  there are finite derivatives

$$\partial^l f^{(i)}(x; \theta) / \partial \theta^l, \quad l = 1, 2; \quad i = \overline{1, k},$$

and

$$\int_{-\infty}^{\infty} |\partial^l f^{(i)}(x; \theta) / \partial \theta^l| dx < \infty, \quad l = 1, 2; \quad i = \overline{1, k};$$

(C4) Functions  $\left\{ \frac{\partial \log f^{(i)}(x; \theta)}{\partial \theta}, \frac{\partial \log h^{(i)}(x; \theta)}{\partial \theta}, i = \overline{1, k} \right\}$  are of bounded variations;

(C5) Fisher information

$$\mathbb{J}(\theta) = \sum_{i=1}^k \int_{-\infty}^{\infty} \left( \frac{\partial \log h^{(i)}(x; \theta)}{\partial \theta} \right)^2 dT^{(i)}(x; \theta) +$$

$$+ \int_{-\infty}^{\infty} \left( \frac{\partial \log \gamma(x; \theta)}{\partial \theta} \right)^2 dT^{(-1)}(x; \theta) +$$

$$+ \int_{-\infty}^{\infty} \left( \frac{\partial \log(1 - H(x; \theta))}{\partial \theta} \right)^2 dT^{(0)}(x; \theta),$$

is finite and positive at point  $\theta = \theta_0$ . Let  $\hat{\theta}_n$  is MLE, i.e. solution of equality

$$\frac{\partial}{\partial \theta} \log p_n(\tilde{Z}^{(n)}; \theta) = 0. \quad (5)$$

Hence, in the regularity conditions (C1)-(C5),

$$\frac{1}{n} \frac{\partial \log p_n(\tilde{Z}^{(n)}; \theta)}{\partial \theta} = \sum_{i=1}^k \int_{-\infty}^{\infty} \frac{\partial \log h^{(i)}(x; \theta)}{\partial \theta} dT_n^{(i)}(x) +$$

$$+ \int_{-\infty}^{\infty} \frac{\partial \log \gamma(x; \theta)}{\partial \theta} dT_n^{(-1)}(x) +$$

$$+ \int_{-\infty}^{\infty} \frac{\partial \log(1 - H(x; \theta))}{\partial \theta} dT_n^{(0)}(x). \quad (6)$$

We have

**Theorem.** Let  $\hat{\theta}_n$  is the unique solution of equation (6) and we have the regularity conditions (C1)-(C5). Then MLE  $\hat{\theta}_n$  is an asymptotically minimax efficient for the parameter  $\theta$ .

## References

- [1] Khas'minskii R.Z. bragimov I.A. *Asymptotic Theory of Estimation*. Nauka, M., 1979 (in Russian).
- [2] Hajek J. Local asymptotic minimax and admissibility in estimation. *Proc. Sixth. Berkeley Symp. on Math. Statist. and Prob.*, 1:175–194, 1972.
- [3] Nurmuhamedova N.S. Abdushukurov A.A. Approximation of the likelihood ratio statistics in competing risks model under random censorship from both sides. *ACTA NUUZ*, 4:162–172, 2011 (in Russian).
- [4] Abdushukurov A.A. *Estimators of Unknown Distributions from Incomplete Observations and its Properties*. LAMBERT Academic Publishing, 2011 (in Russian).



# Hedge Fund Strategies, Financial Instability, and Stock Prices Stabilization

Val Demotchkin

Portal Trading Limited  
London, United Kingdom  
valdemotch@gmail.com

**Аннотация.** *The paper discusses hedge funds in connection to the latest financial crisis.*

**Ключевые слова.** *Hedge Funds, Performance fees, Hedge Fund Redemptions, Capital Raising, traders, speculators, risk, volatility, financial markets, Currency speculations, Bank Run, Global Financial Crisis, eventology, Financial Instability, Nash Equilibrium, Government intervention.*

## 1 Introduction

In [1], we discussed Global Macro, Directional and Delta Neutral hedge fund strategies. Market neutral funds have a low correlation with the overall market return. Directional funds specifically take bets on market movements, and so their returns are often strongly correlated with the market. Still there are similarities between all hedge fund strategies as more flexibility comparing with traditional funds.

In [2], we discussed the non-trivial fact that the success of any investment strategy is highly dependent on the skills and implementation ability of individual managers. A good event-driven investment manager is a useful addition to a diversified portfolio of investments, capable of producing absolute non-market correlated returns throughout the economic cycle. The practical application for eventological [3] approach in Decision Making might be the optimization of performance of the event-driven hedge funds.

In [4], we discussed a Performance Fees structure of the hedge funds in general, and also provide a brief overview of the Energy Funds (Energy, Oil and Gas) and the Real Estate Investment Funds.

In [5] we discussed a connection between the hedge funds and the latest financial crisis.

In this article we will discuss the role of hedge funds in the stabilization of stock prices.

## 2 The Financial Cycles

The Hyman Minsky's Financial Instability Theory presumes pro-cyclical human nature. In prosperous

times, when corporate cash flow rises beyond what is needed to pay off debt, a speculative euphoria develops, and soon thereafter debts exceed what borrowers can pay off from their incoming revenues, which in turn produces a financial crisis. As a result of such speculative borrowing bubbles, banks and lenders tighten credit availability, even to companies that can afford loans, and the economy subsequently contracts. Government intervention required.

A key mechanism that pushes an economy towards a crisis is the accumulation of debt by the non-government sector. He identified three types of borrowers that contribute to the accumulation of insolvent debt: hedge borrowers, speculative borrowers, and Ponzi borrowers.

Both Minsky and Soros agree that people are momentum investors by nature, not valued investors.

## 3 Hedge Funds and the Technology Bubble

Technology stocks on Nasdaq rose to unprecedented levels during the two years leading up to March 2000. Ofek and Richardson (2002) estimate that at the peak, the entire internet sector, comprising several hundred stocks, was priced as if the average future earnings growth rate across all these firms would exceed the growth rates experienced by some of the fastest growing individual firms in the past, and, at the same time, the required rate of return would be 0% for the next few decades. By almost any standard, these valuation levels are so extreme that this period appears to be another episode in the history of asset price bubbles.

Brunnermeier and Nagel argue that hedge funds did not exert a correcting force on stock prices during the technology bubble. Instead, they were heavily invested in technology stocks. This does not seem to be the result of unawareness of the bubble: Hedge funds captured the upturn, but, by reducing their positions in stocks that were about to decline, avoided much of the downturn. Their findings question the efficient markets notion that rational speculators always stabilize prices. They are consistent with models in which rational investors may prefer to ride bubbles because of predictable investor sentiment and limits to arbitrage.

Proponents of the efficient markets hypothesis

(Friedman 1953 and Fama 1965) argue that rational speculative activity would eliminate not only riskless arbitrage opportunities, but also other forms of mispricing whose exploitation may require imperfectly hedged and therefore risky trades. The latter case clearly applies to the technology bubble, as there does not exist a close substitute that could be used to hedge a short position in the technology sector. In contrast, the literature on limits to arbitrage points out that various factors such as noise trader risk, agency problems, and synchronization risk may constrain arbitrageurs and allow mispricing to persist. Moreover, some models indicate that rational investors might find it optimal to ride bubbles for a while before attacking them, making the actions of rational investors destabilizing rather than stabilizing.

To shed some light on these issues, Brunnermeier and Nagel examined empirically the response of hedge funds to the growth of the technology bubble. Hedge funds are among the most sophisticated investors — probably closer to the ideal of Yrational arbitrageursY than any other class of investors. Their aim was to find out whether sophisticated speculators were indeed a correcting force during the bubble period. Their study is unusual in that we look directly at hedge fund holdings. In general, data on the hedge funds are difficult to obtain, because hedge funds are not regulated by the SEC. However, like other institutional investors, hedge funds with large holdings in U.S. equities do have to report their quarterly equity long positions to the SEC on Form 13F. They extract hedge fund holdings from these data, including those of well-known managers such as Soros, Tiger, Tudor, and others.

The fact that highly sophisticated investors were riding the bubble and profited from doing so is consistent with some recent theoretical results on limits of arbitrage. Our findings relate to two main insights in this literature. First, rational investors may be reluctant to trade against mispricing. Risk aversion limits their aggressiveness if close substitutes are unavailable and arbitrageurs therefore have to bear fundamental risk, as is the case, e.g., in Wurgler and Zhuravskaya (2002). Abreu and Brunnermeier (AB) (2002, 2003) argue that synchronization risk can prevent arbitrageurs from attacking mispricing, because each trader is uncertain about when other traders will sell out. Since a single investor cannot bring the market down by himself, coordination among rational traders is required and a synchronization problem arises. This allows the mispricing to persist for some time.

This evidence casts doubt on the presumption underlying the efficient markets hypothesis that it is always optimal for rational speculators to attack a bubble. While the exact implications of our results for the mechanism limiting the forces of arbitrage may be open to different interpretations, two points seem

clear: First, there is no evidence that hedge funds as a whole exerted a correcting force on prices during the technology bubble. Among the few large hedge funds that did, the manager with the least exposure to technology stocks — Tiger Management — did not survive until the bubble burst. Second, it appears that aversion to arbitrage risk and frictions such as short-sales constraints alone are not sufficient to understand the failure of rational speculative activity to contain the bubble. While they may explain the unwillingness of professional investors to short overpriced technology stocks, they do not explain our finding that hedge funds held long positions in these stocks.

## 4 Current Global Financial Crisis

The late-2000s financial crisis, also known as the Global Financial Crisis (GFC), is considered by many economists to be the worst financial crisis since the Great Depression of the 1930s. It resulted in the collapse of large financial institutions, the bailout of banks by national governments and downturns in stock markets around the world. In many areas, the housing market also suffered, resulting in numerous evictions, foreclosures and prolonged unemployment. It contributed to the failure of key businesses, declines in consumer wealth estimated in the trillions of U.S. dollars, and a significant decline in economic activity, leading to a severe global economic recession in 2008.

The financial crisis was triggered by a complex interplay of valuation and liquidity problems in the United States banking system in 2008 [4, ?]. The bursting of the U.S. housing bubble, which peaked in 2007, caused the values of securities tied to U.S. real estate pricing to plummet, damaging financial institutions globally. Questions regarding bank solvency, declines in credit availability and damaged investor confidence had an impact on global stock markets, where securities suffered large losses during 2008 and early 2009. Economies worldwide slowed during this period, as credit tightened and international trade declined. Governments and central banks responded with unprecedented fiscal stimulus, monetary policy expansion and institutional bailouts. Although there have been aftershocks, the financial crisis itself ended sometime between late-2008 and mid-2009.

Many causes for the financial crisis have been suggested, with varying weight assigned by experts. The United States Senate issued the Levin—Coburn Report, which found “that the crisis was not a natural disaster, but the result of high risk, complex financial products; undisclosed conflicts of interest; and the failure of regulators, the credit rating agencies, and the market itself to rein in the excesses of Wall Street.” Critics argued that credit rating agencies and investors failed to accurately price the risk involved with mortgage-

related financial products, and that governments did not adjust their regulatory practices to address 21st-century financial markets. The 1999 repeal of the Glass—Steagall Act of 1933 effectively removed the separation that previously existed between Wall Street investment banks and depository banks. In response to the financial crisis, both market-based and regulatory solutions have been implemented or are under consideration.

#### 4.1 Causes of the late-2000s financial crisis

U.S. housing prices had declined by over 20% from their mid-2006 peak. Easy credit, and a belief that house prices would continue to appreciate, had encouraged many subprime borrowers to obtain adjustable-rate mortgages. These mortgages enticed borrowers with a below market interest rate for some predetermined period, followed by market interest rates for the remainder of the mortgage's term. Borrowers who could not make the higher payments once the initial grace period ended would try to refinance their mortgages. Refinancing became more difficult, once house prices began to decline in many parts of the USA. Borrowers who found themselves unable to escape higher monthly payments by refinancing began to default. During 2007, lenders had begun foreclosure proceedings on nearly 1.3 million properties, a 79% increase over 2006. This increased to 2.3 million in 2008, an 81% increase vs. 2007. As of August 2008, 9.2% of all mortgages outstanding were either delinquent or in foreclosure.

The Economist described the issue this way: “No part of the financial crisis has received so much attention, with so little to show for it, as the tidal wave of home foreclosures sweeping over America. Government programmes have been ineffectual, and private efforts not much better.” Up to 9 million homes may enter foreclosure over the 2009-2011 period, versus one million in a typical year. At roughly U.S. \$50,000 per foreclosure according to a 2006 study by the Chicago Federal Reserve Bank, 9 million foreclosures represents \$450 billion in losses.

#### 4.2 Mortgage underwriting

In addition to considering higher-risk borrowers, lenders offered increasingly risky loan options and borrowing incentives. Mortgage underwriting standards declined gradually during the boom period, particularly from 2004 to 2007 [2]. The use of automated loan approvals allowed loans to be made without appropriate review and documentation. In 2007, 40% of all subprime loans resulted from automated underwriting. The chairman of the Mortgage Bankers Association claimed that mortgage brokers, while profiting from the home loan boom, did not do enough to examine whether borrowers could repay. Mortgage fraud by lenders and borrowers increased enormously.

A study by analysts at the Federal Reserve Bank of Cleveland found that the average difference between subprime and prime mortgage interest rates (the “subprime markup”) declined significantly between 2001 and 2007. The quality of loans originated also worsened gradually during that period. The combination of declining risk premium and credit standards is common to boom and bust credit cycles.

The authors also concluded that the decline in underwriting standards did not directly trigger the crisis, because the gradual changes in standards did not statistically account for the large difference in default rates for subprime mortgages issued between 2001-2005 (which had a 10% default rate within one year of origination) and 2006-2007 (which had a 20% rate). In other words, standards gradually declined but defaults suddenly jumped.

Further, the authors argued that the trend in worsening loan quality was harder to detect with rising housing prices, as more refinancing options were available, keeping the default rate lower.

#### 4.3 Mortgage fraud

In 2004, the Federal Bureau of Investigation warned of an “epidemic” in mortgage fraud, an important credit risk of nonprime mortgage lending, which, they said, could lead to “a problem that could have as much impact as the S&L crisis”.

#### 4.4 Down payments and negative equity

A down payment refers to the cash paid to the lender for the home and represents the initial homeowners' equity or financial interest in the home. A low down payment means that a home represents a highly leveraged investment for the homeowner, with little equity relative to debt. In such circumstances, only small declines in the value of the home result in negative equity, a situation in which the value of the home is less than the mortgage amount owed. In 2005, the median down payment for first-time home buyers was 2%, with 43% of those buyers making no down payment whatsoever. By comparison, China has down payment requirements that exceed 20%, with higher amounts for non-primary residences.

Economist Nouriel Roubini wrote in *Forbes* in July 2009: “Home prices have already fallen from their peak by about 30%. Based on my analysis, they are going to fall by at least 40% from their peak, and more likely 45%, before they bottom out. They are still falling at an annualized rate of over 18%. That fall of at least 40%-45% percent of home prices from their peak is going to imply that about half of all households that have a mortgage — about 25 million of the 51 million that have mortgages — are going to be underwater with negative equity and will have a significant incentive to walk away from their homes.” Economist Stan Leibowitz argued

in the Wall Street Journal that the extent of equity in the home was the key factor in foreclosure, rather than the type of loan, credit worthiness of the borrower, or ability to pay. Although only 12% of homes had negative equity (meaning the property was worth less than the mortgage obligation), they comprised 47% of foreclosures during the second half of 2008. Homeowners with negative equity have less financial incentive to stay in the home.

The L.A. Times reported the results of a study that found homeowners with high credit scores at the time of entering the mortgage are 50% more likely to “strategically default” – abruptly and intentionally pull the plug and abandon the mortgage – compared with lower-scoring borrowers. Such strategic defaults were heavily concentrated in markets with the highest price declines. An estimated 588,000 strategic defaults occurred nationwide during 2008, more than double the total in 2007. They represented 18% of all serious delinquencies that extended for more than 60 days in the fourth quarter of 2008.

#### 4.5 Predatory lending

Predatory lending refers to the practice of unscrupulous lenders, to enter into “unsafe” or “unsound” secured loans for inappropriate purposes. A classic bait-and-switch method was used by Countrywide, advertising low interest rates for home refinancing. Such loans were written into mind-numbingly detailed contracts, and swapped for more expensive loan products on the day of closing. Whereas the advertisement might state that 1% or 1.5% interest would be charged, the consumer would be put into an adjustable rate mortgage (ARM) in which the interest charged would be greater than the amount of interest paid. This created negative amortization, which the credit consumer might not notice until long after the loan transaction had been consummated.

Countrywide, sued by California Attorney General Jerry Brown for “Unfair Business Practices” and “False Advertising” was making high cost mortgages “to homeowners with weak credit, adjustable rate mortgages (ARMs) that allowed homeowners to make interest-only payments”. When housing prices decreased, homeowners in ARMs then had little incentive to pay their monthly payments, since their home equity had disappeared. This caused Countrywide’s financial condition to deteriorate, ultimately resulting in a decision by the Office of Thrift Supervision to seize the lender.

Countrywide, according to Republican Lawmakers, had involved itself in making low-cost loans to politicians, for purposes of gaining political favors.

Former employees from Ameriquest, which was United States’s leading wholesale lender, described a system in which they were pushed to falsify mortgage documents

and then sell the mortgages to Wall Street banks eager to make fast profits. There is growing evidence that such mortgage frauds may be a cause of the crisis.

#### 4.6 Risk-taking behavior

In a June 2009 speech, U.S. President Barack Obama argued that a “culture of irresponsibility” was an important cause of the crisis. He criticized executive compensation that “rewarded recklessness rather than responsibility” and Americans who bought homes “without accepting the responsibilities.” He continued that there “was far too much debt and not nearly enough capital in the system. And a growing economy bred complacency.” A key theme of the crisis is that many large financial institutions did not have a sufficient financial cushion to absorb the losses they sustained or to support the commitments made to others. Using technical terms, these firms were highly leveraged (i.e., they maintained a high ratio of debt to equity) or had insufficient capital to post as collateral for their borrowing. A key to a stable financial system is that firms have the financial capacity to support their commitments. Michael Lewis and David Einhorn argued: “The most critical role for regulation is to make sure that the sellers of risk have the capital to support their bets.” Pro-cyclical human nature

Keynesian economist Hyman Minsky described how speculative borrowing contributed to rising debt and an eventual collapse of asset values. Economist Paul McCulley described how Minsky’s hypothesis translates to the current crisis, using Minsky’s words:

“...from time to time, capitalist economies exhibit inflations and debt deflations which seem to have the potential to spin out of control. In such processes, the economic system’s reactions to a movement of the economy amplify the movement—inflation feeds upon inflation and debt-deflation feeds upon debt deflation.” In other words, people are momentum investors by nature, not value investors. People naturally take actions that expand the apex and nadir of cycles. One implication for policymakers and regulators is the implementation of counter-cyclical policies, such as contingent capital requirements for banks that increase during boom periods and are reduced during busts.

Minsky proposed theories linking financial market fragility, in the normal life cycle of an economy, with speculative investment bubbles endogenous to financial markets. Minsky claimed that in prosperous times, when corporate cash flow rises beyond what is needed to pay off debt, a speculative euphoria develops, and soon thereafter debts exceed what borrowers can pay off from their incoming revenues, which in turn produces a financial crisis. As a result of such speculative borrowing bubbles, banks and lenders tighten credit availability, even to companies that can afford loans, and the



economy subsequently contracts.

This slow movement of the financial system from stability to fragility, followed by crisis, is something for which Minsky is best known, and the phrase “Minsky moment” refers to this aspect of Minsky’s academic work.

“He offered very good insights in the ’60s and ’70s when linkages between the financial markets and the economy were not as well understood as they are now”, said Henry Kaufman, a Wall Street money manager and economist. “He showed us that financial markets could move frequently to excess. And he underscored the importance of the Federal Reserve as a lender of last resort.” Minsky’s model of the credit system, which he dubbed the “financial instability hypothesis” (FIH), incorporated many ideas already circulated by John Stuart Mill, Alfred Marshall, Knut Wicksell and Irving Fisher. “A fundamental characteristic of our economy,” Minsky wrote in 1974, “is that the financial system swings between robustness and fragility and these swings are an integral part of the process that generates business cycles.” Disagreeing with many mainstream economists of the day, he argued that these swings, and the booms and busts that can accompany them, are inevitable in a so-called free market economy — unless government steps in to control them, through regulation, central bank action and other tools. Such mechanisms did in fact come into existence in response to crises such as the Panic of 1907 and the Great Depression. Minsky opposed the deregulation that characterized the 1980s.

#### 4.7 Understanding Minsky’s financial instability hypothesis

Hyman Minsky’s theories about debt accumulation received revived attention in the media during the subprime mortgage crisis of the late 2000s.

Minsky argued that a key mechanism that pushes an economy towards a crisis is the accumulation of debt by the non-government sector. He identified three types of borrowers that contribute to the accumulation of insolvent debt: hedge borrowers, speculative borrowers, and Ponzi borrowers.

The “hedge borrower” can make debt payments (covering interest and principal) from current cash flows from investments. For the “speculative borrower”, the cash flow from investments can service the debt, i.e., cover the interest due, but the borrower must regularly roll over, or re-borrow, the principal.

The “Ponzi borrower” (named for Charles Ponzi, see also Ponzi scheme) borrows based on the belief that the appreciation of the value of the asset will be sufficient to refinance the debt but could not make sufficient payments on interest or principal with the cash flow from investments; only the appreciating asset value

can keep the Ponzi borrower afloat. Because of the unlikelihood of most investments’ capital gains being enough to pay interest and principal, much of this type of finance is fraudulent.

If the use of Ponzi finance is general enough in the financial system, then the inevitable disillusionment of the Ponzi borrower can cause the system to seize up: when the bubble pops, i.e., when the asset prices stop increasing, the speculative borrower can no longer refinance (roll over) the principal even if able to cover interest payments. As with a line of dominoes, collapse of the speculative borrowers can then bring down even hedge borrowers, who are unable to find loans despite the apparent soundness of the underlying investments.

#### 4.8 Applying the hypothesis to the subprime mortgage crisis

Economist Paul McCulley described how Minsky’s hypothesis translates to the subprime mortgage crisis.

McCulley illustrated the three types of borrowing categories using an analogy from the mortgage market: a hedge borrower would have a traditional mortgage loan and is paying back both the principal and interest; the speculative borrower would have an interest-only loan, meaning they are paying back only the interest and must refinance later to pay back the principal; and the ponzi borrower would have a negative amortization loan, meaning the payments do not cover the interest amount and the principal is actually increasing. Lenders only provided funds to ponzi borrowers due to a belief that housing values would continue to increase.

McCulley writes that the progression through Minsky’s three borrowing stages was evident as the credit and housing bubbles built through approximately August 2007. Demand for housing was both a cause and effect of the rapidly-expanding shadow banking system, which helped fund the shift to more lending of the speculative and ponzi types, through ever-riskier mortgage loans at higher levels of leverage.

This helped drive the housing bubble, as the availability of credit encouraged higher home prices. Since the bubble burst, we are seeing the progression in reverse, as businesses de-leverage, lending standards are raised and the share of borrowers in the three stages shifts back towards the hedge borrower.

McCulley also points out that human nature is inherently pro-cyclical, meaning, in Minsky’s words, that “from time to time, capitalist economies exhibit inflations and debt deflations which seem to have the potential to spin out of control. In such processes, the economic system’s reactions to a movement of the economy amplify the movement — inflation feeds upon inflation and debt-deflation feeds upon debt deflation.” In other words, people are momentum investors by

nature, not value investors. People naturally take actions that expand the high and low points of cycles. One implication for policymakers and regulators is the implementation of counter-cyclical policies, such as contingent capital requirements for banks that increase during boom periods and are reduced during busts.

## 5 Eventology and Event-Driven hedge funds

In [2] we discussed that the success of any investment strategy is highly dependent on the skills and implementation ability of individual managers. A good event-driven investment manager is a useful addition to a diversified portfolio of investments, capable of producing absolute non-market correlated returns throughout the economic cycle.

Event Driven Trading is any strategy that seeks to exploit pricing inefficiencies, occurring when companies are involved in corporate events such as mergers, takeovers, restructures (including share buy-backs, spin-offs and capital returns), de-mergers and lock-up expiries.

Traders, who follow event driven strategies, attempt to predict the outcome of a particular corporate event on the security price as well as the optimal time to commit capital.

Another event driven trading strategy involves corporate de-mergers and share lock-up periods. A recent trend to emerge, particularly in Europe, is the process of listed companies floating (spinning off) their subsidiaries. This is based upon the theory that the sum of the parts is often worth more than the whole.

When a company lists one of its subsidiaries on the market, but retains an interest, it is common for there to be a lock-up arrangement over the remaining stake in an attempt to re-assure potential investors that the market is not about to be flooded with cheap stock (a similar arrangement takes place in IPOs).

Event driven traders can benefit in a number of ways from corporate de-mergers. Firstly, they will buy the parent company in the expectation that the share price will benefit from the de-merger process. If the strategy is correct, once the de-merger takes place, they will dispose their holding at a higher price than the original entry price. Then, a short time later, they will seek to exploit the lock-up period (usually founder shareholders/private equity (insiders) can sell down their stock holdings 180 days after the initial share offering).

They do this by short selling stock, often using CFDs, ahead of a lock-up expiry and then buying back their short position at a lower price once the insiders have off-loaded their shares.

Here we will discuss the connection between the hedge funds expansion and the global financial crisis. We also will discuss the financial instability theory.

## 6 Conclusion

The efficient market hypothesis is based on the presumption that rational investors prevent price bubbles by trading against mispricing. In this paper, we study the behavior of some of the most sophisticated investors during a bubble period. Specifically, we analyze stock holdings of hedge funds during the technology bubble, 1998 to 2000. We establish two main facts. First, hedge funds were riding the technology bubble, not attacking it. On average, hedge fund stock portfolios were heavily tilted towards technology stocks. This suggests that short-sales constraints, emphasized in recent work on the technology bubble (Ofek and Richardson 2003; Cochrane 2002), are not sufficient to explain the failure of rational speculative activity to contain the technology bubble. Short-sale constraints and arbitrage risks alone can rationalize reluctance to take short positions, but do not explain why sophisticated investors would buy into the overpriced technology sector.

Second, on a stock-by-stock basis, hedge funds reduced their holdings before prices collapsed. Within the technology segment — and only there — they outperformed standard characteristics-matched benchmarks. This suggests that hedge fund managers understood that prices of these stocks would eventually deflate. Our findings are consistent with the view that the investor sentiment driving the technology bubble was predictable to some extent, and that hedge funds were exploiting this opportunity. Under these conditions, riding a price bubble for a while can be the optimal strategy for rational investors, as, for example, in Abreu and Brunnermeier (2003).

The non-linear Financial Instability model of Hyman Minsky has been considered as well as the causes and timeline of the current global financial crisis. Also, we have continued to discuss hedge funds strategies.

The success of any investment strategy is highly dependent on the skills and implementation ability of individual managers. A good event-driven investment manager is a useful addition to a diversified portfolio of investments, capable of producing absolute non-market correlated returns throughout the economic cycle.

Hedge funds focus on strategies that traditional funds cannot undertake, or cannot do well. The three defining characteristics of hedge funds are very strong incentive structures for managers, light regulation, and investment opacity.

Hedge funds undertake investment strategies which play to one or more of these strengths. Long-short strategies



rely on hedge funds' freedom to short equities and take derivatives positions, both of which are often forbidden for traditional funds, or restricted by funds' investment mandates.

Hedge fund strategies requiring fast decisions in risky environments, such as event-driven strategies and tactical trading strategies, are helped by the very strong management incentives used by hedge funds. The investment opacity of hedge funds is crucial in many relative-value strategies since it prevents other investors from uncovering the fund's positions and squeezing its transactions in illiquid markets.

In contrast to hedge funds, traditional funds are characterized by conservative incentive structures for managers, tight regulation, and investment clarity. This gives traditional funds an advantage for strategies designed to target an asset-class benchmark, with a relatively cautious overlay of active positions.

Back in 2004, the biggest Wall Street investment banks have created the instrument of their own destruction, the credit default swap on the subprime mortgage bond.

The combination of debt-financing and mark-to-market accounting conspire to give price movements in the asset markets an unstable positive feedback.

The self-reinforcing asset-debt cycles are the essential element of Hyman Minsky's Financial Instability Hypothesis.

**Special thanks** are owed to Professor Oleg Yu. Vorobyev.

## References

- [1] Val Demotchkin. A comparison of hedge fund strategies. In. *Proc. of the VIII Intern. FAM Conf. on Financial and Actuarial Mathematics and Related Fields*, Krasnoyarsk, SFU (Oleg Vorobyev ed.)(2):14–18, 2009.
- [2] Val Demotchkin. A comparison of event driven hedge fund strategies. In. *Proc. of the XIII Intern. EM conference on eventological mathematics and related fields*, Krasnoyarsk: SFU (Oleg Vorobyev ed.):13–18, 2009.
- [3] O. Yu. Vorobyev, J. J. Goldblatt, and Finkel R. Eventological theory of decision-making. *Journal of Siberian Federal University. Mathematics & Physics*, 2(1):3–16, 2009.
- [4] Val Demotchkin. A comparison of hedge fund strategies. part 3. In. *Proc. of the XV Intern. EM conference on eventological mathematics and related fields*, Krasnoyarsk: SFU (Oleg Vorobyev ed.):13–19, 2011.
- [5] Val Demotchkin. A comparison of hedge fund strategies. In. *Proc. of the XI Intern. FAM Conf. on Financial and Actuarial Mathematics and Related Fields*, Krasnoyarsk, SFU (Oleg Vorobyev ed.)(2):14–18, 2012.

# Company management system estimation on the basis of integrated indicators

Sergey Masaev

Institute of Business Management and  
Economics of Siberian Federal University.  
Head of Ltd company "Sistemi Upravleniya"  
www.systemsm.ru  
Krasnoyarsk  
Fabery@rambler.ru

**Abstract.** *The method of the structure and indicators analysis of company business processes based on the calculation of simple correlation between historic series of expenses is offered.*

**Keywords.** *Correlation, adaptation, process, system analysis, management.*

## 1 Introduction

After the global crisis outbreak companies have changed a lot: stronger integration of companies, lead to numerous consolidations, acquisitions and mergers. The Russian companies are not an exception. Holdings (hereinafter referred as to production systems (PS)) will be a general moving force, which diversified scope the activities and will continue their resultative work. A lot financial groups had diversified the spheres of activity and put right cooperation of the productive systems within these groups of companies, therefore they are general role of market. It is by the present moment process is being increased. The majority of PS influenced by this or that holding. Nevertheless, the other process is far to be finished - creating of decision-making system within the holding. It helps to consider Russian holdings to be well-functioning systems and not just a collection of heterodeneous assets [1].

Holding company manager will find the following difficulties, while directing the holding companies:

- intercommunication between the companies is inaccurate measurement;
- intercommunication between accounting system and decision-making system;
- a lot of complex methods of decision-making system [2];
- companies within holding have different development vectors and conflicting purposes.

The decision of problem is completed by the following tasks:

- the model of mathematical formulation the discrete- multivariate system for economic system (ES) of company was made;
- the calculations and analyzed plan/fact of sample correlation graph G was made;
- the distributing of resources (special case is cash) among functions of economic system was made;
- measured integrate indicators of company in six modulated situations were measured [3].

## 2 Making mathematical formulation the discrete- multivariate system for economic system of company

Follow the systems theory representable system  $S$  consider

$$S = T, X, \quad (1)$$

where  $T = \{t/t = 0, 1, 2, \dots\}$  – discrete set of time (window of time);  $X$  – phase space of system,  $x(t) = [x^1(t), x^2(t), \dots, x^n(t)]^T \in X$  –  $n$ -variable phase vector, state vector, and  $x^*(t) = [x^{*1}(t), x^{*2}(t), \dots, x^{*n}(t)]^T \in X$  –  $n$ -variable phase vector, suboptimal estimation by R. Bellman.

The variable phase vector of economic system is  $x^i(t)$  – financial expenditure for the functions (describing all action, business process) of ES. These unification accounts are found by functions in special program [4], [3]. For the research of ES 417 functions were used [5], [4].

We use the variable phase vector  $x(t)$  for  $k$  prior periods. The parameter  $k$  is the depth horizon (in our special case  $k = 6$  months).

$$r_{i,j}(t) = \frac{1}{k-1} \sum_{l=1}^k \dot{x}^i(t-l) \dot{x}^j(t-l), \quad (2)$$

where  $i, j = 1, \dots, n$ .

$r_{i,j}(t)$  is a sample correlation coefficient (Pearson coefficient) between variables  $x^i(t)$  and  $x^j(t)$  of window time  $t$ .

On the basis of sample correlation coefficients (2) we create correlation graph of the system  $G_i(t)$ . It shows the relationship among the functions of the system (Fig. 2). Then we single out different correlation indicators. Indicators are graphs  $G_i^{sum-total}(t)$  – absolute sum of sample correlation coefficients  $i$ 's function,  $G_i^{sum-neg}(t)$  – sum of negative samples correlation coefficients and  $G_i^{sum-plus}(t)$  – sum of sample correlation coefficients is more than 0,  $G_i^{difference}(t)$  – sum of sample correlation coefficients is taken meaning the signs.

$$G_i^{sum-total}(t) = \sum_{j=1}^n |r_{i,j}(t)| : (|r_{i,j}(t)| \geq r_{sign}), \quad (3)$$

$$G_i^{sum-plus}(t) = \sum_{j=1}^n r_{i,j}(t) : (|r_{i,j}(t)| \geq r_{sign}) \cap (r_{i,j} > 0), \quad (4)$$

$$G_i^{sum-neg}(t) = \sum_{j=1}^n r_{i,j}(t) : (|r_{i,j}(t)| \geq r_{sign}) \cap (r_{i,j} < 0), \quad (5)$$

$$G_i^{difference}(t) = G_i^{sum-plus}(t) + G_i^{sum-neg}(t), \quad (6)$$

where  $r_{sign}$  – significance of sample correlation coefficient for data of sample correlation matrix with parameter  $k$  which is depth of time.

All graphs  $G_i(t)$  account for window time  $T$  including the relationship between functions of the system. Fig. 1 shows twenty functions of system. For the investigating the whole system we're calculating integrate indicators.

### 3 Making calculations and analyses of plan/fact of sample correlation graph $G$

Our calculation shows the important events in our economic system:

1. Drawing up/making up an estimation – planning action and the budget of construction of the district.
2. Getting a permission for building – the document for permission of the start of project of construction necessary for obtaining a bank loan;
3. Creating the TQM - creating regulations and standards for the company;
4. Getting finance- getting a loan/credit;
5. Starting the project construction - starting the building of the district;
6. Creating of the identical hereinafter production system-companies.

We have opportunity to the analyze the behavior of multivariate system, so we find and control tendencies of the company operation as caused by management and outside influence.

The larger the value of  $G_i^{sum-total}(t)$ , the larger is crisis our system experiences.

The functions involved in economic activity of the production system are interrelated (Fig. 2).

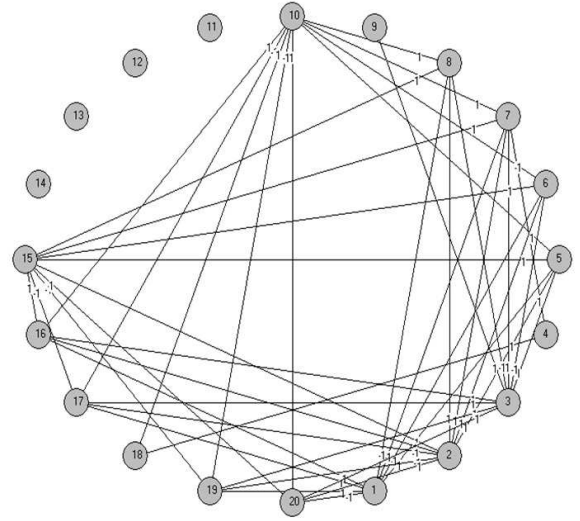


Рис. 1: Correlation graph of the economic system

Within the reported/determined period you can analyze separate functions and determine their influence on the system (Fig. 1). After you have found principal functions You may allocate money in the company by the method of R. Bellman.

The allocation of resources (special case is cash) between functions of economic system If the function has high correlation, it gets more money. The method of R. Bellman is used for financial allocation among corporate objectives, strategies and functional systems [5]. Distribution of money resources on carried according to functions by a method of Bellman is a standard and was described in the last article [4].

Measuring of the integrate indicator of the company in six modulated situations was measured.

For measuring of integral indicator activity of company that is organized from the very beginning is modeled, in six different situations. It model include plans of six different budgets of financial expenditure  $x^i(t)$  (variable phase vector). Each budget has general parameters: credit for company, purchasing assets, personnel, building and installation works on an infrastructure and networks of object, raw material stored. If company does first stages correctly, it will produce goods in 1,5 year. Six variants of development of company have basic

data (Fig. 2). On the integral index on a picture 3 it is evident, that a company is influenced by the crisis less in the first variant development, in comparison with the second and third variants.

The fourth, fifth and sixth variants from 32 to 38 period have brightly expressed the stress situation at the start of the project. The main factor, which results in this stress situation, is the later term of acquisition of the fixed assets and start of production of goods.

Parameters	V1	V2	V3	V4	V5	V6*
Loan/Credit	100%	63%	63%	126%	112%	112%
Rate on a loan/credit	10%	10%	10%	13%	13%	13%
Own facilities	no	37%	37%	no	no	no
Subsidies of the state	33%	100%	100%	100%	100%	100%
Sale of raw material from the moment of realization of project	from 5 month	from 10 month	from 10 month	from 10 month	from 10 month	from 10 month
Sale of products from the moment of realization of project	from 21 month	from 21 month	from 21 month	from 27 month	from 27 month	from 27 month
Displacement of acquisition of assets from the moment of realization of project			on 5 months	on 5 months	on 5 months	on 5 months

Рис. 2: Initial parameters of economic activity of company.

\* The variant 6 differs from the variant 5 in size purveyances of supplies of raw material for a productive process. On a Fig. 3 the correlation graph is represented in period 18. On a Fig. 4 the correlation graph is represented in period 35. On the amount of intercommunications on Fig. 3 and 4 it is possible to say that period 35 is more crisising for a company, than period 18, because an integral indicator  $G_i^{difference}(t)$  in period 35 has the most value.

## 4 Conclusion

For creation of complex and modern control system of company, not only unidimensional approach had been used simple on one index, but also the decision for integration of management in different functions of different enterprises had been found. The method of integral indicators was applied in different companies and industries: service in the real estate, building, purveyance and processing of wood, etc.[4]. Such decision from the one hand is the effective mechanism of management of holding, and from the other hand it has the real money value for every company and for holding in general. The system is the guarantee for investors of managing of projects of this holding company, and also

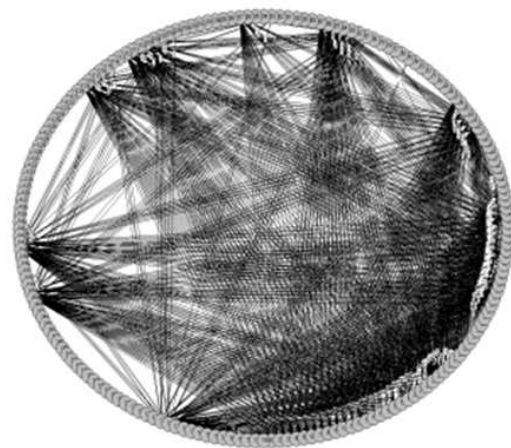


Рис. 3: The correlation graph is represented in period 18

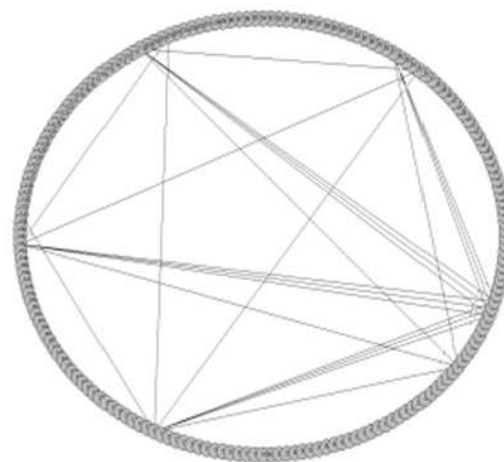


Рис. 4: The correlation graph is represented in period 35.

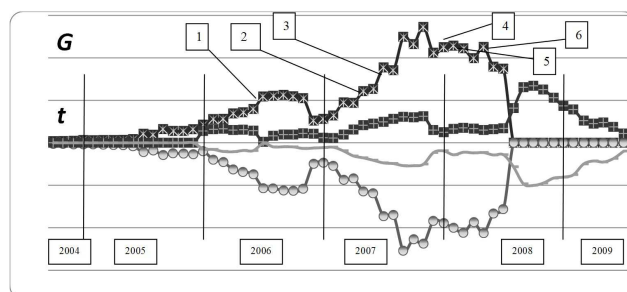


Рис. 5:

for the minimization of administrative risks. You can use the method of integral indicators:

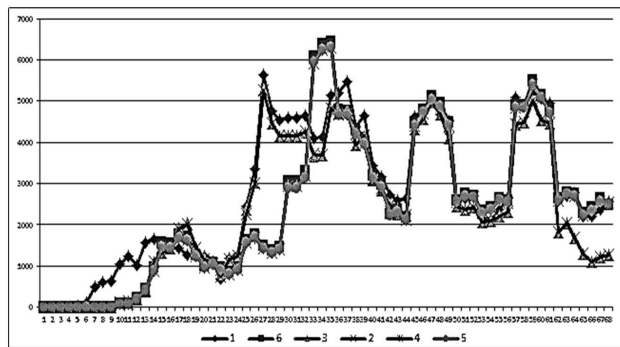


Рис. 6: Dynamics of integral indicator on six modeled variants of development of company.

- to determine and fix the period of stress in economic system of company;
- to optimize the process of management;
- to allocate resources (special case is cash) between functions of economic system
- to measure the integrate indicator of company in six situations [3].

A presence of such system is a valuable intangible asset.

## References

- [1] A.P. Dugelnyj, V.F. Komarov. *Budgeting operation of business*. Ucheb.-prakt. The study. - 2 edi. - m: Business, 2004.
- [2] Nils-Goran O., Roy G., Magnus B. *Estimation of efficiency of activity of company / Practical guidance on the use of the balanced system of indexes*. M.: Williams, 2004.
- [3] S.N. Masaev, M.G. Dorrer, G.A. Dorrer. *Programmatic complex of forming of cross-correlation counts of functions of the productive system(economic, social system) on the basis of plan, actual budget of acuestss and charges, budget of cash of enterprise flow, in the cut of types of activity, reasons and functions and cross-correlation adaptometry*. Testifying of RosPatent to registration of the program for COMPUTER No2009615872.2009.
- [4] S.N. Masaev. *Methodology of complex estimation of administrative decisions in the productive systems with the use of cross-correlation adaptometry: diss. PhD: 05.13.06: 25.03.11 is protected: sign. 25.11.11*. M., 2011.
- [5] S. Masaev, G. Dorrer. Company management system estimation on the basis of adaptive correlation to the environment. *The Issues of Administration and Management*, 3:45–50, 2010.
- [6] K. Druri. *Introduction to the administrative and productive account*. M: Audit of UNITI, 1998.



# In search of a primary source:

## remaking the paper (1975) where at the first time a definition of lattice (Vorob'ev) expectation of a random set was given

Oleg Yu. Vorobyev

Siberian Federal University  
Institute of Mathematics and Computer Science  
Trade-Economic Institute  
Krasnoyarsk  
oleg.yu.vorobyev@gmail.com

**Abstract.** *Remaking the primary source of an old good idea of the “lattice expectation”, published by me in 1975 [1, 2] at the first time and subsequently, especially in the western literature on stochastic geometry and the theory of random sets, named from a light hand of Dietrich Stoyan [3, 1994] “Vorob'ev expectation”<sup>1</sup>; had an only historical and methodic value for eventology and probability theory once more reminding how and “...from what rubbish flowers grow, not knowing shame”.*

**Keywords.** *Probability theory, random set, mean measure set, lattice expectation, Vorob'ev expectation, eventology.*

## 1 Forward to the original source

*You might not know from what rubbish  
flowers grow, not knowing shame.*

Anna Akhmatova, 1940.

Any new finds itself on the old-fashioned way — always on its own. Any invention isn't someone clever fiction, but the discovery of previously hidden, cut off as unnecessary, deliverance from captivity oblivion, and release for public viewing.

New in mathematics is no exception. Strong mathematical definitions given so much that they are more than enough for all the sciences together. But only a rare few of them are definitions of new mathematical concepts. The remaining dissolved completely and forever lost in the maze of mathematical thought.

Why is it, and not otherwise, no one knows. Where, among a rubbish look for something new, that is, being open times will last forever? That can not be avoided? What can not be squeamish?

© 2013 O.Yu.Vorobyev

Oleg Vorobyev (ed.), Proc. XII FAMES'2013, Krasnoyarsk: SFU

<sup>1</sup>I always tried to name this notion so that the name hints at its sense a little, in the beginning (1975) as the “lattice expectation” of a random finite set to emphasize the finiteness of situation (which has allowed for the first time to open it), and then (1984) as the “mean measure set” of a random (measurable, closed, compact) set. But probably “against the nature you will not trample”, and now I'd remember English variant “Vorob'ev expectation” though a special desire to become an eponym behind itself I did not notice.

No answer.

You can only try once more to climb, step by step to recover as it was; to climb again by the same footpath, forward to the source.

## 2 Remake and/и Римейк<sup>2</sup>

### 2.1 Estimating mean contours of forest fire from the probability model of fire spread [2, 1, 1975]

In our study [4, 5, 1973] the probabilistic model of fire spread on a large forest which consists of the separate sites located in units of a *plane square lattice* was considered. This probabilistic model, *Random Spread Process (RSP)*, is defined as the set of discrete random variables

$$U = \{u_{rt}, r \in R_2, t \in T\}, \quad (2.2.1)$$

where  $R_2 \subset \mathbb{R}^2$  is a plane lattice,  $T$  is a space of time parameter,  $u_{rt}$  is a discrete random variable with possible values:

$$u_{rt} = \begin{cases} 1, & \text{if the site } r \text{ burns during the moment } t; \\ 0, & \text{if the site } r \text{ didn't burn till the moment } t; \\ -1, & \text{if the site } r \text{ has burned down by the moment } t. \end{cases} \quad (2.2.2)$$

Besides it is supposedly, that joint probabilistic distribution of random variables  $u_{rt}$  for all  $t \in T$  is completely defined by the field

$$p = \{p(r)\}_{r \in R_2}, \quad (2.2.3)$$

of so called *spread probabilities*, where

$$p(r) = [p_1(r); p_2(r); p_3(r); p_4(r)] \quad (2.2.3')$$

is the *vector of spread probabilities* from the unit  $r$  and  $p_k(r)$  corresponds to the spread probability from the unit  $r$  to the  $k$ -th *adjacent*<sup>3</sup> unit.

<sup>2</sup>The following is the text of paper (1975) (first in the original Russian, then in English translation), which has become the primary source of the idea of *expectation of a random set*. Being the only stylistic changes that facilitate perception, but no loss of *train of thought*; including one of its branches erroneous.

<sup>3</sup>Rules of *neighbourhood* on  $R_2$  is stipulated specially.



In the study a more general definition of RSP was offered which will allow enough room for simple and effective estimation of mean contours of forest fire.

From the definition (2.2.1) and (2.2.2) it follows that a state of RSP during the moment  $t$  can be defined by allocating such two subsets of units  $K_t^1$  and  $K_t^2$  of the square lattice  $R_2$ , that

$$\begin{cases} K_t^1 = \{r : t \in R_2, u_{rt} \neq 0\}, \\ K_t^2 = \{r : t \in R_2, u_{rt} = -1\}. \end{cases} \quad (2.2.4)$$

It is clear, that  $K_t^2 \subseteq K_t^1$  and

$$K_t^1 \setminus K_t^2 = \{r : t \in R_2, u_{rt} = 1\}. \quad (2.2.5)$$

Thus the RSP can be considered as a sequence of pairs  $[K_t^1, K_t^2]$  for all  $t \in T$ . It is necessary to note, that subsets  $K_t^1$  and  $K_t^2$  have, generally speaking, a *random form*<sup>4</sup> with probabilistic characteristics which are defined by the field  $p$  of spread probabilities (2.2.3).

The more general definition of RSP based on this approach, demands an introduction of strictly concept corresponding to our intuitive representation about “*random subset of units*” of a lattice.

It is a question about a *random set* which is possible to consider as a generalization of a discrete random variable and “values” of which are not (unlike a discrete random variable) real numbers, but subsets of units of a lattice.

In this research we'll be limited to only elementary equivalent of a random set definition.

**Definition.** The *random set*  $K$  is no more than countable set of pairs  $(B_i, p_i)_{i \in I}$ , where  $B_i \subseteq R_2$  is some finite subset of units of a lattice, and  $p_i$  is the probability of concurrence of random set  $K$  with the subset  $B_i$ :

$$p_i = \mathbf{P}(K = B_i). \quad (2.2.6)$$

Besides it is required, that

$$\sum_{i \in I} p_i = 1. \quad (2.2.7)$$

A *random vector-set* is the ordered set of random sets.

$$\overline{K} = [K^1, K^2, \dots, K^n]. \quad (2.2.8)$$

The concept of a random set allows to define a *lattice random process* as a sequence of random sets. The Random Spread Process is defined as a special case of the lattice random process in which random sets of the lattice process should be satisfied to (2.2.4). And as we already marked, probabilistic distribution of random variables  $u_{rt}$  is completely defined by a field of spread probabilities  $p$  and the RSP is completely defined by this field.

<sup>4</sup>on the square lattice  $R_2$ .

The basic practical result of the given paper is the offered method of an estimation of a mean contour of forest fire from probabilistic model of its spread, or (that the same) a method of estimation “mean values” of the random sets forming RSP.

Let  $\overline{K} = \{\overline{K}_t, t \in T\}$  be a random spread process where  $\overline{K}_t = [K_t^1, K_t^2]$  is a random vector-set. We'll put the problem of definition of “mean value” of random sets from  $\overline{K}_t$  for any  $t \in T$ .

Abundantly clear, that the problem put thus, is incorrect, as the concept of a “mean value” of a random set the solution of the problem will be essentially depended on a choice of this definition will be not certain yet.

Hence, first of all it is necessary to agree what subset of units of a lattice will be considered as “mean value” of a random set. It would be desirable also that the concept of the “mean value” in a certain sense was similar to the concept of an expectation of a discrete random variable.

First we'll notice, that if  $|K|$  is a number of units in random set  $K$ , i.e.  $|K|$  is a discrete random variable with the given distribution, it is possible to speak about its “mean value”, an expectation  $\mathbf{E}|K|$ .

**Definition.** The *lattice expectation of random set*  $K$  is the subset  $\mathcal{E}K \subseteq R_2$ , satisfying to two conditions:

$$|\mathbf{E}|K| - |\mathcal{E}K|| = \min_{B \subseteq R_2} |\mathbf{E}|K| - |B||, \quad (2.2, A)$$

$$\mathbf{E}|K \Delta \mathcal{E}K| = \min_{B \subseteq R_2} \mathbf{E}|K \Delta B|, \quad (2.2, B)$$

which order to choose from all subsets which satisfy to a condition  $A$  such subset which on the mean least evades from the random set  $K$ , i.e. satisfies to a condition  $B$ .

Certain thus the *lattice expectation of a random set* possesses a number of the properties similar to properties of an *expectation of a random variable*.<sup>5</sup>

Let's stop on the most important characteristic property which allows to find the *lattice expectation of random set*.

Let's enter into consideration probability

$$\pi(r) = \mathbf{P}(r \in K). \quad (2.2.11)$$

It appears, that **in the lattice expectation  $\mathcal{E}K$  enter for those, in round figures,  $\mathbf{E}|K|$  units of a lattice,**

<sup>5</sup> For example, for a random variable  $\xi$  and its expectation  $\mathbf{E}\xi$  it is always fair

$$\mathbf{E}(\xi - \mathbf{E}\xi) = 0. \quad (2.2.9)$$

Similar property was carried out for any random set  $K$  and its lattice expectation  $\mathcal{E}K$ :

$$\mathbf{E}|K \setminus \mathcal{E}K| = \mathbf{E}|\mathcal{E}K \setminus K|. \quad (2.2.10)$$

This analogy and the formula (2.2.10) have appeared erroneous and consequently a remake were omitted in a footnote.

for which probabilities  $\pi(r)$  would accept more values than for other units.

This important characteristic property allows to develop algorithm and the computer program "M-222" which calculate estimations for the *lattice expectation of a random set*.

The same algorithm is applied to an estimation of lattices expectations of random sets forming Random Spread Process.

Modelling RSP, defined by various *vectors of spread probabilities*

$$p(r) = [p_1; p_2; p_3; p_4], \quad (2.2.3'')$$

is lead which are assumed by constants for all  $r \in R_2$ , and for each of them lattice expectations of random sets forming this PCP ( $p_1; p_2; p_3; p_4$ ) are estimated.

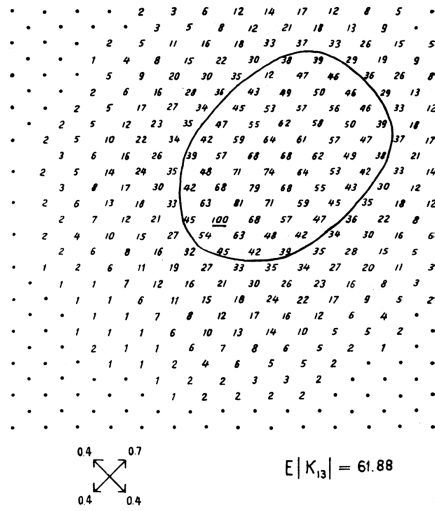


Рис. 1: The part of a square lattice  $R_2$  on which the random set  $K_{13} \subset \mathfrak{X}$ , set-condition of the RSP(0.4; 0.4; 0.4; 0.7) during the moment  $t = 13$ , is statistically modeled. For each unit of the lattice  $r \in \mathfrak{X}$  the quantity of realizations of random set from 100 lead, in which  $r \in K_{13}$  is shown; the contour contains the *lattice expectation*  $\mathcal{E}K_{13}$  of random set  $K_{13}$ . — Часть квадратной решетки  $R_2$ , на которой статистически моделируется случайное множество  $K_{13} \subset \mathfrak{X}$  — сет-состояние ПСР(0.4; 0.4; 0.4; 0.7) в момент  $t = 13$ . Для каждого узла решетки  $r \in \mathfrak{X}$  показано количество реализаций случайного множества из 100 проведенных, в которых  $r \in K_{13}$ ; контур содержит *решетчатое ожидание*  $\mathcal{E}K_{13}$  случайного множества  $K_{13}$ .

In fig. 2 the part of a lattice  $R_2$  on which the RSP is modelling, the aspect shown in print by COMPUTER. On a place of each unit  $r \in R_2$  the quantity  $n(r)$  realizations from 100 in which the unit  $r$  belongs to the random set  $K_t$  at  $t = 13$  settles down, and the estimation  $\mathbf{P}(r \in K_{13})$ , probabilities of that  $r \in K_{13}$ , — looks like:  $\pi(r) \approx n(r)/100$ . The contour on a plane  $\mathbb{R}^2$ , containing the lattice expectation  $\mathcal{E}K_{13}$  of the random set  $K_{13}$  was spent **manually** on the basis of an estimation of the mean  $\mathbf{E}|K_{13}|$  of the random variable  $|K_{13}|$ , calculated using COMPUTER.

In fig. 3 a number of *lattice expectations of random sets*

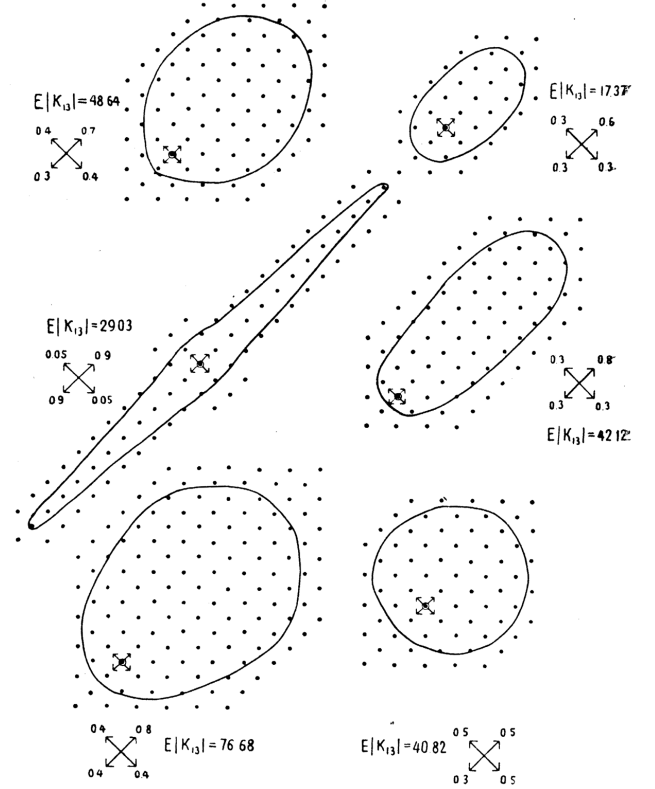


Рис. 2: *Lattice expectations*  $\mathcal{E}K_{13}$  of the random set  $K_{13}$  for various RSPs. — *Решетчатые ожидания*  $\mathcal{E}K_{13}$  случайного множества  $K_{13}$  для различных ПСР.

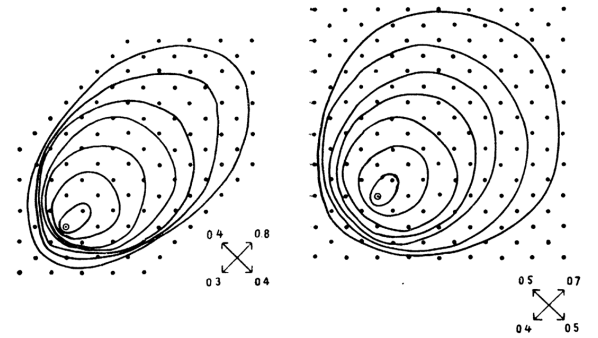


Рис. 3: Two embedded sequences  $\{\mathcal{E}K_t, t = 1, 3, 5, 7, 9, 11, 13\}$  of *lattice expectations of random sets* for the RSP(0.3; 0.4; 0.4; 0.8) and the RSP(0.4; 0.5; 0.5; 0.7). — Две вложенные последовательности  $\{\mathcal{E}K_t, t = 1, 3, 5, 7, 9, 11, 13\}$  *решетчатых ожиданий случайных множеств* для ПСР(0.3; 0.4; 0.4; 0.8) и ПСР(0.4; 0.5; 0.5; 0.7).

for various RSPs were shown.

While in fig. 4 two sequences of lattice expectations of random sets  $\{\mathcal{E}K_t, t = 1, 3, 5, 7, 9, 11, 13\}$  for two RSPs were observed.

In summary we'll consider possible applications of the offered method for an estimation of mean contours of forest fire.

An available program on expert estimations<sup>6</sup> of fields of spread probabilities of fire for concrete wood territories and the program on statistical estimations *lattice expectations of random sets*, forming RSP, allow to model mean contours of a fire and their development in time for various wood conditions. Thus, it enables us to solve the problems of identifications of probabilistic models of forest fires in questions (i.e. definitions and specifications of their parameters). First of all probabilities and speeds of spread of a fire on a large forest mean.

Therefore there is no need to speak of the importance of this problem since there were no satisfactory methods for calculating dependent parameters models for forest fires from all varieties of existing wood conditions.

## 2.2 Оценка средних контуров лесного пожара по вероятностной модели его распространения [1, 2, 1975]

В нашей работе [4, 5, 1973] рассматривалась вероятностная модель распространения пожара по лесному массиву, который состоит из отдельных участков, расположенных в узлах *плоской квадратной решетки*. Эта вероятностная модель — *процесс случайного распространения (ПСР)* — определяется как множество дискретных случайных величин

$$U = \{u_{rt}, r \in R_2, t \in T\}, \quad (2.1.1)$$

где  $R_2 \subset \mathbb{R}^2$  — плоская решетка,  $T$  — пространство временного параметра,  $u_{rt}$  — дискретная случайная величина с возможными значениями:

$$u_{rt} = \begin{cases} 1, & \text{если участок } r \text{ горит в момент } t; \\ 0, & \text{если участок } r \text{ не горел до момента } t; \\ -1, & \text{если участок } r \text{ сгорел к моменту } t. \end{cases} \quad (2.1.2)$$

Кроме того, предполагается, что совместное вероятностное распределение случайных величин  $u_{rt}$  для всех  $t \in T$  полностью определяется полем

$$p = \{p(r)\}_{r \in R_2}, \quad (2.1.3)$$

так называемых *вероятностей распространения*, где

$$p(r) = [p_1(r); p_2(r); p_3(r); p_4(r)] \quad (2.1.3')$$

— *вектор вероятностей распространения* из узла  $r$ , а  $p_k(r)$  соответствует вероятности распространения огня из участка  $r$  на  $k$ -тый *соседний*<sup>7</sup> с ним участок.

В настоящей работе предлагается более общее определение ПСР, которое позволит провести достаточно простую и эффективную оценку средних контуров лесного пожара.

<sup>6</sup>Look the first part of my paper [1, 2], not entered into this remake.

<sup>7</sup>Правила *соседства* на  $R_2$  оговариваются специально.

Из определения (2.1.1) и (2.1.2) следует, что состояние ПСР в момент  $t$  можно задавать, выделяя два таких подмножества узлов  $K_t^1$  и  $K_t^2$  квадратной решетки  $R_2$ , что

$$\begin{cases} K_t^1 = \{r : t \in R_2, u_{rt} \neq 0\}, \\ K_t^2 = \{r : t \in R_2, u_{rt} = -1\}. \end{cases} \quad (2.1.4)$$

Ясно, что  $K_t^2 \subseteq K_t^1$  и

$$K_t^1 \setminus K_t^2 = \{r : t \in R_2, u_{rt} = 1\}. \quad (2.1.5)$$

Таким образом, ПСР можно рассматривать как последовательность пар  $[K_t^1, K_t^2]$  для всех  $t \in T$ . При этом необходимо отметить, что подмножества  $K_t^1$  и  $K_t^2$  имеют, вообще говоря, *случайную форму*<sup>8</sup>, вероятностные характеристики которой определяются полем  $p$  вероятностей распространения (2.1.3).

Более общее определение ПСР, основанное на этом подходе, требует введения строго понятия, соответствующего нашему интуитивному представлению о “случайном подмножестве узлов” решетки.

Речь идет о *случайном множестве*, которое можно считать обобщением дискретной случайной величины и “значениями” которого являются (в отличие от дискретной случайной величины) не действительные числа, а подмножества узлов решетки.

В этой работе мы ограничимся лишь элементарным эквивалентом определения случайного множества.

**Определение.** *Случайным множеством*  $K$  называется не более чем счетный набор пар  $(B_i, p_i)_{i \in I}$ , где  $B_i \subseteq R_2$  — некоторое конечное подмножество узлов решетки, а  $p_i$  — вероятность совпадения случайного множества  $K$  с подмножеством  $B_i$ :

$$p_i = \mathbf{P}(K = B_i). \quad (2.1.6)$$

Кроме того, требуется, чтобы

$$\sum_{i \in I} p_i = 1. \quad (2.1.7)$$

*Случайным вектор-множеством* называется упорядоченный набор случайных множеств

$$\overline{K} = [K^1, K^2, \dots, K^n]. \quad (2.1.8)$$

Понятие случайного множества позволяет определить *решетчатый случайный процесс*, как последовательность случайных множеств. Процесс же случайного распространения определяется как частный случай решетчатого случайного процесса, в котором случайные множества решетчатого процесса должны удовлетворять соотношению (2.1.4). А поскольку, как мы уже отмечали, вероятностное распределение случайных величин  $u_{rt}$  полностью определяется полем

<sup>8</sup>на квадратной решетке  $R_2$ .

вероятностей распространения  $p$ , то и ПСР полностью определяется этим полем.

Основным практическим результатом данной работы является предлагаемый метод оценки среднего контура лесного пожара по вероятностной модели его распространения, или (что то же самое) метод оценки “средних значений” случайных множеств, образующих ПСР.

Пусть  $\bar{K} = \{\bar{K}_t, t \in T\}$  — процесс случайного распространения, где  $\bar{K}_t = [K_t^1, K_t^2]$  — случайное векторное множество. Поставим задачу определения “среднего значения” случайных множеств из  $\bar{K}_t$  для произвольного  $t \in T$ .

Совершенно очевидно, что задача, поставленная таким образом, некорректна, поскольку пока не определено понятие “среднего значения” случайного множества решение задачи будет существенно зависеть от выбора этого определения.

Следовательно, прежде всего нам необходимо условиться о том, какое подмножество узлов решетки мы будем считать “средним значением” случайного множества. Хотелось бы также, чтобы вновь вводимое понятие “среднего значения” в определенном смысле было аналогично понятию математического ожидания дискретной случайной величины.

Сначала заметим, что если  $|K|$  — число узлов в случайном множестве  $K$ , т.е.  $|K|$  — дискретная случайная величина с заданным распределением, то можно говорить о его среднем значении — математическом ожидании  $E|K|$ .

**Определение.** Решетчатым ожиданием случайного множества  $K$  называется подмножество  $\mathcal{E}K \subseteq R_2$ , удовлетворяющее двум условиям:

$$|E|K| - |\mathcal{E}K|| = \min_{B \subseteq R_2} |E|K| - |B||, \quad (2.1, A)$$

$$E|K \Delta \mathcal{E}K| = \min_{B \subseteq R_2} E|K \Delta B|, \quad (2.1, B)$$

которые предписывают выбрать из всех подмножеств, которые удовлетворяют условию A, такое подмножество, которое в среднем наименее уклоняется от случайного множества  $K$ , т.е. удовлетворяет условию B.

Определенное таким образом решетчатое ожидание случайного множества обладает рядом свойств, аналогичных свойствам математического ожидания случайной величины.<sup>9</sup>

<sup>9</sup> Например, для случайной величины  $\xi$  и ее математического ожидания  $E\xi$  всегда справедливо

$$E(\xi - E\xi) = 0. \quad (2.1.9)$$

Аналогичное свойство выполняется и для произвольного случайного множества  $K$  и его решетчатого ожидания  $\mathcal{E}K$

$$E|K \setminus \mathcal{E}K| = E|\mathcal{E}K \setminus K|. \quad (2.1.10)$$

Остановимся на самом важном характеристическом свойстве, которое позволяет найти решетчатое ожидание случайного множества.

Введем в рассмотрение вероятность

$$\pi(r) = P(r \in K). \quad (2.1.11)$$

Оказывается, что в решетчатое ожидание  $\mathcal{E}K$  входят те, округленно,  $E|K|$  узлов решетки, для которых вероятности  $\pi(r)$  принимают значения большие, чем у остальных узлов.

Это важное характеристическое свойство позволяет разработать алгоритм и программу для ЭВМ “М-222”, которые вычисляют оценки решетчатого ожидания случайного множества.

Этот же алгоритм применен для оценки решетчатых ожиданий случайных множеств, образующих процесс случайного распространения.

Проведено моделирование ПСР, определяемых различными векторами вероятностей распространения

$$p(r) = [p_1; p_2; p_3; p_4], \quad (2.1.3'')$$

которые предполагаются постоянными для всех  $r \in R_2$ , и для каждого из них оценены решетчатые ожидания случайных множеств, образующих этот ПСР( $p_1; p_2; p_3; p_4$ ).

На рис. 2 показана часть решетки  $R_2$ , на которой моделируется ПСР, в том виде, в котором его печатает ЭВМ. На месте каждого узла  $r \in R_2$  располагается количество  $n(r)$  реализаций из 100, в которых узел  $r$  принадлежит случайному множеству  $K_t$  при  $t = 13$ , а оценка  $P(r \in K_{13})$  — вероятности того, что  $r \in K_{13}$ , — имеет вид:  $\pi(r) \approx n(r)/100$ . Контур на плоскости  $\mathbb{R}^2$ , содержащий решетчатое ожидание  $\mathcal{E}K_{13}$  случайного множества  $K_{13}$  проводился вручную на основе оценки среднего  $E|K_{13}|$  случайной величины  $|K_{13}|$ , вычисленной ЭВМ.

На рис. 3 показан ряд решетчатых ожиданий случайных множеств для различных ПСР.

На рис. 4 приведены две последовательности решетчатых ожиданий случайных множеств  $\{\mathcal{E}K_t, t = 1, 3, 5, 7, 9, 11, 13\}$  для двух ПСР.

В заключение рассмотрим возможные применения предложенного метода оценки средних контуров лесного пожара.

Имеющаяся программа экспертных оценок<sup>10</sup> поля вероятностей распространения огня для конкретных лесных территорий и программа статистических оценок решетчатых ожиданий случайных множеств, образующих ПСР, позволяют моделировать

Эта аналогия и формула (2.1.10) оказались ошибочными и потому в римейке опущены в сноску.

<sup>10</sup>См. первую часть моей работы [1, 2], не вошедшую в этот римейк.



средние контуры пожара и их развитие во времени для различных лесотаксационных условий. Это дает возможность решать задачи, касающиеся вопросов идентификации вероятностных моделей лесных пожаров (т.е. определения и уточнения их параметров). В первую очередь имеются в виду вероятности и скорости распространения пожара по лесному массиву.

Нет необходимости говорить о важности этой задачи, тем более, что до сих пор не существовало удовлетворительных методов расчета зависимости параметров моделей лесных пожаров от всего многообразия существующих лесотаксационных условий.

### 3 Back to the fuss

*In the fuss of cities and in heavy traffic  
Returning we have nowhere to go,  
And climb down from the conquered peaks  
Leaving the mountains  
Leaving the mountains of the heart.*

Vladimir Vysotsky, 1966.

They said that to be informed, is to be forearmed. Deeply rational thought. Suitable for anything, but not to the discovery of a new one. It's no secret that too much information sweeps clean all the way, besides the already well-worn ruts. Excessive? This is in excess of what? From what the pull down shipping costs to get rid of, to see no one has trodden the top, cling to it and hold it for a moment.

No answer.

\*\*\*

Here you are on top. And where to now?

Back? Down? No lifts? Just walk down?

In the hustle and bustle? Nowhere else to go?

No, well, you can certainly look for more ... Higher.

No, but it is quite something to look for the latest ...

Yeah, well, look, of course, you can.

Yes, but then nothing will be found there?

...

Most have nowhere to go!

Or, as always, in the obscurity of a fuss

Or back to the top of the unknown.

\*\*\*

### 4 Help: mean probable event

In [6, 7, 1984] as well as in [8, 9, 1999, стр. 644] you can find the definition of the *mean measure set*<sup>11</sup>,

<sup>11</sup>Recently, examples of the use of the term multiplying in different areas of mathematics and applications, and in many Western works on the theory of random sets (see, for example, [10, 2012], [11, 2013], [12, 2005] etc.) it is usually called *Vorob'ev expectation* with a light hand Dietrich Stoyan [3, стр. 113-115, 1994] relying on outdated English transliteration of my name, which I'm writing now with slightly different.

which was first introduced in 1973 by me and published in [1, 2, 1975] and in [13, 14, 1977], as well as used by other authors, for example in [15, 16, 1985, 1986], and in [3, 1994]. The *mean measure set* is the mean set characteristics of a random set whose values are measurable subsets of a measure space, playing for a random set of the same role played by the *expectation*, or the *mean value*, of a *random element* with the values of the linear space<sup>12</sup>. In eventology a notion of the mean set of events has long existed [17, 18]. This was the result of literal applying the definition of mean measure set for a *random set of events*. More recently in [19, 20] a general idea of the long-standing definition of mean measure set was again taken by me on the arms so much else to define another new concept for eventology called the *mean probable event*.

**Definition (mean probable event for a finite set of events).** Let  $(\Omega, \mathcal{A}, \mathbf{P})$  be the universal probability space. The mean probable event for a finite set of events  $\mathfrak{X} \subset \mathcal{A}$  is defined as any universal event  $\hat{x}_{\mathfrak{X}} \in \mathcal{A}$  satisfying the inclusions

$$\sum_{|X| > m} \text{ter}(X // \mathfrak{X}) \subset \hat{x}_{\mathfrak{X}} \subseteq \sum_{|X| \geq m} \text{ter}(X // \mathfrak{X}), \quad (4.2)$$

which occurs with probability

$$\mathbf{P}(\hat{x}_{\mathfrak{X}}) = \frac{1}{|\mathfrak{X}|} \sum_{x \in \mathfrak{X}} \mathbf{P}(x) \quad (4.3)$$

just in time of the events of  $\mathfrak{X}$  occurs at least  $m$  events, where  $m \in \{0, 1, \dots, |\mathfrak{X}|\}$  satisfies

$$\sum_{|X| > m} p(X // \mathfrak{X}) < \mathbf{P}(\hat{x}_{\mathfrak{X}}) \leq \sum_{|X| \geq m} p(X // \mathfrak{X}). \quad (4.4)$$

**Remark.** Of course, the *mean probable event* has known extreme properties, which are formulated in terms of probabilistic distance [19, 20].

### Список литературы

- [1] О. Ю. Воробьев. Определение вероятностей распространения горения и оценка развития средних контуров лесного пожара. *Охрана лесных ресурсов Сибири. Красноярск, ИЛД СО АН СССР*, 1:43–67, 1975, URL: <http://eventology-theory.com/0-lec/remake-1975-full-GSFR-43-67.pdf>.
- [2] О. Ю. Vorobyev. Definition of probabilities of fire spread and estimating mean forest fire spread sets. *The Guarding of Siberia Forest Resources. Krasnoyarsk, The Sukachev Institute of Forest and Wood, USSR AS, SB*, 1:43–67, 1975 (in Russian), URL: <http://eventology-theory.com/0-lec/remake-1975-full-GSFR-43-67.pdf>.
- [3] D Stoyan and H. Stoyan. *Fractals, Random Shapes and Point Fields. Methods of Geometrical Statistics. XIV.* John Wiley & Sons, Chichester etc., 1994, 389p.

<sup>12</sup>For example, the expectation of a random variable, of a random vector matrix, of a random functions, etc.

- [4] О. Ю. Воробьев. Математическое описание процессов случайного распространения и управление ими. *Известия СО АН СССР*, 13(3):146–152, 1973.
- [5] O. Yu. Vorobyev. Mathematical description of random spread processes and its control. *Izvestia of SB AS USSR*, 13(3):146–152, 1973 (in Russian).
- [6] О. Ю. Воробьев. *Среднемерное моделирование*. Наука, Москва, 1984, 133с.
- [7] O. Yu. Vorobyev. *Mean Measure Modeling*. Nauka, Moscow, 1984 (in Russian).
- [8] *Вероятность и математическая статистика. Энциклопедия*. Научное издательство Большая Российская Энциклопедия, Москва, 1999.
- [9] *Probability and Mathematical Statistics. Encyclopedia*. Science publisher Great Russian Encyclopedia, Moscow, 1999 (in Russian).
- [10] P. Heinrich, R. C. Stoica, and V. C. Tran. Level sets estimation and Vorob'ev expectation of random compact sets. *Spatial Statistics*, 2:47–61, December 2012.
- [11] C. Chevalier, D. Ginsbourger, J. Bect, and Molchanov I. Estimating and Quantifying Uncertainties on Level Sets Using the Vorob'ev Expectation and Deviation with Gaussian Process Models. *Contributions to Statistics. Advances in Model-Oriented Design and Analysis — mODa 10. D. Ucin'ski et al. (eds.)*, pages 35–43. Springer International Publishing, Switzerland, 2013.
- [12] I. Molchanov. *Theory of Random Sets*. Springer-Verlag, London etc., 2005.
- [13] О. Ю. Воробьев. О множественных характеристиках состояний распределенных вероятностных процессов. *Известия СО АН СССР*, 3(3):3–7, 1977.
- [14] O. Yu. Vorobyev. On set characteristics of states of distributed probability processes. *Izvestia of SB AS USSR*, 3(3):3–7, 1977 (in Russian).
- [15] С. А. Ковязин. О предельном поведении одного класса эмпирических средних случайного множества. *Теория вероятностей и ее применения*, 30(4):767–772, 1985.
- [16] S. Kovyazin. On the limit behavior of a class of empirical means of a random set. *Theory of Probability and its Applications*, 30(4):814–820. Translated from Russian by J. Malek., 1986.
- [17] О. Ю. Воробьев. *Эвентология*. Сибирский федеральный университет, Красноярск, 2007, 435с., <http://eventology-theory.ru/0-books/1-VorobyevOleg~2007~Eventology~435p.pdf>.
- [18] O. Yu. Vorobyev. *Eventology*. Siberian Federal University, Krasnoyarsk, Russia, 2007, 435p. (in Russian, abstract in English), <http://eventology-theory.com/0-books/1-VorobyevOleg~2007~Eventology~435p.pdf>.
- [19] О. Ю. Воробьев. Средневероятное событие для множества событий. *Труды XI Международ. ФАМЭБ конференции по финансово-актуарной математике и эвентологии безопасности*, Красноярск: НИИП-ПБ, СФУ (под ред. Олега Воробьева):139–147, 2012.
- [20] O. Yu. Vorobyev. A mean probability event for a set of events. In. *Proc. of the XI Intern. FAMES Conf. on Financial and Actuarial Mathematics and Eventology of Safety*, Krasnoyarsk, SFU (Oleg Vorobyev ed.):139–147, 2012 (in Russian, abstract in English).



# Applicable eventology of safety: inconclusive totals

Oleg Vorobyev

Siberian Federal University  
Institute of Mathematics and Computer Science  
Trade Economy Institute  
Krasnoyarsk  
oleg.yu.vorobyev@gmail.com

**Abstract.** *Totals of the eventological [1] safety system modeling [2, 3, 4] is considered for examples and illustrations, which are intended to demonstrate the main features of the algorithm for calculating the risk of a dangerous event at the company under established the event-related circumstances based on the portfolio of identification indicators of company safety; inter alia the examples and illustrations show the role and functions (in calculating the risk) of the three main event-based figurants in the safety eventological system: the total subject, the total object and the total barrier; and most importantly they reveal the key of eventological approach applicability for the field of safety in the methods for selecting the optimal portfolio of identification indicators of safety providing specified accuracy of estimating risk of the dangerous event for this company by minimal expert costs.*

**Keywords.** *Eventology, applicable eventology, probability theory, event, probability, set of events, algebra of events, mean probable event, value of an event, Gibbsean event-based model, event identification, total subject, total object, total barrier, portfolio of identification indicators, accuracy of estimating risk, minimum cost of expert.*

*Here you come up with the idea.  
Like, it is quite insane  
to match this crazy world.  
And then it comes up to you,  
that the world is crazier  
than you thought.  
And to understand it,  
it needs to come up with another idea,  
absolutely insane.  
The science is growing  
with a series of that follies  
roughly.  
Sir Roger Penrose, 2013.*

In this paper I intend to very briefly summarize two years of *eventological* [1] research in the field of safety, which have been published in my works [2, 3, 4, 5, 6, 7, 8], stopping only at the nodal eventological concepts and models to provide a compact and transparent picture again eventological prevailing approach in the development of practical methods in the field of safety

and insurance, which led eventually to the establishment of a new *eventological safety system of enterprise*.

Among other things I'm going to sum up the work of this still new and unfamiliar to most researchers field in the examples that illustrate the characteristics of *eventological safety system of enterprise*, for if you want to clarify something new, then to make it clearer in the examples, ruthlessly eliminating unnecessary parts and sending them to the original sources.

The nature of this final work forces me to engage in no small measure self-citations. Some short but important sections cited are unchanged as they appear in the original. In the other - I have made minor editorial changes.

In addition, this review includes some illustrations of my previous works, equipped with self-contained full captions that are not duplicated in the text. However, this work, along with the inevitable new synergies for the first time gathered together the results also contains some new ideas, has never been published, and therefore, I hope, is an independent scientific value for the applicable eventology safety and insurance.

Finally, the pattern of recent eventological research in the field of safety would be far from complete without the results of Arcady Novosyolov [9, 10] for controlling and optimizing the risk of dangerous events, references to which I am pleased to be included in these totals.

## 1 Eventology of safety

Recently, the paper [2] proposes a new eventological world in the field of safety, which is relative to the subject, event, probability and value.

### 1.1 Safety: subject, event, probability, value [2, pp. 92-93]

The meaning of the concept of danger<sup>1</sup> is investing only a *subject* (or a *set of subjects*), which defines and describes the danger. No subject is devoid of any concept of danger, it would not make sense. Danger is always a danger for the subject, the *subject danger*. As well as a

<sup>1</sup>Or it is safe as there is no danger.

safety is always a *subject safety*. Thus, the danger and the safety are *subject categories*.

It has long been noted more than once stated [1] that a *subject* does not exist without the *event*. The event is a *being of the subject*, a *subject being*, a *co-being*, and only in the event subject realizes its existence.

In eventology [1] every event has a *probability*, and the concept of probability does not make sense as long as the event is defined, the probability of which is at stake. Thus, *events do not happen without a probability as the probability does not happen without the event*. The second essential characteristic of subject events in eventology is its value. Just as in the case of probability, the eventological concept of value is meaningless as long as the event is not specified, the value of which is pointed: *event is not without value, as the value does not exist without the event*.

Eventological theory characterizes the subject by this or that *set of events* that are directly or indirectly connected with the subject, reflecting either side of his being. Each such set of events occurs in the form of a *situation* of an *set of circumstances*, of a *combinations* of occurrence or non-occurrence of events from this set of a *terraced events* generated by this set. And each such terraced event occurs with appropriate probability. A *set of probabilities of all the events terraced* generated by a given set of events is called the *probability distribution* of a given set of events, and a *set of values of all the terraced events* is called its *value distribution*<sup>2</sup>.

Here we are interested in those aspects of a subject being, those *sets of events*, along with their *E-distributions* that give rise to its *danger* or its *safety*. It goes without saying that the *subject danger* and the *subject safety* is *relative categories* defined as to the *set of events* that each time selects the subject himself.

Conclusions of the subject of a danger or safety at the current juncture is always preceded by a conscious or unconscious *estimating* by subject to the *probability* and *value* of a causing or not causing damage to the subject by a coincidence, *estimating* by subject the *probability* and *value* of the terraced event. Thus, both the danger and the safety are not only *subject* and *relative*, but also *probabilistic* and *value-categories*.

\*\*\*

It is appropriate to once again after [2] and [3] emphasize the important idea of revealing insight eventological *safety* and *insurance*:

- and safety, and insurance is always *subject safety* and *subject insurance*, as measured mathematically

<sup>2</sup>The *probability* and *value distributions* of events together determine the Gibbs characterization of *eventological distribution* (*E-distribution*) of the set of events.

based on *eventological model of the subject* and of *subject methods* of accounting and control *probabilistic* and *value* eventological distribution of sets of events in the field of safety and insurance.

\*\*\*

In [3] it has a method of eventological simulation of safety systems within eventological system analysis, the result of which is the eventological safety model of subjects at the enterprise.

## 1.2 Eventological system of safety

In [3] it presents mathematical models *safety*<sup>3</sup> that are based on the eventological system theory [1] and the latest developments in the field of eventology safety [2]. Systems of (fire) safety exist in every corner of the world, in every industry and in every enterprise. Current approaches to the development of (fire) safety systems [11, 12, 13] regardless of the specific country, sector and individual features of the system have a common event-based system basis. Eventological systems theory allows the development of a mathematical model which makes it possible not only to express in a unit *eventological safety system* and system basis, and the system shell, but also to explain and to measure the structure of the system event-related interactions between them.

To express, to explain and to measure eventologically safety systems as *systems of events* we must first agree that in eventological theory it means by a *system of events*, and, in particular, than the notion of a *system of events* must be different from still central to this theory the concept of the *set of events*.

In eventology mathematical models of safety systems are considered as a part of *eventological system theory* (*eventological system analysis*) [1]. Since our work [5], a *eventological system* (*system of events*) proposed a set of events, which is composed of *free set of events* (*system basis*), and events operationally related to events from the basis (*system shell*).

The main landmarks in the development of safety systems, we selected from a rather impressive list of two works. The one [12] can rightly be considered the most famous domestic achievement in the field of fire safety, and the other [13] — exemplary performance of an international project of safety system.

The first work [12] allowed a useful comparative analysis. The second [13], in which the concepts of preventive and reactive *barriers*, prompted by the thought put into the eventological system safety analysis new concepts and terms: events that are related to the activities of providing safety, called *barrier event*, and the eventological model of a set of barrier events — the

<sup>3</sup>including *fire* safety systems.

*total barrier*, which, together with the *total subject* and *total object* is one of the three main *event-figurants* in the *eventological model of safety system of subjects at the enterprise* proposed.

## 2 Glossary of terms and problems of the applicable eventology of safety

1. **Gibbs model** of eventological system of safety is an eventological model describing the event-based behavior of three total figurants of safety: *the subject, barrier and the object*, and assessing the risk (the probability of a danger event) in the enterprise as a result of expert review of event state of enterprise safety, carried out within the *expert portfolio of I-signs*.
2. **Identification of regulatory parameters of Gibbs event model** — maximum, medium and minimum risk (the probability of a danger event) — for this enterprise together historical, expert and model statistics (see Section 4.3).
3. **“I-sign”** is a *identification sign of safety* of the enterprise, the values of which are assess by the expert and define execution/non-execution of regulatory safety requirements; synonym for **“I-event”**, *identification event of safety* of this enterprise, the occurrence of which characterizes the performance of the regulatory requirements for safety and is assessed by expert;
4. **“Portfolio of I-signs”** is a set of I-signs  $\mathfrak{A}$ , used in eventological safety model to assess the risk (probability) of the dangerous event in the field of safety.
5. **Figurant portfolio of “I-signs”** is one of the three subsets of I-signs that characterize each of the three figurants individually: *total subject, the total barrier and total object*, and used in the eventological safety model for risk (probability) assessment of a dangerous event in the enterprise for the appropriate figurants (**subject:  $\mathfrak{M}$ , barrier:  $\mathfrak{B}$ , object:  $\mathfrak{X}$** ).
6. **Internal figurant subportfolio of “I-signs”** is one of two subsets of I-signs that characterize the two-set state of each of the three figurants: *total subject, total barrier and total object*, and used in eventological safety model to assess the risk (probability) of a dangerous event in the enterprise for the appropriate figurant (**internal subject:  $\mathfrak{M}_1, \mathfrak{M}_2$ , internal barrier:  $\mathfrak{B}_1, \mathfrak{B}_2$ , internal object:  $\mathfrak{X}_1, \mathfrak{X}_2$** ).
7. **Mean probable portfolio of “I-signs”** is the portfolio *widehat* $\mathfrak{A}$ , which approximates the portfolio of I-signs *frakA* in the mean probable, composed of mean probable I-signs, each of which approximates in the mean probable to one subportfolio of portfolio  $\mathfrak{A}$  respectively.
8. **Expert portfolio of “I-signs”** is a set of I-signs  $\mathfrak{A}^{(e)}$ , selected from a total portfolio of I-signs  $\mathfrak{A}$  and the proposed for expert review the safety of the enterprise; and similarly defined *expert figurant portfolios and internal subportfolios*.
9. **Minimum expert portfolio of “I-signs”** is an expert portfolio in which each I-sign belongs to the *only* one of the six subportfolios and each of the six subportfolios contains the *only* one I-sign.
10. **Weighted minimum expert portfolio of “I-signs”** is an expert portfolio in which each I-sign belongs to the *only* one of the six subportfolios, and each of the six subportfolios contains *not less than one* I-sign.
11. **Assessment of portfolio weights of I-sign** (based on statistical surveys of experts and based on the *Gibbs model*, from which the log-dependence of portfolio weight of I-sign of the probability of its value), which is characterized by its effect on the risk (probability) of a dangerous event among other I-signs of the portfolio (see Section 4.5).
12. **Assessment of information capacity of I-sign** in the portfolio (based on statistical surveys of experts and evaluation of information obtained during testing of its value), which characterizes the importance of expert review information value of this I-sign.
13. **Assessment of accuracy of the calculation of risk** (the probability of a dangerous event) in the enterprise and its dependence on the number of I-signs in the expert portfolio (see Section 4.6).

14. **The optimal expert portfolio of I-signs** for the enterprise is a portfolio of I-signs  $\mathfrak{A}^{(e)}(\delta)$ , provides a given degree of *accuracy* of risk assessment (the probability of a dangerous event) at the minimum cost of expert (see section 4.5).

## 3 An event hierarchy of eventological safety system in pictures

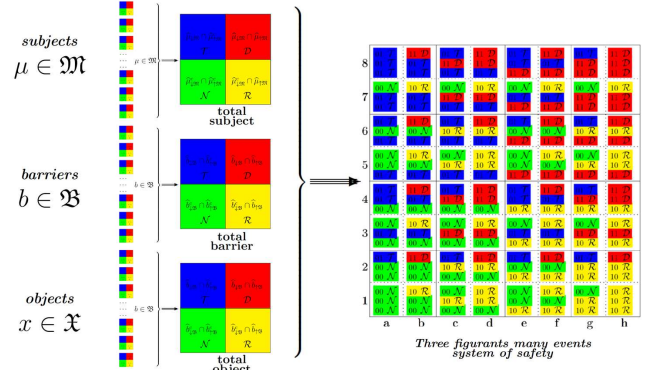


Fig. 1: An event hierarchy of eventological system of safety consists of three levels — *Left*: the set of subjects  $\mathfrak{M}$ , the set of barriers  $\mathfrak{B}$  and the set of objects  $\mathfrak{X}$ ; *Center*: three system figurants; *Right*: three figurants eventological system of safety.

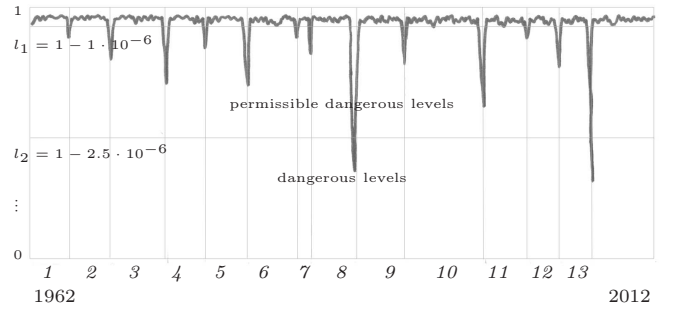


Fig. 2: **Trajectory of safety.** Schedule of probability of safety  $0 < l < 1$  (vertical axis in the *nonlinear* scale) for an 50-year sequence of 13 hyper-scenario cycles of safety system. The frequency of fire dangerous events is 0.04 (2 events in 50 years). Safe levels:  $l_1 < l < 1$ , allowed dangerous levels:  $l_2 < l < l_1$ ; dangerous levels of  $0 < l < l_2$ .

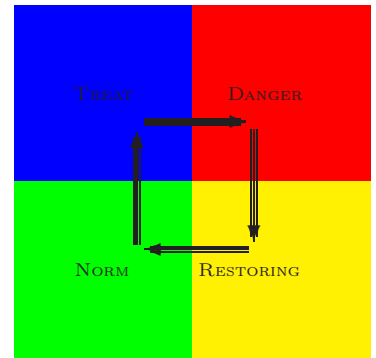


Fig. 3: **Many events cycle.** Venn diagram illustrating many events cycle of safety system, characterized by a succession of mean probable states of its total figurants:  $\dots \rightarrow \text{Norm (N)} \rightarrow \text{Threat (T)} \rightarrow \text{Danger (D)} \rightarrow \text{Restoring (R)} \rightarrow \text{Norm (N)} \rightarrow \dots$ .



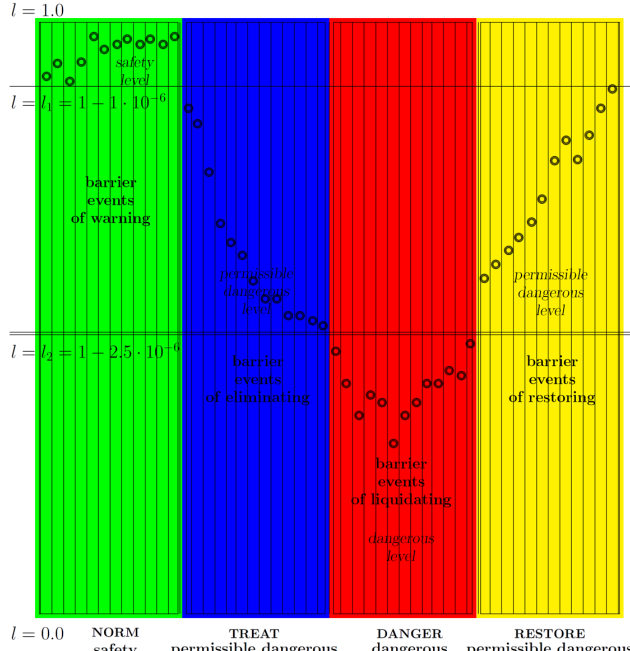


Fig. 4: **Many events cycle in time.** Eventological many events model of the main cycle (*Norm*, *Threat*, *Danger*, *Restore*) of safety system of subjects in the enterprise. Venn diagram of a succession (left to right) of confluence of *mean probable states* of system figurants: the total subjects, barriers and objects. Against the background of *mean probable state* of the total object, the total subject implemented *barrier event*: **warning** (in safe state: NORM (N)), **eliminating** (in permissible dangerous state: TREAT (T)), **liquidating** (in dangerous state: DANGER (D)) and **restoring** (in permissible dangerous state: RESTORE (R)). The vertical axis — *safety* ( $0 \leq l \leq 1$  in *nonlinear scale*) of the subjects, the horizontal axis is the time sequence (left to right) of terraced events.

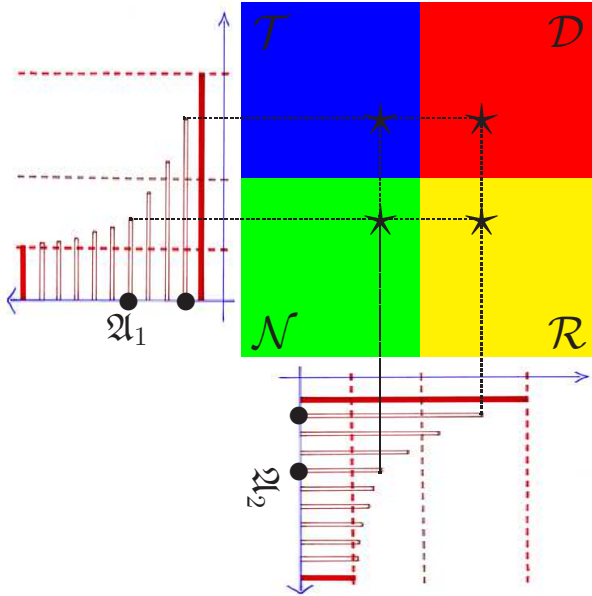


Fig. 5: **The identification of a state of one of the three figurants**  $\mathfrak{A} = \mathfrak{A}_1 \cup \mathfrak{A}_2 = \mathfrak{M} = \mathfrak{M}_1 \cup \mathfrak{M}_2$ ,  $\mathfrak{B} = \mathfrak{B}_1 \cup \mathfrak{B}_2$  and  $\mathfrak{X} = \mathfrak{X}_1 \cup \mathfrak{X}_2$  of safety system of the enterprise based on expert review, which identifies ( $\star$ ) mean probable state *Treat*, *Danger*, *Restoring* or *Norm* for the figurant  $\mathfrak{A} = \mathfrak{A}_1 \cup \mathfrak{A}_2$ .

#### 4 Four stages of the assessment of risk of a dangerous event by eventological safety system

The procedure for assessment of the current risk of a dangerous event by eventological safety system of the

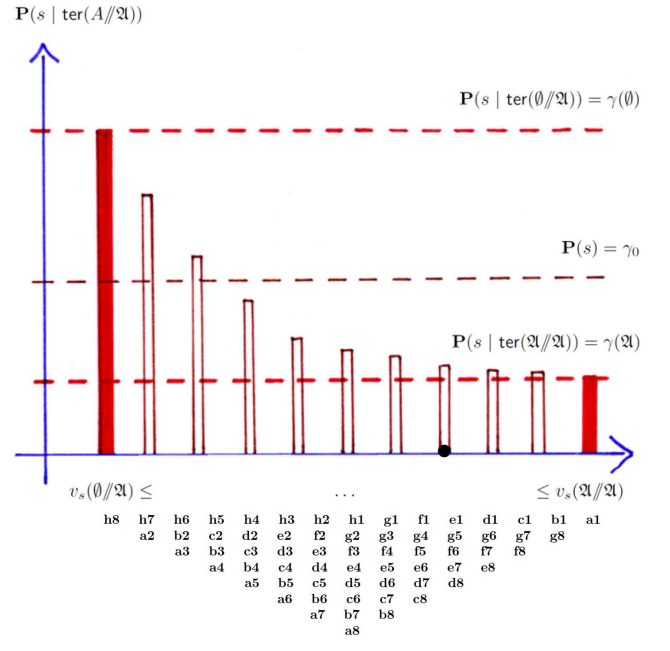


Fig. 6: **Model of Gibbs dependence of risk**  $P(s | \text{ter}(A//\mathfrak{A})) = \gamma(\emptyset) \exp \left( \frac{|A|v_s}{|\mathfrak{A}|v_s} \ln \frac{\gamma(\mathfrak{A})}{\gamma(\emptyset)} \right)$  of dangerous event  $s$  on a confluence of circumstances of safety event  $\text{ter}(A//\mathfrak{A})$  (combinations of values of I-signs of safety) for each figurants  $\mathfrak{A}_1, \mathfrak{A}_2, \mathfrak{A}_3$  or for  $\mathfrak{A}$ , the entire enterprise. On Gibbs characterization of a set of I-events  $\mathfrak{A}$  is based regulatory identification risk of danger event on the known standard values *worst*:  $\gamma(\mathfrak{A}//\mathfrak{A})$ , *best*:  $\gamma(\emptyset//\mathfrak{A})$  and *average risk*:  $\gamma_0$ . The horizontal axis represents the portfolio weight (relative to the danger event  $s$ )  $v_s(A//\mathfrak{A}) = |A|v_s, A \subseteq \mathfrak{A}$ , of terraced I-events  $\text{ter}(A//\mathfrak{A})$ , arranged in ascending order. Under the horizontal axis in accordance with the values of the risk of a dangerous event has “chess notation” of 64 mean probable safety states.

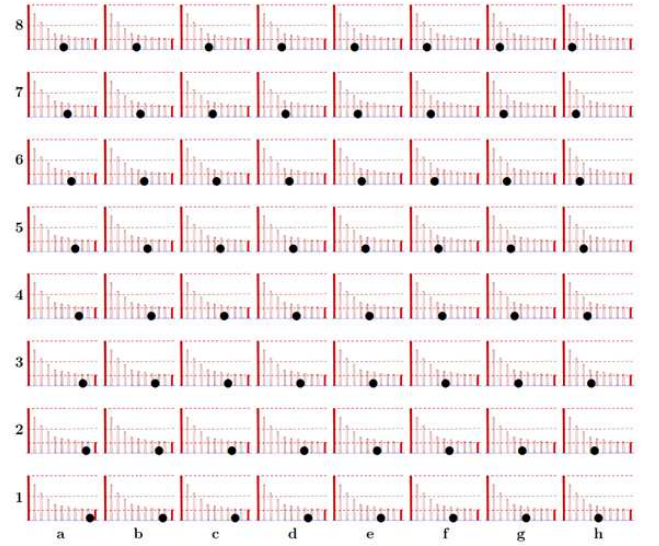


Fig. 7: **Gibbs dependence of risk** of a danger event on the value of the current mean probable state of safety of typical enterprise. Black dots on the horizontal axis shows the values of each of the 64 mean probable states of safety, which on the basis of Gibbs event model the conditional probability of a hazardous event is defined under the current state of safety. The safest state — **a1**, the least safe state — **h8**.

enterprise consists of four stages.

##### I Preparation of the portfolio of I-signs (identification

*signs*) of safety optimal for the enterprise (see Section 4.1).

II *Expertise of I-signs* of optimally prepared the portfolio (see Section 4.2).

III *Identification of the current state of safety* on the results of this expertise of I-signs (see Section 4.3).

IV *The calculation of the risk of a dangerous event* at the enterprise for the current state of its safety, identified on the basis of expertise of the I-signs (see Section 4.4).

#### 4.1 Preparation of the optimal portfolio of I-signs

Preparation of the portfolio of I-signs must meet two criteria of optimality:

1. Portfolio of I-signs of safety should provide the minimum permissible expert cost of checking the I-signs (see Section 4.5);
2. The accuracy of risk assessment of dangerous events on the basis of the portfolio of I-signs should be no worse than a specified level (see Section 4.6).

*Preparation of the optimal portfolio I-signs* is a solution of eventological optimization problem, which optimizes the cost of expert test-portfolio of I-signs that provide a given accuracy of the risk assessment of a dangerous event (see Section 4.6), where under the expert optimization of costs means the search for optimal volume, structure (see section 5) and the information capacity of the portfolio of I-signs (see section 4.5).

#### 4.2 Expertise of I-signs

Expertise of I-signs of the portfolio optimally prepared is an expert assessment of the binary values of I-signs: present or absent, according to experts, this I-sign in the characterization of the current state of enterprise safety.

#### 4.3 Identification of the current state of safety

The results of the expertise of I-signs of optimally prepared portfolio are the initial data for the identification of the current state of safety of enterprise. By the same standard algorithm [3, 4] based on the Gibbs event model (4.4.1) for  $\mathfrak{A} = \mathfrak{M}, \mathfrak{B}$ , or  $\mathfrak{X}$ , the identification is carried out of the current *mean probable state* each of the three figurants in the safety system (see Fig. 5 in Section 3): *the total subject*  $\mathfrak{M}$ , *the total barrier*  $\mathfrak{B}$  and *the total object*  $\mathfrak{X}$ , thereby identifying the current *mean probable state* of safety system (see Fig. 1 in Section 3). Identification of the current state of the system is based on the same Gibbs event model (4.4.1) for  $\mathfrak{A} = \mathfrak{M} \cup \mathfrak{B} \cup \mathfrak{X}$ .

#### 4.4 Calculation of the risk of a dangerous event

On the basis of the Gibbs event model [3, 4] it is offered to calculate the risk of a dangerous event  $s$  at the enterprise under the current state of its safety  $\text{ter}(A//\mathfrak{A})$  using the following formula for conditional probability:

$$\mathbf{P}(s|\text{ter}(A//\mathfrak{A})) = \gamma(\emptyset) \exp \left( \frac{|A|_{v_s}}{|\mathfrak{A}|_{v_s}} \ln \frac{\gamma(\mathfrak{A})}{\gamma(\emptyset)} \right), \quad (4.4.1)$$

where  $\gamma(\emptyset)$  is the worst risk,  $\gamma(\mathfrak{A})$  is the best risk,  $\gamma_0 = \mathbf{P}(s)$  is the average risk of dangerous event (see Fig. 6 in Section 3).

#### 4.5 Optimizing the expert costs for expert review of portfolio of I-signs

To solve the problem of optimizing the costs of expert for checking the portfolio of I-signs it is required the notions of a *portfolio weight* and an *information capacity of the I-signs*.

The *portfolio weight*  $w_a = \mathbf{V}(a)$  of I-sign/I-event  $a \in \mathfrak{A}$  is based on the Gibbs model which connects it with the probability of  $p_a = \mathbf{P}(a)$  of occurrence of the I-event  $a \in \mathfrak{A}$  by known formula [1]

$$p_a = \frac{1}{\mathcal{Z}} \exp\{\alpha w_a\}, \quad (4.5.1)$$

where  $\alpha$  and  $\mathcal{Z}$  are Gibbs model parameters of portfolio of I-signs.

The formula (4.5.1) can solve two mutually inverse probability problems:

- from known portfolio weight  $w_a$  of I-event  $a \in \mathfrak{A}$  to seek the probability  $p_a$ ;
- from known probability  $p_a$  of I-event  $a \in \mathfrak{A}$  to seek the its portfolio weight

$$w_a = \frac{1}{\alpha} \ln(p_a \mathcal{Z}), \quad (4.5.1')$$

as well as the control problems for the risk of a dangerous event under various restrictions on expert costs [9, 10].

The *information capacity of I-sign/I-event* is measured by entropy of I-event  $a \in \mathfrak{A}$  by the known formula [1] (see Fig. 8)

$$\mathcal{I}_a = -p_a \ln p_a - (1 - p_a) \ln(1 - p_a), \quad (4.5.2)$$

where  $p_a = \mathbf{P}(a)$  is the probability of I-event  $a \in \mathfrak{A}$ .

The interpretation of this information specifications of I-event lies in the fact that if the probability of I-event is close to 0 or 1, the expert review of its occurrence provides little additional information and its inclusion in the expert portfolio is not too justified. The criteria for selection of I-sign  $a \in \mathfrak{A}$  of the portfolio  $\mathfrak{A}$  in *optimal expert portfolio*

$$\mathfrak{A}^{(e)}(\delta) = \{a \in \mathfrak{A} : \mathcal{I}_a \geq \delta\} \subseteq \mathfrak{A}, \quad (4.5.3)$$



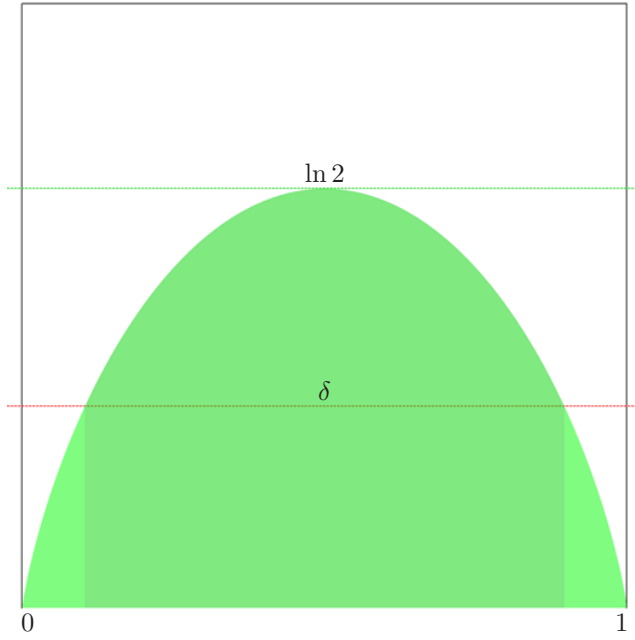


Fig. 8: On the unit square  $[0, 1]^2$  hypograph shown (green) of the information capacity  $\mathcal{I}_a$  of I-sign  $a$ , depending on the probabilities  $p_a$  (horizontal axis) (4.5 .2), used for the selection of the best expertise in the portfolio of the I-signs, the probability of which is not less than the threshold  $\delta$ .

is a significant amount of its information capacity  $\mathcal{I}_a$  surpassing the threshold  $\delta \in [0, \ln 2]$  which depends on the desired accuracy of the risk assessment of a dangerous event.

#### 4.6 Optimizing the accuracy of assessment of the risk of a dangerous event

Error  $\varepsilon$  of the assess  $\hat{P}(s)$  of the risk  $\mathbf{P}(s|\text{ter}(A//\mathfrak{A}))$  of dangerous event  $s$  based on the selected expert portfolio  $\mathfrak{A}^{(e)}(\delta)$  can be computed using the formula:

$$\varepsilon = t_{(1-\alpha)/2, N-1} \frac{\sigma}{\sqrt{N}},$$

where  $t_{(1-\alpha)/2, N-1}$  is the  $(1 - \alpha)/2$ -quantile of the Student  $t$ -distribution with  $N - 1$  degrees of freedom,  $\sigma$  is a standard deviation,  $\alpha$  is the required level of confidence usually taken equal 0.95, and  $N = |\mathfrak{A}^{(e)}(\delta)|$  is the power of expert portfolio, i.e. amount of I-signs including in  $\mathfrak{A}^{(e)}(\delta)$ . Thus the error  $\varepsilon$ , which is the smaller, more square root of its power  $N$ , determines the confidence interval

$$\hat{P}(s) - \varepsilon \leq \mathbf{P}(s|\text{ter}(A//\mathfrak{A})) \leq \hat{P}(s) + \varepsilon,$$

where the value of risk of a dangerous event falls with the probability which is not less than  $\alpha$ .

In addition to the power of expert portfolio  $N$ , the error  $\varepsilon$  of risk assessment greatly depends also on how the relationship between the I-signs characterized by the portfolio of selected I-signs. To control interdependence between I-signs in the portfolio may be, controlling

its probability-event structure. Control principles of probability-event structure of the expert portfolio that lead to a reduction of error set out in section 5.

### 5 Help: eventological model of expert portfolio of I-signs of safety

The key to the applicability of eventological approach in safety is the eventological model of the structure of expert answers to a set of normative questions about the safety of the enterprise, in other words, the structure of *expert portfolio of identifying signs/events (I-signs/I-events) of enterprise safety*<sup>4</sup>.

Such an event-probability structure of *I-signs of safety* should be organized optimally so that at minimum expert cost to provide the required accuracy of risk assessment (probability) of a dangerous event. In theory, it is clear that a large body of regulatory questions is due to the greater accuracy of risk assessment. However, with the increase of the totality of questions, first, rising costs of obtaining expert responses, and secondly, the information capacity of each answer added decreases. The proposed optimally organized structure of *expert portfolio of I-signs* solves this dilemma by providing the required accuracy by the minimum cost.

#### 5.1 Eventological model of the portfolio of I-signs

Eventological safety model proposed in [2, 3, 4, 5, 6, 7, 8], is an event-probability model of the safety enterprise system, which is formed by event-driven reaction of three system total figurants, each of whom is responsible for the event-behavior of the corresponding set of:  $\mathfrak{M}$  — *subjects*,  $\mathfrak{B}$  — *barriers* and  $\mathfrak{X}$  — *objects* of the enterprise, joint event-state of which form the safety state of the enterprise as a whole.

Eventological risk (probability) of a dangerous event at the enterprise depends on the event-state in which there is an enterprise safety. This event-state is characterized by portfolio of I-signs of safety of this enterprise, which is usually a set of answers to the specially selected regulatory questions. The composition of the portfolio of I-signs and organization structure of these I-signs in the portfolio largely determine the accuracy of the expert assessment of risk and the amount of the cost of the expertise of state of enterprise safety.

The general structure of the portfolio of I-signs is defined by the hierarchical structure three-figurants eventological safety model, where each of the three total figurants has its own two-set-structure. This hierarchical

<sup>4</sup>The definitions of the basic concepts and concise formulations of the problems (*usually highlighted in italics*) associated with the optimal choice of the expert portfolio of I-signs of enterprise safety listed in the reference section 2 on page 106.

structure is formed by three *figurant portfolios I-signs*

$$\mathfrak{M}, \mathfrak{B}, \mathfrak{X},$$

each of which is formed by the union<sup>5</sup> of two *inner figurant subportfolios of I-signs*<sup>6</sup>:

$$\mathfrak{M} = \mathfrak{M}_1 \cup \mathfrak{M}_2, \quad \mathfrak{B} = \mathfrak{B}_1 \cup \mathfrak{B}_2, \quad \mathfrak{X} = \mathfrak{X}_1 \cup \mathfrak{X}_2. \quad (5.1.1)$$

As a result, the structure of *portfolio of I-signs* of enterprise safety is characterized by the totality

$$\mathfrak{A} = \mathfrak{M}_1 \cup \mathfrak{M}_2 \cup \mathfrak{B}_1 \cup \mathfrak{B}_2 \cup \mathfrak{X}_1 \cup \mathfrak{X}_2, \quad (5.1.2)$$

which is formed by the union<sup>6</sup> of six *figurant subportfolios of I-signs*.

The problem of choosing the optimal structure of the expert portfolio is to choose six *figurant expert subportfolios of I-signs* (5.1.2), forming a general *expert portfolio of I-signs of the enterprise*  $\mathfrak{A}^{(e)} \subseteq \mathfrak{A}$ , which provides the required accuracy of the risk assessment by minimal expert costs.

## 5.2 Full expert portfolio of I-signs

The Table 1 illustrates the minimum *full expert portfolio of I-signs*

$$\mathfrak{A} = \mathfrak{M}_1 \cup \mathfrak{M}_2 \cup \mathfrak{B}_1 \cup \mathfrak{B}_2 \cup \mathfrak{X}_1 \cup \mathfrak{X}_2 \subset \mathcal{A}, \quad (5.2.1)$$

compiled by combining six subportfolios, each of which is defined by the relevant *subset of I-events*

$$\mathfrak{M}_1, \mathfrak{M}_2, \mathfrak{B}_1, \mathfrak{B}_2, \mathfrak{X}_1 \text{ и } \mathfrak{X}_2 \subset \mathcal{A}. \quad (5.2.2)$$

This full expert portfolio *frakA* is an interesting event-based interpretation, perhaps the only correct only when using the concept of a kind of mean event-based characteristics of *subsets of I-events*. The role of the mean characteristics can be successfully implemented, for example, recently introduced to eventology the concept of the *mean probable event* [7, 8]. In this interpretation, the entire expert portfolio is approximated by a set of just six *mean probable I-events*

$$\widehat{\mathfrak{M}}_1, \widehat{\mathfrak{M}}_2, \widehat{\mathfrak{B}}_1, \widehat{\mathfrak{B}}_2, \widehat{\mathfrak{X}}_1 \text{ и } \widehat{\mathfrak{X}}_2 \in \mathcal{A}, \quad (5.2.3)$$

each of which approximates one of the six *subsets of I-events* (5.2.2), which, for example, for *subset of I-events*  $\mathfrak{M}_1 \subset \mathcal{A}$  is interpreted as follows:

$$\widehat{\mathfrak{M}}_1 = \mathcal{E}(\mu/\mathfrak{M}_1) = \widehat{\mu}_{\mathfrak{M}_1} \in \mathcal{A}.$$

<sup>5</sup>Not necessarily disjoint.

<sup>6</sup>Each pair of inner figurant subportfolios of I-signs is required mainly in order to assess the condition of the 4 states of the safety cycle (norm, threat, danger, restore) is one of three figurants involved and, consequently, in which condition of 64 is the whole safety system of the enterprise. In addition, inner figurant subportfolios of I-signs can be used to estimate the *private risks* of dangerous events for each figurant involved in isolation, or for each of a pair of sides<sup>7</sup> of the figurant separately.

$\mathfrak{A} \overset{mp}{\sim} \widehat{\mathfrak{A}}$	$\widehat{\mathfrak{M}}_1$	$\widehat{\mathfrak{M}}_2$	$\widehat{\mathfrak{B}}_1$	$\widehat{\mathfrak{B}}_2$	$\widehat{\mathfrak{X}}_1$	$\widehat{\mathfrak{X}}_2$
ter--T	o	o	o	o	o	•
ter--R	o	o	o	o	•	o
ter--N	o	o	o	o	•	•
ter-T-	o	o	o	•	o	o
ter-TT	o	o	o	•	o	•
ter-TR	o	o	o	•	•	o
ter-TN	o	o	o	•	•	•
ter-R-	o	o	•	o	o	o
ter-RT	o	o	•	o	o	•
ter-RR	o	o	•	o	•	o
ter-RN	o	o	•	o	•	•
ter-N-	o	o	•	•	o	o
ter-NT	o	o	•	•	o	•
ter-NR	o	o	•	•	•	o
ter-NN	o	o	•	•	•	•
terT--	o	•	o	o	o	o
terT-T	o	•	o	o	o	•
terT-R	o	•	o	o	•	o
terT-N	o	•	o	o	•	•
terTT-	o	•	o	•	o	o
terTTT	o	•	o	•	o	•
terTTR	o	•	o	•	•	o
terTTN	o	•	o	•	•	•
terTR-	o	•	•	o	o	o
terTRT	o	•	•	o	o	•
terTRR	o	•	•	o	•	o
terTRH	o	•	•	o	•	•
terTN-	o	•	•	•	o	o
terTNY	o	•	•	•	o	•
terTNR	o	•	•	•	•	o
terTNN	o	•	•	•	•	•
terR--	•	o	o	o	o	o
terR-T	•	o	o	o	o	•
terR-R	•	o	o	o	•	o
terR-N	•	o	o	o	•	•
terRT-	•	o	o	•	o	o
terRTT	•	o	o	•	o	•
terRTR	•	o	o	•	•	o
terRTN	•	o	o	•	•	•
terRR-	•	o	•	o	o	o
terRRT	•	o	•	o	o	•
terRRR	•	o	•	o	•	o
terRRN	•	o	•	o	•	•
terRN-	•	o	•	•	o	o
terRNT	•	o	•	•	o	•
terRNR	•	o	•	•	•	o
terRNN	•	o	•	•	•	•
terN--	•	•	o	o	o	o
terN-Y	•	•	o	o	o	•
terN-R	•	•	o	o	•	o
terN-H	•	•	o	o	•	•
terNT-	•	•	o	•	o	o
terNTT	•	•	o	•	o	•
terNTR	•	•	o	•	•	o
terNTN	•	•	o	•	•	•
terNR-	•	•	•	o	o	o
terNRT	•	•	•	o	o	•
terNRR	•	•	•	o	•	o
terNRN	•	•	•	o	•	•
terNN-	•	•	•	•	o	o
terNNT	•	•	•	•	o	•
terNNR	•	•	•	•	•	o
terNNN	•	•	•	•	•	•

Tab. 1: Six subportfolios of I-events:  $\mathfrak{M}_1, \mathfrak{M}_2, \mathfrak{B}_1, \mathfrak{B}_2$  and  $\mathfrak{X}_1, \mathfrak{X}_2$  forming a *minimum full portfolio*  $\mathfrak{A} = \{\text{ter}_{A/\widehat{\mathfrak{A}}}, A \subseteq \widehat{\mathfrak{A}}\}$  of all possible (for a portfolio of  $\mathfrak{A}$ ) 63 *terraced I-events of the II kind* generated by six mean probable events (5.2.3) of the sextet  $\widehat{\mathfrak{A}}$ , approximating  $\mathfrak{A}$  in the mean probable:  $\widehat{\mathfrak{A}} \overset{mp}{\sim} \mathfrak{A}$  (on abbreviation of NNN, ..., TTT, ..., RRR for subsets of  $A \subseteq \widehat{\mathfrak{A}}$  see Remark 2 on page 112).

I will outline

$$\widehat{\mathfrak{A}} = \{\widehat{\mathfrak{M}}_1, \widehat{\mathfrak{M}}_2, \widehat{\mathfrak{B}}_1, \widehat{\mathfrak{B}}_2, \widehat{\mathfrak{X}}_1, \widehat{\mathfrak{X}}_2\} \subset \mathcal{A} \quad (5.2.4)$$

the set made as of the elements of the six mean probable events (5.2.3), and *approximating in the mean probable* a whole expert portfolio  $\mathfrak{A}$  and  $A \subseteq \widehat{\mathfrak{A}}$  generated by  $64 = 2^6$  terraced I-events of the second kind

$$\text{ter}_{A//\widehat{\mathfrak{A}}} = \bigcap_{a \in A} a \in \mathcal{A},$$

the probability of which form *probability distribution of II-type*

$$\left\{ p_{A//\widehat{\mathfrak{A}}}, A \subseteq \widehat{\mathfrak{A}} \right\}$$

of sextet of I-events  $\widehat{\mathfrak{A}}$ , where

$$p_{A//\widehat{\mathfrak{A}}} = \mathbf{P} \left( \text{ter}_{A//\widehat{\mathfrak{A}}} \right).$$

**Remark 1.** All I-events from the full portfolio  $\mathfrak{A}$  are *joint* and are generated by six *mean probable* I-events of the sextet  $\widehat{\mathfrak{A}}$  as the different results of *terraced*<sup>8</sup> operation of the II-kind over sextet  $\widehat{\mathfrak{A}}$ , indexed by its different subsets<sup>9</sup>  $A \subseteq \widehat{\mathfrak{A}}$ .

**Remark 2.** In Table 1 abbreviation used to denote subsets of the sextet  $A \subseteq \widehat{\mathfrak{A}}$  refers to the states of Norm (N), Threats (T), Restore (R) and Danger (D) of each of the three total figurants of safety system as follows. For any total figurant involved  $\mathfrak{F} = \mathfrak{M}, \mathfrak{B}$ , or  $\mathfrak{X}$ , approximated by doublet of mean probable events  $\{\widehat{\mathfrak{F}}_1, \widehat{\mathfrak{F}}_2\}$ , its empty subset  $\emptyset \subseteq \{\widehat{\mathfrak{F}}_1, \widehat{\mathfrak{F}}_2\}$  corresponds to the *mean probable state* of Danger (D), the subset  $\{\widehat{\mathfrak{F}}_1\}$  — *mean probable state* of Threat (T), the subset  $\{\widehat{\mathfrak{F}}_2\}$  — *mean probable state* of Restore (R), and the subset  $\{\widehat{\mathfrak{F}}_1, \widehat{\mathfrak{F}}_2\}$  — *mean probable state* of Norm (N). So, for example, the abbreviation “-TN” corresponds to *mean probable state* of safety when the *total subject*  $\mathfrak{M}$  is in an arbitrary *mean probable state*, the *total barrier*  $\mathfrak{B}$  is in the *mean probable state* of Threat (T), and the *total object*  $\mathfrak{X}$  is in the *mean probable state* of Norm (N). Examples of the use of abbreviations are:

$$\text{ter}_{- - -} = \text{ter}_{\emptyset//\widehat{\mathfrak{A}}},$$

$$\text{ter}_{-T-} = \text{ter}_{\{\widehat{\mathfrak{B}}_2\}//\widehat{\mathfrak{A}}},$$

$$\text{ter}_{-TN} = \text{ter}_{\{\widehat{\mathfrak{B}}_2, \widehat{\mathfrak{X}}_1, \widehat{\mathfrak{X}}_2\}//\widehat{\mathfrak{A}}},$$

$$\text{ter}_{-RN} = \text{ter}_{\{\widehat{\mathfrak{B}}_1, \widehat{\mathfrak{X}}_1, \widehat{\mathfrak{X}}_2\}//\widehat{\mathfrak{A}}},$$

$$\text{ter}_{T-T} = \text{ter}_{\{\widehat{\mathfrak{M}}_2, \widehat{\mathfrak{X}}_2\}//\widehat{\mathfrak{A}}},$$

$$\text{ter}_{NNN} = \text{ter}_{\widehat{\mathfrak{A}}//\widehat{\mathfrak{A}}}.$$

<sup>8</sup>Terraced called such a set-theoretic operations on sets of events which resulted is the terraced event of one of six standard kinds [1], generated by this set of events.

<sup>9</sup>The results of terraced operations of II-kind on the events of  $\widehat{\mathfrak{A}}$  are terraced events of II-kind  $\text{ter}(A//\widehat{\mathfrak{A}}) = \bigcap_{a \in A} a \subseteq \Omega$ , indexed by subsets of  $A \subseteq \widehat{\mathfrak{A}}$ .

### 5.3 Minimum expert portfolio of I-signs

Minimum expert portfolio of I-signs (see Tab. 2), though allowing to solve all the problems facing the eventological safety system, but does not guarantee the accuracy required for risk assessment.

I-signs	★★★	Filling subportfolios					
		$\widehat{\mathfrak{M}}$		$\widehat{\mathfrak{B}}$		$\widehat{\mathfrak{X}}$	
		$\widehat{\mathfrak{M}}_1$	$\widehat{\mathfrak{M}}_2$	$\widehat{\mathfrak{B}}_1$	$\widehat{\mathfrak{B}}_2$	$\widehat{\mathfrak{X}}_1$	$\widehat{\mathfrak{X}}_2$
D-opposition of the object?	-T-	-	-	-	-	-	•
D-resistance of the object?	-R-	-	-	-	-	-	•
D-opposition of the barrier?	-T-	-	-	-	•	•	-
D-resistance of the barrier?	-R-	-	-	•	-	-	-
D-opposition of the subject?	T--	-	•	-	-	-	-
D-resistance of the subject?	R--	•	-	-	-	-	-

Tab. 2: Example of filling figurant subportfolios of I-signs in the *minimum* expert portfolio, which consists of six I-signs with weights: 1,1,1,1,1,1 (top-down). In the column ★★★ subsets of the abbreviation (see Remark 2 on page 112), by which the corresponding I-event (terraced event of II-kind) are numbered.

### 5.4 Minimum weighted expert portfolio of I-signs

Minimum weighted expert portfolio of I-signs (see Tab. 3) also solves all the problems facing the safety system, can provide the required accuracy of the risk assessment by varying the weights of I-signs in the portfolio, but unable to consider interconnections between I-signs in the portfolio.

I-signs	★★★	Filling subportfolios					
		$\widehat{\mathfrak{M}}$		$\widehat{\mathfrak{B}}$		$\widehat{\mathfrak{X}}$	
		$\widehat{\mathfrak{M}}_1$	$\widehat{\mathfrak{M}}_2$	$\widehat{\mathfrak{B}}_1$	$\widehat{\mathfrak{B}}_2$	$\widehat{\mathfrak{X}}_1$	$\widehat{\mathfrak{X}}_2$
Amount of entrances-driveways in SS?	-T-	-	-	-	-	-	•
Level of duty in SS?	-T-	-	-	-	-	-	•
Dates of power plants in SS?	-T-	-	-	-	-	-	•
Level of danger zones in SS?	-T-	-	-	-	-	-	•
Systems of inner safety in SS?	-T-	-	-	-	-	-	•
Potential source of danger in SS?	-T-	-	-	-	-	-	•
Distance to SAR?	-T-	-	-	-	-	-	•
Amount of primary-protect. means?	-T-	-	-	-	-	-	•
Maintenance of inner-safety svst.?	-T-	-	-	-	-	-	•
D-opposition of the object?	-T-	-	-	-	-	-	•
Auto alarm in SS?	-T-	-	-	-	-	-	•
Life-support systems in SS?	-T-	-	-	-	-	-	•
D-resistance of the object?	-R-	-	-	-	-	•	-
Safety-guard on duty?	-R-	-	-	-	-	•	-
Power plants in SS?	-R-	-	-	-	-	•	-
Protection of energy communications?	-R-	-	-	-	-	•	-
Carrying out dangerous work in SS?	-R-	-	-	-	-	•	-
No danger zones?	-R-	-	-	-	-	•	-
System of inner safety?	-R-	-	-	-	-	•	-
No potential sources of danger?	-R-	-	-	-	-	•	-
Service against the risk?	-R-	-	-	-	-	•	-
Level of safety-guard on duty?	-R-	-	-	-	-	•	-
Dangerous materials in SS?	-R-	-	-	-	-	•	-
Kind of life-support systems in SS?	-R-	-	-	-	-	•	-
Municipality - City?	-R-	-	-	-	-	•	-
Temporary dangerous devices in SS?	-R-	-	-	-	-	•	-
No transport. of danger materials?	-R-	-	-	-	-	•	-
No substances that may be D?	-R-	-	-	-	-	•	-
D-opposition of the barrier?	-T-	-	-	-	•	-	-
Protection system is working?	-T-	-	-	-	•	-	-
Primary-protection means in SS?	-T-	-	-	-	•	-	-
Primary-protection means?	-R-	-	-	•	-	-	-
Protection svstem?	-R-	-	-	•	-	-	-
D-resistance of the barrier?	-R-	-	-	•	-	-	-
Protect of energv-communicat. in SS?	-R-	-	-	•	-	-	-
Amount entrances in SS?	-R-	-	-	•	-	-	-
D-opposition of the subject?	T--	-	•	-	-	-	-
D-resistance of the subject?	R--	•	-	-	-	-	-

Tab. 3: Example of filling figurant subportfolios by I-signs in the *minimum weighted* expert portfolio, which consists of 38 I-signs corresponding to 6 terraced I-events of the II-kind with weights: 12,16,3,5,1,1 (top-down). In the column ★★★ the abbreviation of subsets shown (see Remark 2 on page 112), by which the corresponding I-event of II-kind are numbered. SS — safety standards, D — dangerous.

### 5.5 Arbitrary expert portfolio of I-signs

The arbitrary expert portfolio of I-signs (see Tab. 2), containing a minimum weighted portfolio, allows us to

solve all the problems facing the eventological safety system that can provide the required accuracy of the risk assessment, and has the option to account for the relationship of I-signs in portfolio with the help of I-signs contained in more than one of the six subportfolios.

I-signs	★★★	Filling subportfolios					
		M		B		X	
		M <sub>1</sub>	M <sub>2</sub>	B <sub>1</sub>	B <sub>2</sub>	X <sub>1</sub>	X <sub>2</sub>
Amount of entrances-driveways in SS?	--T	-	-	-	-	-	•
Level of duty in SS?	--T	-	-	-	-	-	•
Dates of power plants in SS?	--T	-	-	-	-	-	•
Level of danger zones in SS?	--T	-	-	-	-	-	•
Systems of inner safety in SS?	--T	-	-	-	-	-	•
Potential source of danger in SS?	--T	-	-	-	-	-	•
Distance to SAR?	--T	-	-	-	-	-	•
Amount of primary-protect. means?	--T	-	-	-	-	-	•
Maintenance of inner-safety syst.?	--T	-	-	-	-	-	•
D-opposition of the object?	--T	-	-	-	-	-	•
Auto alarm system in SS?	--T	-	-	-	-	-	•
Life-support systems in SS?	--T	-	-	-	-	-	•
D-resistance of the object?	--R	-	-	-	-	-	•
Safety-guard on duty?	--R	-	-	-	-	-	•
Power plants in SS?	--R	-	-	-	-	-	•
Protection of energy communications?	--R	-	-	-	-	-	•
Carrying out dangerous work in SS?	--R	-	-	-	-	-	•
No danger zones?	--R	-	-	-	-	-	•
System of inner safety?	--R	-	-	-	-	-	•
No potential sources of danger?	--R	-	-	-	-	-	•
Service against the risk?	--R	-	-	-	-	-	•
Level of safety-guard on duty?	--R	-	-	-	-	-	•
Dangerous materials in SS?	--R	-	-	-	-	-	•
Kind of life-support systems in SS?	--R	-	-	-	-	-	•
Municipality - City?	--R	-	-	-	-	-	•
Temporary dangerous devices in SS?	--R	-	-	-	-	-	•
No transport. of danger materials?	--R	-	-	-	-	-	•
No substances that may be D?	--R	-	-	-	-	-	•
D-opposition of the barrier?	-T-	-	-	-	-	-	•
Protection system is working?	-T-	-	-	-	-	-	•
Primary-protection means in SS?	-T-	-	-	-	-	-	•
Primary-protection means?	-R-	-	-	-	-	-	•
Protection system?	-R-	-	-	-	-	-	•
D-resistance of the barrier?	-R-	-	-	-	-	-	•
Protect of energy-communicat. in SS?	-R-	-	-	-	-	-	•
Amount entrances in SS?	-R-	-	-	-	-	-	•
Anti dangerous obstacles?	-R-	-	-	-	-	-	•
Stairwells?	-RR	-	-	-	-	-	•
Materials of the object?	-RR	-	-	-	-	-	•
Emergency exits?	-RR	-	-	-	-	-	•
D-opposition of the subject?	T--	-	-	-	-	-	•
Personal safety?	TT-	-	-	-	-	-	•
Evacuation routes are OK?	TT-	-	-	-	-	-	•
Emergency exits are OK?	TT-	-	-	-	-	-	•
Evacuation systems are OK?	TT-	-	-	-	-	-	•
Groups of low mobility are trained?	TT-	-	-	-	-	-	•
Level of education in SS?	TT-	-	-	-	-	-	•
Level of evacuation drills in SS?	TT-	-	-	-	-	-	•
Instructions at workplace in SS?	TT-	-	-	-	-	-	•
Working with D-substances in SS?	TT-	-	-	-	-	-	•
Evac. routes are equipped with SS?	TT-	-	-	-	-	-	•
Sudden Danger is possible?	TTT	-	-	-	-	-	•
ETW training about safety?	YR-	-	-	-	-	-	•
ETW training on workplace safety?	YR-	-	-	-	-	-	•
D-resistance of the subject?	R--	-	-	-	-	-	•
Personal protection?	RR-	-	-	-	-	-	•
Evacuation routes are OK?	RR-	-	-	-	-	-	•
Emergency entrances?	RR-	-	-	-	-	-	•
Evacuation systems?	RR-	-	-	-	-	-	•
No group of low mobility?	RR-	-	-	-	-	-	•
ETW education on safety?	RR-	-	-	-	-	-	•
ETW education on workplace safety?	RR-	-	-	-	-	-	•
Evacuation training?	RR-	-	-	-	-	-	•
Production discipline is enforced?	RR-	-	-	-	-	-	•
Workplace instructions?	RR-	-	-	-	-	-	•
Plans involving emergency services?	RR-	-	-	-	-	-	•
Automatic safety is present?	RRR	-	-	-	-	-	•
Social situation in village in SS?	RRR	-	-	-	-	-	•
Automatic safety is OK?	NNN	-	-	-	-	-	•

Tab. 4: Example of filling figurant subportfolios by I-signs in the *incomplete* expert portfolio, which contains a minimal portfolio (see Tab. 2) and the minimum weighted portfolio (see Tab. 3) consists of 69 I-signs corresponding 13 (instead possible 63) terraced I-events of II-kind with weights: 12,16,3,5,4,1,10,1, 2,1,11,2,1 (top-down). In the column ★★★ the abbreviation of the subsets shown (see Remark 2 on page 112), by which the corresponding I-event of II-kind are numbered. SS — safety standards, D — dangerous; ETW — engineers and technical workers.

To match all kinds of relationships between I-signs in the portfolio can be using the expert portfolio, including the *weighted full portfolio of I-signs of safety*, that includes I-signs corresponding to all possible terraced events of II-kind generated by the corresponding *mean probable expert portfolio*.

## 6 Totals

The proposed succinct statement of the results of eventological studies on the safety turned out, though far from exhaustive, but rewarding venture, during

which opened stripped of unnecessary details, the overall design of eventological safety system that produced new ideas related to key safety procedure of applicable eventology — the preparation of optimal portfolio of I-signs. Detailed research and development of these eventological ideas will be discussed in subsequent papers.

## Acknowledgements

Thank to *Arcady Novosyolov* for reliable resistance, which allowed sharpen almost axiomatic rigor to new theoretical ideas in eventology safety, and *Sergey Amelchugov* for constant advancement of new fire safety problems and their useful discussion, which invariably led to new approaches in eventological modeling and safety control.

## References

- [1] O. Yu. Vorobyev. *Eventology*. Siberian Federal University, Krasnoyarsk, Russia, 2007, 435p. (in Russian, abstract in English), <http://eventology-theory.com/0-books/1-VorobyevOleg~2007~Eventology~435p.pdf>.
- [2] O. Yu. Vorobyev. Eventology of safety. *Proc. of the X Intern. FAMES Conf. on Financial and Actuarial Mathematics and Eventology of Safety*, Krasnoyarsk, SFU (Oleg Vorobyev ed.):92–101, 2011 (in Russian, abstract in English).
- [3] O. Yu. Vorobyev. Eventological system analysis of safety in pictures. In. *Proc. of the XI Intern. FAMES Conf. on Financial and Actuarial Mathematics and Eventology of Safety*, Krasnoyarsk, SFU (Oleg Vorobyev ed.):113–125, 2012 (in Russian, abstract in English).
- [4] O. Yu. Vorobyev. To express and to measure: successes of applicable eventology. In. *Proc. of the XVI Intern. EM conference on eventological mathematics and related fields*, Krasnoyarsk: SFU (Oleg Vorobyev ed.):44–49, 2012 (in Russian, abstract in English).
- [5] O. Yu. Vorobyev. Eventological analysis of systems: an event system under the off-system circumstances. In. *Proc. of the XI Intern. FAMES Conf. on Financial and Actuarial Mathematics and Eventology of Safety*, Krasnoyarsk, SFU (Oleg Vorobyev ed.):126–130, 2012 (in Russian, abstract in English).
- [6] O. Yu. Vorobyev. A total system and a totality of systems: eventological similarity and distinction. In. *Proc. of the XI Intern. FAMES Conf. on Financial and Actuarial Mathematics and Eventology of Safety*, Krasnoyarsk, SFU (Oleg Vorobyev ed.):131–138, 2012 (in Russian, abstract in English).
- [7] O. Yu. Vorobyev. A mean probability event for a set of events. In. *Proc. of the XI Intern. FAMES Conf. on Financial and Actuarial Mathematics and Eventology of Safety*, Krasnoyarsk, SFU (Oleg Vorobyev ed.):139–147, 2012 (in Russian, abstract in English).
- [8] O. Yu. Vorobyev. Event means in eventology, its asymptotic properties, interpretations and visualization. In. *Proc. of the XVI Intern. EM*

*conference on eventological mathematics and related fields*, Krasnoyarsk: SFU (Oleg Vorobyev ed.):50–56, 2012 (in Russian, abstract in English).

- [9] A. A. Novosyolov. Fire risk control and optimization under natural conditions and in technical objects. In. *Proc. of the XI Intern. FAMES Conf. on Financial and Actuarial Mathematics and Eventology of Safety*, Krasnoyarsk, SFU (Oleg Vorobyev ed.):285–287, 2012 (in Russian, abstract in English).
- [10] A. A. Novosyolov. Some problems of fire risk control. In. *Proc. of the XII Intern. FAMES Conf. on Financial and Actuarial Mathematics and Eventology of Safety*, Krasnoyarsk, SFU (Oleg Vorobyev ed.):259–265, 2013 (in Russian, abstract in English).
- [11] Amelchugov S. P., Andreev Yu. A., Komarov S. Yu. Assessment of forest fire risks. In. *Proc. of the X Intern. FAMES Conf. on Financial and Actuarial Mathematics and Eventology of Safety*, Krasnoyarsk, SFU (Oleg Vorobyev ed.):33–38, 2011 (in Russian, abstract in English).
- [12] Методика определения расчетных величин пожарного риска в зданиях, сооружениях и строениях различных классов функциональной пожарной опасности. *Приложение к приказу МЧС России от 30.06.2009, №382*, 2009, 14с.
- [13] Оценка международных стандартов для безопасной разведки, добычи и транспортировки нефти и газа в Баренцевом море. Гармонизация стандартов по охране труда, промышленной безопасности и охране окружающей среды для работ в Баренцевом море. *Баренц-2020. Российско-норвежский проект. Окончательный отчет*, 2009, 139с.



## Не пора ли думать об универсальных методах решения для олимпиадных задач

Рагиб Зейнал оглы Абасов

Азербайджанская государственная нефтяная Академия  
Баку  
rahib-55@mail.ru

Зохра Аббасали кызы Сафарова

Азербайджанская государственная нефтяная Академия  
Баку

**Аннотация.** В данной работе предлагается универсальный метод конструктивного характера для эффективного решения некоторых классов олимпиадных задач по теме: «Треугольники». Предлагаемый материал может быть использован на занятиях математического кружка, а также при подготовке к различным олимпиадам.

**Ключевые слова.** Треугольник, универсальный метод, конструктивный метод, олимпиадные задачи, рациональное решение.

Идея решить несколько олимпиадных задач одним и тем же способом, и притом конструктивным методом, выглядит насколько эффективным, настолько и не убедительным. Все те, которые знакомы со школьными олимпиадными задачами в той или другой степени, могут подтверждать, что олимпиадные задачи — это отдельные математические исследования в определенном масштабе. И, конечно, каждая такая задача требует найти индивидуальный подход, который связан с сущностью предлагаемой задачи. Но, если учесть, что в современной школьной математике число и разнообразие, а также уровень олимпиадных задач значительно возрос, то, все-таки стоит сделать попытку в названном направлении, если даже потерпим неудачу в такой полезной и содержательной работе.

Итак, все решено - выходим на этот путь.

Сначала отметим следующую задачу, замеченную нами случайно, которая фактически является основой для наших дальнейших рассуждений о поставленной цели [1, задача 15.9, с.9].

**Задача.** Докажите, что числа  $a, b, c$  являются длинами сторон некоторого треугольника тогда и только тогда, когда  $a = y + z$ ,  $b = x + z$ ,  $c = x + y$ , где  $x, y, z$  — положительные числа.

Решение [1, с.24]. «Решая систему уравнений

$$\begin{cases} a = y + z, \\ b = x + z, \\ c = x + y, \end{cases} \quad (1)$$

получаем

$$\begin{cases} x = (b + c - a)/2, \\ y = (a + c - b)/2, \\ z = (a + b - c)/2. \end{cases} \quad (2)$$

Положительность чисел  $x, y, z$  следует из неравенства треугольника для  $\triangle ABC$ .

Очевидно, что можно рассуждать и обратно, т.е. из (2) можно получить систему (1). Таким образом, системы (1) и (2) эквивалентны.

Итак, вместе с положительными числами  $a, b, c$  — длинами сторон треугольника  $ABC$ , введены новые величины  $x, y, z$  — тоже положительные. Легко понять, что положительные величины  $x, y, z$  совершенно произвольные, т.е. в отличие от  $a, b, c$  нет необходимости, чтобы они удовлетворяли неравенствам треугольника. Например, пусть  $y$  и  $z$  выбраны произвольно как  $y = 8$ ,  $z = 10$ . Тогда можно проверить, что третья величина  $x$  может принимать любое положительное значение как  $x < y + z$ , так и  $x = y + z$ , и даже  $x > y + z$ . Для всех этих случаев, существует реальный треугольник  $ABC$ , стороны которого вычисляются по равенствам системы (1). Конкретно рассмотрим два примера:

1.  $y = 8$ ,  $z = 10$ ,  $x = 18$  (т.е.  $x = y + z$ ). Тогда из (1) находим реальный треугольник  $ABC$ , со сторонами  $a = y + z = 18$ ,  $b = x + z = 28$ ,  $c = x + y = 26$ , которые удовлетворяют неравенствам треугольника (хотя  $x = y + z$ );
2.  $y = 8$ ,  $z = 10$ ,  $x = 20$  (т.е.  $x > y + z$ ). И в этом случае из (1) найдем числа  $a = 18$ ,  $b = 30$ ,  $c = 18$ , которые могут быть сторонами какого-то треугольника.

Итак, положительные числа  $x, y, z$  не подчиняются неравенствам треугольника.

Как можно объяснить причину этого явления? А причина ясна. Что означает, например величина  $x = (b + c - a)/2$ ? Эта величина фактически определяет насколько сумма длин двух сторон треугольника больше третьей стороны. А для этой разности нет верхней границы. Таким образом, величина  $x$  (а также и  $z$ ) может принимать любое положительное значение. И еще,  $x, y, z$  независимы друг от друга величины, поскольку нет необходимости на то, чтобы они подчинялись неравенствам треугольника.

Учитывая эти качества величин  $x, y, z$ , вносим идею: что будет, если заданную задачу по теме «Треугольники», выраженной на «языке  $a, b, c$ », перенести на «язык  $x, y, z$ »? Учитывая вышесказанное о величинах  $x, y, z$ , естественно ожидать какую-то пользу в этом переводе.

Исследование этого вопроса будем проводить по следующей последовательности.

## 1 Геометрический смысл переменных $x, y, z$

Невольно вспоминаем формулу  $r = (a + b - c)/2$ , которая напоминает нам третье равенство системы (2), где  $a$  и  $b$  – катеты,  $c$  – гипотенуза некоторого прямоугольного треугольника  $ABC (\angle C = 90^\circ)$ , а  $r$  – радиус вписанной окружности.

Пусть имеем произвольный треугольник  $ABC$ . Точки касания вписанной окружности со сторонами этого треугольника обозначим через  $M, N, P$  (рис.1).  $AO, BO$  и  $CO$  – биссектрисы соответствующих углов  $\triangle ABC$ . Обозначим  $AM = x$ , тогда  $AP = x$ ,  $BP = BN = c - x$ ,  $CM = CN = b - x$ .

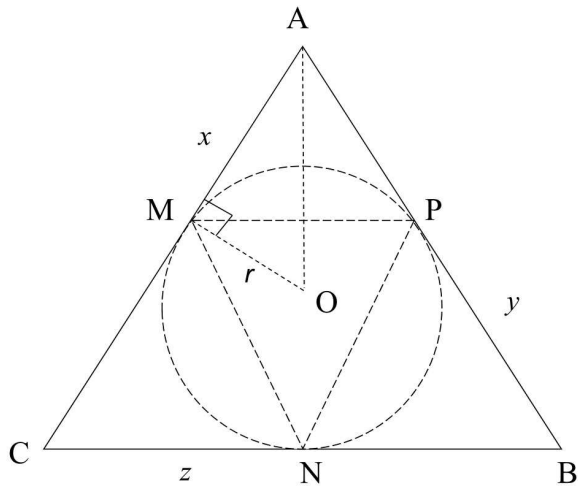


Рис. 1: Произвольный треугольник  $ABC$ .  $M, N, P$  – точки касания вписанной окружности со сторонами этого треугольника.  $AO, BO$  и  $CO$  – биссектрисы соответствующих углов  $\triangle ABC$ . Обозначим  $AM = x$ ,  $BP = y$  и  $CN = z$ .

Отсюда имеем  $2x + 2(b - x) + 2(c - x) = a + b + c$  или  $x = (b + c - a)/2$ . Таким же путем можно проверить, что  $BP = y$  и  $CN = z$  (см.(2)).

Значит величины  $x, y, z$  есть некоторые части сторон  $a, b, c$   $\triangle ABC$ . То, что они определяются с вписанной окружностью, все это обещает полезную работу с этими переменными. Из рис. 1 легко найти формулы

$$\operatorname{tg} \frac{A}{2} = \frac{r}{x}, \operatorname{tg} \frac{B}{2} = \frac{r}{y}, \operatorname{tg} \frac{C}{2} = \frac{r}{z}.$$

С помощью этих формул можно легко и эффективно

доказывать очень много тригонометрических соотношений разного характера.

## 2 Основные формулы на языке переменных $x, y, z$

Здесь будем принимать привычные обозначения для треугольника  $\triangle ABC$ :  $a, b, c$  – длины сторон,  $p$  – полупериметр,  $S$  – площадь,  $h_a, h_b, h_c$  – соответствующие высоты,  $m_a, m_b, m_c$  – соответствующие медианы, а  $l_a, l_b, l_c$  – биссектрисы.  $R$  и  $r$  – радиусы описанной и вписанной окружностей, соответственно, а  $r_a, r_b, r_c$  – радиусы внеписанной окружности соответственно.

Разными способами можно доказать следующие формулы, которые являются рабочими формулами предложенной методики.

$$1. \quad p = \frac{1}{2}(a + b + c) = x + y + z.$$

$$2. \quad S = \sqrt{xyzp} \quad (S = xy, \angle C = 90^\circ).$$

$$3. \quad r = \frac{S}{p} = \sqrt{\frac{xyz}{p}}, \quad (r = z, \angle C = 90^\circ).$$

$$4. \quad R = \frac{abc}{4S}, \quad (R = \frac{c}{2}, \angle C = 90^\circ).$$

$$5. \quad \begin{cases} \cos A = 1 - \frac{2yz}{bc}, \\ \cos B = 1 - \frac{2xz}{ac}, \\ \cos C = 1 - \frac{2xy}{ab}. \end{cases}$$

$$6. \quad \begin{cases} \sin A = \frac{2\sqrt{xyzp}}{bc}, \\ \sin B = \frac{2\sqrt{xyzp}}{ac}, \\ \sin C = \frac{2\sqrt{xyzp}}{ab}. \end{cases}$$

$$7. \quad \begin{cases} \cos A + \cos B = \frac{2xy(a+b)}{abc} = \frac{2xy(z+p)}{abc}, \\ \cos A - \cos B = \frac{2zp(x-y)}{abc}, \\ \sin A - \sin B = \frac{2S(y-x)}{abc}, \\ \sin A + \sin B = \frac{2S(a+b)}{abc} = \frac{2S(z+p)}{abc}. \end{cases}$$

$$8. \begin{cases} \cos A \cdot \cos B = 1 - \frac{2zp(x^2 + y^2)}{abc^2}, \\ \sin A \cdot \sin B = \frac{4xyzp}{abc^2}, \\ \cos A \cdot \sin B = \frac{2\sqrt{xyzp}(bc - 2yz)}{abc^2}, \\ \sin A \cdot \cos B = \frac{2\sqrt{xyzp}(ac - 2zx)}{abc^2}. \end{cases}$$

$$9. \begin{cases} \cos(A + B) = 1 - \frac{2zp}{ab} = \frac{xy - zp}{ab}, \\ \sin(A + B) = \frac{2S}{ab} = \frac{2\sqrt{xyzp}}{ab}, \\ \cos(A - B) = 1 - \frac{2zp(x - y)^2}{abc^2}, \\ \sin(A - B) = \frac{2S(y - x)(z + p)}{abc^2}. \end{cases}$$

$$10. \begin{cases} \cos 2A = \frac{x^2 - zp}{a^2} = \frac{y^2 - zp}{b^2}, \\ \sin 2A = \frac{2x\sqrt{zp}}{a^2} = \frac{2y\sqrt{zp}}{b^2}. \end{cases}$$

$$11. \begin{cases} h_a = 2\sqrt{xyzp}/a, \\ h_b = 2\sqrt{xyzp}/b, \\ h_c = 2\sqrt{xyzp}/c. \end{cases}$$

$$12. \begin{cases} m_a = \frac{1}{2}\sqrt{(y - z)^2 + 4xp}, \\ m_b = \frac{1}{2}\sqrt{(z - x)^2 + 4yp}, \\ m_c = \frac{1}{2}\sqrt{(x - y)^2 + 4zp}. \end{cases}$$

$$13. \begin{cases} l_a = \frac{2\sqrt{bcxp}}{x + p}, \\ l_b = \frac{2\sqrt{acyp}}{y + p}, \\ l_c = \frac{2\sqrt{abzp}}{z + p}. \end{cases}$$

$$\begin{cases} l_a = \frac{2y\sqrt{x^2 + z^2}}{a}, \\ l_b = \frac{2x\sqrt{y^2 + z^2}}{b}, C = 90^\circ, \\ l_c = \frac{2\sqrt{2}xyz}{xy + z^2}. \end{cases}$$

$$14. \begin{cases} \operatorname{tg} \frac{A}{2} = \frac{r}{x}, \\ \operatorname{tg} \frac{B}{2} = \frac{r}{y}, \\ \operatorname{tg} \frac{C}{2} = \frac{r}{z}. \end{cases}$$

$$15. \begin{cases} \sin \frac{A}{2} = \sqrt{\frac{yz}{bc}}, \\ \sin \frac{B}{2} = \sqrt{\frac{xz}{ac}}, \\ \sin \frac{C}{2} = \sqrt{\frac{xy}{ab}}. \end{cases}$$

$$16. \begin{cases} \cos \frac{A}{2} = \sqrt{\frac{xp}{bc}}, \\ \cos \frac{B}{2} = \sqrt{\frac{yp}{ac}}, \\ \cos \frac{C}{2} = \sqrt{\frac{zp}{ab}}. \end{cases}$$

$$17. \begin{cases} r_a = p \operatorname{tg} \frac{A}{2} = \frac{pr}{x}, \\ r_b = p \operatorname{tg} \frac{B}{2} = \frac{pr}{y}, \\ r_c = p \operatorname{tg} \frac{C}{2} = \frac{pr}{z}. \end{cases}$$

Некоторые олимпиадные задачи связаны с треугольником  $MNP$ . С применением теоремы косинусов легко проверить, что

$$\begin{cases} MP = 2x\sqrt{\frac{yz}{bc}} = 2x \sin \frac{A}{2}, \\ PN = 2y\sqrt{\frac{xz}{ac}} = 2y \sin \frac{B}{2}, \\ MN = 2z\sqrt{\frac{xy}{ab}} = 2z \sin \frac{C}{2}. \end{cases}$$

С помощью этих формул легко доказать, что  $\triangle MNP$  всегда является остроугольным треугольником. Для этого, например, неравенство  $MP^2 + PN^2 > MN^2$  легко приводится к неравенству  $xy > 0$  (отметим, что эта задача была предложена на XV Московской олимпиаде в 1952 г. [2, задача 1, с.40]).

А некоторые олимпиадные задачи связаны с описанной окружностью, где  $AK = l_a, BP = l_b, CF = l_c$  – биссектрисы  $\triangle ABC$  (рис.2).

Легко проверить, что

$$AP = bc/(a + c), \quad KC = ab/(b + c),$$

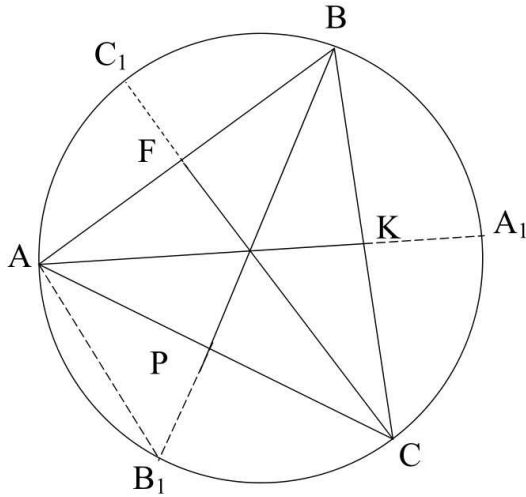


Рис. 2: Некоторые олимпиадные задачи связаны с описанной окружностью, где  $AK, BP, CF$  – биссектрисы  $\triangle ABC$ .

$$PC = ab/(a+c), \quad AF = bc/(a+b), \\ BK = ac/(b+c), \quad BF = ac/(a+b),$$

а также

$$\left\{ \begin{array}{l} PB_1 = \frac{b^2 \sqrt{ac}}{2(a+c)\sqrt{yp}}, \quad BB_1 = \frac{\sqrt{ac}(a+c)}{2\sqrt{yp}}, \\ FC_1 = \frac{c^2 \sqrt{ab}}{2(a+b)\sqrt{zp}}, \quad CC_1 = \frac{\sqrt{ab}(a+b)}{2\sqrt{zp}}, \\ KA_1 = \frac{a^2 \sqrt{bc}}{2(b+c)\sqrt{xp}}, \quad AA_1 = \frac{\sqrt{bc}(b+c)}{2\sqrt{xp}}. \end{array} \right.$$

### 3 Некоторые применения указанной методики

**Задача 1.** [3, задача 17, с.192]. Докажите, что если  $a, b, c$  – длины сторон произвольного треугольника, то

$$(b+c-a) \cdot (a+c-b) \cdot (a+b-c) \leq abc. \quad (3)$$

**Решение** [3, с.332]. «Ясно, что

$$\begin{aligned} a^2 &\geq a^2 - (b-c)^2 = (a+b-c)(a-b+c), \\ b^2 &\geq b^2 - (a-c)^2 = (b+a-c)(b-a+c), \\ c^2 &\geq c^2 - (a-b)^2 = (c+a-b)(c-a+b), \end{aligned}$$

перемножая эти неравенства, сразу получим (3).»

Как видно, здесь учитываются неравенства треугольника и поэтому, нелегко угадать такой способ решения.

Если перейдем на язык  $x, y, z$ , то неравенство (3) сразу принимает вид

$$(x+y)(y+z)(x+z) \geq 8xyz. \quad (4)$$

А это неравенство доказывается очень просто. Для этого достаточно перемножить известные неравенства как

$$\begin{cases} x+y \geq 2\sqrt{xy}, \\ x+z \geq 2\sqrt{xz}, \\ y+z \geq 2\sqrt{yz}. \end{cases} \quad (5)$$

Из систем (1) и (2) ясно, что неравенства (3) и (4) эквивалентны. То есть, достаточно доказать одно из этих двух неравенств. Но, несмотря на это, эти неравенства в [3] доказываются отдельно, разными способами (с.332, 333).

Неравенство (3) в [1] доказывается еще двумя разными способами (задача 15.15, с.25). Один из этих способов выглядит так:

«Так как  $S = \frac{1}{2}ab \sin \gamma$  и  $\sin \gamma = \frac{c}{2R}$ , то  $abc = 4SR$ . По формуле Герона

$$(a+b-c)(a+c-b)(b+c-a) = \frac{8S^2}{p}.$$

Поэтому нужно доказать, что

$$\frac{8S^2}{p} \leq 4SR,$$

т.е.  $2S \leq pR$ . Так как  $S = pr$ , приходим к неравенству  $2r \leq R$ .»

А это неравенство в [1] доказывается совершенно другим способом как независимая задача.

Сделав небольшой анализ этих методов доказательства, приходим к выводу, что переход на переменные  $x, y, z$  в доказательстве неравенства (3) является более целесообразным.

**Задача 2.** Доказать неравенство

$$\begin{aligned} a+b+c &\geq \sqrt{(b+c-a)(a+c-b)} + \\ &+ \sqrt{(a+b-c)(b+c-a)} + \\ &+ \sqrt{(a+c-b)(a+b-c)}, \end{aligned} \quad (6)$$

где  $a, b, c$  – длины сторон произвольного треугольника.

**Решение.** Неравенство достаточно сложное и поэтому найти нужный способ решения будет непросто. Переходим на переменные  $x, y, z$  и сразу получим неравенство  $x+y+z \geq \sqrt{xy} + \sqrt{zx} + \sqrt{zy}$ , которое эквивалентно с неравенством (6). А для его доказательства достаточно сложить неравенства (5).

**Задача 3.** Доказать неравенство [1, задача 15.12, с.9]

$$\begin{aligned} a^3 + b^3 + c^3 &< a(b-c)^2 + \\ &+ b(c-a)^2 + c(a-b)^2 + 4abc, \end{aligned} \quad (7)$$

где  $a, b, c$  – длины сторон произвольного треугольника.

**Решение** [1, с.24]. “Так как  $c(a-b)^2 + 4abc = c(a+b)^2$ , то

$$\begin{aligned} a(b-c)^2 + b(c-a)^2 + c(a-b)^2 + 4abc - a^3 - b^3 - c^3 &= \\ = a[(b-c)^2 - a^2] + b[(c-a)^2 - b^2] + c[(a+b)^2 - c^2] &= \\ = (a+b-c)(a-b+c)(b+c-a) > 0. \end{aligned}$$

Последнее равенство проверяется простым вычислением. Все три сомножителя являются положительными, в силу неравенства треугольника.” Вот и все, задача решена?!

Как видно, предложенный метод нельзя считать убедительным и конструктивным. Он, скорее всего, выглядит случайным, если учесть особенно последнее равенство.

Эта задача была предложена в 1972 г. в Венгрии на олимпиаде по математике.

В [4, с.73] предлагается другой способ для доказательства неравенства (7). Но, для этого сначала проверяется истинность равенства (!)

$$\cos A + \cos B + \cos C - 1 = 4 \sin \frac{A}{2} \cdot \sin \frac{B}{2} \cdot \sin \frac{C}{2},$$

так как все последующие рассуждения основаны на этом равенстве.

Если мы с самого начала перешли бы на переменные  $x, y, z$  в (7), то после несложных вычислений получили бы  $xyz > 0$ .

Ну, что можно сказать о таких способах решения олимпиадных задач? Как видно, такие «методы» не имеют полезную научную основу. Они, скорее всего, имеют случайный характер и трудно поверить, что такие подходы будут вдохновлять ученика к олимпиадным задачам. А как показал наш анализ, олимпиадных задач с такими необоснованными решениями достаточно много.

Правда и то, что олимпиадные задачи, кроме научных рассуждений, любят поиск и интуицию. Но, когда ученик подходит к олимпиадным задачам, он обязательно должен иметь хотя бы минимум элементарный шанс для решения этой задачи. Иначе, над чем, и как думать школьнику?

Учитывая возможности данной статьи, коротко обсудим еще несколько таких олимпиадных задач.

**Задача 4.** (VII Международная олимпиада, 1964 г. Москва). Обозначим через  $a, b, c$  длины сторон некоторого треугольника. Доказать, что

$$\begin{aligned} a^2(b+c-a) + b^2(c+a-b) + \\ + c^2(a+b-c) \leq 3abc. \end{aligned} \quad (8)$$

**Решение** [5, с.95]. “ $(a-b)^2 \geq 0$ ,  $(b-c)^2 \geq 0$ ,  $(c-a)^2 \geq 0$  при любых  $a, b, c$ . Из того, что  $a, b, c$  — стороны треугольника, следует, что  $(b+c-a) > 0$ ,  $(c+$

$a-b) > 0$ ,  $(a+b-c) > 0$ . Значит,  $(b-c)^2(b+c-a) \geq 0$ ,  $(c-a)^2(c+a-b) \geq 0$ ,  $(a-b)^2(a+b-c) \geq 0$ . Сложив почленно эти неравенства, можно получить (8).”

Опять тот же самый вопрос: как можно додуматься до такого решения?

Если перейдем на  $x, y, z$ , то легко получим

$$\frac{x}{z} + \frac{z}{x} + \frac{x}{y} + \frac{y}{x} + \frac{z}{y} + \frac{y}{z} \geq 6.$$

И, наконец

**Задача 5.** (XIV Московская олимпиада, 1982).  $a, b, c$  — длины сторон треугольника. Доказать, что

$$\begin{aligned} a^4 + b^4 + c^4 - 2(a^2b^2 + a^2c^2 + b^2c^2) + \\ + a^2bc + b^2ac + c^2ab \geq 0. \end{aligned} \quad (9)$$

**Решение** [2, с.271]. «Обозначим левую часть неравенства (9) через  $f(a, b, c)$ , заметим, что (!)

$$f(a, b, c) = (a+b+c)[abc - (b+c-a)(a+c-b)(b+a-c)].$$

... » и дальше некоторое продолжение этого решения.

Думаем, что всякие комментарии здесь излишни.

При переходе на  $x, y, z$  можно получить

$$\begin{aligned} f(a, b, c) = \\ = 4 [(xy - xz)^2 + (xy - yz)^2 + (xz - yz)^2] \geq 0. \end{aligned}$$

Невольно напрашивается вопрос, сколько таких олимпиадных задач и сколько таких «случайных» решений. Ответим сразу: их очень много. Поэтому, мы предлагаем уважаемым читателям, ознакомиться с книгой [6] автора этих строк. Книга состоит из 22 параграфов. Там собраны около 250 таких задач. Задачи по характеру разные и тем самым раскрыты все возможности перехода на переменные  $x, y, z$ .

Действительно, не пора ли нам подумать об универсальных методах решения олимпиадных задач?

## Список литературы

- [1] В.В. Прасолов. *Задачи по планиметрии. Ч II*. М.: “Наука”, 1986.
- [2] Г. А. Гальперин, А. К. Толпыго. *Московские математические олимпиады*. М.: “Просвещение”, 1986.
- [3] С.Е. Ляпин, И.В. Баранова, З.Г. Борчугова. *Сборник задач по элементарной алгебре*. М.: “Просвещение”, 1973.
- [4] И. Кюршак, Д. Нейкомм, Д. Хайош, Я. Шурани. *Венгерские математические олимпиады*. М.: “Мир”, 1976.
- [5] Е.А. Морозова, И.С. Петраков, В.А. Скворцов. *Международные математические олимпиады*. М.: “Просвещение”, 1976.
- [6] Р. З. Абасов. *Рациональные решения олимпиадных задач по теме “Треугольники” (на азерб. языке)*. Баку: изд.-во АГНА, 1998.



## О некоторых свойствах одного класса эмпирических процессов

Абдушукуров Абдурахим Ахмедович

Национальный Университет Узбекистана

Ташкент

a\_abdushukurov@rambler.ru

Какаджанова Лейла Рэшитовна

Национальный Университет Узбекистана

Ташкент

egoza13@bk.ru

**Аннотация.** В данной работе обсуждаются асимптотические свойства обобщенных эмпирических процессов по определенным классам событий и интегрируемых функций, а также рассмотрена слабая сходимость к гауссовскому процессу.

**Ключевые слова.** эмпирические процессы, классы Гливленко-Кантелли, метрическая энтропия.

Данная работа является продолжением [1]. Рассмотрим последовательность статистических экспериментов, в которой наблюдаются пары  $\{(X_k, A_k), k \geq 1\}$ , где  $\{X_k, k \geq 1\}$  — независимые и одинаково распределенные случайные величины (с.в.), определенные на вероятностном пространстве  $(\Omega, \mathcal{A}, P)$  и со значениями в измеримом пространстве  $(\mathcal{X}, \mathcal{B})$  с распределением  $Q(\cdot)$  на  $\mathcal{B}$ . Каждая с.в.  $X_k$  сопровождается событием  $A_k$  с вероятностью  $p = P(A_k)$ . Интерес представляют совместные свойства пар  $(X_k, A_k)$ . Пусть за  $n$  шагов наблюдается выборка  $S^{(n)} = \{(X_k, \delta_k), k = \overline{1, n}\}$ , где  $\delta_k = I(A_k)$  — индикатор события  $A_k$ . Пусть  $Q_n(\cdot)$  и  $p_n$  — эмпирические оценки  $Q(\cdot)$  и  $p$ , построенные по выборке  $S^{(n)}$ :

$$Q_n(B) = \frac{1}{n} \sum_{k=1}^n I(X_k \in B), B \in \mathcal{B},$$

$$p_n = \frac{1}{n} \sum_{k=1}^n \delta_k.$$

Введем также и субмеру  $\tilde{Q}(B) = P(X_k \in B, \delta_k = 1)$  и ее эмпирический аналог

$$\tilde{Q}_n(B) = \frac{1}{n} \sum_{k=1}^n \delta_k I(X_k \in B), B \in \mathcal{B}.$$

Легко видеть, что  $p = \tilde{Q}(\mathcal{X})$  и  $p_n = \tilde{Q}_n(\mathcal{X})$ . Далее, пусть  $0 < p < 1$ . Введем эмпирические меры, как случайные процессы на  $\mathcal{B}$ :

$$\mathbb{A}_n(B) = n^{1/2} (Q_n(B) - Q(B)),$$

$$\tilde{\mathbb{A}}_n(B) = n^{1/2} (\tilde{Q}_n(B) - \tilde{Q}(B)),$$

$$\mu_n = n^{1/2} (p_n - p) = \tilde{\mathbb{A}}_n(\mathcal{X}).$$

Одним из важнейших свойств пар  $(X_k, A_k)$  является независимость  $X_k$  и  $A_k$ . Пусть  $\lambda(B) = \tilde{Q}(B) - pQ(B)$ ,  $B \in \mathcal{B}$ . Тогда независимость  $X_k$  и  $A_k$  определяется равенством  $\lambda(B) = 0$  для  $B \in \mathcal{B}$ . Пусть  $\lambda_n(B) = \tilde{Q}_n(B) - p_n Q_n(B)$  — эмпирический аналог  $\lambda(B)$ . Через  $H_o$  обозначим гипотезу о независимости  $X_k$  и  $A_k$  для всех  $k \geq 1$ . Тогда при справедливости гипотезы  $H_o$  при каждом  $B \in \mathcal{B}$ ,  $M\lambda(B) = 0$  и согласно усиленному закону больших чисел,  $\lambda_n(B) \xrightarrow{п.н.} 0$  при  $n \rightarrow \infty$ . Таким образом, эмпирический аналог для  $\lambda_n(B)$  будучи желаемой оценкой для  $\lambda(B)$  может быть полезной для построения состоятельных критериев для проверки  $H_o$ . Рассмотрим нормированный процесс

$$\Delta_n(B) = \left( \frac{n}{p_n(1-p_n)} \right)^{1/2} \cdot \lambda_n(B), B \in \mathcal{B}.$$

Исследуем свойства процессов  $\lambda_n(\cdot)$  и  $\Delta_n(\cdot)$ . Для измеримой на  $\mathcal{X}$  функции  $f$  и знакопеременной меры  $G$ , пусть  $Gf = \int f dG$ . Выделив некоторый класс  $\chi \subset \mathcal{B}$  и множество  $\mathcal{F}$  функций  $f$ , определим величины

$$\mathbb{D}_n(\lambda_n; \chi) = \|\lambda_n\|_\chi = \sup_{B \in \chi} |\lambda_n(B)|, \quad (1)$$

$$\mathbb{D}_n(\lambda_n; \mathcal{F}) = \|\lambda_n\|_\mathcal{F} = \sup_{f \in \mathcal{F}} |\lambda_n f|. \quad (2)$$

Рассмотрим также и нормированный  $f$  — процесс

$$\Delta_n f = \int_{\mathcal{X}} f d\Delta_n = \left( \frac{n}{p_n(1-p_n)} \right)^{1/2} \cdot \lambda_n f.$$

Пусть  $\tilde{\tilde{Q}}(B) = Q(B) - \tilde{Q}(B)$ ,  $\tilde{\tilde{Q}}_n(B) = Q_n(B) - \tilde{Q}_n(B)$  — субмеры, соответствующие событиям  $\delta_k = 0$ . Важным специальным случаем  $\mathcal{F}$  является множество индикаторов некоторого класса множеств  $\chi \subset \mathcal{B}$ :  $\mathcal{F} = \{I(B) : B \in \chi \subset \mathcal{B}\}$ . При  $\mathcal{X} = \mathbb{R}^m$  и  $\chi = \{(-\infty; x] : x \in \mathbb{R}^m\}$ ,  $H(x) = Q((-\infty; x])$  и  $H_n(x) = Q_n((-\infty; x])$  — функция распределения (ф.р.) и эмпирическая ф.р.  $m$ - мерного вектора  $X_k$ , а также  $\mathbb{A}_n((-\infty; x])$  — соответствующий эмпирический процесс, для которого имеет место классическая теорема Гливленко-Кантелли

$$\sup_{x \in \mathbb{R}^m} \left| n^{-1/2} \mathbb{A}_n((-\infty; x]) \right| \xrightarrow[n \rightarrow \infty]{п.н.} 0. \quad (3)$$

Для доказательства аналогичных результатов для величин (1) и (2) рассмотрим обобщения теорем

Гливенко-Кантелли для величин

$$\left\| \tilde{Q}_n - \tilde{Q} \right\|_{\chi}, \left\| \tilde{Q}_n - \tilde{Q} \right\|_{\chi}, \left\| \tilde{Q}_n - \tilde{Q} \right\|_{\mathcal{F}}, \left\| \tilde{Q}_n - \tilde{Q} \right\|_{\mathcal{F}}. \quad (4)$$

В литературе [2, 3, 4] имеются различные обобщенные аналоги теоремы Гливенко-Кантелли для специальных классов  $\chi$  и множеств  $\mathcal{F}$ . Используем подход Вапника-Червоненкиса, который основывается на выборе класса  $\chi$  (класса Вапника-Червоненкиса или ВЧ-класса) для обеспечения сходимости к нулю величины (1). Для формулировки соответствующего результата введем некоторые определения. Поскольку элементы выборки  $S^{(n)}$  состоят из пар  $(X_k, \delta_k)$  со значениями в пространстве

$$\mathcal{X} \times \{0; 1\} = (\mathcal{X} \times \{0\}) \cup (\mathcal{X} \times \{1\}) = \mathcal{X}_0 \cup \mathcal{X}_1,$$

то представим класс  $\chi$  как  $\chi = \chi_0 \cup \chi_1$ , где  $\chi_m \subset \mathcal{B}_m = \sigma(\mathcal{X}_m)$ ,  $m = 0, 1$ ,  $\mathcal{X}_0 \cap \mathcal{X}_1 = \emptyset$  и  $\chi_0 \cap \chi_1 = \emptyset$ . Пусть  $v_m = v(\chi_m)$  индексы Вапника-Червоненкиса классов  $\chi_m$ ,  $m = 0, 1$ , (определение ВЧ-индекса см. [?, 4]) и  $v = \max(v_1, v_2)$ .

**Теорема 1.** Пусть  $v < \infty$  и  $\left\| \tilde{Q}_n - \tilde{Q} \right\|_{\chi}, \left\| \tilde{Q}_n - \tilde{Q} \right\|_{\chi}$  - измеримы. Тогда при справедливости гипотезы  $H_0$ ,  $\mathbb{D}_n(\lambda_n; \chi) \xrightarrow{\text{п.н.}} 0$  при  $n \rightarrow \infty$ .

Теперь рассмотрим величину (2). Теоремы Гливенко-Кантелли подсказывают условия, при которых последние две величины (4) стремятся к нулю либо по вероятности (в этом случае говорят, что  $\mathcal{F}$  является слабым Гливенко-Кантелли классом для  $\lambda_n$  или же почти наверное (в этом случае говорят, что  $\mathcal{F}$  является сильным Гливенко-Кантелли классом для  $\lambda_n$ ). Полезное достаточное условие для класса  $\mathcal{F}$ , чтобы он был сильным Гливенко-Кантелли классом, является требование, чтобы выполнялись следующие условия для метрических энтропий класса  $\mathcal{F}$  в  $\mathcal{L}_q(\tilde{Q})$  и  $\mathcal{L}_q(\tilde{Q})$  для всех  $\varepsilon > 0$ :

$$\begin{aligned} N_{[]}(\varepsilon, \mathcal{F}, \mathcal{L}_q(\tilde{Q})) &< \infty, \\ N_{[]}(\varepsilon, \mathcal{F}, \mathcal{L}_q(\tilde{Q})) &< \infty, \end{aligned} \quad (5)$$

где для  $0 < q \leq \infty$

$$\begin{aligned} N_{[]}(\varepsilon, \mathbb{F}, \mathcal{L}_q(Q)) &= \\ &= \min \left\{ k : \text{для некоторых } f_1, \dots, f_k \in \mathcal{L}_q(Q), \right. \\ &\quad \left. \min_{1 \leq i \leq k} \|f - f_i\|_q \leq \varepsilon, \text{ для всех } f \in \mathcal{F}, \right\} \end{aligned}$$

и  $\mathcal{L}_q(Q)$  — пространство с нормой  $\|f\|_q = \left\{ \int |f|^q dQ \right\}^{1/q}$ .

В следующей теореме установлено, что класс  $\mathcal{F}$  является сильным Гливенко-Кантелли классом для меры  $\lambda_n$ .

**Теорема 2.** Пусть для каждого класса функций  $\mathcal{F}$  выполнены условия (5) при  $q = 1$  и каждом  $\varepsilon > 0$ . Тогда при справедливости гипотезы  $H_0$ ,  $\mathbb{D}_n(\lambda_n; \mathcal{F}) \xrightarrow{\text{п.н.}} 0$ , при  $n \rightarrow \infty$ .

Пусть  $\{\Delta_n f, f \in \mathcal{F}\}$  — последовательность  $f$  - процессов, получаемых из  $\Delta_n(B)$  заменой  $\lambda_n(B)$  на  $\lambda_n f$ . Очевидно, в  $\Delta_n f$  содержится и  $\Delta_n(B)$ . Теперь сформулируем основное утверждение о сходимости процесса  $\Delta_n f$  к  $Q$  - броуновскому мосту. Определим характеристики (интеграл-скобки):

$$\begin{aligned} \tilde{J}_{[]}(\delta) &= J_{[]}(\delta; \mathcal{F}; \mathcal{L}_2(\tilde{Q})) = \\ &= \int_0^\delta \sqrt{\log N_{[]}(\varepsilon, \mathcal{F}, \mathcal{L}_2(\tilde{Q}))} d\varepsilon, \\ \tilde{\tilde{J}}_{[]}(\delta) &= J_{[]}(\delta; \mathcal{F}; \mathcal{L}_2(\tilde{\tilde{Q}})) = \\ &= \int_0^\delta \sqrt{\log N_{[]}(\varepsilon, \mathcal{F}, \mathcal{L}_2(\tilde{\tilde{Q}}))} d\varepsilon \end{aligned}$$

В следующей теореме утверждается  $Q$  - Донскеровость класса  $\mathcal{F}$  для меры  $\lambda_n$ .

**Теорема 3.** Пусть для каждого класса  $\mathcal{F}$  измеримых функций,  $\tilde{J}_{[]} (1) < \infty$  и  $\tilde{\tilde{J}}_{[]} (1) < \infty$ . Тогда при справедливости гипотезы  $H_0$ ,  $f$  - процесс  $\{\Delta_n f, f \in \mathcal{F}\}$  слабо сходится к  $Q$  - броуновскому мосту, т.е. гауссовскому процессу с нулевым средним и ковариационной функцией  $Q(f_1 f_2) - Q(f_1) Q(f_2)$ .

Этот результат может быть использован для построения статистических критериев для проверки гипотезы  $H_0$ .

## Список литературы

- [1] Абдушукуров А.А., Какаджанова Л.Р. . Об одном классе эмпирических процессов. *Труды XVII Международной конференции по эвентологической математике и смежным вопросам*, Красноярск:300–31, 2012.
- [2] Gaensler P., Stute W. Empirical processes: a survey of results for independent and identically distributed random variables. *Ann. Probab.*, 7, №2:193–243, 1979.
- [3] Wellner J.A. Shorack G.R. *Empirical processes with applications to statistics*. JOHN WILEY & SONS, 1986.
- [4] Van der Vaart A. W., Wellner J.A. . *Weak convergence and empirical processes*. Springer, 1996.

## Оценивание производных функции распределения в модели пропорциональных интенсивностей

Абдушукуров Абдурахим Ахмедович

Национальный Университет Узбекистана  
Ташкент  
a\_abdushukurov@rambler.ru

Сагидуллаев Калмурза Сапарбаевич

Национальный Университет Узбекистана  
Ташкент  
khalmurza@mail.ru

**Аннотация.** В работе вводятся полупараметрические ядерные оценки производных функции распределения в модели пропорциональных интенсивностей случайного цензурирования справа. Доказано утверждение о строгой состоятельности оценок.

**Ключевые слова.** случайное цензурирование, пропорциональные интенсивности, ядерные оценки.

### 1 Введение

Пусть  $X_1, X_2, \dots, X_n$  — независимые и одинаково распределённые (н.о.р.) случайные величины (с.в.) с функцией распределения (ф.р.)  $F$ , у которой существуют непрерывные и ограниченные производные

$$F^{(i)}(t) = \frac{d^i F(t)}{dt^i},$$

$i = 1, 2, \dots, r + 1$ . Через  $f$  обозначим плотность ф.р.  $F$  ( $f = F^{(1)}$ ). Предположим, что с.в.  $X_1, X_2, \dots, X_n$  цензурятся справа соответственно с.в.  $Y_1, Y_2, \dots, Y_n$  с общей непрерывной ф.р.  $G$  и наблюдается выборка

$$C^{(n)} = \{(Z_j, \delta_j), 1 \leq j \leq n\},$$

где  $Z_j = \min(X_j, Y_j)$ ,  $\delta_j = I(Z_j = X_j)$  и  $I(A)$  — индикатор события  $A$ . Предположим, что совокупности  $\{X_j\}$  и  $\{Y_j\}$  взаимонезависимы. Легко видеть, что

$$1 - H(t) = P(Z_j > t) = (1 - F(t))(1 - G(t)), t \in R.$$

Рассмотрим задачу оценивания производных  $F^{(i)}$  по выборке  $C^{(n)}$ , в которой представляющие интерес с.в.  $X_j$  наблюдаемы лишь при  $\delta_j = 1$ . Эту задачу мы рассмотрим в специальном случае информативного цензурирования, когда пара  $(F, G)$  отвечает модели пропорциональных интенсивностей (МПИ). В МПИ предполагается, что существует положительное число  $\beta$  такое, что

$$1 - G(t) = (1 - F(t))^\beta, t \in R. \quad (1)$$

Равенство (1) эквивалентно пропорциональности интегральных функций интенсивностей

$$\Lambda_G(t) = \beta \Lambda_F(t), t \in R, \quad (2)$$

где  $\Lambda_G = -\log(1 - G)$  и  $\Lambda_F = -\log(1 - F)$ . Из (1) также следует, что

$$1 - F(t) = (1 - H(t))^p, t \in R, \quad (3)$$

где

$$p = \frac{1}{1 + \beta} = P(Z_j = X_j) = M\delta_j.$$

Одним из важных свойств МПИ является то, что (1) эквивалентно независимости с.в.  $Z_j$  и  $\delta_j$ . Эта характеристика МПИ играет ключевую роль при исследовании свойств статистик, построенных по выборке  $C^{(n)}$ . Согласно (3) естественной полупараметрической оценкой  $F$  является степенная ACL - оценка (Abdushukurov-Cheng-Lin см.[1]):

$$\begin{aligned} F_n(t) &= 1 - (1 - H_n(t))^{p_n} = \\ &= 1 - \left\{ \prod_{j=1}^n \left( 1 - \frac{I(Z_j \leq t)}{n - j + 1} \right) \right\}^{p_n} = \\ &= \frac{1}{n} \sum_{j=1}^n \omega_{nj} I(Z_{nj} \leq t), \end{aligned} \quad (4)$$

где

$$n^{-1} \omega_{nj} = \left\{ \left( 1 - \frac{j-1}{n} \right)^{p_n} - \left( 1 - \frac{j}{n} \right)^{p_n} \right\}$$

— скачки оценки (4) в точках  $Z_{nj}$ ,  $j = \overline{1, n}$ ;  $Z_{1n} \leq \dots \leq Z_{nn}$  — вариационный ряд, построенный по наблюдениям  $Z_1, \dots, Z_n$ . При этом  $p_n$  и  $H_n(t)$  — эмпирические оценки  $p$  и  $H(t)$ :

$$p_n = \frac{1}{n} \sum_{j=1}^n \delta_j,$$

$$H_n(t) = \frac{1}{n} \sum_{j=1}^n I(Z_j \leq t).$$

Следует отметить, что ACL-оценка (4) является оптимальной оценкой в МПИ по сравнению с остальными известными оценками для  $F$  (см.[1]). Согласно (3) для  $F$  также имеем представление

$$F(t) = p \int_{-\infty}^t (1 - H(s))^{p-1} dH(s), t \in R. \quad (5)$$

Отсюда для  $F$  можно построить также и оценку

$$\begin{aligned} F_n^0(t) &= p_n \int_{-\infty}^t (1 - H_n(s))^{p_n-1} dH_n(s) = \\ &= \frac{1}{n} \sum_{j=1}^n \omega_{nj}^0 I(Z_{nj} \leq t), \end{aligned} \quad (6)$$

— со скачками величины

$$n^{-1} \omega_{nj}^0 = n^{-1} p_n (1 - H_n(Z_{nj}))^{p_n-1}$$

в каждой точке  $Z_{nj}$ ,  $j = \overline{1, n}$ . Легко заметить, что при достаточно больших  $n$ :  $\omega_{nj} \approx \omega_{nj}^0$ ,  $j = 1, \dots, n$ , следовательно оценки (4) и (6) являются асимптотически эквивалентными. Поэтому для оценивания производных  $F^{(i)}$  используем оценку (6), так как она более подходит для этих целей. Пусть  $K(t)$  — абсолютно непрерывная ф.р. (ядро) с симметричной и непрерывно-дифференцируемой плотностью  $k(t)$ . Выберем последовательность "ширины окна"  $\{\mu = \mu(n), n \geq 1\}$  так, чтобы  $\mu \downarrow 0$  и  $n\mu \rightarrow \infty$  при  $n \rightarrow \infty$ . Обозначим

$$K^i(t) = \frac{d^i K(t)}{dt^i}, \quad i = 0, 1, \dots, r+1,$$

где  $K^{(0)} = K$ ,  $K^{(1)} = k$  и  $K^{(i)} = k^{(i-1)}$  при  $i = 1, \dots, r+1$ . Рассмотрим последовательность линейных функционалов от  $F$ :

$$\tilde{F}_n(t) = \int_{-\infty}^{\infty} K\left(\frac{t-y}{\mu}\right) dF(y), \quad t \in R. \quad (7)$$

Далее нами будут использованы условия:

1.  $\mu \downarrow 0$  при  $n \rightarrow \infty$ ;
2.  $n\mu^{2i} \rightarrow \infty$ ,  $i = \overline{1, r+1}$  при  $n \rightarrow \infty$ ;
3.  $v = Var |K^{(i)}| < \infty$ ,  $i = \overline{1, r+1}$ ;
4. Плотность  $k$  ф.р.  $K$  симметрична:  $k(-t) = k(t)$ ;
5.  $\int_{-\infty}^{\infty} t^i dK(t) = 0$ ,  $i = \overline{1, r}$ ;
6.  $\lambda_1 = \int_{-\infty}^{\infty} |t|^{r+1} dK(t) < \infty$ ;
7.  $\lambda_2 = \int_{-\infty}^{\infty} (K^{(i)})^2 dt < \infty$ ,  $i = \overline{1, r+1}$ . Из (7) инте-

грируя по частям при  $n \rightarrow \infty$  имеем

$$\begin{aligned} \sup_{t \in R} |\tilde{F}_n(t) - F(t)| &\leq \\ &\leq \sup_{-\infty}^{\infty} \int_{-\infty}^{\infty} |F(t - \mu u) - F(t)| k(u) du \leq \\ &\leq \sup_{t \in R} \mu \int_{-\infty}^{\infty} f(t + \theta_n(u)) u k(u) du = \\ &= O(\mu(n)), \end{aligned} \quad (8)$$

при условии

$$\lambda_1(K) = \int_{-\infty}^{\infty} |u| dK(u) < \infty. \quad (9)$$

Учитывая (8), функционал (7) используем для оценивания производных ф.р.  $F$ . Легко проверить, что

$$\begin{aligned} \tilde{F}_n^{(i)}(t) &= \frac{d^i \tilde{F}_n(t)}{dt^i} = \\ &= \int_{-\infty}^{\infty} \frac{1}{\mu^i} K^{(i)}\left(\frac{t-y}{\mu}\right) dF(y) = \\ &= \int_{-\infty}^{\infty} \frac{1}{\mu^i} k^{(i-1)}\left(\frac{t-y}{\mu}\right) dF(y), \\ &i = 1, \dots, r+1, \quad t \in R, \end{aligned} \quad (10)$$

где последовательно интегрируя (10) по частям, получаем

$$\tilde{F}_n^{(i)}(t) = \int_{-\infty}^{\infty} F^{(i)}(t - u\mu) dK(u), \quad (11)$$

$i = 1, \dots, r+1$ ,  $t \in R$ .

Теперь, аналогично (8) имеем

$$\sup_{t \in R} |\tilde{F}_n^{(i)}(t) - F^{(i)}(t)| = O(\mu(n)). \quad (12)$$

С учетом (12), оценку для  $F^{(i)}$  построим как

$$F_n^{(i)}(t) = \int_{-\infty}^{\infty} \frac{1}{\mu^i} K^{(i)}\left(\frac{t-y}{\mu}\right) dF_n^0(y) = \quad (13)$$

$$= \frac{p_n}{\mu^i} \int_{-\infty}^{\infty} K^{(i)}\left(\frac{t-y}{\mu}\right) (1 - H_n(y))^{p_n-1} dH_n(y),$$

$i = 1, \dots, r+1$ ,  $t \in R$ .

Пусть  $C(F^{(i)})$  — множество точек непрерывности функции  $F^{(i)}$ ,  $i = \overline{1, r+1}$  и

$$T_F = \inf \{t \in R : F(t) = 1\}.$$

Тогда согласно (1) и (3):  $T_F = T_G = T_H$ . Справедлива теорема о сильной состоятельности оценок производных (13).

**Теорема.** Пусть выполнены условия 1)-6) и при каждом  $\gamma > 0$  и  $i = \overline{1, r+1}$ ,

$$\sum_{n=1}^{\infty} \exp(-\gamma n \mu^{2i}) < \infty.$$

Тогда в каждой точке  $x \in (-\infty, T_F) \cap C(F^{(i)})$  при  $n \rightarrow \infty$ ,

$$P\left(F_n^{(i)}(x) \rightarrow F^{(i)}(x)\right) = 1,$$

$i = 1, \dots, r+1$ . Отметим, что оценка (13) обобщает исследованную в работах [2, 3, 4] оценок производных.

## Список литературы

- [1] А.А. Абдушукуров. *Статистика неполных наблюдений*. Университет, Ташкент, 2009.
- [2] Алексеев В.Г. О допустимых непараметрических оценках плотности вероятности и ее производных. *Проблемы передачи информации*, 30, вып. 2:36–41, 1994.
- [3] Alekseev. V.G. Estimation probability density function and its derivatives. *Мат. заметки*, 12, №5:621–626, 1972.
- [4] Schuster E.F. Estimation of a probability density function and its derivatives. *Ann. Math. Statist.*, 40, №4:1187–1195, 1969.



## Об асимптотических поведениях ветвящихся процессов с неоднородной иммиграцией, зависящей от состояния

Азимов Джахангир Б.

Институт математики при НУУз  
Ташкент, Узбекистан  
jakhongir20@rambler.ru

**Аннотация.** Рассматриваются асимптотические поведения ветвящихся процессов с неоднородной иммиграцией, зависящей от состояния.

**Ключевые слова.** Ветвящиеся случайные процессы.

Пусть на вероятностном пространстве  $(\Omega, \mathfrak{F}, P)$  заданы:

1. множество  $X = \{X_{in}; i = 1, 2, \dots; n = 0, 1, \dots\}$  независимых одинаково распределенных случайных величин (с.в.) с вероятностной производящей функцией (в.п.ф.)

$$F(s) = \sum_{j=0}^{\infty} p_j s^j, \quad |s| \leq 1, \quad p_j \geq 0, \quad \sum_{j=0}^{\infty} p_j = 1;$$

2. независимое от  $X$  множество  $Y = \{Y_n; n = 0, 1, \dots\}$  независимых с.в. с в.п.ф.

$$g_n(s) = \sum_{j=0}^{\infty} q_j(n) s^j, \quad |s| \leq 1, \quad q_j \geq 0,$$

$$\sum_{j=0}^{\infty} q_j(n) = 1; \quad n = 0, 1, 2, \dots$$

Определим процесс  $\{Z_n\}_{n=0}^{\infty}$  следующим образом:

$$Z_0 = 0, \quad Z_{n+1} = \sum_{i=1}^{Z_n} X_{in} + Y_n I_{\{Z_n=0\}},$$

где  $\sum_{i=1}^0 = 0$ , а  $I_{\{Z_n=0\}}$ , как обычно индикатор события  $\{Z_n = 0\}$ .

Как известно, такой процесс Гальтона-Ватсона с иммиграцией впервые изучался в работе Фостера [1]. В последующем такие процессы стали называться ветвящимися случайными процессами с иммиграцией, зависящей от состояния. Отметим, что процессы с убывающей иммиграцией, зависящей от состояния, с дискретным временем рассматривались в работах [2], [3].

Предположим, что п.ф.  $F(s)$  допускает следующее представление

$$F(s) = s + (1-s)^{1+\nu} L(1-s), \quad (1)$$

где  $0 < \nu \leq 1$  и  $L(s)$  – медленно меняющаяся функция (м.м.ф.) при  $s \rightarrow 0$ .

В представлении (1) учтено то, что данный процесс является критическим ( $F'(1) = 1$ ) и, кроме того, из (1) следует, что

$$F'''(1-) = +\infty,$$

т.е. число частиц, порождаемых одной частицей, имеет бесконечную дисперсию.

Известно [4], что если  $0 < F(0) < 1$ , то для процесса Гальтона-Ватсона существует стационарная мера, п.ф.  $U(s)$  которая аналитична в круге  $|s| < q$  ( $q$  – вероятность вырождения) и при  $U(F(0)) = 1$  удовлетворяет функциональному уравнению Абеля:

$$U(F(s)) = 1 + U(s), \quad |s| < q, \quad U(1) = \infty.$$

Как показал Слэк [5], представление (1), влечет

$$U(s) = \frac{1 + o(1)}{\nu(1-s)^{\nu} L(1-s)}, \quad s \rightarrow 1-. \quad (2)$$

Из асимптотического представления (2) следует, что обратная к  $U(1-s)$  функция  $G(x)$ ,  $x > 0$  имеет вид

$$G(x) = \frac{N(x)}{x^{1/\nu}},$$

где  $N(x)$  – м.м.ф. на бесконечности и при  $x \rightarrow \infty$  удовлетворяет условию  $\nu N^{\nu}(x) L(x^{-1/\nu} N(x)) \rightarrow 1$ .

Обозначим

$$\alpha_n = EY_n = g_n(1), \quad \beta_n = DY_n + \alpha_n^2 - \alpha_n,$$

$$Q_1(n) = \alpha_n \sum_{k=0}^n (1 - F_k(0)),$$

$$Q_2(n) = (1 - F_n(0)) \sum_{k=0}^n \alpha_k,$$

где  $F_0(s) = s$ ,  $F_{n+1} = F(F_n(s))$ .

В дальнейшем предположим, что выполнены условия

$$\begin{aligned} \sup_n \alpha_n < \infty, \quad \sup_n \beta_n < \infty, \\ 0 < \alpha_n \rightarrow 0, \quad \beta_n \rightarrow 0, \quad n \rightarrow \infty. \end{aligned} \quad (3)$$

Введем функцию

$$M(n) = \sum_{k=1}^n \frac{N(k)}{k^{1/\nu}}.$$

В работе [3] такой процесс, изучен в случае, когда  $\nu = 1$ ,  $M(n) \rightarrow \infty$ ,  $n \rightarrow \infty$ . Мы рассмотрим случай, когда  $M(n) \rightarrow M < \infty$ ,  $n \rightarrow \infty$ . Отметим, что  $M < \infty$  может иметь место как в случае  $0 < \nu < 1$ , так и в случае  $\nu = 1$ .

Справедливы следующие теоремы.

**Теорема 1.** Пусть  $\alpha_n \sim \frac{l(n)}{n^r}$ ,  $\beta_n = o(Q_1(n))$ ,  $n \rightarrow \infty$ , где  $r \geq 0$  и  $l(n)$  – м.м.ф. на бесконечности, причем  $l(n) = o(1)$ ,  $n \rightarrow \infty$ , когда  $r = 0$ . Кроме того  $\theta_n = \frac{Q_1(n)}{Q_2(n)} \rightarrow \infty$  при  $n \rightarrow \infty$ . Тогда существуют пределы

$$\lim_{n \rightarrow \infty} P\{Z_n = k | Z_n > 0\} = P_k^*, \quad \sum_{k=1}^{\infty} P_k^* = 1,$$

причем для п.ф.  $\varphi(s) = \sum_{k=1}^{\infty} P_k^* s^k$  имеет место

$$\varphi(s) = 1 - \frac{1}{M} \sum_{k=1}^{\infty} G(k + U(s)). \quad (4)$$

**Теорема 2.** Пусть выполнены условия теоремы 1 и  $\theta_n \rightarrow \theta$ ,  $0 < \theta < \infty$  при  $n \rightarrow \infty$ . Тогда существуют пределы

$$\lim_{n \rightarrow \infty} P\{Z_n = k | Z_n > 0\} = \hat{P}_k, \quad \sum_{k=1}^{\infty} \hat{P}_k = \frac{\theta}{1 + \theta} < 1,$$

Кроме того,  $\sum_{k=1}^{\infty} \hat{P}_k s^k = \frac{\theta}{1 + \theta} \varphi(s)$  и  $\varphi(s)$  определяется равенством (4).

**Теорема 3.** Пусть  $\alpha_n \sim \frac{l(n)}{n^r}$ ,  $\beta_n = o(Q_1(n))$ ,  $n \rightarrow \infty$ , где  $r \geq 0$  и  $l(n)$  – м.м.ф. на бесконечности, причем  $l(n) = o(1)$ ,  $n \rightarrow \infty$ , когда  $r = 0$ . Кроме того  $\frac{Q_1(n)}{Q_2(n)} \rightarrow \theta$  при  $n \rightarrow \infty$ ,  $0 \leq \theta < \infty$ . Тогда для  $x \geq 0$

$$\lim_{n \rightarrow \infty} P\{Z_n(1 - F_n(0)) < x | Z_n > 0\} = \frac{\theta + G(x)}{1 + \theta},$$

где

$$\int_0^{\infty} e^{-\tau x} dG(x) = 1 - \tau(1 + \tau^\nu)^{-1/\nu}.$$

## Список литературы

- [1] G. Foster. A limit theorem for a branching processes with state-dependent immigration. *Ann.Math.Stat.*, 42(5):1173–1176, 1971.
- [2] К. Митов, В. Ватулин, Н. Янев. Критические процессы Гальтона-Ватсона с убывающей иммиграцией, зависящей от состояния процесса. *Сердика Българско мат. списание*, 10:412–424, 1984.
- [3] J.B. Azimov. Limit theorems for the critical galton-watson branching process with non-homogeneous immigration. *Abstracts of the international conference "Stochastic analysis and random dynamics", June 14-20, 2009, Lviv, Ukraine*, (~):15–16, 2009.
- [4] Т.Е. Харрис. *Теория ветвящихся процессов*. Мир, М., 1966.
- [5] R.S. Slack. A branching process with mean one and possibly infinite variance. *Z. Wahrsch. Geb.*, 9(2):139–145, 1968.

# Регрессионный анализ двудольных множеств событий

Баранова Ирина Владимировна

Сибирский Федеральный Университет  
Институт математики и фундаментальной информатики  
Красноярск  
irinabar@yandex.ru

**Аннотация.** В работе рассматривается понятие двудольного множества событий, состоящее из множеств событий, первая доля которых определяется случайными величинами, а вторая — случайными множествами, а также — множества двудольных множеств. Приводятся основные понятия классического и эвентологического регрессионного анализа. Изучаются зависимости между двудольными множествами событий и проводится построение условных распределений и функции теоретической эвентологической регрессии для двудольных множеств событий.

**Ключевые слова.** Регрессия, множественная регрессия, условное событие, эвентологическая регрессия, двудольное множество событий

## 1 Введение

Данная работа посвящена изучению зависимости между двудольными множествами событий с помощью инструмента эвентологической регрессии.

Понятие регрессии (от лат. regressio — обратное движение, переход от более сложных форм развития к менее сложным) является одним из основных понятий в теории вероятностей и математической статистике, и выражает зависимость среднего значения случайной величины от значения другой случайной величины или от нескольких случайных величин. Термин «регрессия» был впервые введен Френсисом Гальтоном в 1886 году в результате проведенного исследования соотношения роста родителей и их детей.

**Определение 1.** Регрессионный анализ — раздел математической статистики, объединяющий практические методы исследования регрессионной зависимости между величинами по статистическим данным.

Целями регрессионного анализа данных являются:

- определение общего вида уравнения регрессии — вклада отдельных независимых переменных в вариацию зависимой;
- построение статистических оценок неизвестных параметров, входящих в уравнение регрессии;

- проверка статистических гипотез о регрессии;
- предсказание значения зависимой переменной с помощью независимой(ых).

Регрессионный анализ нельзя использовать для определения наличия связи между переменными, поскольку наличие такой связи и есть предпосылка для применения анализа. Поэтому нахождение регрессии предваряется вычислением ковариации между данными и может дополняться другими методами корреляционного анализа данных.

Регрессионный анализ является одним из наиболее распространенных методов обработки результатов наблюдений при изучении зависимостей в физике, экономике, технике, биологии, химии и других областях знаний. На моделях регрессионного анализа основаны такие разделы математической статистики как дисперсионный анализ и планирование экспериментов. Модели регрессионного анализа широко используются в многомерном статистическом анализе.

В современном регрессионном анализе наиболее изученной является методика исследования влияния одной или нескольких независимых случайных величин  $\xi_1, \xi_2, \dots, \xi_n$  на зависимую случайную величину  $\eta$ . Особенно это касается ситуации линейной регрессии. Случай нелинейной регрессионной зависимости исследован не так полно, зачастую на практике нелинейные модели сводятся к линейной форме с помощью подходящих преобразований.

В статье приводятся понятия основные понятия классического регрессионного анализа: парная и множественная регрессия, регрессия случайных величин, регрессия случайных векторов, а также — функция теоретической регрессии.

Особое внимание уделяется эвентологическому подходу к регрессии, разработанному О. Ю. Воробьевым, в котором изучается распределение случайных событий и случайных множеств событий. В рамках данного подхода О. Ю. Воробьевым было предложено понятие эвентологической регрессии одного множества событий на другое множество событий [1].

Во многих областях современной научной и практической деятельности при проведении различных исследований довольно часто возникает необходимость

решения задач регрессионного анализа для случая сложных статистических систем, поведение которых описывается числовыми и множественными данными. К числу таких областей человеческой деятельности можно отнести здравоохранение, технику, экономику, финансы, страхование, психологию, социологию, экологию, биологию, филологию, лингвистику, педагогику и другие.

Задача выявления зависимости между данными, описывающими поведение подобных систем, является очень важной и актуальной.

Основные трудности при проведении анализа таких систем заключаются в сложной структуре зависимости между элементами системы, большом числе элементов и их разнотипности. Поэтому для устранения этих трудностей автором был предложен метод двудольных множеств случайных событий [2, 3]. Согласно данному методу каждому элементу системы было предложено поставить в соответствие двудольное множество случайных событий, которое представляет собой объединение двух множеств событий, первая доля которого определяется случайными величинами, а вторая — случайными множествами.

В работе [4] были предложены условные вероятности двудольных множеств событий и, затем, они были применены для решения задачи статистической оценки функции теоретической эвентологической регрессии [5].

## 2 Классический подход к регрессии

Рассмотрим основные понятия связанные с нахождением парной и множественной регрессии.

### 2.1 Основные понятия

**Определение 2.** Уравнение  $y = f(x)$  (уравнение регрессии) — формула статистической связи между двумя переменными, которая называется парной регрессией.

**Определение 3.** Если при каждом  $x = x_i$  наблюдается  $n_i$  значений  $y_{i1}, \dots, y_{in_i}$  величины  $y$ , то зависимость средних арифметических

$$\bar{y} = \frac{y_{i1}, \dots, y_{in_i}}{n_i} \text{ от } x_i, i = 1, \dots, N$$

и является регрессией в статистическом понимании этого термина.

**Определение 4.** Множественной регрессией называется модель, которая включает несколько предсказывающих или объясняющих переменных.

Общее назначение множественной регрессии состоит в анализе связи между несколькими независимыми переменными (называемыми также **регрессорами** или **предикторами**) и зависимой переменной.

Модель парной линейной регрессии представляет собой зависимость результирующего признака от одного фактора и имеет вид

$$y = \alpha_0 + \alpha_1 x + \varepsilon,$$

$$M\varepsilon = 0, D\varepsilon = \sigma^2, \text{cov}(\varepsilon_i, \varepsilon_j) = 0, i \neq j,$$

где  $\alpha_0$  и  $\alpha_1$  — неизвестные параметры регрессии;  $(\alpha_0 + \alpha_1 x)$  образует детерминированную составляющую модели.

На основе имеющихся наблюдений необходимо подобрать наилучшие по определенному критерию оценки  $a$  и  $b$  неизвестных параметров модели  $\alpha_0$  и  $\alpha_1$ . Уравнение регрессии с оценками параметров запишем как

$$\hat{y} = a + bx,$$

где  $\hat{y}$  — выровненная или расчетная величина  $y$ .

Оценки параметров  $a$  и  $b$  определяются с помощью метода наименьших квадратов.

### 2.2 Метод наименьших квадратов

Рассмотрим метод наименьших квадратов (МНК) для парной линейной регрессии. Предположим, имеются  $n$  пар наблюдений над переменными  $x$  и  $y$ :  $(x_i, y_i)$ ,  $i = 1, 2, \dots, n$ . Для отдельных наблюдений модель парной линейной регрессии можно записать как

$$y_i = \alpha_0 + \alpha_1 x_i + \varepsilon_i.$$

Оценки  $a$  и  $b$  неизвестных параметров  $\alpha_0$  и  $\alpha_1$  определяются из условия минимизации суммы квадратов отклонений наблюдаемых значений  $y_i$  от расчетных  $\hat{y}_i$ , где  $\hat{y}_i = a + bx_i$ .

Условие минимизации суммы квадратов отклонений  $y_i$  от  $\hat{y}_i$  выглядит следующим образом:

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2 \rightarrow \min$$

Для уравнений, приводимых к линейному виду, решается следующая система линейных уравнений:

$$\begin{cases} na + b \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \\ a + \sum_{i=1}^n x_i + b \sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i x_i \end{cases}$$

Необходимо найти решение системы относительно неизвестных оценок параметров  $a$  и  $b$ .

Оценку качества построенной модели дает коэффициент  $R^2 = r_{xy}^2$  (индекс детерминации), а также средняя ошибка аппроксимации:

$$\bar{A} = \frac{1}{n} \sum_{i=1}^n A_i = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \cdot 100\%$$

Коэффициент  $a$  называется *постоянным коэффициентом* линейной регрессии, а  $b$  — *переменным коэффициентом* линейной регрессии.

### 2.3 Регрессия для случайных величин

Пусть задано  $(\Omega, \mathcal{F}, \mathbf{P})$  — вероятностное пространство, где  $\Omega$  — пространство элементарных событий,  $\mathcal{F}$  — алгебра событий,  $\mathbf{P}$  — вероятностная мера, заданная на  $\mathcal{F}$ .

**Определение 5.** Случайной величиной  $\xi$  называется измеримая функция  $\xi = \xi(\omega)$ , отображающая  $\Omega$  в  $\mathcal{R}$ .

**Определение 6.** Регрессия величины  $\eta$  по величине  $\xi$  определяется условным математическим ожиданием  $\eta$ , вычисленным при условии, что  $\xi = x$ :

$$\mathbf{E}(\eta | x) = y(x).$$

Уравнение  $y = y(x)$ , в котором  $x$  играет роль «независимой» переменной, называется уравнением регрессии, а соответствующий график — линией или кривой регрессии  $\eta$  по  $\xi$ , переменная  $x$  называется *регрессионной переменной* или *регрессором*. Точность, с которой уравнение регрессии  $\eta$  по  $\xi$  отражает изменение  $\eta$  в среднем при изменении  $x$ , измеряется *условной дисперсией* величины  $\eta$ , вычисленной для каждого значения  $\xi = x$ :

$$\mathbf{D}(\eta | x) = \sigma^2(x).$$

Наиболее простым является тот случай, когда регрессия  $\eta$  по  $\xi$  линейна:

$$\mathbf{E}(\eta | x) = \beta_0 + \beta_1 x$$

Коэффициенты  $\beta_0$  и  $\beta_1$  называются *коэффициентами регрессии* и определяются равенствами:

$$\beta_0 = m_\eta - \rho_{\xi, \eta} \frac{\sigma_\eta}{\sigma_\xi} m_\xi, \quad \beta_1 = \rho_{\xi, \eta} \frac{\sigma_\eta}{\sigma_\xi},$$

где  $m_\xi$  и  $m_\eta$  — математические ожидания  $\xi$  и  $\eta$ ,  $\sigma_\xi^2$  и  $\sigma_\eta^2$  — дисперсии  $\xi$  и  $\eta$ , а  $\rho_{\xi, \eta}$  — коэффициент корреляции между  $\xi$  и  $\eta$ . Уравнение регрессии при этом выражается формулой

$$y = m_\eta + \rho_{\xi, \eta} \frac{\sigma_\eta}{\sigma_\xi} (x - m_\xi).$$

В случае, когда совместное распределение  $\xi$  и  $\eta$  нормально, обе линии регрессии  $y = y(x)$  и  $x = x(y)$  являются прямыми.

В условиях статистических приложений, когда для точного определения регрессии нет достаточных сведений о форме совместного распределения вероятностей, возникает задача приближенного нахождения регрессии. Решению этой задачи может служить выбор из всех функций  $f(x)$ , принадлежащих заданному классу, такой функции, которая дает наилучшее

представление величины  $\eta$  в том смысле, что она минимизирует математическое ожидание  $\mathbf{E}[\eta - f(x)]^2$ . Найденная функция называется *средней квадратической регрессией*. Простейшим будет случай линейной средней квадратической регрессии. Если регрессия  $\eta$  по  $\xi$  отлична от линейной, то уравнение

$$y = m_\eta + \rho_{\xi, \eta} \frac{\sigma_\eta}{\sigma_\xi} (x - m_\xi)$$

есть линейная аппроксимация истинного уравнения регрессии: математическое ожидание  $\mathbf{E}(\eta - b_0 - b_1 \xi)^2$  достигает минимума по  $b_0$  и  $b_1$  при  $b_0 = \beta_0$  и  $b_1 = \beta_1$ .

В общем случае функция регрессии выражается линейной комбинацией тех или иных заданных функций:

$$y(x) = \beta_0 \varphi_0(x) + \beta_1 \varphi_1(x) + \dots + \beta_m \varphi_m(x).$$

Наиболее важное значение имеет *полиномиальная (параболическая) регрессия*, при которой

$$\varphi_0(x) = 1, \varphi_1(x) = x, \dots, \varphi_m(x) = x^m.$$

### 2.4 Линейная регрессионная модель

Модель множественной регрессии, включающая  $k$  объясняющих переменных  $x_1, x_2, \dots, x_k$ , имеет вид

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_k x_k + \varepsilon,$$

$$\mathbf{E}\varepsilon = 0, \quad \mathbf{D}\varepsilon = \sigma^2, \quad \text{cov}(\varepsilon_i, \varepsilon_j) = 0, \quad i \neq j,$$

где  $\alpha_0, \alpha_1, \dots, \alpha_k$  — неизвестные параметры регрессии ( $\alpha_1, \alpha_2, \dots, \alpha_k$  соответствуют включенным в уравнение регрессии факторам,  $\alpha_0$  — свободному члену регрессии);  $\varepsilon$  — случайная составляющая модели,  $i$  и  $j$  — индексы наблюдений.

Уравнение регрессии с оценками параметров можно записать как

$$\hat{y} = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_k x_k,$$

где  $a_0, a_1, \dots, a_k$  — оценки параметров регрессии.

Как и в случае парной регрессии, задача оценивания заключается в том, чтобы с помощью метода наименьших квадратов найти такие оценки  $a_0, a_1, \dots, a_k$  неизвестных параметров  $\alpha_0, \alpha_1, \dots, \alpha_k$ , которые минимизировали бы квадраты отклонений наблюдаемых значений зависимой переменной  $y_i$  от вычисленных с помощью уравнения регрессии  $\hat{y}_i$ . Расчетные значения  $\hat{y}_i$  определяются как

$$\hat{y}_i = a_0 + a_1 x_{i1} + a_2 x_{i2} + \dots + a_k x_{ik}, \quad i = 1, \dots, n,$$

где  $x_{i1}, x_{i2}, \dots, x_{ik}$ ,  $i = 1, \dots, n$  — значения факторов, полученные при  $i$ -том наблюдении.

Функция, значение которой минимизируют, имеет вид

$$Q = \sum_{i=1}^n (y_i - (a_0 + a_1 x_{i1} + a_2 x_{i2} + \dots + a_k x_{ik}))^2 \rightarrow \min$$



### 3 Условные события и условные Э-распределения

В рамках эвентологического подхода О. Ю. Воробьёвым были введены понятия условного события, условного эвентологического распределения и эвентологической регрессии одного множества событий на другое множество событий [1].

Рассмотрим вероятностное пространство  $(\Omega, \mathcal{F}, \mathbf{P})$ , множество избранных  $\mathcal{F}$ -измеримых событий  $\mathfrak{X} \subseteq \mathcal{F}$ , некоторое  $\mathcal{F}$ -измеримое событие  $s \subseteq \Omega$ , имеющее ненулевую неединичную вероятность:  $0 < \mathbf{P}(s) < 1$  и событие  $x \in \mathfrak{X}$ .

**Определение 7.** Условное событие (*у|событие*)  $x|s$  — это событие, означающее, что наступление события  $x$  происходит в соответствии с наступлением события  $s$ . Вероятность наступления условного события  $x|s$  соответствует понятию **условной вероятности**  $\mathbf{P}(x|s)$  и определяется следующей формулой:

$$\mathbf{P}(x|s) = \frac{\mathbf{P}(x \cap s)}{\mathbf{P}(s)}.$$

Рассмотрим два множества событий  $\mathfrak{X}_s = \{x \cap s, x \in \mathfrak{X}\}$ ,  $\mathfrak{X}_{s^c} = \{x \cap s^c, x \in \mathfrak{X}\}$ , на которые событие  $s$  разбивает по Минковскому множество событий  $\mathfrak{X}$ :  $\mathfrak{X} = \mathfrak{X}_s (+) \mathfrak{X}_{s^c} = \{x \cap s + x \cap s^c, x \in \mathfrak{X}\}$ . Множество событий  $\mathfrak{X}$  разбивает пространство э-событий  $\Omega = \sum_{X \subseteq \mathfrak{X}} \text{ter}(X)$  на события-терраски

$$\text{ter}(X) = \bigcap_{x \in X} x \bigcap_{x \in X^c} x^c, \quad X \subseteq \mathfrak{X}.$$

Множества событий  $\mathfrak{X}_s$  и  $\mathfrak{X}_{s^c}$  также разбивают, хотя и каждое по своему, пространство э-событий

$$\Omega = \sum_{X \subseteq \mathfrak{X}} \text{ter}_s(X), \quad \Omega = \sum_{X \subseteq \mathfrak{X}} \text{ter}_{s^c}(X)$$

на собственные события-терраски

$$\text{ter}_s(X) = \bigcap_{x \in X} (x \cap s) \bigcap_{x \in X^c} (x \cap s)^c$$

$$\text{ter}_{s^c}(X) = \bigcap_{x \in X} (x \cap s^c) \bigcap_{x \in X^c} (x \cap s^c)^c$$

Поскольку все события из  $\mathfrak{X}_s$  содержатся в событии  $s$ , а все события из  $\mathfrak{X}_{s^c}$  — в его дополнении  $s^c = \Omega - s$ , то для любых  $X \neq \emptyset$  и  $Y \neq \emptyset$  — события-терраски попарно  $\text{ter}_s(X)$  и  $\text{ter}_{s^c}(Y)$  не пересекаются:  $\text{ter}_s(X) \cap \text{ter}_{s^c}(Y) = \emptyset$ . Поэтому

$$s \cap \text{ter}(X) = \begin{cases} \text{ter}_s(X), & X \neq \emptyset, \\ s \cap \text{ter}_s(\emptyset), & X = \emptyset, \end{cases}$$

$$s^c \cap \text{ter}(X) = \begin{cases} \text{ter}_{s^c}(X), & X \neq \emptyset, \\ s^c \cap \text{ter}_{s^c}(\emptyset), & X = \emptyset. \end{cases}$$

А отсюда

$$\text{ter}(X) = \begin{cases} \text{ter}_s(X) + \text{ter}_{s^c}(X), & X \neq \emptyset, \\ s \cap \text{ter}_s(\emptyset) + s^c \cap \text{ter}_{s^c}(\emptyset), & X = \emptyset. \end{cases}$$

Обозначим

$$\mathbf{p} = \{p(X), X \subseteq \mathfrak{X}\},$$

$$\mathbf{p}_s = \{p_s(X), X \subseteq \mathfrak{X}\}, \quad \mathbf{p}_{s^c} = \{p_{s^c}(X), X \subseteq \mathfrak{X}\}$$

— Э-распределения множеств событий  $\mathfrak{X}$ ,  $\mathfrak{X}_s$  и  $\mathfrak{X}_{s^c}$ , соответственно, где

$$p(X) = \mathbf{P}(\text{ter}(X)),$$

$$p_s(X) = \mathbf{P}(\text{ter}_s(X)), \quad p_{s^c}(X) = \mathbf{P}(\text{ter}_{s^c}(X)).$$

В силу (1)

$$p(X) = \begin{cases} p_s(X) + p_{s^c}(X), & X \neq \emptyset, \\ \mathbf{P}(s \cap \text{ter}_s(\emptyset)) + \mathbf{P}(s^c \cap \text{ter}_{s^c}(\emptyset)), & X = \emptyset. \end{cases}$$

Причем, очевидно, что

$$\mathbf{P}(s \cap \text{ter}_s(\emptyset)) = \mathbf{P}(s) - \sum_{X \neq \emptyset} p_s(X),$$

$$\mathbf{P}(s^c \cap \text{ter}_{s^c}(\emptyset)) = \mathbf{P}(s^c) - \sum_{X \neq \emptyset} p_{s^c}(X).$$

**Определение 8.** Условное Э-распределение множества событий  $\mathfrak{X}$  при условии, что наступило событие  $s \in \mathcal{F}$  определяется как набор вероятностей  $\mathbf{p}_{|s} = \{p_{|s}(X), X \subseteq \mathfrak{X}\}$ , где

$$p_{|s}(X) = \begin{cases} \frac{p_s(X)}{\mathbf{P}(s)}, & X \neq \emptyset, \\ 1 - \frac{1 - p_s(\emptyset)}{\mathbf{P}(s)}, & X = \emptyset. \end{cases}$$

Очевидно, что  $\mathbf{p} = \mathbf{p}_{|s} \mathbf{P}(s) + \mathbf{p}_{|s^c} \mathbf{P}(s^c)$ . Более того, если множество событий  $\mathcal{S} \subseteq \mathcal{F}$  образует разбиение  $\Omega = \sum_{s \in \mathcal{S}} s$ , то справедлива общая «формула полной вероятности»:  $\mathbf{p} = \sum_{s \in \mathcal{S}} \mathbf{p}_{|s} \mathbf{P}(s)$ , позволяющая представить Э-распределение  $\mathbf{p}$  в виде взвешенной суммы условных Э-распределений  $\mathbf{p}_{|s}$  с весами, равными  $\mathbf{P}(s)$  — вероятностям соответствующих событий из  $\mathcal{S}$ .

#### 4 Первая террасная формулировка теоремы Байеса

Теорема Байеса на практике применяется часто, например, в диагностике, а также в других областях, где имеется множество  $\mathcal{S} \subseteq \mathcal{F}$  взаимоисключающих *событий-гипотез* и требуется определить, как меняются вероятности этих событий-гипотез при условии, что произошло некоторое *событие-обстоятельство*  $x \in \mathfrak{X}$ . Формулы Байеса для условных вероятностей событий-гипотез  $s \in \mathcal{S}$  имеют вид:

$$\mathbf{P}(s | x) = \frac{\mathbf{P}(s | x)\mathbf{P}(s)}{\mathbf{P}(x)},$$

где

$$\mathbf{P}(x) = \sum_{s \in \mathcal{S}} \mathbf{P}(x | s)\mathbf{P}(s),$$

хотя наиболее популярна запись:

$$\mathbf{P}(x | s) = \frac{\mathbf{P}(x | s)\mathbf{P}(s)}{\sum_{s \in \mathcal{S}} \mathbf{P}(x | s)\mathbf{P}(s)}, \quad s \in \mathcal{S}.$$

Эвентологическая теория позволяет «переформулировать» теорему Байеса в существенно более общей форме, во-первых, сняв несущественное и чисто техническое требование *взаимоисключаемости* и предположив, что множество событий-гипотез  $\mathcal{S}$  имеет произвольную структуру, и во-вторых, заменив одно событие-обстоятельство  $x \in \mathfrak{X}$  всем множеством событий-обстоятельств  $\mathfrak{X}$ .

**Первая террасная формулировка.** Множество событий-гипотез  $\mathcal{S}$  разбивает  $\Omega$  на взаимоисключающие события-терраски

$$\text{ter}(S) = \bigcap_{s \in S} s \bigcap_{s \in S^c} s^c, \quad S \in \mathcal{S},$$

где  $s^c = \Omega - s$ ,  $S^c = \mathcal{S} - S$ , и формулы Байеса принимают более общий *террасный вид*:

$$\mathbf{P}(\text{ter}(S) | x) = \frac{\mathbf{P}(x | \text{ter}(S))\mathbf{P}(\text{ter}(S))}{\mathbf{P}(x)}, \quad S \subseteq \mathcal{S},$$

где

$$\mathbf{P}(x) = \sum_{S \subseteq \mathcal{S}} \mathbf{P}(x | \text{ter}(S))\mathbf{P}(\text{ter}(S)),$$

и оказывается возможным записать их в виде формул

$$p_{|x}(S) = \frac{\mathbf{P}(x | \text{ter}(S))}{\mathbf{P}(x)} p(S), \quad S \subseteq \mathcal{S},$$

где

$$\mathbf{P}(x) = \sum_{S \subseteq \mathcal{S}} \mathbf{P}(x | \text{ter}(S)) p(S),$$

связывающих  $\mathfrak{X}$ -распределение  $\mathbf{p} = \{p(S), S \subseteq \mathcal{S}\}$  множества событий-гипотез  $\mathcal{S}$  с условным  $\mathfrak{X}$ -распределением  $\mathbf{p}_{|x} = \{p_{|x}(S), S \subseteq \mathcal{S}\}$  множества событий-гипотез  $\mathcal{S}$  при условии наступления события-обстоятельства  $x \in \mathfrak{X}$ .

#### 5 Эвентологическая регрессия

В работе [1] О.Ю. Воробьевым была выведена формула эвентологической регрессии.

**Определение 9.** Эвентологическая регрессия *одного множества титульных событий*  $\mathfrak{X}$  на *другое множество титульных событий*  $\mathcal{Y}$  — это зависимость  $\varphi$ , связывающая безымянные события-терраски  $\text{ter}(X), X \subseteq \mathfrak{X}$  с событиями-террасками  $\text{ter}(Y), Y \subseteq \mathcal{Y}$ :

$$\text{ter}(Y) = \varphi(\text{ter}(X)), \quad X \subseteq \mathfrak{X}.$$

Поскольку собственных имен у событий-террасок нет, и разумный субъект имеет к ним лишь косвенный «доступ» через логические комбинации собственных имен титульных событий, которые их порождают, то роль собственных имен для событий-террасок  $\text{ter}(X)$  исполняют подмножества  $X \subseteq \mathfrak{X}$  титульных событий, которыми события-терраски занумерованы. Таким образом, для разумного субъекта безымянное событие-терраска

$$\text{ter}(X)$$

«скрывается» за подмножеством титульных событий

$$X \subseteq \mathfrak{X},$$

а зависимость  $\varphi$  между безымянными событиями-террасками

$$\text{ter}(Y) = \varphi(\text{ter}(X))$$

«скрывается» за зависимостью между соответствующими подмножествами титульных событий:

$$Y = \varphi(X)$$

##### 5.1 Функция теоретической эвентологической регрессии

При фиксированной метрике  $\rho$  на  $2^{\mathfrak{X}} \times 2^{\mathfrak{X}}$  функция теоретической эвентологической регрессии определяется способом, вполне аналогичным классическому:

$$E\rho(Y, \varphi(X)) \rightarrow \min_{\varphi}.$$

Если

$$\rho(Y, \varphi(X)) = \rho_\alpha(Y, \varphi(X)),$$

то функция теоретической эвентологической регрессии имеет вид

$$\varphi(X) = Q_\alpha(Y \mid \text{ter}(X))$$

условно-эвентологического сет-квантиля порядка  $\alpha$ , где

$$Q_\alpha(Y \mid \text{ter}(X)) = \{y : P(y \mid \text{ter}(X)) \geq \alpha\}$$

подмножества титульных событий  $y \in \mathcal{Y}$ , условные вероятности, которые не меньше  $\alpha$ . В этих обозначениях

$$\text{ter}(Q_\alpha(Y \mid \text{ter}(X))) = \varphi(\text{ter}(X)).$$

## 6 Двудольное множество случайных элементов и двудольное множество случайных событий

В ситуации, когда поведение каждого элемента сложной системы характеризуется данными, одна часть которых — числовая, а другая — множественная, объект, порождающий данную статистику, может быть представим как объединение двух долей: случайных величин и случайных множеств событий.

Пусть  $(\Omega, \mathcal{F}, \mathbf{P})$  — вероятностное пространство, где  $\mathcal{F}$  — алгебра событий, случайные события  $x, y, \dots \in \mathcal{F}$  — элементы алгебры  $\mathcal{F}$ , на которой определена вероятность  $\mathbf{P}$ . Множество избранных случайных событий  $\mathfrak{X} \in \mathcal{F}$ .

**Определение 10.** Случайное множество событий под  $\mathfrak{X}$  определяется как измеримое отображение

$$K : (\Omega, \mathcal{F}, \mathbf{P}) \rightarrow (2^{\mathfrak{X}}, 2^{2^{\mathfrak{X}}}),$$

где  $2^{\mathfrak{X}}$  — множество всех подмножеств множества  $\mathfrak{X}$ .

В методе было предложено случайную величину и случайное конечное множество событий рассматривать как реализацию случайного элемента, и объект, порождающий статистику, — рассматривать как двудольное множество случайных элементов.

**Определение 11.** Двудольным множеством случайных элементов называется множество  $\{\xi, \mathbf{K}\}$ , представимое в следующем виде:

$$\{\xi, \mathbf{K}\} = \xi \cup \mathbf{K} = \{\xi_a, a \in A, K_\beta, \beta \in B\}, \quad (1)$$

здесь первая доля — это случайные величины  $\xi = \{\xi_a, a \in A\}$ , вторая — случайные множества событий  $\mathbf{K} = \{K_\beta, \beta \in B\}$ ,  $A$  — множество индексов случайных величин,  $B$  — множество индексов случайных множеств событий.

События, которые порождаются двудольным множеством случайных элементов, будем называть множеством случайных событий.

**Определение 12.** Двудольное множество случайных событий представляет собой объединение двух множеств — множества событий, которое определяется случайными величинами, и множества событий, которое определяется случайными множествами событий:

$$\{\mathcal{Y}, \mathfrak{X}\} = \{\mathcal{Y}_a, \mathfrak{X}_\beta, a \in A, \beta \in B\}. \quad (2)$$

Полной характеристикой двудольного множества случайных событий служит его эвентологическое распределение. В работе [1] были найдены все формы эвентологического распределения двудольного множества случайных событий. Приведем ниже вид для одного из них.

Распределение вероятностей событий двудольного множества  $\{\mathcal{Y}, \mathfrak{X}\}$ :

$$p(\mathbf{r}, \mathbf{X}) = \mathbf{P} \left( \bigcap_{a \in A} \left\{ \bigcap_{y \neq r} \{\xi = y\}^c \right\} \bigcap_{x_\beta \in X} x_\beta \bigcap_{x_\beta \in X^c} x_\beta^c, a \in A, \beta \in B \right), \quad (3)$$

$$x_\beta \in \mathfrak{X}, \beta \in B,$$

$$\mathbf{r} = \{r_a, r_a \in \mathbb{R}\},$$

$$\mathbf{X} = \{X_\beta, X_\beta \subseteq \mathfrak{X}_\beta\}.$$

На практике работать с такими распределения достаточно сложно. Поэтому в [2] было предложено использовать для их оценивания сет-операции по Минковскому.

## 7 Сет-операции по Минковскому над двудольными множествами случайных событий

В работе [2] было введено понятие произвольной операции по Минковскому над двумя двудольными множествами событий и ее вероятность.

Пусть имеются два двудольных множества  $s^1 = \{\mathcal{Y}_a^1, \mathfrak{X}_\beta^1, a \in A, \beta \in B\}$  и  $s^2 = \{\mathcal{Y}_a^2, \mathfrak{X}_\beta^2, a \in A, \beta \in B\}$ .

**Определение 13.** Произвольной операцией по Минковскому над двумя двудольными множествами событий  $s^1$  и  $s^2$  называется теоретико-множественная операция (сет-операция), которая представляется как множество событий, полученных с помощью операций по Минковскому над соответствующими событиями из каждой доли:

$$s1(\mathcal{O})s2 = \left\{ \mathcal{Y}_a^1(\mathcal{O})\mathcal{Y}_a^2, \mathfrak{X}_\beta^1(\mathcal{O})\mathfrak{X}_\beta^2, a \in A, \right. \\ \left. \beta \in B \right\} = \left\{ \mathcal{Y}_a^1(r_a)\mathcal{O}\mathcal{Y}_a^2(r_a), X_\beta^1\mathcal{O}X_\beta^2, X_\beta \subseteq \mathfrak{X}_\beta, \right. \\ \left. r_a \in \mathcal{R}_a, a \in A, \beta \in B \right\}. \quad (4)$$

Здесь  $\mathcal{Y}_a^1, \mathcal{Y}_a^2 \subseteq \mathcal{Y}_a, \mathfrak{X}_\beta^1, \mathfrak{X}_\beta^2 \subseteq \mathfrak{X}_\beta$ .

Примером такой операции является сет-операция симметрической разности по Минковскому двух двудольных множеств событий.

Сет-операция симметрической разности по Минковскому двух двудольных множеств  $s^1$  и  $s^2$  имеет следующий вид:

$$s1(\Delta)s2 = \left\{ \mathcal{Y}_a^1(r_a)\Delta\mathcal{Y}_a^2(r_a), X_\beta^1\Delta X_\beta^2, \right. \\ \left. X_\beta \subseteq \mathfrak{X}_\beta, r_a \in \mathcal{R}_a, a \in A, \beta \in B \right\}. \quad (5)$$

Вероятность симметрической разности равна

$$\mathbf{P}(s^1(\Delta)s^2) = \frac{1}{|A|} \sum_{a \in A} \frac{1}{|\mathcal{Y}_a|} \sum_{r_a \in \mathbb{R}} \mathbf{P}(\mathcal{Y}_a^1(r_a)\Delta \\ \mathcal{Y}_a^2(r_a)) + \frac{1}{|B|} \sum_{\beta \in B} \frac{1}{|\mathfrak{X}_\beta|} \sum_{X_\beta \subseteq \mathfrak{X}_\beta} \mathbf{P}(X_\beta^1\Delta X_\beta^2). \quad (6)$$

В [1] было доказано, что вероятность сет-операции симметрической разности по Минковскому двудольных множеств событий является псевдометрикой между ними. Поэтому было предложено использовать ее для измерения расстояния между двудольными множествами событий.

## 8 Понятие событие-терраски двудольного множества событий

Обозначим двудольное множество случайных событий, представленное в формуле (2), следующим образом:

$$Z = \{\mathcal{Y}, \mathfrak{X}\} = \{\mathcal{Y}_a, \mathfrak{X}_\beta, a \in A, \beta \in B\}.$$

Пусть задано двудольное множество случайных событий  $s$ , являющееся подмножеством двудольного множества событий  $Z$  (т.е.  $s \subseteq Z$ ):

$$s = \{\mathcal{Y}_{s_A}, \mathfrak{X}_{s_B}, s_A \subseteq A, s_B \subseteq B\}.$$

**Определение 14.** Событие-терраска двудольного множества случайных событий  $s$  представляет собой набор непересекающихся событий, где каждое событие является подмножеством соответствующего

множества событий  $\mathcal{Y}_a$  или  $\mathfrak{X}_\beta$ :

$$\text{ter}(s) = \text{ter}\{\mathcal{Y}_{s_A}, \mathfrak{X}_{s_B}\} = \\ = \bigcap_{a \in s_A} \text{ter}(\mathcal{Y}_a) \bigcap_{\beta \in s_B} \text{ter}(\mathfrak{X}_\beta) = \\ = \bigcap_{a \in s_A} \mathcal{Y}_a(r_a) \bigcap_{\beta \in s_B} \left( \bigcap_{x_\beta \in X_\beta} x_\beta \bigcap_{x_\beta \in X_\beta^c} x_\beta^c \right), \quad (7)$$

где  $s_A \subseteq A, s_B \subseteq B, r_a \in \mathcal{R}_a, X_\beta \subseteq \mathfrak{X}_\beta$ . Здесь множество  $\mathcal{R}_a$  представляет собой множество возможных значений случайной величины  $\xi_a$ :

$$\mathcal{R}_a = \{r_{a_1}, \dots, r_{a_{N_a}}\} \subset \mathbb{R}, a \in A.$$

## 9 Регрессия двудольных множеств случайных событий

В работе [4] были предложены условные вероятности двудольных множеств событий и, затем, они были применены для решения задачи статистической оценки функции теоретической эвентологической регрессии [5].

Пусть заданы двудольные множества случайных событий  $s$  и  $t$ , являющиеся подмножествами двудольного множества событий  $Z$  (т.е.  $s, t \subseteq Z$ ):

$$s = \{\mathcal{Y}_{s_A}, \mathfrak{X}_{s_B}, s_A \subseteq A, s_B \subseteq B\},$$

$$t = \{\mathcal{Y}_{t_A}, \mathfrak{X}_{t_B}, t_A \subseteq A, t_B \subseteq B\}.$$

**Определение 15.** Условная вероятность наступления событий из двудольного множества  $t$  при условии наступления событий из двудольного множества  $s$  имеет следующий вид:

$$\mathbf{P}(\text{ter}(t) | \text{ter}(s)) = p_{|\text{ter}(s)}(\text{ter}(t)) = \\ = \frac{\mathbf{P}(\text{ter}(t) \cap \text{ter}(s))}{p(\text{ter}(s))}, \quad (8)$$

где  $p(\text{ter}(s)) = \sum_{\text{ter}(s) \subseteq \hat{Z} \subseteq Z} P(\hat{Z})$ ,  $\text{ter}(s)$  и  $\text{ter}(t)$  задаются по вышеприведенной формуле (4).

**Определение 16.** Эвентологическая регрессия одного двудольного множества событий  $s \subseteq Z$  на другое двудольное множество событий  $t \subseteq Z$ , связывающая события-терраски  $\text{ter}(s)$ ,  $s \subseteq Z$  с событиями-террасками  $\text{ter}(t)$ ,  $t \subseteq Z$  имеет вид

$$\text{ter}(t) = \varphi(\text{ter}(s)), s \subseteq Z.$$

Поставим задачу статистической оценки функции теоретической эвентологической регрессии на уровне  $\alpha$ , значение которой на  $s \subseteq Z$  определяется как условный  $\alpha$ -квантиль случайного множества  $t$  по следующей формуле:

$$Q_\alpha(t | \text{ter}(s)) = \{\text{ter}(t) \subseteq Z : P(\text{ter}(t) | \text{ter}(s)) \geq \alpha\}$$

Условные вероятности, определяющие, какие элементы  $\text{ter}(t) \subseteq Z$  принадлежат значению функции Э-регрессии на  $s \subseteq Z$ , определяются формулами (4).

Как уже говорилось ранее, построение регрессии возможно только в случае наличия зависимости между величинами (в нашем случае между двудольными множествами). Наличие зависимости исследуется с помощью ковариации или корреляции.

В работе [6] было дано определение ковариации для двух событий-террасок и решена задача ее нахождения. Таким образом, был рассмотрен случай, когда множество состоит всего из двух событий-террасок двудольных множеств. Приведем это определение.

**Определение 17.** Ковариацией событий-террасок двудольных множеств случайных событий  $s$  и  $t$  называется

$$\text{Kov}_{\text{ter}(s), \text{ter}(t)} = \mathbf{P}(\text{ter}(s) \cap \text{ter}(t)) - \mathbf{P}(\text{ter}(s)) \mathbf{P}(\text{ter}(t)).$$

Обозначим границы Фреше для ковариации событий-террасок двудольных множеств событий следующим образом:

$$F_{\text{ter}(s), \text{ter}(t)}^+ = \min\{\mathbf{P}(\text{ter}(s)), \mathbf{P}(\text{ter}(t))\} - \mathbf{P}(\text{ter}(s))\mathbf{P}(\text{ter}(t))$$

— правая граница Фреше для ковариации событий-террасок двудольных множеств случайных событий,

$$F_{\text{ter}(s), \text{ter}(t)}^- = \max\{0, \mathbf{P}(\text{ter}(s)) + \mathbf{P}(\text{ter}(t)) - 1\} - \mathbf{P}(\text{ter}(s))\mathbf{P}(\text{ter}(t))$$

— левая граница Фреше для ковариации событий-террасок двудольных множеств случайных событий.

**Определение 18.** Корреляция событий-террасок двудольных множеств случайных событий  $s, t \subseteq Z$  определяется как соответствующая ковариация событий-террасок двудольных множеств событий  $\text{Kov}_{\text{ter}(s), \text{ter}(t)}$ , нормированная границами Фреше для ковариации двудольных множеств:

$$\text{Cor}_{\text{ter}(s), \text{ter}(t)} = \begin{cases} \frac{\text{Kov}_{\text{ter}(s), \text{ter}(t)}}{|F_{\text{ter}(s), \text{ter}(t)}^-|}, & \text{если } \text{Kov} \leq 0, \\ \frac{\text{Kov}_{\text{ter}(s), \text{ter}(t)}}{|F_{\text{ter}(s), \text{ter}(t)}^+|}, & \text{если } \text{Kov} > 0, \end{cases}$$

$F_{\text{ter}(s), \text{ter}(t)}^-$  и  $F_{\text{ter}(s), \text{ter}(t)}^+$  приведены выше.

## 10 Множество двудольных множеств событий

В работе [7] было предложено понятие множества двудольных множеств событий.

**Определение 5.** Множество двудольных множеств случайных событий представляет собой набор множеств событий, каждое из которых состоит из двух долей, причем их первая доля определяется случайными величинами, а вторая — случайными множествами:

$$S = \{s^1, \dots, s^n\}. \quad (9)$$

Возможны следующие разновидности структуры данного множества:

1. Случай однотипных двудольных множеств событий (все множества порождаются одним и тем же двудольным множеством случайных элементов, они описываются одним и тем же двудольным множеством  $Z$  и, следовательно, их терраски-события одинаковы. А различаются они для каждого множества  $s^i$  только вероятностями их наступления):

$$S = \{s^1 = \{\mathcal{Y}_a^1, \mathfrak{X}_\beta^1, a \in A, \beta \in B\}, \dots, s^n = \{\mathcal{Y}_a^n, \mathfrak{X}_\beta^n, a \in A, \beta \in B\}\}.$$

Этот случай является простейшим. Именно эта ситуация была рассмотрена в работах [1] и [7].

2. Случай, когда двудольные множества событий  $s^i$  являются подмножествами одного и того же двудольного множества событий  $Z$ :

$$S = \{s^1 = \{\mathcal{Y}_{s_A^1}, \mathfrak{X}_{s_B^1}, s_A^1 \subseteq A, s_B^1 \subseteq B\}, \dots, s^n = \{\mathcal{Y}_{s_A^n}, \mathfrak{X}_{s_B^n}, s_A^n \subseteq A, s_B^n \subseteq B\}\}.$$

3. Случай, когда двудольные множества событий  $s^i$  являются подмножествами различных двудольных множеств событий  $Z^i = \{\mathcal{Y}^i, \mathfrak{X}^i\} = \{\mathcal{Y}_a^i, \mathfrak{X}_\beta^i, a \in A^i, \beta \in B^i\}$ :

$$S = \{s^1 = \{\mathcal{Y}_{k_A^1}, \mathfrak{X}_{k_B^1}, k_A^1 \subseteq A^1, k_B^1 \subseteq B^1\}, \dots, s^n = \{\mathcal{Y}_{k_A^n}, \mathfrak{X}_{k_B^n}, k_A^n \subseteq A^n, k_B^n \subseteq B^n\}\}.$$

Данный случай является самым общим и самым сложным по своей структуре.

В работе [8] были подробно разобраны все перечисленные случаи, найдены их Э-распределения, приведены их практические примеры и набор возможных задач, приводящих к описываемым случаям.



## 11 Различные виды множества двудольных множеств и их распределения

### 11.1 Множество однотипных двудольных множеств событий

Любая задача системного или факторного анализа приводит нас к изучению такого множества.

Пусть имеется сложная система, которая состоит из  $n$  элементов. Каждый элемент характеризуется показателями, одни из которых являются числовыми, а другие — множественными. Согласно методу двудольных множеств событий, поведение каждого элемента системы можно охарактеризовать двудольным множеством событий  $\{\mathcal{Y}, \mathfrak{X}\}$ . Таким образом, сложной системе ставится в соответствие множество

$$S = \left\{ s^1 = \{ \mathcal{Y}_a^1, \mathfrak{X}_\beta^1, a \in A, \beta \in B \}, \dots, \right. \\ \left. s^n = \{ \mathcal{Y}_a^n, \mathfrak{X}_\beta^n, a \in A, \beta \in B \} \right\}.$$

Пусть задано распределение вероятностей событий двудольного множества  $S$ :

$$p(s) = p(\mathcal{Y}_a, \mathfrak{X}_\beta) = p(\mathbf{r}, \mathbf{X}) = \\ \mathbf{P} \left( \bigcap_{a \in A} \{ \mathcal{Y}_a(r_a) \} \right. \\ \left. \bigcap_{x_\beta \in X} x_\beta \bigcap_{x_\beta \in X^c} x_\beta^c, a \in A, \beta \in B \right), \\ x_\beta \in \mathfrak{X}, \beta \in B, \\ \mathbf{r} = \{r_a, r_a \in \mathbb{R}\}, \\ \mathbf{X} = \{X_\beta, X_\beta \subseteq \mathfrak{X}_\beta\}. \quad (10)$$

Тогда распределение вероятностей событий множества двудольных множеств событий  $S = \{s^1, \dots, s^n\}$   $p(S^t), S^t \subseteq S$  имеет вид:

$$p(S^t) = \mathbf{P} \left( \bigcap_{s^i \in S^t} s^i \right) = \\ = \mathbf{P} \left( \bigcap_{s^i \in S^t} \left\{ \bigcap_{a \in A} \mathcal{Y}_a^i(r_a) \right. \right. \\ \left. \left. \bigcap_{x_\beta \in X} x_\beta \bigcap_{x_\beta \in X^c} x_\beta^c \right\} \right), S^t \subseteq S. \quad (11)$$

Очевидно, что для практических задач достаточно сложно оценить вероятности распределения  $p(S^t), S^t \subseteq S$ . Поэтому для работы с ними лучше всего использовать сет-операцию по Минковскому над множеством двудольных множеств событий и ее вероятности.

**Определение 19.** Произвольную сет-операцию по Минковскому над множеством двудольных множеств событий можно определить как множество событий, полученных с помощью операций по Минковскому над соответствующими событиями из каждой доли:

$$\left( \mathcal{O} \right)_{s \in S} = \left\{ \left( \mathcal{O} \right)_{s \in S} \{ \mathcal{Y}_a \}, \left( \mathcal{O} \right)_{s \in S} \{ \mathfrak{X}_\beta \}, a \in A, \right. \\ \left. \beta \in B \right\} = \left\{ \mathcal{O}_{s \in S} \{ \mathcal{Y}_a(r_a) \}, \mathcal{O}_{s \in S} \{ X_\beta \}, r_a \in \mathcal{R}_a, \right. \\ \left. a \in A, X_\beta \subseteq \mathfrak{X}_\beta, \beta \in B \right\}. \quad (12)$$

Приведем определение и вероятность для сет-операции симметрической разности по Минковскому над множеством двудольных множеств.

**Определение 20.** Сет-операция симметрической разности по Минковскому над множеством двудольных множеств событий определяется как множество событий, полученных с помощью операций по Минковскому над соответствующими событиями из каждой доли:

$$\left( \Delta \right)_{s \in S} = \left\{ \left( \Delta \right)_{s \in S} \{ \mathcal{Y}_a \}, \left( \Delta \right)_{s \in S} \{ \mathfrak{X}_\beta \} \right\} = \\ \left\{ \Delta_{s \in S} \{ \mathcal{Y}_a(r_a) \}, \Delta_{s \in S} \{ X_\beta \}, r_a \in \mathcal{R}_a, a \in A, \right. \\ \left. X_\beta \subseteq \mathfrak{X}_\beta, \beta \in B \right\}.$$

**Определение 21.** Под оператором общей симметрической разности  $\Delta_{s \in S}$  над множеством событий  $S$  понимается среднее событие по множеству, вероятность которого определяется следующим образом:

$$\mathbf{P} \left( \Delta_{s \in S} s \right) = \frac{1}{|S|} \sum_{s, t \in S} P(s \Delta t). \quad (13)$$

Вероятность общей симметрической разности по Минковскому над множеством двудольных множеств событий выглядит следующим образом:

$$\mathbf{P} \left( \left( \Delta \right)_{s \in S} s \right) = \\ \frac{1}{|A|} \sum_{a \in A} \frac{1}{|\mathcal{Y}_a|} \sum_{r_a \in \mathcal{R}_a} \mathbf{P} \left( \Delta_{r_a \in \mathcal{R}_a} \{ \mathcal{Y}_a(r_a) \} \right) + \\ + \frac{1}{|B|} \sum_{\beta \in B} \frac{1}{|\mathfrak{X}_\beta|} \sum_{X_\beta \subseteq \mathfrak{X}_\beta} \mathbf{P} \left( \Delta_{s \in S} \{ X_\beta \} \right).$$

### 11.2 Множество подмножеств двудольных множеств событий

Теперь рассмотрим второй случай, когда двудольные множества событий  $s^i$  являются подмножествами одного и того же двудольного множества событий  $Z$ :

$$S = \left\{ s^1 = \left\{ \mathcal{Y}_{s_A^1}, \mathfrak{X}_{s_B^1}, s_A^1 \subseteq A, s_B^1 \subseteq B \right\}, \dots, \right. \\ \left. s^n = \left\{ \mathcal{Y}_{s_A^n}, \mathfrak{X}_{s_B^n}, s_A^n \subseteq A, s_B^n \subseteq B \right\} \right\}.$$

Распределение вероятностей событий множества двудольных множеств событий  $S = \{s^1, \dots, s^n\}$   $p(S^t), S^t \subseteq S$  для данного случая имеет следующий вид:

$$\begin{aligned} p(S^t) &= \mathbf{P} \left( \bigcap_{s^i \in S^t} s^i \right) = \\ &= \mathbf{P} \left( \bigcap_{s^i \in S^t} \left\{ \bigcap_{a \in s_A^i} \text{ter}(\mathcal{Y}_a) \bigcap_{\beta \in s_B^i} \text{ter}(\mathfrak{X}_\beta) \right\} \right) = \\ &= \mathbf{P} \left( \bigcap_{a \in s_A^i} \mathcal{Y}_a(r_a) \bigcap_{\beta \in s_B^i} \left( \bigcap_{x_\beta \in X_\beta} x_\beta \bigcap_{x_\beta \in X_\beta^c} x_\beta^c \right) \right), \\ s_A^i &\subseteq A, s_B^i \subseteq B, S^t \subseteq S. \end{aligned} \quad (14)$$

### 11.3 Множество подмножеств различных двудольных множеств событий

Теперь рассмотрим третий случай, когда двудольные множества событий  $s^i$  являются подмножествами различных двудольных множеств событий  $Z^i = \{\mathcal{Y}^i, \mathfrak{X}^i\} = \{\mathcal{Y}_a^i, \mathfrak{X}_\beta^i, a \in A^i, \beta \in B^i\}$ :

$$\begin{aligned} S &= \{s^1 = \{\mathcal{Y}_{k_A^1}, \mathfrak{X}_{k_B^1}, k_A^1 \subseteq A^1, k_B^1 \subseteq B^1\}, \dots, \\ s^n &= \{\mathcal{Y}_{k_A^n}, \mathfrak{X}_{k_B^n}, k_A^n \subseteq A^n, k_B^n \subseteq B^n\}\}, \end{aligned}$$

здесь под  $A^j$  понимается множество индексов для числовой доли двудольного множества  $s^j \in S$ , под  $B^j$  понимается множество индексов для множественной доли этого же двудольного множества. Обозначение  $k_A^j$  принято для индекса, который пробегает соответствующее множество индексов  $A^j$ . Аналогичным образом, обозначение  $k_B^j$  — для индекса, который пробегает соответствующее множество индексов  $B^j$ . Таким образом, множество событий  $\mathcal{Y}_{k_A^1} \subseteq \mathcal{Y}^1$ , а  $\mathfrak{X}_{k_B^1} \subseteq \mathfrak{X}^1$ .

Распределение вероятностей событий множества двудольных множеств событий  $S = \{s^1, \dots, s^n\}$   $p(S^t), S^t \subseteq S$  для данного случая имеет следующий вид:

$$\begin{aligned} p(S^t) &= \mathbf{P} \left( \bigcap_{s^i \in S^t} s^i \right) = \\ &= \mathbf{P} \left( \text{ter}(\mathcal{Y}_{k_A^1}) \bigcap \text{ter}(\mathcal{Y}_{k_A^2}) \dots \bigcap \text{ter}(\mathcal{Y}_{k_A^t}) \right. \\ &\quad \left. \text{ter}(\mathfrak{X}_{k_B^1}) \bigcap \text{ter}(\mathfrak{X}_{k_B^2}) \dots \bigcap \text{ter}(\mathfrak{X}_{k_B^t}) \right) = \\ &= \mathbf{P} \left( \bigcap_{k_A^1 \in A^1} \mathcal{Y}_{k_A^1} \dots \bigcap_{k_A^t \in A^t} \mathcal{Y}_{k_A^t} \right. \\ &\quad \bigcap_{k_B^1 \in B^1} \left( \bigcap_{x_\beta \in X_{k_B^1}} x_\beta \bigcap_{x_\beta \in X_{k_B^1}^c} x_\beta^c \right) \dots \\ &\quad \bigcap_{k_B^t \in B^t} \left( \bigcap_{x_\beta \in X_{k_B^t}} x_\beta \bigcap_{x_\beta \in X_{k_B^t}^c} x_\beta^c \right), \\ &\quad \mathcal{Y}_{k_A^1} \subseteq \mathcal{Y}^1, \mathfrak{X}_{k_B^1} \subseteq \mathfrak{X}^1, S^t \subseteq S. \end{aligned} \quad (15)$$

В данной формуле для простоты записи для подмножества  $S^t$  указано подмножество  $\{s^1, \dots, s^t\}$  со множествами индексов для числовой доли  $A^1, \dots, A^t$  и множествами индексов для множественной доли  $B^1, \dots, B^t$ , а также с соответствующими им индексами  $k_A^1, \dots, k_A^t$  и  $k_B^1, \dots, k_B^t$ . На самом деле речь идет не о первых  $t$  двудольных множествах (с первого по  $t$ -ый включительно), а о произвольном наборе из  $t$  двудольных множеств (т.е. о любом подмножестве  $S^t \subseteq S$ , с мощностью равной  $t$  (т.е.  $|S^t| = t$ )).

Выше было дано понятие эвентологической регрессии одного двудольного множества событий  $s \subseteq Z$  на другое двудольное множество событий  $t \subseteq Z$  и приведено подробное описание алгоритма решения задачи нахождения статистической оценки функции теоретической эвентологической регрессии данных двудольных множеств.

Очевидно, что по аналогии с классическим подходом к регрессионному анализу можно обобщить ситуацию и перейти к изучению ситуации множественной регрессии двудольных множеств. Когда исследуется регрессионная зависимость между некоторым набором двудольных множеств событий  $S = \{s^1, \dots, s^n\}$  и некоторым зависимым двудольным множеством  $t \in Z$ .

Как уже говорилось ранее, построение регрессии возможно только в случае наличия зависимости между величинами (в нашем случае между множеством двудольных множеств и другим множеством). Наличие зависимости исследуется с помощью ковариации или корреляции.

Понятие ковариации двудольных множеств можно распространить и на множество событий-террасок двудольных множеств.

**Определение 22.** Ковариацией множества  $S^t \subseteq S$ , где

$$\begin{aligned} S &= \{s^1 = \{\mathcal{Y}_{s_A^1}, \mathfrak{X}_{s_B^1}, s_A^1 \subseteq A, s_B^1 \subseteq B\}, \dots, \\ s^n &= \{\mathcal{Y}_{s_A^n}, \mathfrak{X}_{s_B^n}, s_A^n \subseteq A, s_B^n \subseteq B\}\}. \end{aligned}$$

называется

$$Kov_{S^t} = p(S^t) - \prod_{s^i \in S^t} p(s^i), \quad (16)$$

где  $p(s^i)$  — распределение вероятностей событий двудольного множества  $s^i$ , а распределение множества двудольных множеств  $p(S^t)$  находится по формуле (14).

Распределения  $p(s^i)$  имеют вид, представленный выше формулами (11), (14) или (15).

**Определение 23.** Множественная эвентологическая регрессия множества двудольных множеств событий  $S \subseteq Z$  на двудольное множество событий

$t \in Z$ , имеет вид

$$\text{ter}(t) = \varphi(\text{ter}(s)), \quad s \subseteq Z.$$

Вид уравнения регрессии находится путем решения задачи статистической оценки функции теоретической эвентологической регрессии на уровне  $\alpha$ , значение которой на  $S \subseteq Z$  определяется как условный  $\alpha$ -квантиль случайного множества  $t$  по следующей формуле:

$$Q_\alpha(t | \text{ter}(s)) = \{\text{ter}(t) \subseteq Z : P(\text{ter}(t) | \text{ter}(S)) \geq \alpha\}.$$

Наибольшую сложность в данной задаче вызывает определение вида условных вероятностей, определяющие, какие элементы  $\text{ter}(t) \subseteq Z$  принадлежат значению функции Э-регрессии на  $S \subseteq Z$ , что является следствием сложной структуры зависимостей элементов множества  $S$  и различных видов распределения данного множества.

Нахождение распределений условных вероятностей для данной ситуации и их изучение планируется выполнить в последующих работах.

## Список литературы

- [1] О. Ю. Воробьев. *Эвентология*. СФУ, Красноярск, 2007.
- [2] И. В. Баранова. Метод двудольных множеств событий в эвентологическом анализе социально-экономических систем. *Вестник Красноярского государственного университета: физико-математические науки*, 1:142–152, 2006.
- [3] О. Ю. Воробьев и И. В. Баранова. *Метод двудольных множеств событий в эвентологическом анализе сложных систем*. Ин-т естеств. и гуманит. наук, Красноярск, 2007.
- [4] И. В. Баранова. Условное распределение двудольного множества событий. In *Труды IX международной конференции по финансово-актуарной математике и эвентоконвергенции*, pages 51–59. СФУ, Красноярск, 2010.
- [5] И. В. Баранова. Метод двудольных множеств событий в факторном анализе разнотипных данных. In *Интеллектуальные системы принятия решений и проблемы вычислительного интеллекта: материалы международной научной конференции*, volume т.1, pages 19–23. ХНТУ, Херсон, 2009.
- [6] И. В. Баранова. Коварияция и регрессия двудольных множеств событий. In *Труды VIII Международной конференции по финансово-актуарной математике и смежным вопросам*, pages 62–67. СФУ, Красноярск, 2009.
- [7] I. V. Baranova. Bipartite set of random events method application in factorial analysis of complex systems. In *Proceedings of the Second International Conference «Problems of Cybernetics and Informatics»*, volume volume 1, pages 259–262. Azerbaijan National Academy of Sciences, Baku, Azerbaijan, 2008.
- [8] И. В. Баранова. Множества двудольных множеств событий и их применение. In *Труды XI международной конференции по финансово-актуарной математике и эвентологии безопасности*, pages 72–79. НИИШПБ,СФУ, Красноярск, 2012.

# Применение энтропии для исследования распределений двудольных множеств событий

Ирина Владимировна Баранова

Сибирский Федеральный Университет  
Институт математики и фундаментальной информатики  
Красноярск  
irinabar@yandex.ru

Елена Николаевна Темерова

Сибирский Федеральный Университет  
Институт математики и фундаментальной информатики  
Красноярск  
TemerovaL@yandex.ru

**Аннотация.** В работе рассматриваются различные подходы к понятию энтропии, в том числе и эвентологический подход. Особое внимание уделяется понятию двудольного множества событий, состоящего из множеств событий, первая доля которых определяется случайными величинами, а вторая — случайными множествами, и его энтропии.

**Ключевые слова.** Термодинамическая энтропия, статистическая энтропия, информационная энтропия, информация, энтропия в теории вероятности, эвентологическая энтропия, относительная энтропия, случайное событие, двудольное множество событий.

## 1 Введение

Энтропия (от др.-греч. — Ч поворот, превращение) понимается в естественных науках как мера беспорядка системы, состоящей из многих элементов.

Понятие энтропии впервые возникло в термодинамике и получило распространение в различных областях знаний: статистике, лингвистике, математике, химии, биологии и других [1, 2]. Соответственно, понятие энтропия имеет множество трактовок.

Так, в термодинамике энтропия является мерой необратимого рассеивания энергии, в теории информации энтропия понимается как мера неопределенности некоторой ситуации, в теории вероятностей — как мера неопределенности испытаний с различными исходами.

В рамках эвентологического подхода, применяемого для изучения множества событий, О.Ю. Воробьевым было предложено понятие эвентологической энтропии [3, 4], которая рассматривается как мера неопределенности эвентологического распределения множества событий. Для более сложных систем, поведение которых описывается числовыми и множественными данными, И. В. Барановой было построено обобщение понятия энтропии для случая двудольных множеств событий [5].

В данной работе особое внимание уделяется понятию

энтропии двудольных множеств событий.

## 2 Энтропия в термодинамике

Впервые понятие энтропии было введено Р.Ю. Клаузиусом в 1865 г. Введение этого понятия связано с поиском координаты теплообмена, т.е. физической величины, неизбежно изменяющейся в процессе теплообмена и остающейся неизменной в его отсутствие. Клаузиус нашел эту координату для частного случая равновесного (обратимого) теплообмена путем разбиения произвольного цикла тепловой машины серией адиабат и изотерм на ряд элементарных обратимых циклов Карно.

Название параметра  $S$ , которое дал ему Клаузиус, подчеркивало важное свойство энтропии возрастать и в отсутствие теплообмена (“энтропия” в переводе с греческого означает “внутреннее превращение”). Это было нечто необычное для науки того времени.

Первый закон термодинамики гласит, что энергию невозможно создать или уничтожить. Следовательно, количество энергии во вселенной постоянно, в любой момент времени оно такое же, как и было при ее создании.

На основе второго начала термодинамики, которое можно математически представить в виде неравенства Краузиуса:

$$\oint \frac{\partial Q}{T} \geq 0,$$

было введено понятие энтропии. Второй закон термодинамики гласит, что изменение энтропии изолированной системы всегда положительно  $dS > 0$  или равно нулю в случае достижения энтропией своего максимального значения, т.е. энтропия изолированной системы не может убывать. Поэтому формула для термодинамической энтропии имеет вид:

$$\Delta S = \frac{\Delta Q}{T}.$$

## 3 Статистическая энтропия

Поиски физического смысла энтропии и попытки найти альтернативу неизбежному, казалось бы, выводу о “тепловой смерти Вселенной” привели к статистическому толкованию второго начала термодинамики. Л. Больцман, полагая, что возрастание энтропии

в необратимых процессах отражает стремление природы к более вероятному состоянию, пришел к выводу, что зависимость между энтропией  $S$  и термодинамической вероятностью  $\Omega$  имеет вид:

$$S = k \cdot \ln |W|,$$

где  $k$  — константа, названная впоследствии его именем.

## 4 Информационная энтропия

В теории информации энтропия — мера хаотичности информации или мера внутренней неупорядоченности системы. Энтропия увеличивается при хаотическом распределении информационных ресурсов и уменьшается при их упорядочении.

В 1948 году, исследуя проблему рациональной передачи информации через зашумленный коммуникационный канал, Клод Шеннон предложил революционный вероятностный подход к пониманию коммуникаций и создал первую, истинно математическую, теорию энтропии. Его сенсационные идеи быстро послужили основой разработки двух основных направлений:

- теории информации, которая использует понятие вероятности, и эргодической теории, изучающей статистические характеристики данных и коммуникационных систем,
- теории кодирования, в которой используются главным образом алгебраические и геометрические инструменты для разработки эффективных кодов.

Взаимосвязь энтропии и информации нашла отражение в формуле:

$$H + I = 1,$$

где  $H$  — энтропия,  $I$  — информация. Этот вывод количественно был обоснован Бриллюэном.

Информационная энтропия для независимых случайных событий  $x$  с  $n$  возможными состояниями (от 1 до  $n$ ) рассчитывается по формуле:

$$H(x) = - \sum_{i=1}^n \log p(i)$$

Эта величина также называется средней энтропией сообщения.

Величина

$$\log \frac{1}{p(i)}$$

называется частной энтропией, характеризующей только  $i$ -е состояние. От основания логарифма зависит единица измерения информации и энтропии: бит, нат, хартли.

Итак, информационная энтропия — мера хаотичности информации или мера внутренней неупорядоченности информационной системы. Энтропия увеличивается при хаотичном распределении информации и уменьшается при упорядоченности.

## 5 Эвентологическая энтропия

Пусть  $(\Omega, \mathcal{F}, P)$  — эвентологическое пространство, а  $\mathfrak{X} \subseteq \mathcal{F}$  — конечное множество событий, выбранных из алгебры событий  $\mathcal{F}$  этого пространства. Каждое  $N$ -множество характеризуется набором из  $2^N$  вероятностей. Множество  $\mathfrak{X}$  характеризуется набором из вероятностей:

$$p(X) = P(\text{ter}(X)), X \subseteq \mathfrak{X}$$

— вероятности событий-террасок

$$\text{ter}(X) = \bigcap_{x \subseteq X} x \bigcap_{x \subseteq X^c} x^c.$$

Данный набор называется эвентологическим распределением множества событий  $\mathfrak{X}$  мощности  $N$ , т.е.  $N = |\mathfrak{X}|$ .

Энтропией множества случайных событий  $\mathfrak{X} \subseteq \mathcal{F}$  (энтропией эвентологического распределения множества событий  $\mathfrak{X}$ ) называется величина

$$S(\mathfrak{X}) = - \sum_{X \subseteq \mathfrak{X}} p(X) \log p(X).$$

Известно, что абсолютного максимума энтропия  $S$  достигает на равномерном  $\mathfrak{X}$ -распределении:

$$p(X) = \frac{1}{2^{|\mathfrak{X}|}}, X \subseteq \mathfrak{X}.$$

Энтропия — это теоретико-информационная мера степени неопределенности случайного множества. Хорошо известна интерпретация этой энтропии: математическое ожидание информации, полученной в результате соответствующего случайного эксперимента.

Относительная энтропия  $H_{p/q}$  — это энтропия эвентологического распределения  $p(X)$  множества событий  $\mathfrak{X}$  относительно эвентологического распределения  $q(X)$ :

$$H_{p/q} = \sum_{X \subseteq \mathfrak{X}} p(X) \log \frac{p(X)}{q(X)}$$

## 6 Энтропия двудольных множеств

В [5] И. В. Барановой было введено понятие двудольного множества случайных событий, описывающее поведение элемента системы, и показана его



связь с двудольным множеством случайных элементов. Возникновение данного понятия является следствием применения эвентологического подхода к анализу сложных систем, поведение которых характеризуется разнотипными данными: числовыми и множественными.

В ситуации, когда поведение каждого элемента сложной системы характеризуется данными, одна часть которых является числовой, а другая часть — множественной, объект, порождающий данную двудольную статистику, было предложено представить как двудольное множество случайных элементов.

Случайное множество событий под  $\mathfrak{X}$  определяется как измеримое отображение

$$K : (\Omega, \mathcal{F}, \mathbf{P}) \rightarrow (2^{\mathfrak{X}}, 2^{2^{\mathfrak{X}}}),$$

где  $2^{\mathfrak{X}}$  — множество всех подмножеств множества  $\mathfrak{X}$ .

**Определение 24.** Двудольным множеством случайных элементов называется множество  $\{\xi, \mathbf{K}\}$ , представимое в следующем виде:

$$\{\xi, \mathbf{K}\} = \xi \cup \mathbf{K} = \{\xi_a, a \in A, K_\beta, \beta \in B\}, \quad (1)$$

здесь первая доля — это случайные величины  $\xi = \{\xi_a, a \in A\}$ , вторая — случайные множества событий  $\mathbf{K} = \{K_\beta, \beta \in B\}$ ,  $A$  — множество индексов случайных величин,  $B$  — множество индексов случайных множеств событий.

**Определение 25.** Двудольное множество случайных событий представляет собой объединение двух множеств — множества событий, которое определяется случайными величинами, и множества событий, которое определяется случайными множествами событий:

$$\{\mathcal{Y}, \mathfrak{X}\} = \{\mathcal{Y}_a, \mathfrak{X}_\beta, a \in A, \beta \in B\}, \quad (2)$$

здесь случайным величинам ставится в соответствие множество событий  $\mathcal{Y}$ :

$$\xi \implies \mathcal{Y} = \sum_{a \in A} \mathcal{Y}_a, \mathcal{Y}_a = \{\mathcal{Y}_a(r_a), r_a \in \mathcal{R}_a\}.$$

а случайным множествам ставится в соответствие множество всех возможных случайных событий  $\mathfrak{X}$ :

$$\mathbf{K} \iff \mathfrak{X} = \sum_{\beta \in B} \mathfrak{X}_\beta, \beta \in B.$$

Здесь множество  $\mathcal{R}_a$  представляет собой множество возможных значений случайной величины  $\xi_a$ :

$$\mathcal{R}_a = \{r_{a_1}, \dots, r_{a_{N_a}}\} \subset \mathbb{R}, a \in A.$$

Полной характеристикой двудольного множества случайных событий служит его эвентологическое распределение. В работе [5] приведены все формы

эвентологического распределения двудольного множества случайных событий. В данной работе покажем вид одной из них — распределение вероятностей событий двудольного множества  $s$ :

$$\begin{aligned} p(\mathcal{Y}_a, \mathfrak{X}_\beta) &= p(\mathbf{r}, \mathbf{X}) = \\ &\mathbf{P}\left(\bigcap_{a \in A} \{\mathcal{Y}_a(r_a)\} \right. \\ &\left. \bigcap_{x_\beta \in X} x_\beta \bigcap_{x_\beta \in X^c} x_\beta^c, a \in A, \beta \in B\right), \end{aligned} \quad (3)$$

$$x \in \mathfrak{X}', x_\beta \in \mathfrak{X}, \beta \in B,$$

$$\mathbf{r} = \{r_a, r_a \in \mathbb{R}\},$$

$$\mathbf{X} = \{X_\beta, X_\beta \subseteq \mathfrak{X}_\beta\}.$$

Обозначим двудольное множество случайных событий, представленное в формуле (2), следующим образом:

$$Z = \{\mathcal{Y}, \mathfrak{X}\} = \{\mathcal{Y}_a, \mathfrak{X}_\beta, a \in A, \beta \in B\}.$$

Пусть задано двудольное множество случайных событий  $z$ , являющееся подмножеством двудольного множества событий  $Z$  (т.е.  $z \subseteq Z$ ):

$$z = \{\mathcal{Y}_{s_A}, \mathfrak{X}_{s_B}, s_A \subseteq A, s_B \subseteq B\}.$$

**Определение 26.** Событие-терраска двудольного множества случайных событий  $z$  представляет собой набор непересекающихся событий, где каждое событие является подмножеством соответствующего множества событий  $\mathcal{Y}_a$  или  $\mathfrak{X}_\beta$ :

$$\begin{aligned} \text{ter}(z) &= \text{ter}\{\mathcal{Y}_{s_A}, \mathfrak{X}_{s_B}\} = \\ &= \bigcap_{a \in s_A} \text{ter}(\mathcal{Y}_a) \bigcap_{\beta \in s_B} \text{ter}(\mathfrak{X}_\beta) = \\ &= \bigcap_{a \in s_A} \mathcal{Y}_a(r_a) \bigcap_{\beta \in s_B} \left( \bigcap_{x_\beta \in X_\beta} x_\beta \bigcap_{x_\beta \in X_\beta^c} x_\beta^c \right), \end{aligned} \quad (4)$$

где  $s_A \subseteq A, s_B \subseteq B, r_a \in \mathcal{R}_a, X_\beta \subseteq \mathfrak{X}_\beta$ . Здесь множество  $\mathcal{R}_a$  представляет собой множество возможных значений случайной величины  $\xi_a$ :

$$\mathcal{R}_a = \{r_{a_1}, \dots, r_{a_{N_a}}\} \subset \mathbb{R}, a \in A.$$

Вероятность события-терраски двудольного множества  $z$  вычисляется по следующей формуле:

$$\begin{aligned} \mathbf{P}(\text{ter}(z)) &= \mathbf{P}\left(\bigcap_{a \in s_A} \text{ter}(\mathcal{Y}_a) \bigcap_{\beta \in s_B} \text{ter}(\mathfrak{X}_\beta)\right) = \\ &= \mathbf{P}\left(\bigcap_{a \in s_A} \mathcal{Y}_a(r_a) \bigcap_{\beta \in s_B} \left( \bigcap_{x_\beta \in X_\beta} x_\beta \bigcap_{x_\beta \in X_\beta^c} x_\beta^c \right)\right). \end{aligned}$$

Обобщая вышеприведенное понятие эвентологической энтропии множества событий, было введено понятие **энтропии для двудольного множества случайных событий**  $z \subseteq Z$  по следующей формуле:

$$H(z) = - \sum_{\text{ter}(z)} p(\text{ter}(z)) \ln p(\text{ter}(z)), \quad (5)$$

где  $p(\text{ter}(z))$  — вероятности событий-террасок двудольного множества событий  $z$ , вычисляемые по формуле (5).

Энтропия здесь понимается как мера хаотичности множества  $z$  (неопределенность появления какого-либо события из множества  $\mathfrak{X}_B$  и  $\mathfrak{Y}_A$ ). Таким образом, энтропию можно интерпретировать как меру неопределенности двудольного множества случайных событий.

Своего максимума энтропия  $H(z)$  достигает в случае равномерного  $\mathfrak{E}$ -распределения. И чем больше система стремиться к равновероятностной, тем ее энтропия соответственно больше.

Определим число подмножеств двудольного множества  $S$ :

$$2^{|S|} = 2^{|\mathfrak{X}_1| + \dots + |\mathfrak{X}_{S_B}|} * |\mathcal{R}_1| * \dots * |\mathcal{R}_a|. \quad (6)$$

Максимальная энтропия будет вычисляться следующим образом:

$$H_{\max}(S) = - \sum_{n=1}^{2^{|S|}} \frac{1}{2^{|S|}} * \ln \frac{1}{2^{|S|}} = - \ln \frac{1}{2^{|S|}}. \quad (7)$$

Определим теперь минимум энтропии. Это случай, когда система находится в одном состоянии, т.е. вероятность только одного из  $2^{|S|}$  подмножеств двудольного множества  $S$  равна 1.

$$H_{\min}(S) = -(1 \cdot \ln 1 + \sum_{n=1}^{2^{|S|}-1} 0 \cdot \ln 0) = 0. \quad (8)$$

Минимальная энтропия равна 0, т.о. неопределенности нет, система находится в одном состоянии с максимальной вероятностью.

В работе [6] было рассмотрено понятие множества двудольных множеств событий.

**Определение 27.** Множество двудольных множеств случайных событий представляет собой набор множеств событий, каждое из которых состоит из двух долей, причем их первая доля определяется случайными величинами, а вторая — случайными множествами:

$$S = \{s^1, \dots, s^n\}. \quad (9)$$

В этой же работе было подробно рассмотрено эвентологическое распределение вероятностей множества двудольных множеств событий.

Можно ввести понятие энтропии и для множества двудольных множеств. Данное понятие будет иметь смысл хаотичности системы — множества элементов (т.е. меру разброса двудольных множеств относительно друг друга). В последующих работах будет произведен вывод формулы энтропии для множества двудольных множеств событий.

## Список литературы

- [1] М. В. Волькенштейн. *Энтропия и информация*. Наука, М., 1986.
- [2] А. Я. Хинчин. Понятие энтропии в теории вероятностей. *Успехи математических наук*, т.VIII, вып. 3:3–20, 1953.
- [3] О. Ю. Воробьев. *Эвентология*. Сибирский Федеральный Университет, Красноярск, 2007.
- [4] О.Ю. Воробьев и Н.А. Лукьянова. Энтропийные свойства мультипликативно-усеченных аппроксимаций эвентологических распределений. *Вестник Сибирского федерального университета. Математика и физика*, 2:319–331, 2009.
- [5] О.Ю. Воробьев и И. В. Баранова. *Метод двудольных множеств событий в эвентологическом анализе сложных систем*. Ин-т естеств. и гуманитар. наук, Красноярск, 2007.
- [6] И. В. Баранова. Множества двудольных множеств событий и их применение. In *Труды XI международной конференции по финансово-актуарной математике и эвентологии безопасности*, pages 72–79. НИИППБ,СФУ, Красноярск, 2012.

## Численная идентификация правой части специального вида системы двух уравнений параболического типа

**Белосохова Анна Владимировна**

Сибирский федеральный университет  
Институт математики и фундаментальной информатики  
Красноярск  
Honey\_m00n@mail.ru

**Распопов Виталий Евгеньевич**

Сибирский федеральный университет  
Институт математики и фундаментальной информатики  
Красноярск

**Аннотация.** В данной работе с помощью двух математических моделей численно исследуется предполагаемое распространение загрязняющих веществ, выбрасываемых в атмосферу Енисейским ферросплавным заводом (ЕФЗ). Разработан программный продукт с дружественным интерфейсом, проведены вычислительные эксперименты.

**Ключевые слова.** Модель гаусса, уравнение конвекции-диффузии, Енисейский ферросплавный завод, схема предиктор — корректор, вычислительный эксперимент, концентрация загрязняющего вещества.

В связи с предстоящим строительством вблизи Красноярска Енисейского ферросплавного завода, возникает вопрос о влиянии выбросов этого завода на экологически значимую зону г. Красноярска и ближайших населённых пунктов.

Размещение Енисейского ферросплавного завода предполагается на территории бывшего ООО «Завод тяжелого машиностроения» в Емельяновском районе на расстоянии 1,5 км от пос. Придорожный, 3,5 км от д. Старцево, 5 км от с.Шуваево и порядка 10 км от г. Красноярска.

Предполагаемое воздействие на окружающую среду Енисейского ферросплавного завода: выбросы загрязняющих веществ (далее — ЗВ) в атмосферный воздух от предприятия составят 4207 т/год по 37 ингрдиентам, в том числе:

ЗВ 1 класса опасности: свинец и его неорганические соединения — 0,00004 т/год;

ЗВ 2 класса опасности: марганец и его соединения — 184,7 т/год, оксид кальция - 26,4 т/год, триоксид диАлюминия — 11,5 т/год;

ЗВ 3 класса опасности: серы диоксид — 2695,0 т/год, азота диоксид 344,3 т/год, пыль неорганическая — 83,8 т/год, азота оксид — 9,3 т/год, сажа - 8,1 т/год, оксид магния — 6,4 т/год;

ЗВ 4 класса опасности - оксид углерода — 815,7 т/год.

В работе с помощью двух математических моделей

численно исследуется распространение агрегированной примеси в атмосфере.

Пусть  $q(t, x, y, z)$  — функция, значения которой в момент времени  $t$  в точке  $(x, y, z)$  трёхмерного евклидова пространства  $E^3$  совпадают со значениями концентрации некоторого вещества (примеси) в этой точке. Сначала для расчёта средних значений концентрации примеси в экологически значимой зоне в работе применяется [1, 2, 3, 4, 5] гауссова модель распространения примеси (или модель Сеттона). Согласно этой модели, изменения концентрации примеси от непрерывного точечного источника примеси подчиняются нормальному закону распределения [3]:

$$q(t, x, y, z) = \int_0^{t_k} Q e^{\left( \frac{(x-x_0-V_x t)^2}{(\sigma_x)^2 t} + \frac{(y-y_0)^2}{(\sigma_y)^2 t} + \frac{(z-z_0)^2}{(\sigma_z)^2 t} \right)} \frac{1}{(\sqrt{2\pi t})^3 \sigma_x \sigma_y \sigma_z} dt, \quad (1)$$

где

$q(t, x, y, z)$  - концентрация вещества;  $(x_0, y_0, z_0)$  — координаты источника примеси;  $Q$  — коэффициент, характеризующий мощность источника примеси;

$V_x$  — коэффициент, характеризующий скорость ветра (предполагается, что направление ветра совпадает с осью  $OX$ );

$\sigma_x, \sigma_y, \sigma_z$  — средние квадратичные отклонения частиц примеси в момент времени  $t$  соответственно вдоль координатных осей  $OX, OY, OZ$  [1, 2, 3]:

$$\begin{aligned} \sigma_x^2 &= \frac{2}{h} \int_0^h K_x(z) dz, \\ \sigma_y^2 &= \frac{2}{h} \int_0^h K_y(z) dz, \\ \sigma_z^2 &= \frac{2}{h} \int_0^h K_z(z) dz, \end{aligned} \quad (2)$$

здесь  $K_x, K_y, K_z$  - коэффициенты турбулентной диффузии вдоль координатных осей  $OX, OY, OZ$ ;

$h$  - высота приземного слоя.

Для проведения численных расчетов по формуле (1), покрываем расчетную область равномерной сеткой  $(x_i, y_j, z_k)$  с шагами  $\Delta x, \Delta y, \Delta z$  вдоль осей  $OX, OY, OZ$  соответственно.

Средние квадратичные отклонения частиц примеси в момент времени  $t$  вдоль координатных осей  $OX, OY, OZ$  пользуясь формулами (2) можно записать в виде [1, 2]:

$$\sigma_x^2 \approx \sigma_y^2 = \frac{2}{h} \int_0^h K_0 V_{x1} z^n dz = \frac{2K_0 V_{x1} h^n}{n+1}, \quad (3)$$

$$K_0 \approx 0,1 \div 1, n \approx 0, 15.$$

$$\sigma_z^2 = \frac{2}{h} \int_0^h K_z(z) dz = \frac{2K_1 h^m}{m+1}, \quad (4)$$

$$K_1 \approx 0,1 \div 3, m \approx 1.$$

где  $V_{x1}$  — коэффициент, характеризующий скорость ветра при  $z = 1$ .

Определив по формулам (3), (4) во всех узлах сетки  $(x_i, y_j, z_k)$  вспомогательные параметры  $\Delta x(z_k)$ ,  $\Delta y(z_k)$ ,  $\Delta z(z_k)$ , аппроксимируем интеграл суммой [3]:

$$q(t, x_i, y_j, z_k) = \sum_{l=0}^T \frac{Q e^{-\frac{1}{2} \left( \frac{(x_i - x_0 - V_x(x_k) \Delta t)^2}{(\sigma_x(z_k))^2 \Delta t} + \frac{(y_j - y_0)^2}{(\sigma_y(z_k))^2 \Delta t} + \frac{(z_k - z_0)^2}{(\sigma_z(z_k))^2 \Delta t} \right)}}{(\sqrt{2\pi t})^3 \sigma_x \sigma_y \sigma_z} \cdot \Delta t, \quad (5)$$

где

$$V_x(z) = V_{x1} \left( \frac{z}{z_1} \right)^n,$$

где  $z_1 = 1$ ;

$$T = \left\lceil \frac{t}{\Delta t} \right\rceil,$$

где  $t$  — время действия соответствующего источника,  $\Delta t$  — шаг дискретизации по времени,  $t > 0$ .

Для реализации гауссовой модели рассеяния примеси в атмосфере была составлена компьютерная программа на языке C++, в среде разработки приложений code gear от Borland. В программе есть многофункциональное меню, в котором пользователь может сам задать параметры необходимые для расчёта концентрации в экологически значимой зоне, тем самым расчёты проводились для различных высот источника загрязнений и высотах приземного слоя, при различных метеорологических условиях и временных параметрах. Полученные результаты расчетов представлены в виде таблиц, а также визуально отражены на карте г. Красноярска. Примеры результатов расчётов представлены на следующих рисунках.

На рисунках, черным цветом обозначены наиболее высокие значения концентрации загрязняющей примеси, более светлым цветом средние значения концентрации и самым светлым цветом наименьшие значения.

Заметим, что гауссова модель рассеяния примеси дает результаты, хорошо согласующиеся с экспериментальными данными, но только для источников примеси, расположенных на небольшой высоте и действующих в течение

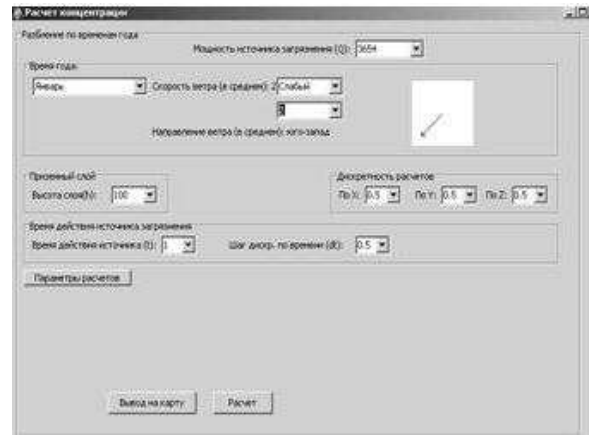


Рис. 1: Многопользовательское меню



Рис. 2: Расчетный период: апрель,  $h = 25$ ,  $Q = 1564$ ,  $\Delta t = 0,5$ ,  $t = 2$ ,  $V = 7/$ .

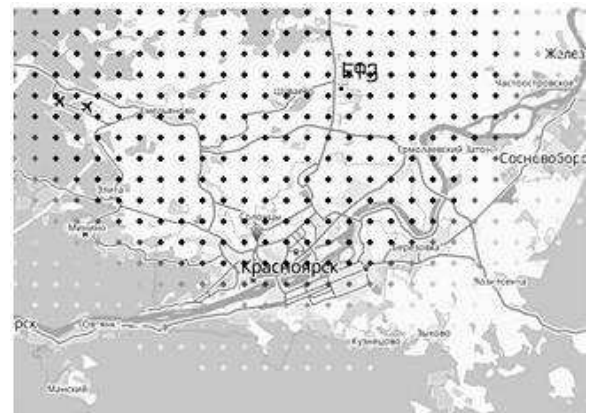


Рис. 3: Расчетный период: осень,  $h = 100$ ,  $Q = 4200$ ,  $\Delta t = 1$ ,  $t = 3$ ,  $V = 4/$ .



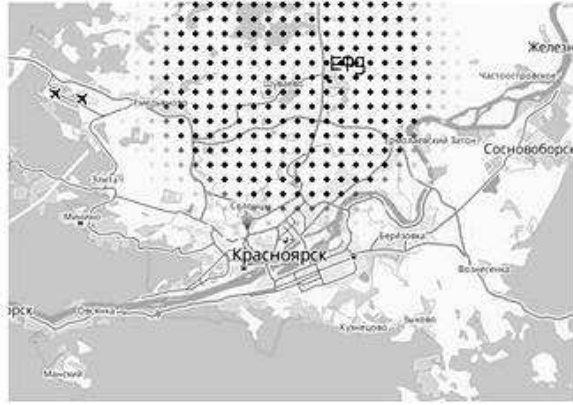


Рис. 4: Расчетный период: лето,  $h = 4$ ,  $Q = 313$ ,  $\Delta t = 0,25$ ,  $t = 1$ ,  $V = 2$ .

небольшого промежутка времени [1]. Поэтому наряду с приведенной моделью рассмотрена математическая модель, основанная на уравнении конвекции-диффузии [6]:

$$\begin{aligned} \frac{dq}{dt} + V_x \frac{dq}{dx} + V_y \frac{dq}{dy} + V_z \frac{dq}{dz} + \alpha q = \\ = \frac{d}{dx} K_x \frac{dq}{dx} + \frac{d}{dy} K_y \frac{dq}{dy} + \frac{d}{dz} K_z \frac{dq}{dz} + f(t, x, y, z) \end{aligned} \quad (6)$$

с начальным  $q_0(t_0, x, y, z) = \phi(x, y, z)$  и граничными  $K_z \frac{dq}{dz} = 0$  при  $z = z_0$  условиями, где  $V = \langle V_x, V_y, V_z \rangle$  - вектор, скорости частиц примеси;  $\alpha$  - коэффициент, характеризующий химическую активность примеси;  $f(t, x, y, z)$  - функция источников примеси.

Для численного решения уравнение (6) аппроксимируется схемой предиктор-корректор [7]:

$$\begin{aligned} \frac{u^{n+\frac{1}{6}} - u^n}{\Delta \frac{\tau}{2}} &= \Lambda_1 u^{n+\frac{1}{6}}, \\ \frac{u^{n+\frac{2}{6}} - u^{n+\frac{1}{6}}}{\Delta \frac{\tau}{2}} &= \Lambda_2 u^{n+\frac{2}{6}}, \\ \frac{u^{n+\frac{1}{2}} - u^{n+\frac{2}{6}}}{\Delta \frac{\tau}{2}} &= \Lambda_3 u^{n+\frac{1}{2}}, \\ \frac{u^{n+1} - u^n}{\Delta \tau} &= \Lambda_1 u^{n+\frac{1}{2}} + \Lambda_2 u^{n+\frac{1}{6}} + \Lambda_3 u^{n+\frac{1}{6}} + f^{n+\frac{1}{2}}, \end{aligned} \quad (7)$$

где  $\Lambda_p$ , ( $p = 1, 2, 3$ ) - соответствующие одномерные разностные операторы:

$$\begin{aligned} \Lambda_1 q &= V_x \frac{dq}{dx} - \frac{d}{dx} K_x \frac{dq}{dx} \\ \Lambda_2 q &= V_y \frac{dq}{dy} - \frac{d}{dy} K_y \frac{dq}{dy} \\ \Lambda_3 q &= V_z \frac{dq}{dz} - \frac{d}{dz} K_z \frac{dq}{dz} + \alpha q \end{aligned} \quad (8)$$

Коэффициенты  $i, j, k$  в формуле (8) упущены.

Данная схема имеет второй порядок аппроксимации по всем переменным и абсолютно устойчива.

Проведённые расчёты показывают, что велика вероятность формирования метеорологических условий, способствующих накоплению загрязняющих веществ в приземном слое атмосферы после строительства ЕФЗ. В этих условиях возникает риск повышения загрязнения атмосферного воздуха в г. Красноярске и близлежащих населённых пунктах [8]. Так как в работе ведутся расчёты для агрегированной смеси, то при совместном присутствии в атмосферном воздухе нескольких веществ, обладающих суммацией действия, сумма отношений их концентраций к ПДК не должна превышать 1 при расчёте по формуле:

$$\sum_{i=0}^n \frac{q_i}{PDK_i} \leq 1 \quad (9)$$

По формуле (8) была проверена сумма концентраций трёх веществ: диоксида азота, диоксида серы и оксида углерода. Эти величины взяты нами из полученной концентрации агрегированной примеси, согласно заявленной заводом документации. Проверка показала, что сумма (8) для концентраций этих трёх веществ превышает 1, а следовательно превышает допустимый уровень концентраций загрязнений этих веществ в атмосфере.

## Список литературы

- [1] Берлянд М.Е. *Современные проблемы атмосферной диффузии и загрязнения атмосферы*. Л.: Гидрометеоиздат, 1975.
- [2] Бызова Н.Л., Гаргер Е.К., Иванов В.И. *Экспериментальные исследования атмосферной диффузии и расчеты рассеяния примеси*. Л.: Гидрометеоиздат, 1989.
- [3] Ионисян А.С. *Математическое моделирование процесса распространения примеси активной примеси в свободной и облачной атмосфере*. Ставрополь, 2003.
- [4] Ионисян А.С., Тоторкулов Х.А. *О гауссовой модели распространения примеси в атмосфере*. Ставрополь, Издательство СГУ, 2002.
- [5] Семенчин Е.А. *Аналитические решения краевых задач в математической модели атмосферной диффузии*. Ставрополь, Издательство СКИУУ, 1993.
- [6] Марчук Г.И. *Математическое моделирование в проблеме окружающей среды*. М.: Наука, 2002.
- [7] Яненко Н.Н. *Метод дробных шагов решения многомерных задач математической физики*. Новосибирск, Издательство «Наука», 1967.
- [8] Сакаш Г.В. «Енисейский ферросплавный завод». Доклад на общественных слушаниях по проекту ЗАО «Чек-СУ.В».



# Решение обратной задачи многокритериального выбора в методологии метода анализа иерархий

**Бескорсый Никита Сергеевич**

Сибирский федеральный университет  
Институт математики и фундаментальной информатики  
Красноярск  
besn1989gmail.com

**Олейников Борис Васильевич**

Сибирский федеральный университет  
Институт математики и фундаментальной информатики  
Красноярск  
oleynik48@mail.ru

**Аннотация.** В работе формулируется обратная задача в рамках методологии анализа иерархий Саати и предлагается два способа ее решения. Подробно рассматривается способ основанный на решении матричного уравнения. Описывается разработанное авторами программное обеспечение и приводятся результаты практического применения при обработке медицинских данных.

**Ключевые слова.** Обратная задача, многокритериальный выбор, метод анализа иерархий, Саати, псевдообращение.

## 1 Введение

Задачи принятия решений заключаются в выборе из множества возможных вариантов некоторого оптимального подмножества, в частном случае — одного варианта. Выделение требуемых альтернатив производится на основе представления лиц принимающих решение об их качестве. В случае с задачами многокритериального выбора эти суждения о качестве альтернатив появляются на основе оценки каждой альтернативы по множеству критериев.

В этой работе рассматривается обратная задача. Имея данные о ранжировании объектов необходимо найти оценки важности критериев. Такая задача является не менее актуальной. Область её применения широка и включает в себя анализ результатов выбора, прогнозирование выбора из новых объектов по уже оценённым критериям, валидация результатов полученных другими методами (в том числе и не математическими).

В 1970х года американский математик Томас Саати разработал процедуру поддержки принятия решения, которую назвал «Analytic hierarchy process» (АНР). В России эта процедура известна как «Метод анализа иерархий» (МАИ) [1].

Область применения МАИ — решение задач принятия решения в условиях многокритериального выбора. МАИ базируется на использовании линейной функции полезности [2] и построении иерархии взаимодействия объект-критерий [3]. Допустим необхо-

димо ранжировать  $n$  объектов по  $m$  критериям. Экспертным путем находятся оценки каждого объекта по каждому из критериев и оценки критериев относительно цели принятия решения. На основе этих экспертных оценок строится  $m$  матриц попарного сравнений объектов по каждому из критериев и матрица попарного сравнения критериев. Следующим шагом находится вектор  $V$  — транспонированный доминирующий собственный вектор [4] для матрицы попарного сравнения критериев (вектор весов критериев), и матрица  $A$  — матрица составленная из доминирующих векторов каждой из матриц попарного сравнения. Согласно МАИ вектор  $U$ , являющийся решением уравнения

$$U = AV, \quad (1)$$

будет вектором весов объектов. Используя этот вектор можно ранжировать объекты или выбрать подмножество оптимальных вариантов [5].

Таким образом МАИ позволяет на основе оценок экспертов найти веса каждого объекта, ранжировать их, разместить на шкале приоритетности. Однако, если посмотреть на МАИ с точки зрения поиска решения обратной задачи принятия решения в условиях многокритериального выбора, то можно увидеть новый потенциал для его интерпретации.

## 2 Постановка задачи

**Определение 28.** Прямой задачей Саати назовём задачу нахождения весов объектов исходя из весов критериев и данных о попарном сравнении объектов относительно каждого критерия.

Исходя из этого определения можно сформулировать обратную задачу.

**Определение 29.** Обратной задачей Саати назовём задачу нахождения весов критериев, основываясь на весах объектов и данных о попарном сравнении объектов относительно каждого критерия.

Исходными данными для решения обратной задачи Саати являются множество объектов, множество критериев, множества весов объектов и множества попарных сравнений объектов по каждому критерию.

Авторами найдено два подхода к решению обратной

задачи Саати. Один из них основывается на инвертировании условий прямой задачи и нахождении решения с помощью описанной Томасом Саати схемы МАИ. К недостаткам такого подхода можно отнести то, что экспертам снова придется производить оценку, теперь уже оценивая критерии относительно объектов. Во многих случаях логика такой связи будет не очевидна. Кроме того, найдя подобным образом решение обратной задачи и использовав их при повторном решении прямой задачи возможно получить отличные от первоначальных условий результаты. Это делает подобный подход неприменимым при работ по валидации, прогнозированию и повторному ранжированию.

Другой подход базируется на использовании данных, которые мы можем получить решая прямую задачу Саати. Из уравнения (10) видно, что решением обратной задачи Саати будет вектора  $V$ , являющийся решением уравнения

$$V = A^{-1}U. \quad (2)$$

Этот подход к решению более предпочтителен. К его преимуществам можно отнести то что он более строго формализован и использует те же матрицы попарного сравнения что и МАИ. При единственности матрицы  $A^{-1}$  вектора  $V$  и  $U$  определяются однозначно.

Недостатком описанного метода то, что в общем случае матрица  $A$  не квадратная и ее обращение в классическом понимании невозможно.

### 3 Решение задачи

Для решения проблемы обращения используется псевдообращение Мура-Пенроуза [6][4]. Псевдообратная матрица Мура-Пенроуза отвечает условию единственности [6], что позволяет использовать её в данном случае — вектора  $V$  и  $U$  будут определены однозначно.

Существует несколько методов нахождения псевдообратной матрицы, таких как:

- метод Гревилля;
- метод Бен-Израэля;
- метод основанный на SVD (сингулярное разложение);
- метод наименьших квадратов;

Подробнее с перечисленными методами можно ознакомиться в [6][4][7].

Кроме того практическое применение описанного метода решения обратных задач Саати показало, что не исключены отрицательные значения в векторе весов критериев  $V$ . От них можно избавиться с помощью линейного сдвига.

## 4 Практическое применение. Программное обеспечение

Существует большое количество программного обеспечения для решения задач многокритериального выбора. Примерами программного обеспечения, реализующего МАИ, могут служить:

- SuperDecisions;
- Expert Choice;
- MPRIORITY;
- Император 3.1;

Однако ни один из этих программных продуктов не затрагивает решение обратной задачи Саати. Даже в самом мощном и полном из них — SuperDecisions — нет никаких средств для реализации подобных приведенных выше вычислений [8].

Этот факт побудил авторов разработать программное обеспечение DIPS (Direct and Inverse Problems Saaty), способное решать как прямую так и обратную задачу Саати. Кроме того были поставлены задачи по производительности приложения, удобства и наглядности его графического интерфейса, отсутствию недопустимо жестких ограничений на входные данные, кроссплатформенности.

DIPS было реализовано на основе технологии Java с использованием свободных библиотек Efficient Java Matrix Library (EJML), Swing и Prefuse. DIPS является кроссплатформенным приложением с графическим интерфейсом пользователя. Для нахождения доминирующих собственных векторов используется метод сильной итерации (также известный как метод итерации фон Мизеса) [9]. При решении обратной задачи для нахождения псевдообратной матрицы Мура-Пенроуза используется метод основанный на SVD [6]. На момент публикации статьи работа по усовершенствованию DIPS продолжается.

Для проверок корректности реализуемых алгоритмов на этапе разработки использовался пакет MATLAB 7.12.

В мае-июне 2012 года предложенная технология была апробирована на задаче по выявлению влияния ряда медицинских показателей (72 критерия) на степень детского ожирения на примере 266 обследованных разделенных на 4 группы (исходные данные предоставлены Т. А. Костаревой, Красноярский государственный медицинский университет имени профессора В. Ф. Войно-Ясенецкого). Кроме того нужно было проверить правильность разделения детей по группам.

Были проведены следующие работы.

- Подготовлена усредненная таблица для групп детей разной степени ожирения (с отличающимися весами групп). Для этого была составлена

таблица объект-критерий, где в качестве объектов выступали группы детей (4 объекта). В качестве усредненных числовых показателей по каждому из 72 критериев для каждой из 4-х групп было использовано среднее арифметическое значение соответствующих критериев по всем входящим в конкретную группу детям.

- Решая обратную задачу с помощью DIPS, по этой таблице и известным весам объектов (индекс ожирения группы) были рассчитаны веса всех 72-х критериев. В качестве проверки полученного результата, с помощью разработанного ПО была решена прямая задача Саати для этих же критериев и объектов. Веса объектов совпали с исходными данными.
- Проверена правильность исходного отнесения детей к группам ожирения. Для этого все четыре исходные таблицы соответствующие группам детей, страдающими ожирением разной тяжести, были объединены в одну. На основе полученной таблицы размера 266 на 72 и вектора весов критериев, рассчитанных на шаге 2), с помощью DIPS была решена прямая задача Саати по нахождению весов объектов соответствующих каждому ребенку. Веса объектов были расставлены по числовой шкале.

По результатам анализа были сделаны выводы, что разработанный метод и программное обеспечение DIPS на его основе эффективно решает поставленные задачи. В частности на это указывает тот факт, что при проведении верификации данных распределения детей по степеням ожирения первоначальные диагнозы подтвердились в 93% случаев.

## 5 Заключение

Апробация разработанного метода показывает, что он может эффективно применён для решения многих задач постоянно возникающих перед представителями науки, предпринимательства и политики. Примерами могут служить задачи следующих типов:

- Проверка результатов ранжирования и группировки объектов, когда ранжирование и группировка была проведена какими-либо другими методами (возможно не математическим);
- Решение задач прогнозирования повторного выбора. На основе проведенного выбора можно оценить веса критериев с помощью решения обратной задачи Саати и, используя их, спрогнозировать выбор из новых объектов уже с помощью решения прямой задачи;
- Оценка влияния критерия на успешность объекта. Из уже определенных весов объектов можно посчитать веса критериев и сделать выводы по влиянию каждого критерия на успех объекта. Подобная информация позволит эффективнее перераспределить усилия.

- Создания рейтингов аналогичных уже существующим;

## Список литературы

- [1] Т. Саати. Взаимодействия в иерархических системах. *Техническая кибернетика*, 1:68–84, 1979.
- [2] А. В. Лотов, И. И. Поспелова. *Принятие решений. Метод анализа иерархий*. МАКС Пресс, Москва, 2008.
- [3] И. А. Шуйкова С. Л. Блюмин. *Введение в математические методы принятия решения*. ЛГПИ, Липецк, 1999.
- [4] Ф. Р. Гантмахер. *Теория матриц*. Наука, Москва, 1966.
- [5] Т. Саати. *Принятие решений. Метод анализа иерархий*. Радио и связь, Москва, 1993.
- [6] Г. Стренг. *Линейная алгебра и ее применения*. Мир, Москва, 1980.
- [7] Д. В. Беклемишев. *Дополнительные главы линейной алгебры*. Наука, Москва, 1983.
- [8] R. W. Saaty. *Decision making in complex environments. The Analytic Hierarchy Process (AHP) for Decision Making and The Analytic Network Process (ANP) for Decision Making with Dependence and Feedback*. Creative Decisions Foundation, Pittsburgh, PA, USA, 2003.
- [9] В. М. Вержбицкий. *Численные методы. Линейная алгебра и нелинейные уравнения*. Высшая школа, Москва, 2000.

# Приложение $N$ -граммного анализа в финансовой математике

Игорь Анатольевич Боровиков

Калифорния  
igor.borovikov@gmail.com

Михаил Георгиевич Садовский

Институт вычислительного моделирования СО РАН  
Красноярск  
msad@icm.krasn.ru

**Аннотация.** Представлены предварительные результаты приложений анализа временных рядов с помощью частотных словарей и их информационной ёмкости. Обсуждена проблема выбора оптимального конечного алфавита, используемого для квантизации исходных данных. Затронуты вопросы шумового предела для конечных последовательностей и эффекта квантизации для коротких  $n$ -грамм. В качестве примера рассмотрено поведение акций ВАС (Bank of America Corp.) на рынке.

**Ключевые слова.** Частота, энтропия, условная энтропия, оптимальность.

## 1 Введение

Анализ временных рядов составляет один из центральных разделов финансовой математики [1]. Под временным рядом будем понимать случайный процесс, определённый на интервале целых чисел или конкретную реализацию этого случайного процесса; из контекста всегда будет понятно, о чем идет речь. В этой работе мы будем рассматривать только скалярные вещественные строго положительные временные ряды. Временной ряд будем обозначать  $z_1, z_2, \dots, z_t, z_{t+1}, \dots$ . Мы не предполагаем а priori стационарности, гомоскедастичности или иных свойств кроме того, что существуют конечные математическое ожидание  $E(z)$  и дисперсия  $D(z)$  у рассматриваемого процесса.

Задача анализа временных рядов заключается в том, чтобы (в идеале) научиться предсказывать величину  $z_{t+1}$ , либо извлечь иные содержательно важные характеристики, определяющие поведение ряда. Понятно, что такого рода характеристики целиком определяются задачами того исследования, которое проводится на материале временного ряда.

Несмотря на большой объём исследований по анализу временных рядов, до сих пор очень мало внимания уделялось возможностям анализа рядов, опирающегося на изучение частот  $n$ -грамм (то есть, сравнительно коротких подпоследовательностей длины  $n$ , встречающихся в исходном временном ряде).

В настоящей работе изложены предварительные результаты исследований, основывающиеся на изуче-

нии частот  $n$ -грамм к финансовым данным. Идейно эти исследования восходят к работам по статистическому анализу нуклеотидных последовательностей [2, 3, 4]. Следует подчеркнуть, что данная работа не опирается на предположение о Марковости временного ряда. Поэтому подход, применённый здесь, является более общим, чем изучение временных рядов методами Марковских моделей как простых, так и моделей со скрытыми параметрами.

### 1.1 Некоторые понятия и определения

Приведём здесь некоторые понятия и определения, необходимые для дальнейшей работы.

Без ограничения общности будем называть тикером любые ценные бумаги и т. п., например ETF или другие индексы (такие, как GOOG, YANG или \$QQQ). Термином *символ* будем обозначать элементы (конечного) алфавита, который будет построен. Для начала мы будем рассматривать простейший случай финансовых временных рядов, а именно цену закрытия на тикер. Все данные по тикерам, используемые в настоящей статье, взяты из открытых источников Yahoo!Finance, если не указано иное. Будем рассматриваемый временной ряд как  $z(t)$ , где  $z$  является ценой закрытия (интерпретируемой как вещественное положительное число), а  $t$  — номер торгового дня (понимаемый как целое число).

**Определение 30.** Конечный алфавит  $\Omega$  мощности  $\|\Omega\| = 2L > 0$  называется выходным алфавитом, если он упорядочен биективным отображением на множество целых чисел  $Z_L = \{-L, -(L-1), \dots, -2, -1, 1, 2, \dots, L-1, L\}$  (0 исключён из алфавита). Такое отображение  $X: L_N \rightarrow \Omega$  будет называться индексированием.

**Определение 31.** Частотный словарь  $D(n)$  текста  $T$  — это набор всех пар вида  $\{(w, f_w)\}$ , где  $w$  — это уникальная  $n$ -грамма, а  $f_w$  — её частота. Параметр  $n$  называется толщиной *of* словаря.

Частота  $f_w$   $n$ -граммы  $w$  определяется стандартным образом: она есть отношение числа копий данной  $n$ -граммы к общему числу всех  $n$ -грамм, встречающихся в тексте. При этом мы не замыкаем текст в кольцо, однако будем пользоваться выводами, справедливыми для замкнутых в кольцо текстов. Для целей данной работы мы будем пренебрегать вызванной этой погрешностью и отложим ее анализ до последующих



публикаций.

Всякий словарь  $D(n)$  может быть однозначно преобразован в словарь меньшей толщины  $D(n-l)$ ,  $1 \leq l \leq n-1$ . Это преобразование осуществляется суммированием частот всех слов, отличающихся лишь первыми (либо последними)  $l$  символами.

Обратное преобразование  $D(n) \mapsto D(n+l)$ ,  $l = 1, 2, \dots$  порождает в общем случае не единственный частотный словарь  $D(n+l)$ , а семейство словарей  $\{D(n+l)\}$ , во-первых, не совпадающих друг с другом, а во-вторых, порождающих при обратном преобразовании исходный словарь  $D(n)$ ; случай однозначного преобразования  $D(n) \mapsto D(n+l)$  рассмотрен в [5, 6, 7, 8].

**Определение 32.** Восстановленным частотным словарём  $\tilde{D}(n)$  (по частотному словарю  $D(n-1)$ ) будем называть такой словарь, частоты которого определяются по формуле

$$\tilde{f}_{\tau_1 \tau_2 \dots \tau_{n-1} \tau_n} = \frac{f_{\tau_1 \tau_2 \dots \tau_{n-2} \tau_{n-1}} \times f_{\tau_2 \tau_3 \dots \tau_{n-1} \tau_n}}{f_{\tau_2 \tau_3 \dots \tau_{n-2} \tau_{n-1}}}.$$

Определённый таким образом восстановленный частотный словарь обладает максимальной энтропией среди всех в том семействе, которое может быть порождено словарём  $D(n-1)$  (см. детали в [2, 3, 4]).

**Определение 33.** Информационной ёмкостью  $\bar{S}_j$  частотного словаря  $D_j$  называется условная энтропия

$$\bar{S}_j = \sum_{w \in \Omega} f_w \times \ln \left( \frac{f_w}{\tilde{f}_w} \right) \quad (1)$$

реального частотного словаря, вычисленная относительно восстановленного  $\tilde{D}_j$  (см. определение 32), причём этот последний построен на реальном частотному словарю  $D_{j-1}$ .

Для случая частотного словаря  $\tilde{D}_j$ , восстановленного по словарю  $D_{j-1}$ , информационная ёмкость определяется так<sup>1</sup>:

$$\bar{S}_j = 2S_{j-1} - S_j - S_{j-2} \quad \text{и} \quad \bar{S}_2 = 2S_1 - S_2, \quad (2)$$

где  $S_j$  — абсолютная энтропия (реального) частотного словаря  $D_j$ .

**Определение 34.** Информационно значимым словом  $\hat{w}$  (для заданного словаря  $D_j$ ) называется слово, не удовлетворяющее неравенству

$$\alpha \leq \frac{f_{\hat{w}_\alpha}}{\tilde{f}_{\hat{w}_\alpha}} \leq \frac{1}{\alpha};$$

здесь  $0 < \alpha < 1$  — порог информационной значимости.

<sup>1</sup>Как и раньше, эти формулы становятся точными для замкнутых в кольцо слов.

## 2 Проблема квантизации

Подход к исследованию временных рядов, опирающийся на анализ частот  $n$ -грамм, существенно зависит от того, каким именно способом представлены для анализа исходные данные. Напомним, исходные данные представляют собой последовательность (действительных) чисел, как-то распределённых в интервале  $[0, \infty)$  и представляющих собой значения цены на некоторый продукт. Цена определяется на бирже или ещё где-то и происхождение этих данных нас далее интересовать не будет. Тем не менее, следует заметить, что рынок функционирует через достижение динамического равновесия, что невозможно без обмена информацией между участниками рынка. Такой обмен не происходит напрямую, а опирается на субъективное “настроение рынка” (market sentiment), на текущие и недавние цены (через т. н. технический анализ рынка), а также на т. н. фундаментальные показатели (такие, как состояние экономики в целом или определённого сектора, доходы конкретной компании и т. д. и т. п.). Поведение цен во времени является, таким образом, важной движущей силой в их последующем формировании. Именно обнаружение следов неявного обмена информацией между участниками рынка через временные ряды цен (т. е. наличие в рядах информационно-значимых слов) и является одним из мотивов для данной работы.

Стандартным приёмом в анализе временных рядов в задачах финансовой математики является переход от цены к логарифму ежедневного возврата на инвестиции:

$$p_t = \log(z_t/z_{t-1}).$$

Величина  $p_t$  более удобна тем, что является безразмерной и позволяет сравнивать различные по цене продукты. Учитывая близость величины  $z_t/z_{t-1}$  к единице для ежедневных изменений цены, мы можем ограничиться приближенной формулой

$$p_t = \frac{z_t}{z_{t-1}} - 1.$$

Использование такого приближения не оказывает практического влияния на результаты.

Собственно проблема квантизации заключается в следующем: величина  $p_t$  принимает вещественные значения, в то время, как  $n$ -граммный анализ требует перехода к дискретным данным. Иными словами, непрерывную функцию  $p_t$ , принимающую значения на некотором вещественном интервале (значения  $p_t$  ограничены на практике), требуется приблизить другой функцией

$$s_t, \quad s \in \Omega = \{A, B, C, \dots, X, \},$$

принимающей значения лишь на (конечном) дискретном множестве, элементы которого тут обозначены латинскими буквами. Задача квантизации за-



ключается в том, чтобы подобрать подходящие множество  $\Omega$  и отображение  $\mathfrak{P}: \{p_t\} \mapsto \Omega$  в него. Важнейшей характеристикой  $\Omega$  является его мощность. Построение такого отображения не является простой технической проблемой: не существует простой и очевидной процедуры преобразования значения цен в последовательность символов.

Ниже мы рассмотрим ряд процедур, каждая из которых позволяет построить своё множество  $\Omega$ ; здесь важно соблюсти баланс между сохранением значимого сигнала в таких дискретизованных данных (в том числе, и, возможно, весьма слабого) и вырождения частотного словаря  $n$ -грамм. Слишком большой алфавит приведёт к тому, что даже очень короткие слова будут встречаться в единственном экземпляре (квазиравновесие), а слишком бедный — к потере существенных данных.

## 2.1 Двоичный алфавит

Двоичный  $\Omega = \{0, 1\}$  алфавит является базовым: он абсолютно минимален. Пусть имеется исходный временной ряд  $p(t)$ ; отображение

$$s_t = \begin{cases} 1 & \text{если } p(t) \geq 0, \\ 0 & \text{в противоположном случае} \end{cases} \quad (3)$$

отражает базовое свойство цены: её рост и падение. Тем самым, бинарный алфавит точно отображает самую главную характеристику процесса, хотя и с потерей некоторых деталей.

Простые обобщения двоичного алфавита на большее число значений более или менее очевидны.

## 2.2 Лестница алфавитов

Очевидно, что мощность носителя частотного словаря, т. е. число различных слов в нём, растёт экспоненциально, в то время как длина слов растёт линейно. Тем самым, можно ожидать, что некоторые важные свойства временного ряда, отражаемые в свойствах частотного словаря, получаемого в результате квантизации, будет иметь логарифмический характер роста.

Таким образом, сравнительное исследование тех или иных характеристик частотных словарей, получаемых последовательно с помощью набора квантизаций, в которых мощность алфавита  $\Omega$  будет расти экспоненциально (в геометрической прогрессии в показателе 2) может выявить ряд неочевидных свойств исходного временного ряда. Эта простая и в целом прозрачная мысль приводит нас к идее построения “лестницы алфавитов”: стартуя с двоичного алфавита (раздел 2.1) и переходя затем к четверичному, восьмеричному и т. д. алфавитам, следует пройти по всем  $2^k$ ,  $k = 1, 2, \dots$  алфавитам возрастающей мощности.

Очевидно, здесь возникает проблема остановки. Довольно быстро (например, для  $k = 5$  или 6) может наступить вырождение частотного словаря. Не существует естественных ограничений для выбора  $k$ ; один из способов рассмотрен ниже (см. раздел 2.3), другой может выглядеть следующим образом.

Рост толщины словаря приводит к быстрому увеличению числа слов, представленных в единственном экземпляре в исходной временном ряду. Такие единичные слова не несут никакой необходимой для анализа информации. Соответственно, идея заключается в том, чтобы исключить их из рассмотрения. Это, однако, ещё не до конца определяет процедуру выбора оптимальной толщины словаря.

Окончательно выбор толщины, на которой следует остановиться, основывается на оценке “относительной” энтропии оставшегося частотного (под)словаря, в наборе частотных словарей возрастающей (в геометрической прогрессии) толщины  $\tilde{D}(2^k)$ : найдётся словарь с максимальной энтропией

$$\max_k \left\{ - \sum_{w \in \tilde{D}(2^k)} f_w \cdot \ln f_w \right\}.$$

Здесь словарь  $\tilde{D}(l)$  содержит только те слова, которые встречаются в двух и более копиях.

К сожалению, практические ограничения оставляют относительно небольшой диапазон для построения лестницы алфавитов. Это связано с наличием шумового предела для конечной последовательности (см. раздел 3). Длины последовательностей в случае финансовых рынков имеют порядок величин  $2^{20}$  для минутных данных и  $2^{15}$  для ежедневных данных у более “старых” тикеров. Для недавних IPO эти величины, соответственно, гораздо меньше. Тем не менее, авторы надеются провести практический расчет словарей с максимальной энтропией на рыночных данных в последующих работах.

## 2.3 Информативный алфавит

Возможно, наиболее продуктивной процедурой для определения оптимальной мощности алфавита  $\Omega$  является следующая. Она опирается на идею информационной ёмкости частотного словаря (см. определение 33). Рассмотрим два набора частотных словарей: реальных

$$D_1 \leftrightarrow D_2 \leftrightarrow \dots \leftrightarrow D_j \leftrightarrow D_{j+1} \leftrightarrow \dots \leftrightarrow D_L$$

и восстановленных

$$\tilde{D}_2 \leftrightarrow \tilde{D}_3 \leftrightarrow \dots \leftrightarrow \tilde{D}_j \leftrightarrow \tilde{D}_{j+1} \leftrightarrow \dots \leftrightarrow \tilde{D}_L.$$

Будем считать, что каждый  $\tilde{D}_j$  построен по словарю  $D_{j-1}$ ,  $\forall j$ .

Функция  $\bar{S}_j$ , как функция  $j$ , всегда имеет колоколообразную форму (обсуждение этого факта для анализа временных рядов см. в разделе 3). Тем самым, для любого текста  $T$  и любой квантизации  $\Omega$  всегда найдётся такая длина слов  $j^*$ , что  $\bar{S}_{j^*}$  является наибольшей. Рассмотрим теперь множество всех возможных квантизаций  $\{\Omega\}$ ; оно очевидным образом конечно для конечных алфавитов. Очевидно, что величина  $j^*$  зависит от квантизации.

Информационно оптимальной квантизацией будем называть такую квантизацию  $\Omega^*$ , которая даёт максимальное значение длины

$$j_{\max}^* = \max_{\{\Omega\}} \{j^*\}$$

среди всех возможных квантизаций  $\{\Omega\}$ .

### 3 О конечности временного ряда

Как было отмечено выше, функция (2) всегда имеет колоколообразную форму с максимумом  $\bar{S}$ , наблюдаемым на некоторой толщине  $j_{\max}$ . Это связано с конечностью анализируемого текста  $T$ . Поскольку характерная длина временных рядов, используемых в анализе финансовой динамики, имеет порядок  $10^4$  символов, постольку встаёт задача “исключения” паразитного сигнала, порождаемого эффектами конечности.

Оценка  $j_{\max}$  для пика колокола в общем случае представляется нетривиальной. Следующие простые неформальные рассуждения позволяют получить оценки для значения  $j_{\max}$  в случае бинарного алфавита, т. е. алфавита мощности 2, а затем обобщить их на случай алфавитов большей мощности.

Число различных бинарных  $n$ -грамм длины  $n$  равно  $L = 2^n$ . Если мы рассматриваем текст длины  $L$ , то в нём приблизительно  $L$  слов ( $n$ -грамм) длины  $n$ . То есть, включение каждого из  $L = 2^n$  возможных слов длины  $n$  в словарь является “существенным” в том смысле, что отличие восстановленного словаря толщины  $n$  от реального словаря толщины  $n$  будет “усиливать” случайную природу исходного текста. Более того, при толщине словаря большей, чем  $n$ , будет наблюдаться вырождение информационной ёмкости, т. к. многие  $n$ -граммы не будут представлены в исходном тексте. Таким образом,  $j_{\max} = \log_2 L$  является критическим значением толщины словаря для конечных текстов, выше которого подавляющее большинство слов становятся восстанавливаемыми однозначно из более коротких [2, 3, 4]. Аналогично, для алфавитов мощности  $k$  оценка для критической толщины словаря равна  $j_{\max} = \log_k L$ . Величину  $j_{\max}$  будем также называть *шумовым порогом*. По существу, он имеет природу, близкую к естественным максимальным лагам автокорреляции.

Рассуждения для оценки критической толщины при-

водят к необходимости учета конечности входного текста следующим образом. Сначала вычисляются вероятности символов алфавита из исходного (входного) текста. Затем методом Монте-Карло строится выборка случайных текстов с распределением вероятностей символов, полученным выше. По этой выборке вычисляется информационная ёмкость слов  $\bar{S}'_n$  длин  $n < j_{\max}$  (слов короче критической длины), а также оценивается стандартное отклонение информационных ёмкостей для каждого такого  $n$ . Информационная ёмкость рыночных данных сравнивается с  $\bar{S}'_n$  с учётом уже известной оценки стандартного отклонения для данного  $n$ . Поскольку информационная ёмкость имеет разный порядок величин для словарей разной толщины, на графиках удобнее представлять ее в логарифмической шкале. Результаты такого сравнения приведены на графике 2 для индекса Russel 2000 (тикер RUT). Для индексов, описывающих поведение рынка в целом, трудно ожидать ярко выраженные отклонения от шума. Это хорошо видно на графике 2.

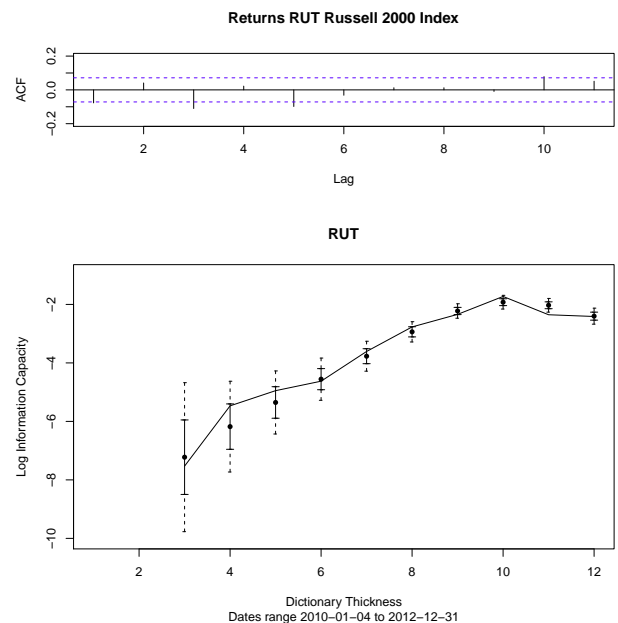


Рис. 1: На верхнем графике — автокорреляция логарифма ежедневного возврата, на нижнем — информационная ёмкость. Сплошная линия соответствует рыночным данным. Точками на графике показано мат. ожидание для случайного сигнала с таким же как и у рынка распределением вероятностей по символам алфавита. Вертикальные диапазоны ошибок соответствуют одному и двум стандартным отклонениям. Все расчёты проведены в логарифмической шкале.

Интересно отметить отсутствие *очевидной* связи между автокорреляцией и информационной ёмкостью на приведенном графике (и других графиках, полученных авторами, но не вошедших в статью из-за ограничения на объема). Несомненно, информационная ёмкость является показателем, принципи-

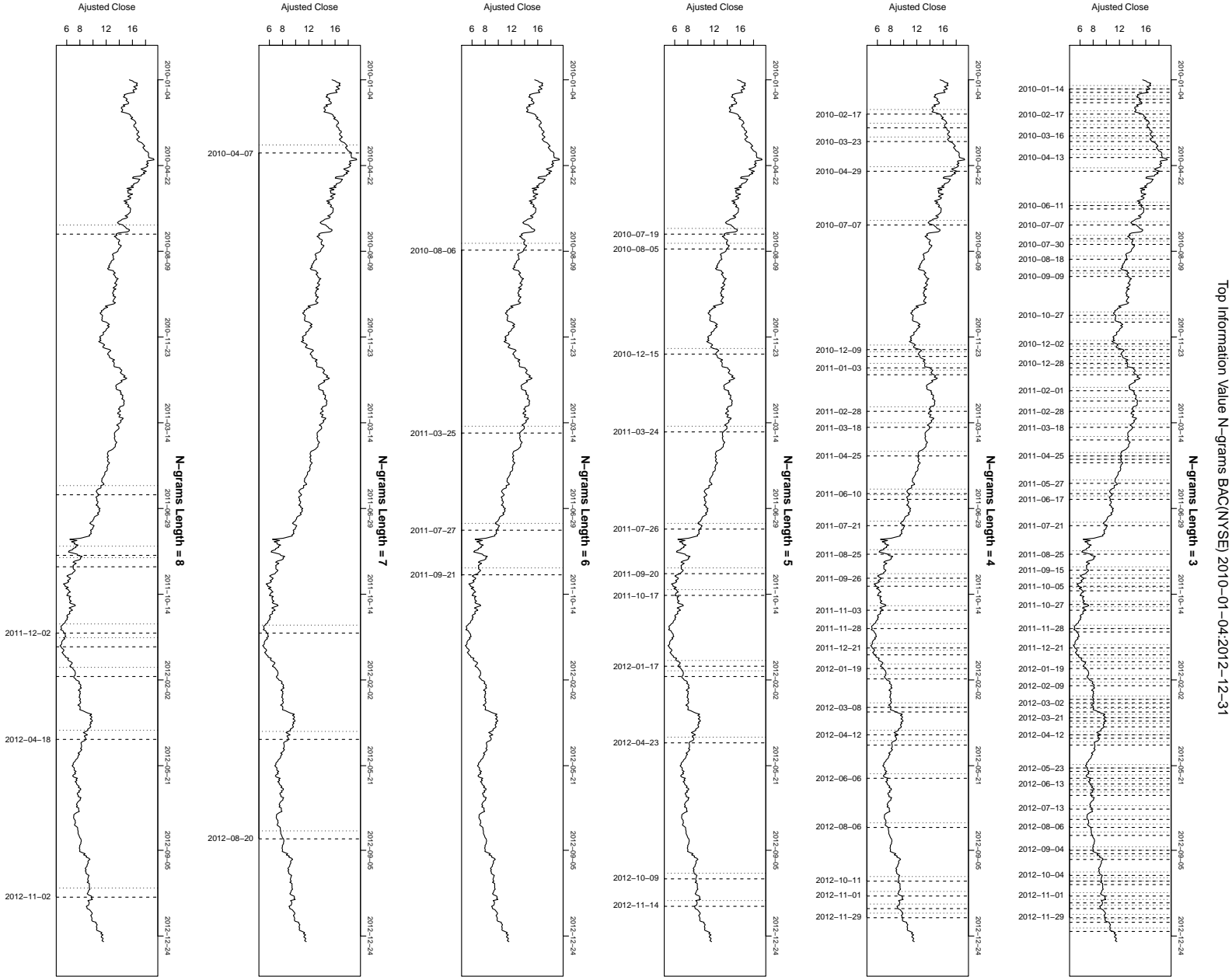


Рис. 2: Акции Bank of America Corp. вместе с положением информационно-значимых  $n$ -грамм разных длин.

ально отличным от автокорреляции<sup>2</sup>. С другой стороны, для периодической последовательности периода  $T$  очевидно информационная ёмкость вырождается (принимая нулевое значение), начиная с толщины словаря  $T$ , в то время как автокорреляция равна 1 на длине периода  $T$ . Таким образом, соотношение автокорреляции и информационной ёмкости существует, но требует более детального анализа.

#### 4 Нормализованная информационная ёмкость

На практике более удобной оказывается нормализованная информационная ёмкость:

$$S_n^* = \frac{\bar{S}_n - E(S'_n)}{\sigma(S'_n)}.$$

Здесь  $\bar{S}_n$  — информационная ёмкость квантизованного временного ряда рыночных данных, а  $E(S'_n)$  и  $\sigma(S'_n)$  — соответственно, математическое ожидание и стандартное отклонение выборки, полученной методом Монте-Карло для серии такой же длины и с тем же распределением вероятностей символов алфавита. На графике 3 показана нормализованная информационная ёмкость для тикера RUT (Russell 2000). Этот график находится в соответствии с гипотезой случайного рынка, т.к. все значения информационной ёмкости до шумового порога расположены в пределах  $\sigma$ .

Более интересный пример представлен на графике 4. В случае тикера SBUX (Starbucks Corp.) гипотеза о случайной природе рынка не выглядит столь же убедительной. Аналогичные, хотя и менее выраженные пики обнаруживаются и в поведении некоторых других тикеров. Например, BAC (Bank of America Corp.) имеет пик более  $2\sigma$  на 7 днях. Более высокая информационная ёмкость тикеров, соответствующих отдельным компаниям, представляется естественной, так как их поведение более подвержено локальным факторам рынка и не может быть так же эффективно усреднено до случайного, как в случае широких рыночных индексов.

#### 5 Распределение информационно-значимых $n$ -грамм по времени

На графиках 2 изображены информационно-значимые слова (ИЗС) для словарей разной толщины в диапазоне от минимума ( $j = 3$ ) до (почти) шумового предела; данные взяты для Bank of America Corp. (тикер BAC). ИЗС были выбрана как  $n$ -граммы с наиболее высокой информационной значимостью, которые встречаются не менее двух раз в исходном тек-

<sup>2</sup>Вопрос о её связи с другими широко используемыми статистическими показателями для временных рядов требует дальнейшего изучения.

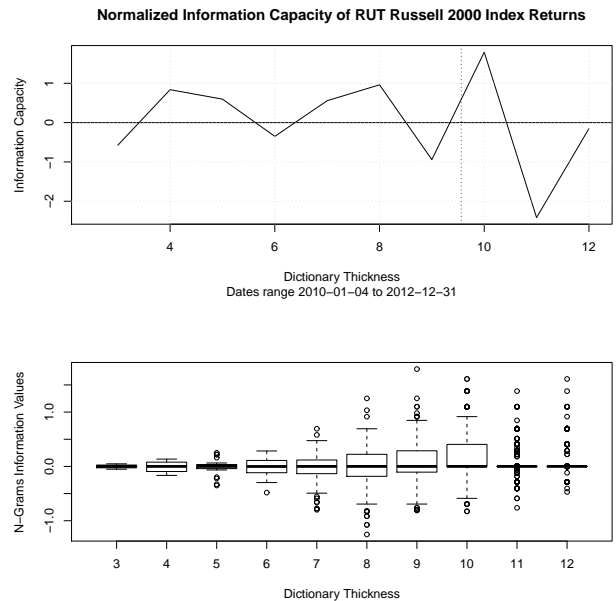


Рис. 3: Верхний график — нормализованная информационная ёмкость, нижний — box-plot для информационной значимости  $n$ -грамм соответствующей длины. Вертикальная точечная линия показывает положение  $d_0$  — “шумового порога”.

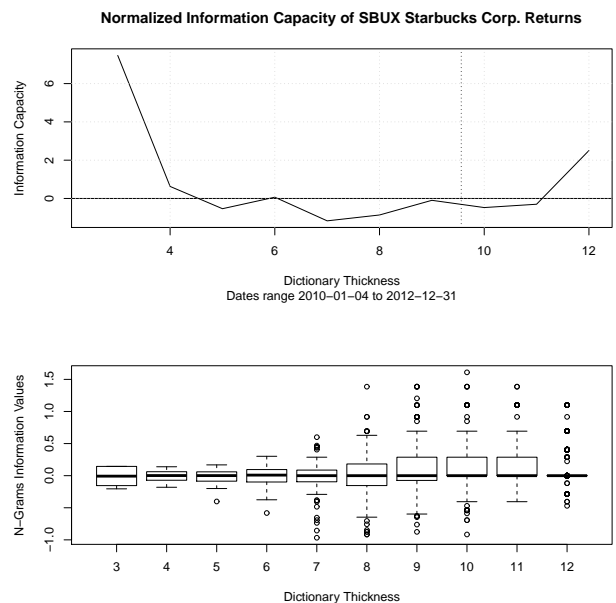


Рис. 4: Акции Starbucks Corp. на верхнем графике нормализованной информационной ёмкости указывают на возможное наличие нетривиального сигнала для  $n$ -грамм длины 3. Нижний график (box-plot для информационной значимости  $n$ -грамм) поддерживает это наблюдение, показывая anomalously высокое количество информационно-богатых  $n$ -грамм для  $n = 3$ . При оценке информационной ёмкости при малых  $n$  следует также иметь в виду квантизационный шум, который обсуждается в разделе 5.

сте. Таким образом, мы игнорировали уникальные ИЗС. Также мы ставили целью выбрать не менее 5% всех  $n$ -грамм для каждого фиксированного  $n$ . Про-

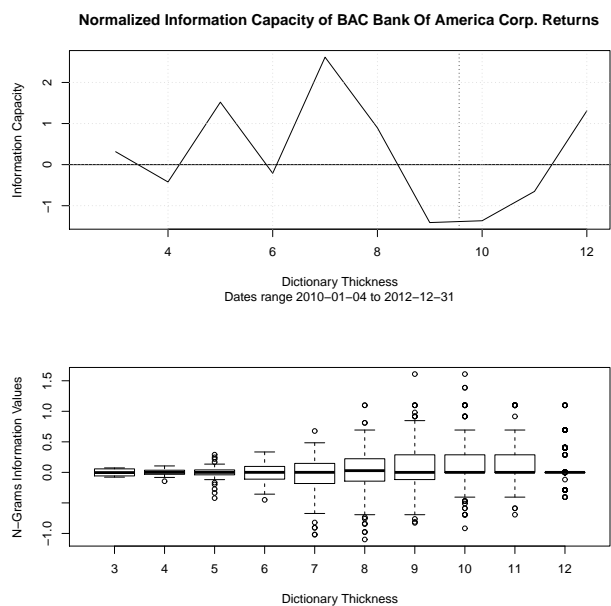


Рис. 5: Пики информационной емкости у тикера ВАС (Bank of America Corp.).

центное количество фактически выбранных  $n$ -грамм было больше в случае малых  $n$  по причине квантизационного шума, который мы обсудим в конце раздела.

Полученные результаты позволяют сделать несколько важных наблюдений. Во-первых, ИЗС для словарей разной толщины  $n$  не обязаны совпадать по положению во времени. Иными словами, отсутствует прямая “фильтрация” ИЗС при переходе между соседними толщинами: наблюдается лишь приблизительное совпадение положений кластеров ИЗС на шкале времени. Отсутствие фильтрации подразумевает уникальность информации, содержащейся в ИЗС для каждого  $n$ . Вероятно, это можно объяснить различным поведением инвесторов, имеющих разные горизонты инвестирования: краткосрочные и спекулятивные стратегии могут быть более связаны с короткими ИЗС. Более длинные ИЗС, вероятно, связаны с поведением более долгосрочных инвесторов. Во-вторых, отрезки времени, не содержащие ИЗС на малых значениях  $n$ , имеют тенденцию оставаться свободными от ИЗС и на более высоких значениях  $n$ . Это наблюдение поддерживает нашу гипотезу о роли ИЗС в неявном обмене информацией между участниками рынка: если такой обмен не происходил на коротких промежутках времени, то вероятность его на более длительных промежутках, включающих данный, также невелика. Следует подчеркнуть, что в настоящий момент приведенные рассуждения носят статус правдоподобных гипотез и требуют более детального исследования.

Наблюдения более технического характера включают в себя, во-первых, шум на значениях  $n$ , прибли-

жающихся к шумовому пределу конечного отрезка временного ряда; и, во-вторых, наличие квантизационных артефактов на малых  $n$ . Это проявление обычного эффекта квантизации (aliasing): небольшое количество возможных  $n$ -грамм для малых  $n$  ведет к зашумлению действительно существенных ИЗС теми, которые попали в диапазон ИЗС только в силу недостатка других возможных значений информационной ёмкости. Например, при  $n = 3$  мы имеем всего  $2^n = 8$  разных  $n$ -грамм, то есть мы можем ожидать не более 8 разных значений информационной ёмкости  $n$ -грамм. Это соображение определяет выбор порога информационной значимости, т. е. для относительно коротких входных слов и  $n$ , близких к 3, существует только небольшое количество существенно разных порогов для определения ИЗС. Это наблюдение также поддержано гистограммами информационной ценности словарей в зависимости от толщины  $n$  (см. рисунок 6). Для  $n = 3$  гистограмма не имеет центрального пика, в то время как в середине диапазона значений  $n$  видны отчетливые максимумы. Для значений, превышающих критическое, гистограммы имеют острый пик в районе 0 и большое количество изолированных значений, разбросанных по шкале информационной ценности. Такой же вывод о поведении в диапазоне больших  $n$  можно сделать из boxplot-ов на графиках 3, 4 и 5.

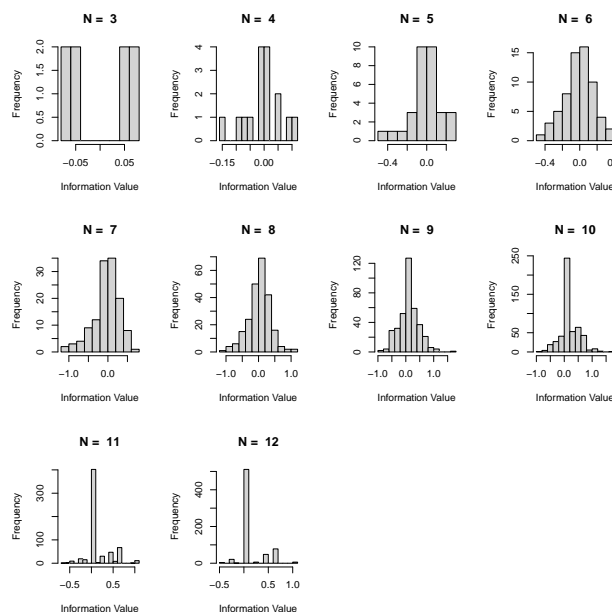


Рис. 6: Гистограммы информационной ценности  $n$ -грамм для Bank of America Corp. для разных  $n$ .

Представляет также интерес поведение набора ИЗС при сдвиге временного интервала исследуемого ряда. Для оценки такого поведения авторы извлекли ИЗС на двух перекрывающихся интервалах [2010-01-01, 2012-12-31] и [2010-03-31, 2013-03-31] для тикера ВАС. Результаты сравнения наборов ИЗС представлены в таблице 1. Эта таблица находится в разуме



ном соответствии с наблюдениями о распределении  $n$ -грамм по информационной значимости (см. гистограммы 6). В частности, ИЗС для малых  $n$  неустойчивы в силу артефактов квантизации. ИЗС на значениях  $n$ , близких к шумовому пределу, относительно устойчивы. Легко видеть, что шумовой барьер делает набор длинных ИЗС более “жестким” и, следовательно, более устойчивым. В середине диапазона  $n$  также наблюдается значительное пересечение или полное совпадение наборов ИЗС.

Таблица 1: Неуникальные ИЗС (5%-процентиль) на двух перекрывающихся отрезках временного ряда.

$n$ -граммы	2010-01:2012-12	2010-03:2013-03
3	011	111
4	0011	0011
5	11000	11000
6	110000	110000
7	1011111 0110111	1011111 0100011
8	00110111 11111000 10000111 10001001	00110111 11111000 10000111
9	111001000 101011100 010111001 011100110	111001000 101011100 010111001 110010111

## 6 О сложности метода Монте-Карло для оценки шума

Метод Монте-Карло для оценки шума, очевидно, имеет экспоненциальную сложность. Поэтому получение более эффективной оценки для сигнала  $S'_n$  случайного конечного ряда представляется важной задачей для последующей работы. В то же время, рост критической длины  $j_{\max}$  носит логарифмический характер в зависимости от входного интервала. Поэтому для практических временных интервалов в несколько лет сложность предложенного метода остается вполне в рамках возможностей доступных вычислительных средств. Так, например, интервал в 4 года содержит приблизительно 1000 торговых дней и позволяет проводить анализ слов длины  $< \log_2(1000) \approx 10$ , что соответствуют двум рыночным неделям.

Экспоненциальная сложность метода в целом делает количество случайных реализаций в вычислениях менее существенным параметром. В примерах, использованных в статье, авторы задали этот параметр равным 1000 случайных реализаций для каждого ти-

кера.

## 7 Связь ИЗС с событиями на рынке

Связь ИЗС с событиями на рынке была оценена методом сравнения дат соответствующих ИЗС и дат появления новостей о компании. В качестве примера такого анализа был выбран Bank of America Corp. (тикер BAC). Учитывая, что ежедневно публикуется большое количество новостей о Bank of America Corp., целесообразно рассматривать два подхода.

Первый поход состоит в попытке охватить достаточно широкую выборку новостей и рассматривать их в совокупности, т.е. их суммарное распределение по времени вместе с упрощенным (взвешенным) представлением как позитивных-негативных новостей для компании (такие новости имеют максимальный вес), финансового сектора (им соответствует “средний” вес) и рынка в целом (новости имеют ещё меньший вес). Такой подход предполагает доступ к обширным интернет-архивам, т.к. не все “вторичные” новости или комментарии остаются в открытом доступе после своего опубликования.

Второй подход может быть основан на количественных данных торговли, инициированной инсайдерами. Такие данные тщательно отслеживаются государственными органами по контролю за оборотом ценных бумаг (например, SEC Form 4). Автоматизированный доступ к их архивам также требует некоторых усилий.

Подход, альтернативный двум предыдущим, состоит в намеренном ограничении только официальными новостями. Очевидно, официально опубликованные новости не исчерпывают всех новостей, так или иначе влияющих на ценообразование акций конкретной компании. Тем не менее, по крайней мере теоретически, они оказывают наиболее существенное влияние. В этом смысле они должны быть достаточно репрезентативными. Мы выбрали последний подход ввиду его сравнительной простоты. Новости и их даты для анализа были извлечены с официальной веб-страницы для инвесторов <http://investor.bankofamerica.com>, раздел Financial Press Releases. Результаты такого сравнения показаны на графике 7.

Можно предположить следующее влияние событий на цены акций. Вероятный сценарий — это быстрое падение или рост цены на акцию в течение нескольких дней непосредственно после события, в зависимости от того, воспринята новость как положительная или как отрицательная. Часто также происходит более мягкое изменение в сторону будущего изменения непосредственно перед событием (в силу вероятной утечки информации или ожиданий рынка). Как следует из графика 7, такие движения цены являются

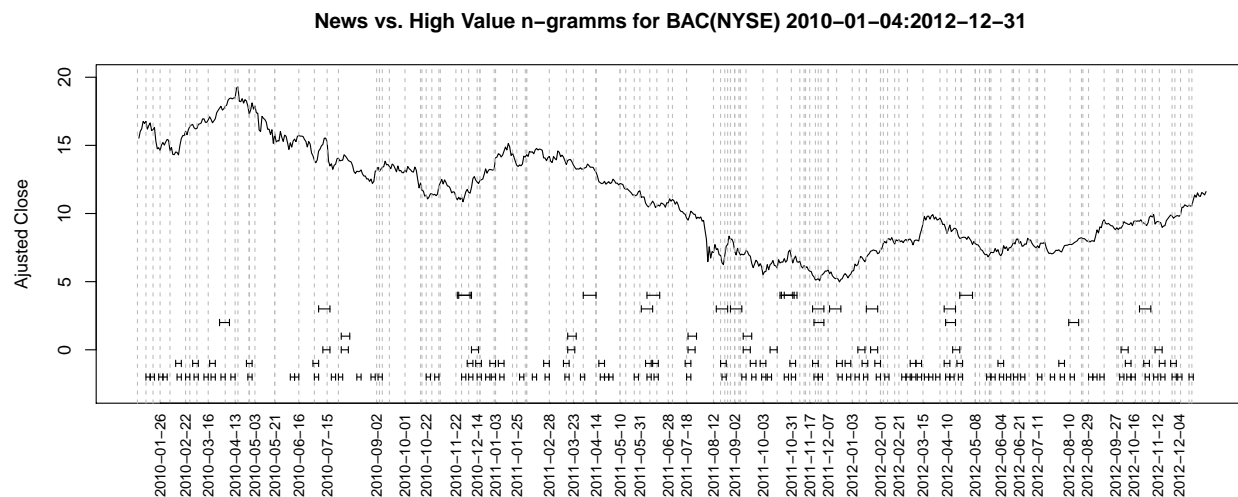


Рис. 7: Акции Bank of America Corp. вместе с положением информационно-значимых  $n$ -грамм разных длин и событий компании.

слишком распространёнными и не порождают ИЗС в  $n$ -граммном анализе.

## 8 Обсуждение

Проведенный предварительный анализ данных показывает, что подход, основанный на относительной энтропии информационной значимости  $n$ -грамм может существенно дополнить уже существующие методы анализа временных рядов, возникающих в финансовых приложениях. Этот вывод основан на отсутствии очевидной связи между событиями на рынке и ИЗС, а также между автокорреляцией и информационной ёмкостью словарей. При наличии такой связи развитие методов  $n$ -граммного анализа не принесло бы новых результатов. При обнаружении гипотетической связи между ИЗС и неявным обменом информацией между участниками рынка возможно построение новых технических показателей, которые могут быть использованы при оценке поведений цен в прошлом и предсказанию возможных изменений цен в недалеком будущем.

## Список литературы

- [1] Tsay Ruye S. *Analysis of Financial Time Series*. John Wiley & Sons, Inc., Chicago, 2001.
- [2] Sadovsky M. G. and Putinzeva Yu. A. and Zajtzeva N. A. System biology on mitochondrion genomes. *Proc. of the 3<sup>rd</sup> Int. Conf. on Bioinformatics Biocomputational Systems and Bioengineering*, pages 61–66, 2011.
- [3] Sadovsky M. G. Information capacity of symbol sequences. *Open Sys. & Information Dyn.*, 9:1–12, 2002.
- [4] Sadovsky M. G. and Putinzeva Yu. A. and Shchepanovsky A. S. Genes, information and sense: complexity and knowledge retrieval. *Theory in Bioscience*, 127:69–78, 2008.

- [5] А. Н. Горбань, Е. М. Миркес, Т. Г. Попова и М. Г. Садовский. Новый подход к изучению статистических свойств генетических последовательностей. *Биофизика*, 38 (5):762–767, 1993.
- [6] А. Н. Горбань, Е. М. Миркес, Т. Г. Попова и М. Г. Садовский. Сравнительная избыточность генов различных организмов и их вирусов. *Генетика*, 29 (9):1413–1419, 1993.
- [7] А. Н. Горбань, Т. Г. Попова и М. Г. Садовский. Избыточность генетических текстов и мозаичная структура генома. *Мол. биология*, 28 (2):313–322, 1994.
- [8] М. Г. Садовский. К вопросу об избыточности геномов вирусов и прокариот. *Генетика*, 38 (5):695–701, 2002.

# Задача построения множества наиболее независимых событий, которые порождают заданное террасное разбиение пространства всеобщих элементарных исходов

Веретнова Ксения Юрьевна

Институт математики и фундаментальной информатики  
Сибирский Федеральный Университет  
Красноярск  
veretnova141189@mail.ru

**Аннотация.** Приводится идея событийного метода главных компонент, цель которого уменьшить размерность множества факторных событий, влияющих на совокупность опасных событий. Рассмотрен первый шаг алгоритма эвентологического метода главных компонент.

**Ключевые слова.** Метод главных компонент, уменьшение размерности, террасное разбиение, независимые события

## 1 Введение

В настоящей статье проводится ознакомление с эвентологическим методом главных компонент, а именно описание основной идеи этой аналогии классического метода главных компонент, основанной на эвентологическом подходе [1]. Данный метод может быть применен в области эвентологии безопасности, и в частности в эвентологии пожарной безопасности [2, 3, 4]. В качестве базовой точки отправления был взят классический метод главных компонент, суть которого заключается в уменьшении размерности показателей, влияющих на систему исследуемых объектов.

Очевидно, что наступление множества исследуемых событий одного рода зависит от множества факторов. В случае, когда требуется оценить вероятность наступления события приходится учитывать все факторы, влияющие на данное событие. Порой количество таких факторных событий велико и приходится иметь дело с большим объемом информации. Эвентологический метод главных компонент позволит выделить такие факторы, которые являются наиболее значимыми.

В статье описан первый шаг алгоритма событийного метода главных компонент, который состоит в переходе от первоначального множества факторных событий к новому множеству факторных событий меньшей размерности.

### 1.1 Классический метод главных компонент: краткий обзор

Во многих физических, статистических, биологических и других исследованиях система объектов описывается множеством факторов, влияющих на совокупность предметов исследования. Уменьшение размерности системы показателей, влияющих на совокупность объектов, является неотъемлемым шагом в исследовании системы, когда количество факторов велико. Причем, стоит задача не только уменьшить размерность задачи, но и сохранить те из признаков, которые вносят существенный вклад в поведение системы. Необходимо отбросить те факторы, которые содержат ненужный “шум”.

Существует множество методов уменьшения размерности данных. Например, факторный анализ, метод упругих карт, многомерное шкалирование. Среди всех методов, как наиболее популярный среди них, можно выделить метод главных компонент. Данный метод изобретен Карлом Пирсоном.

К достоинствам данного метода можно причислить простоту реализации, широту класса объектов, для которых применим метод, возможность визуализации данных с помощью метода главных компонент.

Объектами применения метода главных компонент являются конечные системы материальных объектов, которые обладают сходными свойствами, присущими всему классу. Для каждой такой совокупности объектов необходимо выделить признаки. Признаки – это показатели, которыми можно охарактеризовать каждый объект из множества объектов изучения. Например, если исследуется группа людей в условиях крайнего Севера [5], то такими признаками могут выступать артериальное давление, пульс, количество приседаний в минуту, которое способен сделать человек. Подробнее с оригинальным описанием метода можно ознакомиться в статье [6].

### 1.2 Идея событийного метода главных компонент

Как говорилось выше, класс объектов изучения, для которых применим метод главных компонент – это

конечная совокупность объектов материального мира. Расширим класс изучаемых методом главных компонент объектов и рассмотрим ситуацию, когда исследуемое множество объектов – это множество событий.

В [2] приведена модель, в которой наступление опасного события – пожара – характеризуется лесопожарными рисками. Рассмотрим один из таких рисков – риск возникновения лесного пожара.

Риск возникновения лесного пожара определяется несколькими характеризующими, например погодой, наличием поблизости горючих материалов, возникновением молнии, извержения, фактом нарушения пожарной безопасности. Каждую из таких характеризующих можно представить как событие, или совокупность событий. Таким образом, пожар, как опасное событие, может быть охарактеризовано множеством событий – множеством факторных событий.

Перейдем к формализации событийного метода главных компонент. Пусть задано вероятностное пространство  $(\Omega, \mathcal{A}, \mathbf{P})$ . Пусть  $\mathfrak{Y} \subseteq \mathcal{A}$  – множество событий. Обозначим множество факторных событий  $\mathfrak{X}$ .

Алгебра событий  $\mathcal{A}_{\mathfrak{X}}$ , порожденная множеством событий  $\mathfrak{X}$ , содержит всевозможные террасные события:

$$\text{ter}(X) = \bigcap_{x \in X} x \bigcap_{x \in X^c} x^c,$$

где  $X \subseteq \mathfrak{X}$ .

В настоящей работе будем называть их факторными террасными событиями.

Событийный (эвентологический) метод главных компонент – способ выявления главного событийного фактора (факторного террасного события) среди алгебры событий  $\mathcal{A}_{\mathfrak{X}}$ , порожденной множеством событий  $\mathfrak{X}$ , влияющих на наступление событий  $y \in \mathfrak{Y}$ . Задача эвентологического метода главных компонент – это уменьшение размерности множества факторных террасных событий, отыскание фактора, который более всего влияет на наступление событий из  $\mathfrak{Y}$ .

## 2 Алгоритм событийного метода главных компонент: первый шаг

Необходимо перейти к новому множеству факторных событий, которое имеет мощность меньше, чем первоначальное множество факторных событий.

Классический метод главных компонент заключается в переходе к новому базису, оси которого ортогональны. Известно, что ортогональность эквивалентна независимости векторов. В случае событийного метода главных компонент осуществляется поиск новых факторных событий, множество которых является множеством тотально независимых событий.

**Определение 35.** Пусть задано множество событий  $\mathfrak{X}$ , состоящее из  $x \in \mathfrak{X}$ , имеющих произвольные вероятности  $p_x = \mathbf{P}(x)$ . Если  $\mathfrak{X}$  имеет вероятностное распределение вида:

$$p(X) = \prod_{x \in X} p_x \prod_{x \in X^c} (1 - p_x),$$

где  $X \subseteq \mathfrak{X}$ ,  $p(X) = \mathbf{P}\left(\bigcap_{x \in X} x \bigcap_{x \in X^c} x^c\right)$  – вероятности террасных событий, порожденных  $\mathfrak{X}$ , тогда  $\mathfrak{X}$  – это тотально независимое множество событий.

Необходимо осуществить переход от факторных событий к новым факторным событиям, порождающим то же террасное разбиение, ту же алгебру событий  $\mathcal{A}_{\mathfrak{X}}$ .

**Определение 36.** Пусть задано множество событий  $\mathfrak{X}$ , состоящее из  $x \in \mathfrak{X}$ . Вероятностным распределением множества  $\mathfrak{X}$  будем называть набор вероятностей  $\{p(X) = \mathbf{P}(\text{ter}(X)), X \subseteq \mathfrak{X}\}$ .

То есть стоит задача построения такого множества событий, которые удовлетворяют следующим условиям:

- Разбиение  $\Omega$ , порождаемое данным множеством событий совпадает с изначальным террасным разбиением.
- Построенные события являются максимально независимыми, то есть невязка  $\left| \mathbf{P}\left(\bigcap_{x \in \mathfrak{X}} x\right) - \prod_{x \in \mathfrak{X}} \mathbf{P}(x) \right|$  принимает наименьшее значение, при фиксированном наборе вероятностей террасных событий.

Прежде чем приступить к описанию процедуры построения множества наиболее независимых событий, докажем следующее замечание.

**Замечание 1.** Пусть задано вероятностное пространство  $(\Omega, \mathcal{A}, \mathbf{P})$ . Пусть также задано множество событий  $\mathfrak{X}$  мощности  $N$ . Множество  $\mathfrak{X}$  разбивает  $\Omega$  на  $2^N$  террасок. Тогда любое множество из  $\mathfrak{X}$  содержит в себе ровно половину всех террасок.

*Доказательство.* Занумеруем события, содержащиеся в  $\mathfrak{X}$  индексами от 1 до  $N$ . То есть множество событий  $\mathfrak{X}$  можно представить в виде:  $\mathfrak{X} = \{x_1, x_2, \dots, x_N\}$ . Докажем утверждение замечания для произвольного события из  $\mathfrak{X}$ . Пусть  $x_i \in \mathfrak{X}$ . Не вызывает сомнений тот факт, что событие  $x_i$  содержит в себе террасное событие<sup>1</sup>  $\text{ter}(x_i)$ . Также событие  $x_i$  содержит в себе террасные события вида  $\text{ter}(x_i x_j)$ , где  $j = \overline{1, N}$  и  $j \neq i$ . Всего таких террасных событий  $C_{N-1}^1$ . Также событие  $x_i$  содержит в себе террасные события вида  $\text{ter}(x_i x_j x_k)$ , где  $j = \overline{1, N}$ ,

<sup>1</sup>Здесь и далее под сокращением  $\text{ter}(x_i)$  понимается запись  $\text{ter}(\{x_i\})$ . Запись  $\text{ter}(x_i x_j)$  является сокращением для  $\text{ter}(\{x_i, x_j\})$  и т.д.



$k = \overline{1, N}$ ,  $j \neq i$ ,  $k \neq i$ ,  $j \neq k$ . Всего таких террасных событий  $C_{N-1}^2$ . Продолжая рассуждения, на  $N - 2$ -ом шаге получим, что событие  $x_i$  содержит в себе террасные события вида  $\text{ter}(x_i \dots x_k)_{N-1}$ . Всего

таких событий  $C_{N-1}^{N-2}$ . На  $N - 1$ -ом шаге получим, что событие  $x_i$  содержит в себе террасное событие  $\text{ter}(x_1 \dots x_N)$ . Всего таких событий  $C_{N-1}^{N-1} = 1$ . Таким образом, получаем, что событие  $x_i$  содержит в себе всего  $n = 1 + \sum_{i=1}^{N-1} C_{N-1}^i$ . Истинность выражения  $n = 2^{N-1}$  легко доказывается с помощью метода математической индукции для  $N > 1$ . Замечание доказано.

Нам необходимо построить новые факторные события, порождающие заданное террасное разбиение, заданную алгебру событий  $\mathcal{A}_x$ . Согласно замечанию, каждое такое новое событие должно содержать  $2^{N-1}$  террасных событий, если  $N$  – мощность множества факторных событий.

### 3 Пример

Рассмотрим пример, иллюстрирующий процедуру поиска множества наиболее независимых событий, порождающих заданное террасное разбиение.

Пусть задано  $\mathcal{X}' = \{x', y'\}$ , разбивающее множество  $\Omega$  на четыре терраски. Пусть вероятностное распределение для  $\mathcal{X}'$  равно:  $p(x') = 0.35$ ,  $p(y') = 0.4$ ,  $p(x'y') = 0.15$ ,  $p(\emptyset') = 0.1$ . Стоит задача построения новых событий  $x$  и  $y$ . Согласно замечанию, каждое событие,  $x$  и  $y$ , должно содержать в себе ровно половину террасок, то есть две терраски.

Разбиение, имеющееся изначально, мы можем представить в виде, изображенном на Рис.1 (вверху слева).

Вычислим вероятности событий  $x'$  и  $y'$ :

$$\mathbf{P}(x') = p(\{x'\}) + p(\{x', y'\}) = 0.5$$

$$\mathbf{P}(y') = p(\{y'\}) + p(\{x', y'\}) = 0.55$$

Вычислим невязку по формуле:

$$|\mathbf{P}(x' \cap y') - \mathbf{P}(x')\mathbf{P}(y')| = |0.15 - 0.275| = 0.125.$$

Таким образом, в случае, когда события  $x'$  и  $y'$  имеют террасное наполнение, описанное в данном примере величина невязки равна 0.125.

На Рис. 1 – 3 приведены примеры построения новых дуплетов событий, когда  $x$  и  $y$  по другому составлены из тех же фрагментов разбиения (террасных событий).

Примечание: Здесь и далее положим, что события  $x'$  и  $y'$  – это события определяющие изначально террасное разбиение множества  $\mathcal{X}'$ . Данное разбиение проиллюстрировано на Рис. 1 (вверху слева).

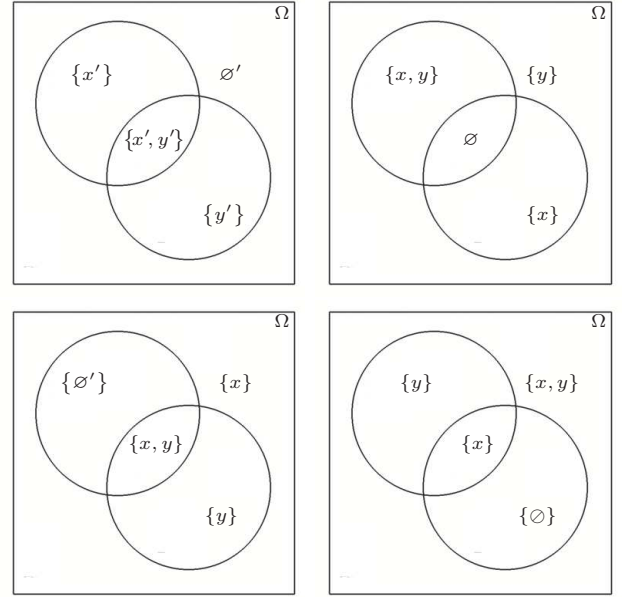


Рис. 1: Варианты построения дуплета новых базисных событий  $\{x, y\}$ , порождающих то же разбиение  $\Omega$ , что и начальный дуплет событий  $\{x', y'\}$ . Слева вверху – начальное террасное разбиение, порожденное  $\{x', y'\}$ ; вероятности террасных событий:  $p(\{x'\}) = 0.35$ ;  $p(\{x', y'\}) = 0.15$ ;  $p(\{y'\}) = 0.4$ ;  $p(\emptyset) = 0.1$ ; вероятности событий:  $\mathbf{P}(x') = p(\{x'\}) + p(\{x', y'\}) = 0.5$ ;  $\mathbf{P}(y') = p(\{x', y'\}) + p(\{y'\}) = 0.55$ ; невязка:  $|\mathbf{P}(x' \cap y') - \mathbf{P}(x')\mathbf{P}(y')| = 0.125$ . Террасные разбиения, порожденные новыми дуплетами  $\{x, y\}$  – справа вверху:  $p(\{x\}) = 0.4$ ;  $p(\{x, y\}) = 0.35$ ;  $p(\{y\}) = 0.1$ ;  $p(\emptyset) = 0.15$ ;  $\mathbf{P}(x) = 0.75$ ;  $\mathbf{P}(y) = 0.45$ ;  $|\mathbf{P}(x \cap y) - \mathbf{P}(x)\mathbf{P}(y)| = 0.0125$ ; слева внизу:  $p(\{x\}) = 0.1$ ;  $p(\{x, y\}) = 0.15$ ;  $p(\{y\}) = 0.4$ ;  $p(\emptyset) = 0.35$ ;  $\mathbf{P}(x) = 0.25$ ;  $\mathbf{P}(y) = 0.55$ ;  $|\mathbf{P}(x \cap y) - \mathbf{P}(x)\mathbf{P}(y)| = 0.0125$ ; справа внизу:  $p(\{x\}) = 0.15$ ;  $p(\{x, y\}) = 0.1$ ;  $p(\{y\}) = 0.35$ ;  $p(\emptyset) = 0.4$ ;  $\mathbf{P}(x) = 0.25$ ;  $\mathbf{P}(y) = 0.45$ ;  $|\mathbf{P}(x \cap y) - \mathbf{P}(x)\mathbf{P}(y)| = 0.0125$ .

В примерах, рассмотренных на Рис. 1, минимум невязки  $|\mathbf{P}(x \cap y) - \mathbf{P}(x)\mathbf{P}(y)|$  достигается при следующем наполнении событий террасками:

$$x = \text{ter}(\{x'\} // \mathcal{X}') + \text{ter}(\{y'\} // \mathcal{X}'),$$

$$y = \text{ter}(\{x'\} // \mathcal{X}') + \text{ter}(\{\emptyset'\} // \mathcal{X}');$$

$$x = \text{ter}(\{x', y'\} // \mathcal{X}') + \text{ter}(\{\emptyset'\} // \mathcal{X}'),$$

$$y = \text{ter}(\{x', y'\} // \mathcal{X}') + \text{ter}(\{y'\} // \mathcal{X}');$$

$$x = \text{ter}(\{x', y'\} // \mathcal{X}') + \text{ter}(\{\emptyset'\} // \mathcal{X}'),$$

$$y = \text{ter}(\{x'\} // \mathcal{X}') + \text{ter}(\{\emptyset'\} // \mathcal{X}').$$

В примерах, рассмотренных на Рис. 2, минимум невязки  $|\mathbf{P}(x' \cap y') - \mathbf{P}(x')\mathbf{P}(y')|$  не достигается.

В примерах, рассмотренных на Рис. 3, минимум невязки  $|\mathbf{P}(xy) - \mathbf{P}(x)\mathbf{P}(y)|$  достигается при следующем наполнении событий террасками:

$$x = \text{ter}(\{x'\} // \mathcal{X}') + \text{ter}(\{y'\} // \mathcal{X}'),$$

$$y = \text{ter}(\{x', y'\} // \mathcal{X}') + \text{ter}(\{y'\} // \mathcal{X}').$$

Выбирая наполнение новых событий  $x$  и  $y$  террасками, при котором невязка  $|\mathbf{P}(x \cap y) - \mathbf{P}(x)\mathbf{P}(y)|$  принимает наименьшее значение, мы получим множество наиболее независимых событий, порождающих



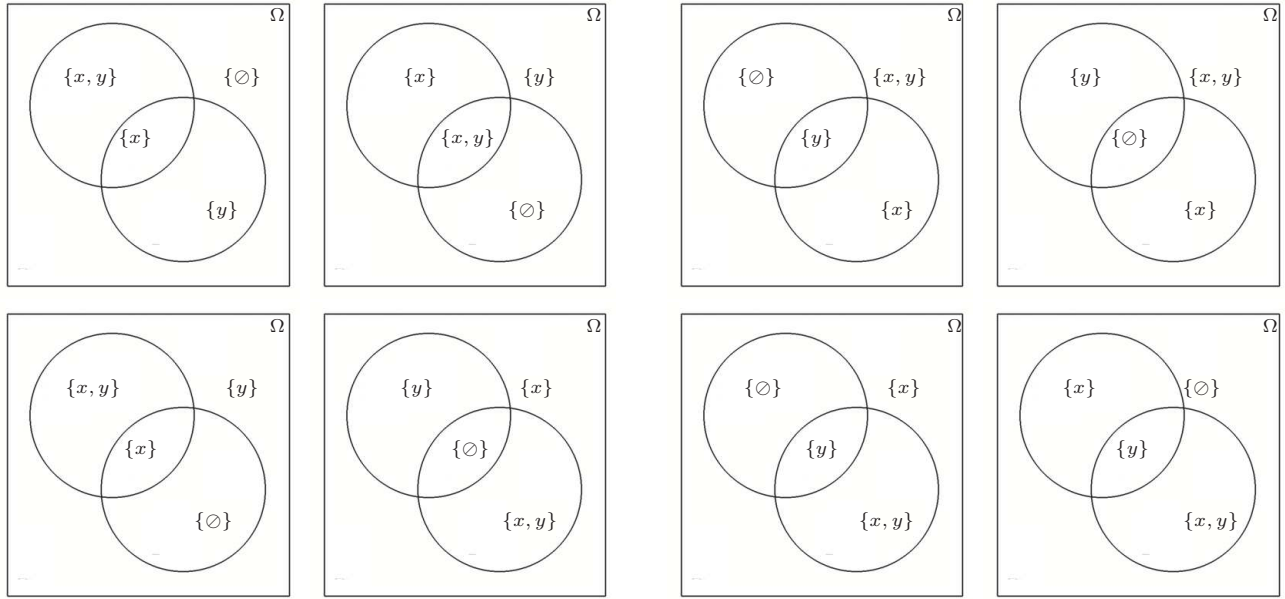


Рис. 2: Варианты построения дуплета новых базисных событий  $\{x, y\}$ , порождающих то же разбиение  $\Omega$ , что и начальный дуплет событий  $\{x', y'\}$ . Слева вверху — вероятности террасных событий:  $p(\{x\}) = 0.15$ ;  $p(\{x, y\}) = 0.35$ ;  $p(\{y\}) = 0.4$ ;  $p(\emptyset) = 0.1$ ; вероятности событий:  $\mathbf{P}(x) = p(\{x\}) + p(\{x, y\}) = 0.5$ ;  $\mathbf{P}(y) = p(\{x, y\}) + p(\{y\}) = 0.75$ ; невязка:  $|\mathbf{P}(x \cap y) - \mathbf{P}(x)\mathbf{P}(y)| = 0.025$ ; справа вверху:  $p(\{x\}) = 0.35$ ;  $p(\{x, y\}) = 0.15$ ;  $p(\{y\}) = 0.1$ ;  $p(\emptyset) = 0.4$ ;  $\mathbf{P}(x) = 0.5$ ;  $\mathbf{P}(y) = 0.75$ ;  $|\mathbf{P}(x \cap y) - \mathbf{P}(x)\mathbf{P}(y)| = 0.025$ ; слева внизу:  $p(\{x\}) = 0.15$ ;  $p(\{x, y\}) = 0.35$ ;  $p(\{y\}) = 0.1$ ;  $p(\emptyset) = 0.4$ ;  $\mathbf{P}(x) = 0.5$ ;  $\mathbf{P}(y) = 0.45$ ;  $|\mathbf{P}(x \cap y) - \mathbf{P}(x)\mathbf{P}(y)| = 0.125$ ; справа внизу:  $p(\{x\}) = 0.1$ ;  $p(\{x, y\}) = 0.4$ ;  $p(\{y\}) = 0.35$ ;  $p(\emptyset) = 0.15$ ;  $\mathbf{P}(x) = 0.5$ ;  $\mathbf{P}(y) = 0.75$ ;  $|\mathbf{P}(x \cap y) - \mathbf{P}(x)\mathbf{P}(y)| = 0.025$ .

заданное террасное разбиение.

Подводя итог, получим, что множество наиболее независимых событий, порождающих заданное террасное разбиение может быть получено при следующих комбинациях:

$$\begin{aligned}
 x &= \text{ter}(\{x'\} // \mathcal{X}') + \text{ter}(\{y'\} // \mathcal{X}'), \\
 y &= \text{ter}(\{x'\} // \mathcal{X}') + \text{ter}(\{\emptyset'\} // \mathcal{X}'); \\
 x &= \text{ter}(\{x', y'\} // \mathcal{X}') + \text{ter}(\{\emptyset'\} // \mathcal{X}'), \\
 y &= \text{ter}(\{x', y'\} // \mathcal{X}') + \text{ter}(\{y'\} // \mathcal{X}'); \\
 x &= \text{ter}(\{x', y'\} // \mathcal{X}') + \text{ter}(\{\emptyset'\} // \mathcal{X}'), \\
 y &= \text{ter}(\{x'\} // \mathcal{X}') + \text{ter}(\{\emptyset'\} // \mathcal{X}'); \\
 x &= \text{ter}(\{x'\} // \mathcal{X}') + \text{ter}(\{y'\} // \mathcal{X}'), \\
 y &= \text{ter}(\{x', y'\} // \mathcal{X}') + \text{ter}(\{y'\} // \mathcal{X}').
 \end{aligned}$$

Выберем среди них то наполнение, при котором сумма вероятностей  $p_x$  и  $p_y$  является наибольшей.

Этому условию удовлетворяет наполнение  $x = \text{ter}(\{x'\} // \mathcal{X}') + \text{ter}(\{y'\} // \mathcal{X}')$ ,  $y = \text{ter}(\{x', y'\} // \mathcal{X}') + \text{ter}(\{y'\} // \mathcal{X}')$ .

Задача метода — уменьшение размерности событий. Поэтому необходимо отбросить одно из событий и оставить то, которое несет в себе большее количество

Рис. 3: Варианты построения дуплета новых базисных событий  $\{x, y\}$ , порождающих то же разбиение  $\Omega$ , что и начальный дуплет событий  $\{x', y'\}$ . Слева вверху — вероятности террасных событий:  $p(\{x\}) = 0.4$ ;  $p(\{x, y\}) = 0.1$ ;  $p(\{y\}) = 0.15$ ;  $p(\emptyset) = 0.35$ ; вероятности событий:  $\mathbf{P}(x) = p(\{x\}) + p(\{x, y\}) = 0.5$ ;  $\mathbf{P}(y) = p(\{x, y\}) + p(\{y\}) = 0.25$ ; невязка:  $|\mathbf{P}(x \cap y) - \mathbf{P}(x)\mathbf{P}(y)| = 0.025$ ; справа вверху:  $p(\{x\}) = 0.4$ ;  $p(\{x, y\}) = 0.1$ ;  $p(\{y\}) = 0.35$ ;  $p(\emptyset) = 0.15$ ;  $\mathbf{P}(x) = 0.5$ ;  $\mathbf{P}(y) = 0.45$ ;  $|\mathbf{P}(x \cap y) - \mathbf{P}(x)\mathbf{P}(y)| = 0.125$ ; слева внизу:  $p(\{x\}) = 0.1$ ;  $p(\{x, y\}) = 0.4$ ;  $p(\{y\}) = 0.15$ ;  $p(\emptyset) = 0.35$ ;  $\mathbf{P}(x) = 0.5$ ;  $\mathbf{P}(y) = 0.55$ ;  $|\mathbf{P}(x \cap y) - \mathbf{P}(x)\mathbf{P}(y)| = 0.125$ ; справа внизу:  $p(\{x\}) = 0.35$ ;  $p(\{x, y\}) = 0.4$ ;  $p(\{y\}) = 0.15$ ;  $p(\emptyset) = 0.1$ ;  $\mathbf{P}(x) = 0.75$ ;  $\mathbf{P}(y) = 0.55$ ;  $|\mathbf{P}(x \cap y) - \mathbf{P}(x)\mathbf{P}(y)| = 0.0125$ .

информации. В настоящей статье предлагается следующий критерий выбора оптимального числа факторных событий:

Пусть дано множество факторных событий  $\mathcal{X}$  мощности  $N$ . Данное множество задает террасное разбиение. Пусть построено новое множество факторных событий  $\mathcal{X}^*$  той же мощности, которое содержит в себе максимально независимые события, порождающие заданное террасное разбиение. События из множества  $\mathcal{X}^*$  имеют вероятности  $p_1, p_2, \dots, p_N$ . Новое факторное событие сохраняется в списке главных факторных событий, если его вероятность больше среднего арифметического вероятностей событий из  $\mathcal{X}^*$ .

Согласно этому правилу, в данном примере следует оставить событие  $x'$ , которое содержит в себе террасные события  $\text{ter}(\{x'\} // \mathcal{X}')$ ,  $\text{ter}(\{y'\} // \mathcal{X}')$  и имеет вероятность 0.75.

## 4 Заключение

В настоящей статье описан первый шаг построения алгоритма событийного метода главных компонент. Построение множества независимых событий, порождающих заданное террасное разбиение, порожденное множеством факторных событий реализовано пока тривиальным способом — путем перебора

всевозможных вариантов наполнения событий террасками, с учетом того, что каждое событие должно содержать в себе ровно половину всех террасок. После построение такого множества осуществляется выбор факторных событий, которые составят новый "базис". Данный шаг является аналогом построения в классическом методе главных компонент векторов главных компонент, которые являются ортогональными. Независимость для событий является аналогом ортогональности для векторов. В классическом методе главных компонент после того, как найдены ортогональные вектора главных компонент происходит выбор числа векторов главных компонент, после чего осуществляется переход к пространству меньшей размерности, построенному на выбранных векторах главных компонент.

## Список литературы

- [1] О. Ю. Воробьев. *Эвентология*. Сибирский федеральный университет, Красноярск, 2007, 435с., <http://eventology-theory.ru/0-books/1-VorobyevOleg~2007~Eventology~435p.pdf>.
- [2] О. Ю. Воробьев. Эвентология безопасности. *Труды X Международ. ФАМЭБ конференции по финансово-актуарной математике и эвентологии безопасности*, Красноярск: НИИППБ, СФУ (под ред. Олега Воробьева):92–101, 2011.
- [3] О. Ю. Воробьев. Эвентологический системный анализ безопасности в картинках. *Труды XI Международ. ФАМЭБ конференции по финансово-актуарной математике и эвентологии безопасности*, Красноярск: НИИППБ, СФУ (под ред. Олега Воробьева):113–125, 2012.
- [4] О. Ю. Воробьев. Выразить и измерить: успехи применимой эвентологии. *Труды XVI Международ. ЭМ конференции по эвентологической математике и смежным вопросам*, Красноярск: СФУ (под ред. Олега Воробьева):44–49, 2012.
- [5] K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, ~:559–572, 1901.
- [6] Покидышева Л.И., Белоусова Р.А., Смирнова Е.В. Метод корреляционной адаптометрии в оценке секреторной функции желудка у детей в условиях Севера. *Вестник РАМН*, 5:42–45, 1996.

## В поисках первоисточника:

римейк статьи [1, 1975], где впервые дается определение  
решетчатого (воробьевского) ожидания случайного множества

Олег Юрьевич Воробьев

Сибирский федеральный университет  
Институт математики и фундаментальной информатики  
Торгово-экономический институт  
Красноярск  
oleg.yu.vorobyev@gmail.com

**Аннотация.** Римейк первоисточника одной хотя и старой, но оказавшейся вполне здоровой идеи “решетчатого ожидания случайного множества”, впервые опубликованной мною в 1975 [1, 2] и впоследствии, особенно в западной литературе по стохастической геометрии и теории случайных множеств, названной с легкой руки Дитриха Штойяна [3, 1994] “Vorob’ev expectation” — “воробьевским ожиданием”, или “ожиданием по Воробьеву”<sup>1</sup>; имеет лишь историческую и методическую ценность для эвентологии и теории вероятностей, еще раз напоминая, как и “...из какого сора растут стихи, не ведая стыда”.

**Ключевые слова.** Эвентология, теория вероятностей, случайное множество, среднемерное множество, решетчатое ожидание, воробьевское ожидание, ожидание по Воробьеву.

## 1 Вперед к первоисточнику

*Когда б вы знали, из какого сора  
Растут стихи, не ведая стыда.*

Анна Ахматова, 1940.

Всякое новое обнаруживает себя по-старинке — всегда по-своему. Любое изобретение — не чья-то искусная выдумка, а открытие ранее сокрытого, отсечение всего излишнего, избавление из плена забвения, высвобождение ради всеобщего обозрения.

Новое в математике — не исключение. Строгих математических определений дается столько, что их с лихвой хватило бы на все науки вместе взятые. Но лишь редким единицам из них становятся определениями новых математических понятий. Остальные

растворяются бесследно и пропадают навсегда в лабиринтах математической мысли.

Почему происходит так, а не иначе, никому неизвестно. Где, среди какого сора искать то новое, которое, будучи открытым раз, останется навсегда? Чего нельзя избегать? Чем нельзя брезговать?

Нет ответа.

Можно лишь пытаться еще раз взобраться, шаг за шагом восстановить, как это было; чтобы снова вскарабкаться всё той же тропинкой, вперед к первоисточнику.

## 2 Римейк и/and Remake<sup>2</sup>

### 2.1 Оценка средних контуров лесного пожара по вероятностной модели его распространения [1, 2, 1975]

В нашей работе [4, 5, 1973] рассматривалась вероятностная модель распространения пожара по лесному массиву, который состоит из отдельных участков, расположенных в узлах *плоской квадратной решетки*. Эта вероятностная модель — *процесс случайного распространения (ПСП)* — определяется как множество дискретных случайных величин

$$U = \{u_{rt}, r \in R_2, t \in T\}, \quad (2.1.1)$$

где  $R_2 \subset \mathbb{R}^2$  — плоская решетка,  $T$  — пространство временного параметра,  $u_{rt}$  — дискретная случайная величина с возможными значениями:

$$u_{rt} = \begin{cases} 1, & \text{если участок } r \text{ горит в момент } t; \\ 0, & \text{если участок } r \text{ не горел до момента } t; \\ -1, & \text{если участок } r \text{ сгорел к моменту } t. \end{cases} \quad (2.1.2)$$

Кроме того, предполагается, что совместное вероятностное распределение случайных величин  $u_{rt}$  для

<sup>2</sup>Ниже приведен текст моей статьи 1975 года, ставшей первоисточником идеи об *ожидании случайного множества* (сначала в русском оригинале, а потом в английском переводе). Допущены лишь стилистические правки, облегчающие восприятие, но не искажающие *того* хода мысли, в том числе, одного ее ошибочного ответвления.

<sup>1</sup>Всегда старался называть это понятие так, чтобы сам термин слегка намекал на свой смысл, вначале (1975) “*решетчатым ожиданием*” *случайного конечного множества*, чтобы подчеркнуть конечность ситуации (кстати, позволившей впервые обнаружить это понятие), а затем (1984) “*среднемерным множеством*” *случайного (измеримого, замкнутого, компактного) множества*. Но видимо “против природы не попрешь”, и теперь придется помнить и об англ. варианте “*Vorob’ev expectation*”, хоть особого желания стать эпонимом за собой не замечал.

всех  $t \in T$  полностью определяется полем

$$p = \{p(r)\}_{r \in R_2}, \quad (2.1.3)$$

так называемых *вероятностей распространения*, где

$$p(r) = [p_1(r); p_2(r); p_3(r); p_4(r)] \quad (2.1.3')$$

— *вектор вероятностей распространения* из узла  $r$ , а  $p_k(r)$  соответствует вероятности распространения огня из участка  $r$  на  $k$ -тый *соседний*<sup>3</sup> с ним участок.

В настоящей работе предлагается более общее определение ПСР, которое позволит провести достаточно простую и эффективную оценку средних контуров лесного пожара.

Из определения (2.1.1) и (2.1.2) следует, что состояние ПСР в момент  $t$  можно задавать, выделяя два таких подмножества узлов  $K_t^1$  и  $K_t^2$  квадратной решетки  $R_2$ , что

$$\begin{cases} K_t^1 = \{r : t \in R_2, u_{rt} \neq 0\}, \\ K_t^2 = \{r : t \in R_2, u_{rt} = -1\}. \end{cases} \quad (2.1.4)$$

Ясно, что  $K_t^2 \subseteq K_t^1$  и

$$K_t^1 \setminus K_t^2 = \{r : t \in R_2, u_{rt} = 1\}. \quad (2.1.5)$$

Таким образом, ПСР можно рассматривать как последовательность пар  $[K_t^1, K_t^2]$  для всех  $t \in T$ . При этом необходимо отметить, что подмножества  $K_t^1$  и  $K_t^2$  имеют, вообще говоря, *случайную форму*<sup>4</sup>, вероятностные характеристики которой определяются полем  $p$  вероятностей распространения (2.1.3).

Более общее определение ПСР, основанное на этом подходе, требует введения строго понятия, соответствующего нашему интуитивному представлению о “случайном подмножестве узлов” решетки.

Речь идет о *случайном множестве*, которое можно считать обобщением дискретной случайной величины и “значениями” которого являются (в отличие от дискретной случайной величины) не действительные числа, а подмножества узлов решетки.

В этой работе мы ограничимся лишь элементарным эквивалентом определения случайного множества.

**Определение.** *Случайным множеством*  $K$  называется не более чем счетный набор пар  $(B_i, p_i)_{i \in I}$ , где  $B_i \subseteq R_2$  — некоторое конечное подмножество узлов решетки, а  $p_i$  — вероятность совпадения случайного множества  $K$  с подмножеством  $B_i$ :

$$p_i = \mathbf{P}(K = B_i). \quad (2.1.6)$$

Кроме того, требуется, чтобы

$$\sum_{i \in I} p_i = 1. \quad (2.1.7)$$

<sup>3</sup>Правила *соседства* на  $R_2$  оговариваются специально.

<sup>4</sup>на квадратной решетке  $R_2$ .

*Случайным вектор-множеством* называется упорядоченный набор случайных множеств

$$\overline{K} = [K^1, K^2, \dots, K^n]. \quad (2.1.8)$$

Понятие случайного множества позволяет определить *решетчатый случайный процесс*, как последовательность случайных множеств. Процесс же случайного распространения определяется как частный случай решетчатого случайного процесса, в котором случайные множества решетчатого процесса должны удовлетворять соотношению (2.1.4). А поскольку, как мы уже отмечали, вероятностное распределение случайных величин  $u_{rt}$  полностью определяется полем вероятностей распространения  $p$ , то и ПСР полностью определяется этим полем.

Основным практическим результатом данной работы является предлагаемый метод оценки среднего контура лесного пожара по вероятностной модели его распространения, или (что то же самое) метод оценки “средних значений” случайных множеств, образующих ПСР.

Пусть  $\overline{K} = \{\overline{K}_t, t \in T\}$  — процесс случайного распространения, где  $\overline{K}_t = [K_t^1, K_t^2]$  — случайное вектор-множество. Поставим задачу определения “среднего значения” случайных множеств из  $\overline{K}_t$  для произвольного  $t \in T$ .

Совершенно очевидно, что задача, поставленная таким образом, некорректна, поскольку пока не определено понятие “среднего значения” случайного множества решение задачи будет существенно зависеть от выбора этого определения.

Следовательно, прежде всего нам необходимо условиться о том, какое подмножество узлов решетки мы будем считать “средним значением” случайного множества. Хотелось бы также, чтобы вновь вводимое понятие “среднего значения” в определенном смысле было аналогично понятию математического ожидания дискретной случайной величины.

Сначала заметим, что если  $|K|$  — число узлов в случайном множестве  $K$ , т.е.  $|K|$  — дискретная случайная величина с заданным распределением, то можно говорить о его среднем значении — математическом ожидании  $\mathbf{E}|K|$ .

**Определение.** *Решетчатым ожиданием случайного множества*  $K$  называется подмножество  $\mathcal{E}K \subseteq R_2$ , удовлетворяющее двум условиям:

$$|\mathbf{E}|K| - |\mathcal{E}K|| = \min_{B \subseteq R_2} |\mathbf{E}|K| - |B||, \quad (2.1, A)$$

$$\mathbf{E}|K \Delta \mathcal{E}K| = \min_{B \subseteq R_2} \mathbf{E}|K \Delta B|, \quad (2.1, B)$$

которые предписывают выбрать из всех подмножеств, которые удовлетворяют условию  $A$ ,

такое подмножество, которое в среднем наименее уклоняется от случайного множества  $K$ , т.е. удовлетворяет условию  $B$ .

Определенное таким образом *решетчатое ожидание* случайного множества обладает рядом свойств, аналогичных свойствам математического ожидания случайной величины.<sup>5</sup>

Остановимся на самом важном характеристическом свойстве, которое позволяет найти *решетчатое ожидание* случайного множества.

Введем в рассмотрение вероятность

$$\pi(r) = \mathbf{P}(r \in K). \quad (2.1.11)$$

Оказывается, что в *решетчатое ожидание*  $\mathcal{E}K$  входят те, округленно,  $\mathbf{E}|K|$  узлов решетки, для которых вероятности  $\pi(r)$  принимают значения больше, чем у остальных узлов.

Это важное характеристическое свойство позволяет разработать алгоритм и программу для ЭВМ “М-222”, которые вычисляют оценки *решетчатого ожидания* случайного множества.

Этот же алгоритм применен для оценки решетчатых ожиданий случайных множеств, образующих процесс случайного распространения.

Проведено моделирование ПСР, определяемых различными векторами вероятностей распространения

$$p(r) = [p_1; p_2; p_3; p_4], \quad (2.1.3'')$$

которые предполагаются постоянными для всех  $r \in R_2$ , и для каждого из них оценены *решетчатые ожидания* случайных множеств, образующих этот ПСР( $p_1; p_2; p_3; p_4$ ).

На рис. 2 показана часть решетки  $R_2$ , на которой моделируется ПСР, в том виде, в котором его печатает ЭВМ. На месте каждого узла  $r \in R_2$  располагается количество  $n(r)$  реализаций из 100, в которых узел  $r$  принадлежит случайному множеству  $K_t$  при  $t = 13$ , а оценка  $\mathbf{P}(r \in K_{13})$  — вероятности того, что  $r \in K_{13}$ , — имеет вид:  $\pi(r) \approx n(r)/100$ . Контур на плоскости  $\mathbb{R}^2$ , содержащий решетчатое ожидание  $\mathcal{E}K_{13}$  случайного множества  $K_{13}$  проводился *вручную* на основе оценки среднего  $\mathbf{E}|K_{13}|$  случайной величины  $|K_{13}|$ , вычисленной ЭВМ.

На рис. 3 показан ряд *решетчатых ожиданий* случайных множеств для различных ПСР.

<sup>5</sup> Например, для случайной величины  $\xi$  и ее математического ожидания  $\mathbf{E}\xi$  всегда справедливо

$$\mathbf{E}(\xi - \mathbf{E}\xi) = 0. \quad (2.1.9)$$

Аналогичное свойство выполняется и для произвольного случайного множества  $K$  и его решетчатого ожидания  $\mathcal{E}K$

$$\mathbf{E}|K \setminus \mathcal{E}K| = \mathbf{E}|\mathcal{E}K \setminus K|. \quad (2.1.10)$$

Эта аналогия и формула (2.1.10) оказались ошибочными и потому в римейке опущены в сноску.

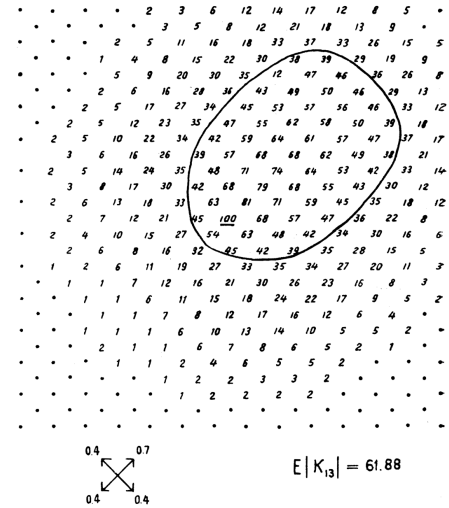


Рис. 1: Часть квадратной решетки  $R_2$ , на которой статистически моделируется случайное множество  $K_{13} \subset \mathcal{X}$  — состояние ПСР(0.4; 0.4; 0.4; 0.7) в момент  $t = 13$ . Для каждого узла решетки  $r \in \mathcal{X}$  показано количество реализаций случайного множества из 100 проведенных, в которых  $r \in K_{13}$ ; контур содержит *решетчатое ожидание*  $\mathcal{E}K_{13}$  случайного множества  $K_{13}$ . — The part of a square lattice  $R_2$  on which the random set  $K_{13} \subset \mathcal{X}$ , set-condition of the RSP(0.4; 0.4; 0.4; 0.7) during the moment  $t = 13$ , is statistically modeled. For each unit of the lattice  $r \in \mathcal{X}$  the quantity of realizations of random set from 100 lead, in which  $r \in K_{13}$  is shown; the contour contains the *lattice expectation*  $\mathcal{E}K_{13}$  of random set  $K_{13}$ .

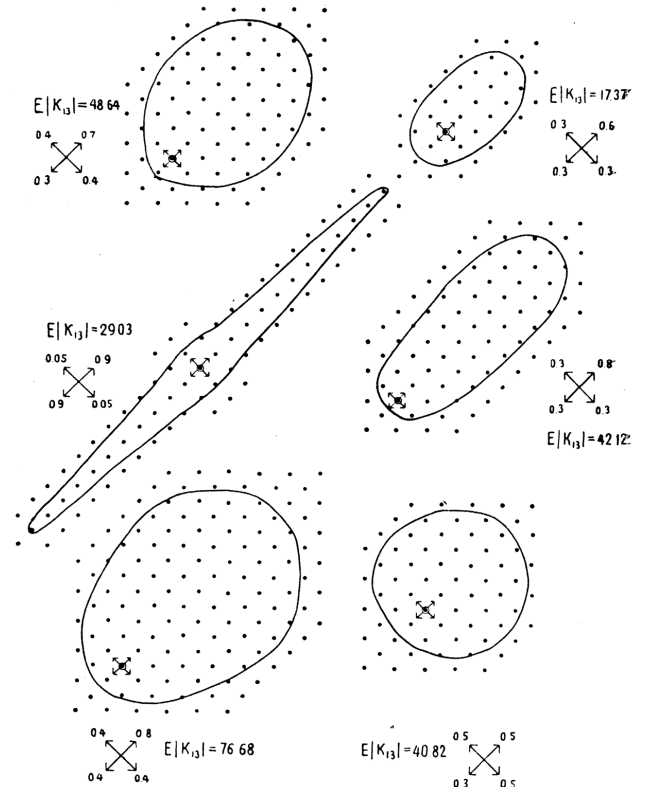


Рис. 2: *Решетчатые ожидания*  $\mathcal{E}K_{13}$  случайного множества  $K_{13}$  для различных ПСР. — *Lattice expectations*  $\mathcal{E}K_{13}$  of the random set  $K_{13}$  for various RSPs.



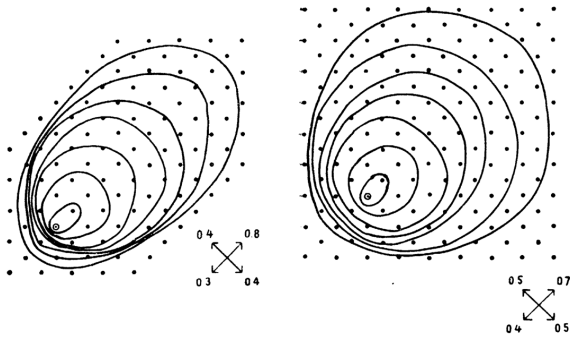


Рис. 3: Две вложенные последовательности  $\{\mathcal{E}K_t, t = 1, 3, 5, 7, 9, 11, 13\}$  решетчатых ожиданий случайных множеств для ПСР(0.3; 0.4; 0.4; 0.8) и ПСР(0.4; 0.5; 0.5; 0.7). — Two embedded sequences  $\{\mathcal{E}K_t, t = 1, 3, 5, 7, 9, 11, 13\}$  of lattice expectations of random sets for the RSP(0.3; 0.4; 0.4; 0.8) and the RSP(0.4; 0.5; 0.5; 0.7).

На рис. 4 приведены две последовательности решетчатых ожиданий случайных множеств  $\{\mathcal{E}K_t, t = 1, 3, 5, 7, 9, 11, 13\}$  для двух ПСР.

В заключение рассмотрим возможные применения предложенного метода оценки средних контуров лесного пожара.

Имеющаяся программа экспертных оценок<sup>6</sup> поля вероятностей распространения огня для конкретных лесных территорий и программа статистических оценок решетчатых ожиданий случайных множеств, образующих ПСР, позволяют моделировать средние контуры пожара и их развитие во времени для различных лесотаксационных условий. Это дает возможность решать задачи, касающиеся вопросов идентификации вероятностных моделей лесных пожаров (т.е. определения и уточнения их параметров). В первую очередь имеются в виду вероятности и скорости распространения пожара по лесному массиву.

Нет необходимости говорить о важности этой задачи, тем более, что до сих пор не существовало удовлетворительных методов расчета зависимости параметров моделей лесных пожаров от всего многообразия существующих лесотаксационных условий.

## 2.2 Estimating mean contours of forest fire from the probability model of fire spread [2, 1, 1975]

In our paper [4, 5, 1973] the probabilistic model of fire spread on a large forest which consists of the separate sites located in units of a plane square lattice was considered. This probabilistic model, *Random Spread Process (RSP)*, is defined as the set of discrete random variables

$$U = \{u_{rt}, r \in R_2, t \in T\}, \quad (2.2.1)$$

<sup>6</sup>См. первую часть моей работы [1, 2], не вошедшую в этот римейк.

where  $R_2 \subset \mathbb{R}^2$  is a plane lattice,  $T$  is a space of time parameter,  $u_{rt}$  is a discrete random variable with possible values:

$$u_{rt} = \begin{cases} 1, & \text{if the site } r \text{ burns during the moment } t; \\ 0, & \text{if the site } r \text{ didn't burn till the moment } t; \\ -1, & \text{if the site } r \text{ has burned down by the moment } t. \end{cases} \quad (2.2.2)$$

Besides it is supposed, that joint probabilistic distribution of random variables  $u_{rt}$  for all  $t \in T$  is completely defined by the field

$$p = \{p(r)\}_{r \in R_2}, \quad (2.2.3)$$

of so called *spread probabilities*, where

$$p(r) = [p_1(r); p_2(r); p_3(r); p_4(r)] \quad (2.2.3')$$

is the *vector of spread probabilities* from the unit  $r$  and  $p_k(r)$  corresponds to the spread probability from the unit  $r$  to the  $k$ -th adjacent<sup>7</sup> unit.

In the paper a more general definition of RSP is offered which will allow to do enough simple and effective estimation of mean contours of forest fire.

From the definition (2.2.1) and (2.2.2) it follows that a state of RSP during the moment  $t$  can be defined by allocating such two subsets of units  $K_t^1$  and  $K_t^2$  of the square lattice  $R_2$ , that

$$\begin{cases} K_t^1 = \{r : t \in R_2, u_{rt} \neq 0\}, \\ K_t^2 = \{r : t \in R_2, u_{rt} = -1\}. \end{cases} \quad (2.2.4)$$

It is clear, that  $K_t^2 \subseteq K_t^1$  and

$$K_t^1 \setminus K_t^2 = \{r : t \in R_2, u_{rt} = 1\}. \quad (2.2.5)$$

Thus the RSP can be considered as a sequence of pairs  $[K_t^1, K_t^2]$  for all  $t \in T$ . It is necessary to note, that subsets  $K_t^1$  and  $K_t^2$  have, generally speaking, a *random form*<sup>8</sup> with probabilistic characteristics which are defined by the field  $p$  of spread probabilities (2.2.3).

The more general definition of RSP based on this approach, demands an introduction of strictly concept corresponding our intuitive representation about “*random subset of units*” of a lattice.

It is a question about a *random set* which is possible to consider as a generalization of a discrete random variable and “values” of which are not (unlike a discrete random variable) real numbers, but subsets of units of a lattice.

In this paper we'll be limited to only elementary equivalent of a random set definition.

**Definition.** The *random set*  $K$  is no more than countable set of pairs  $(B_i, p_i)_{i \in I}$ , where  $B_i \subseteq R_2$  is some

<sup>7</sup>Rules of *neighbourhood* on  $R_2$  is stipulated specially.

<sup>8</sup>on the square lattice  $R_2$ .

finite subset of units of a lattice, and  $p_i$  is the probability of concurrence of random set  $K$  with the subset  $B_i$ :

$$p_i = \mathbf{P}(K = B_i). \quad (2.2.6)$$

Besides it is required, that

$$\sum_{i \in I} p_i = 1. \quad (2.2.7)$$

A *random vector-set* is the ordered set of random sets.

$$\overline{K} = [K^1, K^2, \dots, K^n]. \quad (2.2.8)$$

The concept of a random set allows to define a *lattice random process* as a sequence of random sets. The Random Spread Process is defined as a special case of the lattice random process in which random sets of the lattice process should satisfy to (2.2.4). And as we already marked, probabilistic distribution of random variables  $u_{rt}$  is completely defined by a field of spread probabilities  $p$  and the RSP is completely defined by this field.

The basic practical result of the given paper is the offered method of an estimation of an mean contour of forest fire from probabilistic model of its spread, or (that the same) a method of estimation “mean values” of the random sets forming RSP.

Let  $\overline{K} = \{\overline{K}_t, t \in T\}$  be a random spread process where  $\overline{K}_t = [K_t^1, K_t^2]$  is a random vector-set. We'll put the problem of definition of “mean value” of random sets from  $\overline{K}_t$  for any  $t \in T$ .

Abundantly clear, that the problem put thus, is incorrect, as the concept of an “mean value” of a random set the solution of the problem will be essentially depended on a choice of this definition will be not certain yet.

Hence, first of all it is necessary to agree what subset of units of a lattice will be considered as “mean value” of a random set. It would be desirable also that again entered concept of the “mean value” in the certain sense was similar to concept of an expectation of a discrete random variable.

First we'll notice, that if  $|K|$  is a number of units in random set  $K$ , i.e.  $|K|$  is a discrete random variable with the given distribution, it is possible to speak about its “mean value”, an expectation  $\mathbf{E}|K|$ .

**Definition.** The *lattice expectation of random set*  $K$  is the subset  $\mathcal{E}K \subseteq R_2$ , satisfying to two conditions:

$$|\mathbf{E}|K| - |\mathcal{E}K|| = \min_{B \subseteq R_2} |\mathbf{E}|K| - |B||, \quad (2.2, A)$$

$$\mathbf{E}|K \Delta \mathcal{E}K| = \min_{B \subseteq R_2} \mathbf{E}|K \Delta B|, \quad (2.2, B)$$

which order to choose from all subsets which satisfy to a condition  $A$  such subset which on

the mean least evades from the random set  $K$ , i.e. satisfies to a condition  $B$ .

Certain thus the *lattice expectation of a random set* possesses a number of the properties similar to properties of an *expectation of a random variable*.<sup>9</sup>

Let's stop on the most important characteristic property which allows to find the *lattice expectation of random set*.

Let's enter into consideration probability

$$\pi(r) = \mathbf{P}(r \in K). \quad (2.2.11)$$

It appears, that in the *lattice expectation*  $\mathcal{E}K$  those enter, in round figures,  $\mathbf{E}|K|$  units of a lattice, for which probabilities  $\pi(r)$  would accept values more than for other units.

This important characteristic property allows to develop algorithm and the computer program “M-222” which calculate estimations the *lattice expectation of a random set*.

The same algorithm is applied to an estimation of lattices expectations of random sets forming Random Spread Process.

Modelling RSP, defined by various *vectors of spread probabilities*

$$p(r) = [p_1; p_2; p_3; p_4], \quad (2.2.3'')$$

is lead which are assumed by constants for all  $r \in R_2$ , and for each of them lattice expectations of random sets forming this ПСР ( $p_1; p_2; p_3; p_4$ ) are estimated.

In fig. 2 the part of a lattice  $R_2$  on which the RSP is modelling, in that kind in which it prints by COMPUTER is shown. On a place of each unit  $r \in R_2$  the quantity  $n(r)$  realizations from 100 in which the unit  $r$  belongs to the random set  $K_t$  at  $t = 13$  settles down, and the estimation  $\mathbf{P}(r \in K_{13})$ , probabilities of that  $r \in K_{13}$ , — looks like:  $\pi(r) \approx n(r)/100$ . The contour on a plane  $\mathbb{R}^2$ , containing the lattice expectation  $\mathcal{E}K_{13}$  of the random set  $K_{13}$  was spent **manually** on the basis of an estimation of the mean  $\mathbf{E}|K_{13}|$  of the random variable  $|K_{13}|$ , calculated by COMPUTER.

In fig. 3 a number of *lattice expectations of random sets* for various RSPs is shown.

In fig. 4 two sequences of lattice expectations of random sets  $\{\mathcal{E}K_t, t = 1, 3, 5, 7, 9, 11, 13\}$  for two RSPs are resulted.

<sup>9</sup> For example, for a random variable  $\xi$  and its expectation  $\mathbf{E}\xi$  it is always fair

$$\mathbf{E}(\xi - \mathbf{E}\xi) = 0. \quad (2.2.9)$$

Similar property is carried out for any random set  $K$  and its lattice expectation  $\mathcal{E}K$ :

$$\mathbf{E}|K \setminus \mathcal{E}K| = \mathbf{E}|\mathcal{E}K \setminus K|. \quad (2.2.10)$$

This analogy and the formula (2.2.10) have appeared erroneous and consequently in a remake are omitted in a footnote.

In summary we'll consider possible applications of the offered method of an estimation of mean contours of forest fire.

An available program of expert estimations<sup>10</sup> fields of spread probabilities of fire for concrete wood territories and the program of statistical estimations *lattice expectations of random sets*, forming RSP, allow to model mean contours of a fire and their development in time for various wood conditions. It enables to solve the problems of identifications of probabilistic models of forest fires concerning questions (i.e. definitions and specifications of their parameters). First of all probabilities and speeds of spread of a fire on a large forest mean.

There is no necessity to speak about importance of this problem as till now there were no satisfactory methods of calculation of dependence of parameters of models of forest fires from all variety existing wood conditions.

### 3 Назад в суету

*В суету городов и в потоки машин  
Возвращаемся мы, просто некуда деться,  
И спускаемся вниз с покоренных вершин,  
Оставляя в горах,  
Оставляя в горах свое сердце.*

Владимир Высоцкий, 1966.

Говорят, мол, информирован, значит, вооружен. Глубоко рациональная мысль. Подходит к чему угодно, но только не к открытию нового. Ни для кого не секрет — излишняя информация замечает начисто все пути, кроме уже проторенной колеи. Излишняя, это сверх чего? От какого тянущего вниз груза стоит избавиться, чтобы разглядеть никем не исхоженную вершину, зацепиться за нее и удержаться там на мгновение.

Нет ответа.

\*\*\*

Вот ты на вершине. И куда теперь?

Назад? Вниз? Никаких подъемов? Только спуск?

В суету? Больше некуда деться?

Нет, ну, можно, конечно, поискать еще... Повыше.

Нет, но — чего-нибудь уж совсем наинейшего...

Да, ну, поискать, конечно, можно.

Да, но вдруг ничего такого там не отыщется?

...

Больше некуда деться!

Или, как всегда, в неизвестную суету,

Или снова к неизвестной вершине.

\*\*\*

<sup>10</sup>Look the first part of my paper [1, 2], not entered into this remake.

### 4 Справка: средневероятное событие

В [6, 7, 1984], а также в [8, 9, 1999, стр. 644] можно найти определение понятия *среднемерного множества*<sup>11</sup>, которое впервые было введено мною в 1973 и опубликовано в [1, 2, 1975] и [13, 14, 1977], а затем использовано другими авторами, например, в [15, 16, 1985, 1986] и в [3, 1994]. *Среднемерное множество* — это средняя множественная характеристика *случайного множества*, значениями которого служат подмножества измеримого пространства с мерой, играющая для случайного множества ту же роль, которую для *случайного элемента* со значениями из линейного пространства<sup>12</sup> играет *математическое ожидание*, или *среднее значение*. В эвентологии уже давно существует понятие *среднего множества событий* [17, 18] — результат буквального применения определения среднемерного множества для *случайного множества событий*. Совсем недавно в работе [19, 20] общая идея этого давнего определения *среднемерного множества* была опять взята мною на вооружение, чтобы существенно иначе определить еще одно новое для эвентологии понятие, которое я назвал *средневероятным событием*.

**Определение (средневероятное событие для конечного множества событий).** *Средневероятным событием* для конечного множества событий  $\mathfrak{X} \subset \mathcal{A}$  называется любое такое всеобщее событие  $\hat{x}_{\mathfrak{X}} \in \mathcal{A}$ , удовлетворяющее включениям

$$\sum_{|X| > m} \text{ter}(X // \mathfrak{X}) \subset \hat{x}_{\mathfrak{X}} \subseteq \sum_{|X| \geq m} \text{ter}(X // \mathfrak{X}), \quad (4.2)$$

которое наступает с вероятностью

$$\mathbf{P}(\hat{x}_{\mathfrak{X}}) = \frac{1}{|\mathfrak{X}|} \sum_{x \in \mathfrak{X}} \mathbf{P}(x) \quad (4.3)$$

всякий раз, когда среди событий из  $\mathfrak{X}$  наступает не менее, чем  $m$  событий, где  $m \in \{0, 1, \dots, |\mathfrak{X}|\}$  удовлетворяет неравенствам

$$\sum_{|X| > m} p(X // \mathfrak{X}) < \mathbf{P}(\hat{x}_{\mathfrak{X}}) \leq \sum_{|X| \geq m} p(X // \mathfrak{X}). \quad (4.4)$$

Разумеется, средневероятное событие обладает известными экстремальными свойствами, которые формулируются в терминах вероятностного расстояния [19, 20].

### Список литературы

- [1] О. Ю. Воробьев. Определение вероятностей пространства горения и оценка развития средних контуров лесного пожара. *Охрана лесных ресурсов Сибири. Красноярск, ИЛЧД СО АН СССР*, 1:43–67, 1975, URL: <http://eventology-theory.com/0-1ec/remake-1975-full-GSFR-43-67.pdf>.

<sup>11</sup>В последнее время множатся примеры использования этого понятия в различных областях математики и приложений, и во многих западных работах по теории случайных множеств (см., например, [10, 2012], [11, 2013], [12, 2005] и пр.) его обычно называют с легкой руки Дитриха Штойяна [3, стр. 113–115, 1994] “*Vorob'ev expectation (воробьевское ожидание)*”, опираясь на устаревшую английскую транслитерацию моей фамилии, которую теперь я пишу слегка иначе.

<sup>12</sup>Например, для случайной величины, вектора, матрицы, функции и т.п.

- [2] O. Yu. Vorobyev. Definition of probabilities of fire spread and estimating mean forest fire spread sets. *The Guarding of Siberia Forest Resources. Krasnoyarsk, The Sukachev Institute of Forest and Wood, USSR AS, SB*, 1:43–67, 1975 (in Russian), URL: <http://eventology-theory.com/0-lect/remake-1975-full-GSFR-43-67.pdf>.
- [3] D Stoyan and H. Stoyan. *Fractals, Random Shapes and Point Fields. Methods of Geometrical Statistics. XIV*. John Wiley & Sons, Chichester etc., 1994, 389p.
- [4] О. Ю. Воробьев. Математическое описание процессов случайного распространения и управление ими. *Известия СО АН СССР*, 13(3):146–152, 1973.
- [5] O. Yu. Vorobyev. Mathematical description of random spread processes and its control. *Izvestia of SB AS USSR*, 13(3):146–152, 1973 (in Russian).
- [6] О. Ю. Воробьев. *Среднемерное моделирование*. Наука, Москва, 1984, 133с.
- [7] O. Yu. Vorobyev. *Mean Measure Modeling*. Nauka, Moscow, 1984 (in Russian).
- [8] *Вероятность и математическая статистика. Энциклопедия*. Научное издательство Большая Российская Энциклопедия, Москва, 1999.
- [9] *Probability and Mathematical Statistics. Encyclopedia*. Science publisher Great Russian Encyclopedia, Moscow, 1999 (in Russian).
- [10] P. Heinrich, R. C. Stoica, and V. C. Tran. Level sets esimation and Vorob'ev expectation of random compact sets. *Spatial Statistics*, 2:47–61, December 2012.
- [11] C. Chevalier, D. Ginsbourger, J. Bect, and Molchanov I. Estimating and Quantifying Uncertainties on Level Sets Using the Vorob'ev Expectation and Deviation with Gaussian Process Models. *Contributions to Statistics. Advances in Model-Oriented Design and Analysis — mODa 10. D. Uciński et al. (eds.)*, pages 35–43. Springer International Publishing, Switzerland, 2013.
- [12] I. Molchanov. *Theory of Random Sets*. Springer-Verlag, London etc., 2005.
- [13] О. Ю. Воробьев. О множественных характеристиках состояний распределенных вероятностных процессов. *Известия СО АН СССР*, 3(3):3–7, 1977.
- [14] O. Yu. Vorobyev. On set characteristics of states of distributed probability proesses. *Izvestia of SB AS USSR*, 3(3):3–7, 1977 (in Russian).
- [15] С. А. Ковязин. О предельном поведении одного класса эмпирических средних случайного множества. *Теория вероятностей и ее применения*, 30(4):767–772, 1985.
- [16] S. Kovyazin. On the limit behavior of a class of empirical means of a random set. *Theory of Probability and its Applications*, 30(4):814–820. Translated from Russian by J. Malek., 1986.
- [17] О. Ю. Воробьев. *Эвентология*. Сибирский федеральный университет, Красноярск, 2007, 435с., <http://eventology-theory.ru/0-books/1-VorobyevOleg~2007~Eventology~435p.pdf>.
- [18] O. Yu. Vorobyev. *Eventology*. Siberian Federal University, Krasnoyarsk, Russia, 2007, 435p. (in Russian, abstract in English), <http://eventology-theory.com/0-books/1-VorobyevOleg~2007~Eventology~435p.pdf>.
- [19] О. Ю. Воробьев. Средневероятное событие для множества событий. *Труды XI Международ. ФАМЭБ конференции по финансово-актуарной математике и эвентологии безопасности*, Красноярск: НИИППБ, СФУ (под ред. Олега Воробьева):139–147, 2012.
- [20] O. Yu. Vorobyev. A mean probability event for a set of events. In. *Proc. of the XI Intern. FAMES Conf. on Financial and Actuarial Mathematics and Eventology of Safety*, Krasnoyarsk, SFU (Oleg Vorobyev ed.):139–147, 2012 (in Russian, abstract in English).



## Применимая эвентология безопасности: неокончательные итоги

Олег Юрьевич Воробьев

Сибирский федеральный университет  
Институт математики и фундаментальной информатики  
Торгово-экономический институт  
Красноярск  
oleg.yu.vorobyev@gmail.com

**Аннотация.** Итоги эвентологического [1] моделирования системы безопасности [2, 3, 4, 5, 6, 7, 8] рассматриваются в примерах и иллюстрациях, которые предназначены для демонстрации основных особенностей оценивания риска опасного события на предприятии при сложившихся событийных обстоятельствах на основе портфеля идентификационных признаков безопасности; кроме того, эти примеры и иллюстрации показывают роль и функции в оценивании риска трех основных событийных фигурантов эвентологической системы безопасности: совокупного субъекта, совокупного объекта и совокупного барьера; и, самое главное, обнаруживают ключ применимости эвентологического подхода в области безопасности в методах выбора оптимального портфеля идентификационных признаков безопасности, обеспечивающего заданную точность оценки риска опасного события на данном предприятии минимальными экспертными затратами.

**Ключевые слова.** Эвентология, применимая эвентология, теория вероятностей, событие, вероятность, множество событий, алгебра событий, среднеевероятное событие, ценность события, гиббсовская событийная модель, событийная идентификация, совокупный субъект, совокупный объект, совокупный барьер, портфель идентификационных признаков, точность оценки риска, минимальные экспертные затраты.

*Вот ты придумал идею  
— вроде, она достаточно безумна,  
чтобы соответствовать  
этому безумному миру.  
А потом до тебя доходит,  
что мир еще безумнее,  
чем тебе казалось,  
и, чтобы в нем разобраться,  
нужно придумать еще одну идею,  
совсем безумную.  
Примерно, такой чередой безумств  
растет наука.*  
Сэр Роджер Пенроуз, 2013.

В этой работе я намерен чрезвычайно сжато подвести итоги двухлетних эвентологических [1] исследований в сфере безопасности, которые были опубликованы в моих работах [2, 3, 4, 5, 6, 7, 8]<sup>1</sup>, останавливаясь лишь на узловых эвентологических понятиях и моделях, чтобы представить компактную и прозрачную картину вновь сложившегося эвентологического подхода в разработке практических методик в области безопасности и страхования, приведшего в итоге к созданию новой эвентологической системы безопасности предприятия. Помимо всего прочего я намерен подвести итоги работ в этой всё ещё новой и непривычной для большинства исследователей области на примерах, иллюстрирующих характерные особенности эвентологической системы безопасности предприятия, ибо если и требуется разъяснить что-то новое, то это нагляднее сделать на примерах, беспощадно удаляя излишние детали и отсылая за ними к первоисточникам. Характер этой итоговой работы вынуждает меня в немалой степени заниматься самоцитированием. Одни короткие но важные цитируемые разделы приводятся без изменений так, как они выглядят в оригинале. В другие — мною внесены несущественные редакционные правки. Кроме этого, в этот обзор включены некоторые иллюстрации из моих предыдущих работ, снабженные автономными полными подписями, которые не дублируются в тексте. Вместе с тем, данная работа, наряду с неизбежным новым синергетическим эффектом впервые собранных вместе результатов, содержит также и некоторые новые идеи, еще нигде не опубликованные, и поэтому, надеюсь, имеет самостоятельное научное значение для применимой эвентологии безопасности и страхования. Наконец, картина недавних эвентологических исследований в сфере безопасности была бы далеко неполной без результатов работ Аркадия Новоселова [19, 20]<sup>1</sup> по управлению и оптимизации риска опасного события, ссылки на которые я с удовольствием включаю в эти итоги.

<sup>1</sup>На всякий пожарный случай привожу те же самые ссылки на английском: [9, 10, 11, 12, 13, 14, 15, 16] и [17, 18], надеясь что эвентологическим подходом могут заинтересоваться также и нерусскоговорящие энтузиасты. Just in fire case the same references in English: [9, 10, 11, 12, 13, 14, 15, 16] и [17, 18] are given with the hope that the eventological approach can be interested for non-Russian enthusiasts too.



## 1 Эвентология безопасности

Недавно в [2, 10] предложено новое эвентологическое мировоззрение в области безопасности, которая определяется относительно субъекта, события, вероятности и ценности.

### 1.1 Безопасность: субъект, событие, вероятность, ценность [2, 10, стр. 92-93]

Смысл в понятие *опасности*<sup>2</sup> вкладывает только субъект (или множество субъектов), который определяет и характеризует опасность. Без субъекта понятие опасности лишено какого бы то ни было смысла. Опасность — это всегда опасность для субъекта, *субъектная опасность*. Так же как безопасность — это всегда *субъектная безопасность*. Таким образом, и опасность и безопасность — это *субъектные категории*.

Уже давно было замечено и не раз указано [1], что субъект не существует без событий. Событие есть бытие субъекта, субъектное бытие, со-бытие, и только в событии субъект реализует свое существование.

В эвентологии [1] всякое событие имеет свою *вероятность*, а понятие вероятности не имеет смысла до тех пор, пока не определено событие, о вероятности которого идет речь. Таким образом, *события не бывает без вероятности, как и вероятности не бывает без события*. Второй неотъемлемой субъектной характеристикой события в эвентологии считается его ценность. Так же, как и в случае вероятности, эвентологическое понятие ценности не имеет смысла до тех пор, пока не указано событие, о ценности которого идет речь: *события не бывает без ценности, как и ценности не бывает без события*.

Эвентологическая теория характеризует субъекта тем или иным *множеством событий*, которые прямо или непрямо связаны с данным субъектом, отражая ту или иную сторону его бытия. Каждое такое множество событий наступает в виде той или иной *ситуации*, того или иного *стечения обстоятельств*, той или иной *комбинации* наступления или ненаступления событий из этого множества, того или иного *террасного события*, создаваемого этим множеством. И каждое такое террасное событие наступает с соответствующей вероятностью. *Набор вероятностей всех террасных событий*, порождаемых данным множеством событий, называется *вероятностным распределением* данного множества событий, а *набор ценностей всех террасных событий* называется его *ценностным распределением*<sup>3</sup>.

<sup>2</sup> Или безопасности, как отсутствие опасности.

<sup>3</sup> Вероятностное и ценностное распределение множества события вместе определяют гиббсовскую характеристику эвентологического распределения (*Э-распределения*) данного множества событий.

Здесь нам интересны те стороны бытия субъекта, те *множества событий* вместе с их *Э-распределениями*, которые порождают его *опасность* или его *безопасность*. Само собой разумеется, что *субъектная опасность* и *субъектная безопасность* — это *относительные категории*, определяемые относительно того *множества событий*, которое всякий раз выбирает сам субъект.

Выводам субъекта о наличии опасности или безопасности при текущем стечении обстоятельств всегда предшествует сознательная или бессознательная оценка субъектом *вероятности* и *ценности* причинения или непричинения ущерба субъекту данным стечением обстоятельств — оценка субъектом *вероятности* и *ценности* данного террасного события. Таким образом, и опасность, и безопасность являются не только *субъектными* и *относительными*, но также и *вероятностными*, и *ценностными* категориями.

\*\*\*

Уместно еще раз вслед за [2] и [3] подчеркнуть важную идею, раскрывающую эвентологическое понимание безопасности и страхования:

- и безопасность, и страхование — это всегда *субъектная безопасность* и *субъектное страхование*, измеряемые математически на основе эвентологической модели субъекта и субъектных методов учета и управления вероятностными и ценностными эвентологическими распределениями множеств событий в области безопасности и страхования.

\*\*\*

В [3, 11] предлагается воспользоваться методами эвентологического моделирования систем безопасности в рамках эвентологического системного анализа, результатом которых служит эвентологическая модель системы безопасности субъектов на предприятии.

### 1.2 Эвентологическая система безопасности

В работе [3] предложены математические модели *систем безопасности*<sup>4</sup>, построенные на основе эвентологической теории систем [1] и последних разработок в области эвентологии безопасности [2]. Системы (пожарной) безопасности существуют в каждом уголке мира, в каждой отрасли и на каждом предприятии. Современные подходы к созданию систем (пожарной) безопасности [21, 22, 23], несмотря на специфические страновые, отраслевые и индивидуальные

<sup>4</sup> В том числе, *систем пожарной безопасности*.

системные особенности, имеют общий системный базис. Эвентологическая теория систем позволяет разработать такую математическую модель, которая открывает возможность не только выразить в единой *эвентологической системе безопасности* и системный базис, и системную оболочку, но также объяснить и измерить структуру системных событийных взаимодействий между ними.

Чтобы эвентологически выразить, объяснить и измерить системы безопасности, как *системы событий*, надо сначала договориться, что в эвентологии понимать под *системой событий*, и, в частности, чем понятие *системы событий* должно отличаться от до сих пор центрального в этой теории понятия *множества событий*.

В эвентологии математические модели систем безопасности рассматриваются в рамках *эвентологической теории систем (эвентологического системного анализа)* [1]. Начиная с нашей работы [13], под *эвентологической системой (системой событий)* предложено понимать множество событий, которое складывается из *свободного множества событий (системный базис)* и событий, операционно связанных с событиями из базиса (*системная оболочка*).

В качестве основных ориентиров в области разработки систем безопасности нами были выбраны из довольно внушительного списка две работы. Одну [22] можно по праву считать наиболее известным отечественным достижением в области пожарной безопасности, а другую [23] — образцово выполненным международным проектом системы безопасности.

Первая работа [22] позволила провести полезный сравнительный анализ. А вторая [23], в которой вводятся понятия предупредительных и реагирующих *барьеров*, натолкнула на мысль ввести в эвентологическом системном анализе безопасности новые понятия и термины: события, которые имеют отношение к мероприятиям, обеспечивающим безопасность, названы *барьерными событиями*, а эвентологическая модель совокупности барьерных событий — *совокупным барьером*, который вместе с *совокупным субъектом* и *совокупным объектом* является одним из трех основных *событийных фигурантов* предлагаемой в работе *эвентологической модели системы безопасности субъектов на предприятии*.

## 2 Глоссарий понятий и задач применимой эвентологии безопасности

1. **Гиббсовская модель** эвентологической системы безопасности предприятия — эвентологическая модель, описывающая событийное поведение трех совокупных фигурантов безопасности: *субъекта, барьера и объекта*, и оценивающая риск (вероятность опасного события) на данном предприятии по результатам экспертной проверки событийного состояния безопасности предприятия, проводимой в рамках *экспертного портфеля И-признаков*.

2. **Идентификация нормативных параметров гиббсовской событийной модели** — максимального, среднего и минимального риска (вероятности опасного события) — для данного предприятия по совокупности исторической, экспертной и модельной статистик (см. параграф 4.3).
3. **“И-признак”** — *идентификационный признак безопасности* данного предприятия, значения которого, оцениваемые экспертом, определяют выполнение/невыполнение нормативного требования безопасности; синоним понятия **“И-события”** — *идентификационного события безопасности* данного предприятия, наступление которого характеризует выполнение нормативного требования безопасности и оценивается экспертом;
4. **Портфель “И-признаков”** — множество И-признаков  $\mathcal{A}$ , используемых в эвентологической модели безопасности для оценки риска (вероятности) опасного события в данной области безопасности.
5. **Фигурантный портфель “И-признаков”** — одно из трех подмножеств И-признаков, характеризующих каждого из трех фигурантов в отдельности: *совокупного субъекта, совокупный барьер и совокупный объект*, и используемых в эвентологической модели безопасности для оценки риска (вероятности) опасного события на данном предприятии для соответствующего фигуранта (**субъектный**:  $\mathcal{M}$ , **барьерный**:  $\mathcal{B}$ , **объектный**:  $\mathcal{X}$ ).
6. **Внутренний фигурантный подпортфель “И-признаков”** — одно из двух подмножеств И-признаков, характеризующих двух-множественное состояние каждого из трех фигурантов: *совокупного субъекта, совокупного барьера и совокупного объекта*, и используемых в эвентологической модели безопасности для оценки риска (вероятности) опасного события на данном предприятии для соответствующего фигуранта (**внутренние субъектные**:  $\mathcal{M}_1, \mathcal{M}_2$ , **внутренние барьерные**:  $\mathcal{B}_1, \mathcal{B}_2$ , **внутренние объектные**:  $\mathcal{X}_1, \mathcal{X}_2$ ).
7. **Средневероятный портфель “И-признаков”** — портфель  $\mathcal{A}$ , который аппроксимирует портфель И-признаков  $\mathcal{A}$  в средневероятном; составлен из средневероятных И-признаков, каждый из которых аппроксимирует в средневероятном один из подпортфелей портфеля  $\mathcal{A}$  соответственно.
8. **Экспертный портфель “И-признаков”** — множество И-признаков  $\mathcal{A}^{(e)}$ , выбранное из общего портфеля И-признаков  $\mathcal{A}$  и предлагаемое для экспертной проверки уровня безопасности данного предприятия; аналогично определяются *экспертные фигурантные портфели и внутренние подпортфели*.
9. **Минимальный экспертный портфель “И-признаков”** — экспертный портфель, в котором каждый И-признак принадлежит *только одному* из шести подпортфелей и каждый из шести подпортфелей содержит *только по одному* И-признаку.
10. **Взвешенный минимальный экспертный портфель “И-признаков”** — экспертный портфель, в котором каждый И-признак принадлежит *только одному* из шести подпортфелей, а каждый из шести подпортфелей содержит *не меньше одного* И-признака.
11. **Оценка портфельного веса И-признака** (на основе статистики опросов экспертов и на основе *гиббсовской модели*, из которой следует лог-зависимость портфельного веса И-признака от вероятности его значений), который характеризует его влияние на риск (вероятность) опасного события среди других И-признаков портфеля (см. параграф 4.5).
12. **Оценка информационной емкости И-признака** в портфеле (на основе статистики опросов экспертов и оценки информации, получаемой при проверке его значения), которая характеризует информационную важность экспертной проверки значения этого признака.
13. **Оценка точности расчета риска** (вероятности опасного события) на данном предприятии и ее зависимости от количества И-признаков в экспертном портфеле (см. параграф 4.6).
14. **Оптимальный экспертный портфель И-признаков** для данного предприятия — это портфель И-признаков  $\mathcal{A}^{(e)}(\delta)$ , обеспечивающего заданный уровень *точности оценки риска* (вероятности опасного события) при минимальных экспертных затратах (см. параграф 4.5).

### 3 Событийная иерархия эвентологической системы безопасности в картинках

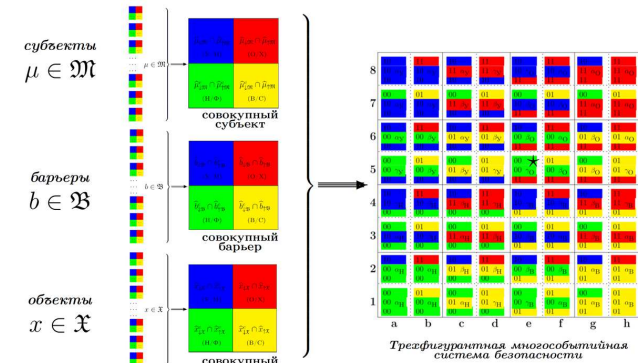


Рис. 1: Событийная иерархия эвентологической системы безопасности состоит из трех уровней — левый: множество субъектов  $\mathcal{M}$ , множество барьеров  $\mathcal{B}$  и множество объектов  $\mathcal{X}$ ; средний: три системных фигуранта; правый: трехфигурная эвентологическая система безопасности.

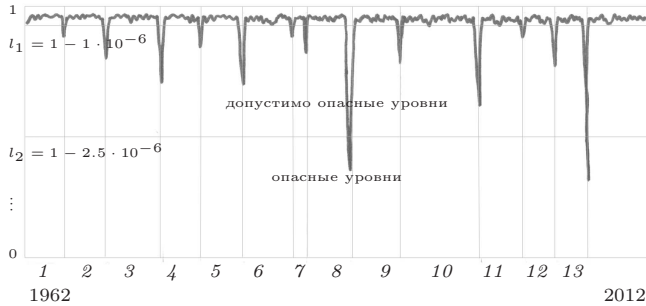


Рис. 2: Траектория безопасности. График вероятности безопасности  $0 < l < 1$  (вертикальная ось в нелинейной шкале) для примерной 50-летней последовательности 13-ти гиперсценарных циклов системы безопасности. Частота опасных пожарных событий равна 0.04 (2 события за 50 лет). Безопасные уровни:  $l_1 < l < 1$ ; допустимо опасные уровни:  $l_2 < l < l_1$ ; опасные уровни:  $0 < l < l_2$ .

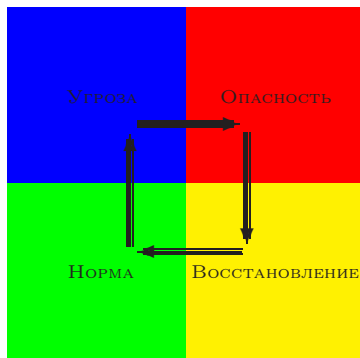


Рис. 3: Многособытийный цикл. Диаграмма Венна, иллюстрирующая многособытийный цикл системы безопасности, характеризующий последовательной сменой средневероятных состояний ее совокупных фигурантов: ... → Норма → Угроза → Опасность → Восстановление → Норма → ...

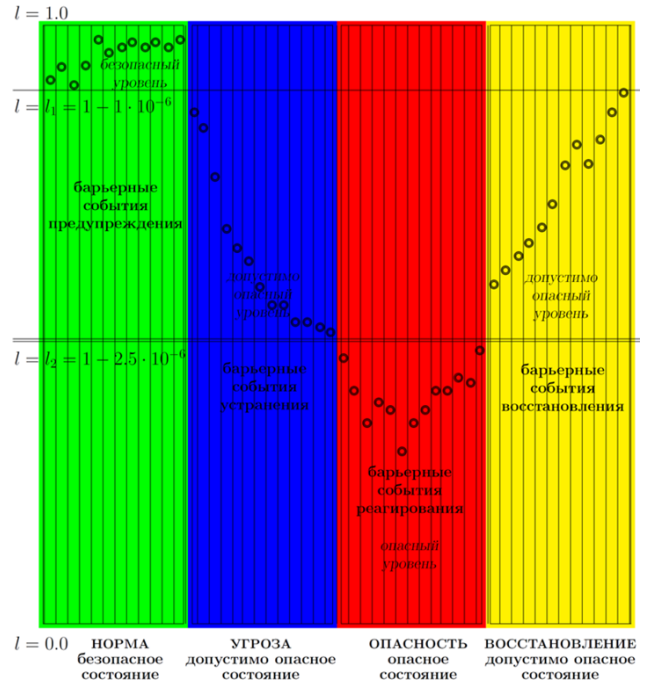


Рис. 4: Многособытийный цикл по времени. Эвентологическая многособытийная модель основного цикла (норма, угроза, опасность, восстановление) системы безопасности субъектов на предприятии. Диаграмма Венна последовательности (слева направо) стечения средневероятных состояний системных фигурантов: совокупных субъектов, барьеров и объектов. На фоне средневероятных состояний совокупного объекта, совокупным субъектом осуществляются барьерные события: предупреждающие (в безопасном состоянии: НОРМА), устраняющие (в допустимо опасном состоянии: УГРОЗА), реагирующие (в опасном состоянии: ОПАСНОСТЬ) и восстанавливающие (в допустимо опасном состоянии: ВОССТАНОВЛЕНИЕ). Вертикальная ось — уровень безопасности ( $0 \leq l \leq 1$  в нелинейной шкале) субъектов, горизонтальная ось — временная последовательность (слева направо) террасных событий.

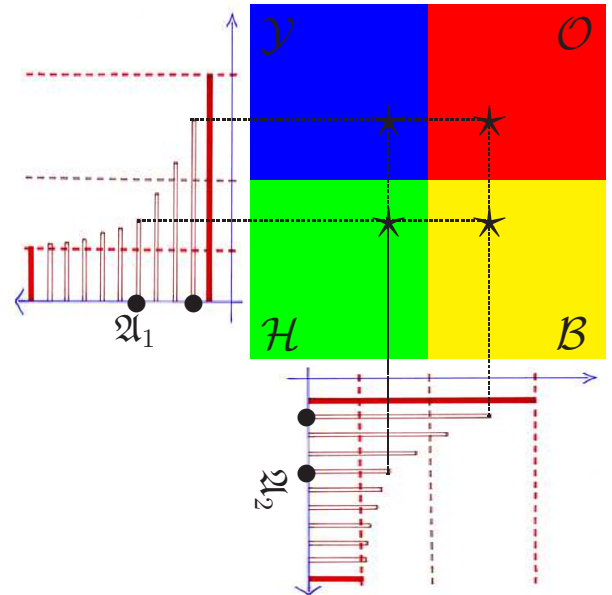


Рис. 5: Идентификация состояния одного из трех фигурантов  $\mathcal{A} = \mathcal{A}_1 \cup \mathcal{A}_2 = \mathcal{M} = \mathcal{M}_1 \cup \mathcal{M}_2$ ,  $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$  и  $\mathcal{X} = \mathcal{X}_1 \cup \mathcal{X}_2$  системы безопасности предприятия на основе данных экспертной проверки, которая идентифицирует (\*) средневероятное состояние Угрозы, Опасности, Восстановления или Нормы у фигуранта  $\mathcal{A} = \mathcal{A}_1 \cup \mathcal{A}_2$ .



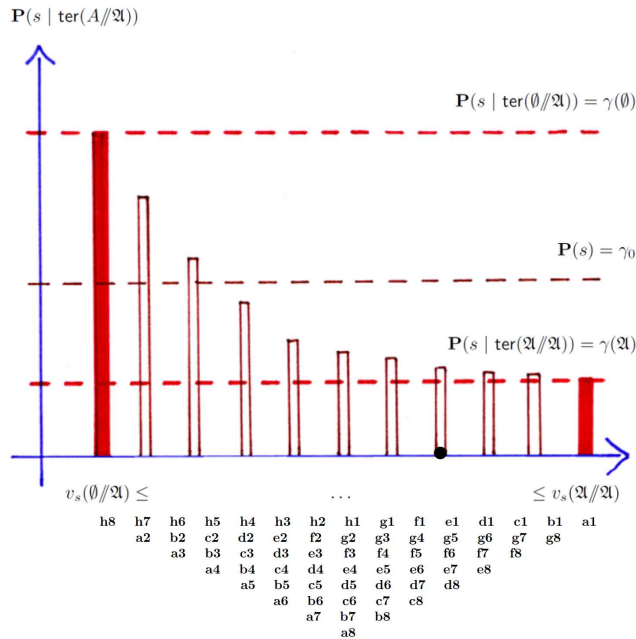


Рис. 6: Модель гиббсовской зависимости риска  $P(s|\text{ter}(A//Q)) = \gamma(\emptyset) \exp\left(\frac{|A|_{v_s} \ln \frac{\gamma(Q)}{\gamma(\emptyset)}}{|Q|_{v_s}}\right)$  опасного события  $s$  от стечения событийных обстоятельств безопасности  $\text{ter}(A//Q)$  (комбинаций значений И-признаков безопасности) для каждого фигуранта  $\mathcal{M}, \mathcal{B}, \mathcal{X}$  или для  $\mathcal{Q}$  — всего предприятия в целом. На гиббсовской модели множества И-событий  $\mathcal{Q}$  основана нормативная идентификация риска опасного события по известным нормативным значениям *наихудшего*:  $\gamma(Q//Q)$ , *наилучшего*:  $\gamma(\emptyset//Q)$  и *среднего риска*:  $\gamma_0$ . По горизонтальной оси отложены портфельные веса (относительно опасного события  $s$ )  $v_s(A//Q) = |A|_{v_s}$ ,  $A \subseteq Q$ , террасных И-событий  $\text{ter}(A//Q)$ , расположенные по возрастанию. Под горизонтальной осью в соответствии со значениями риска опасного события размещены «шахматные» обозначения 64-х средневероятных состояний безопасности.

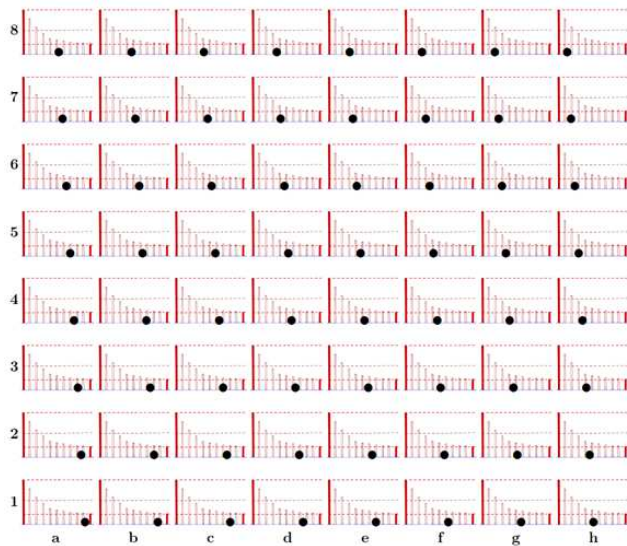


Рис. 7: Гиббсовская зависимость риска опасного события от текущего средневероятного состояния безопасности типичного предприятия. Черными кружками по горизонтальной оси показаны значения ценности каждого из 64-х средневероятных состояний безопасности, по которым на основе гиббсовской событийной модели определяется условная вероятность опасного события при условии текущего состояния безопасности. Наиболее безопасное состояние — **a1**, наименее безопасное состояние — **h8**.

#### 4 Четыре этапа оценивания риска опасного события эвентологической системой безопасности

Процедура оценивания текущего риска опасного события эвентологической системой безопасности данного предприятия состоит из четырех этапов.

- I *Приготовление портфеля И-признаков (идентификационных признаков) безопасности, оптимального для данного предприятия (см. параграф 4.1).*
- II *Экспертиза И-признаков из оптимально приготовленного портфеля (см. параграф 4.2).*
- III *Идентификация текущего состояния безопасности данного предприятия по результатам экспертизы И-признаков (см. параграф 4.3).*
- IV *Расчет риска опасного события на предприятии при текущем состоянии его безопасности, идентифицированном на основе экспертизы И-признаков (см. параграф 4.4).*

##### 4.1 Приготовление оптимального портфеля И-признаков

Приготовление портфеля И-признаков должно удовлетворять двум критериям оптимальности:

1. Портфель И-признаков безопасности должен обеспечивать допустимо минимальные экспертные затраты на проверку И-признаков (см. параграф 4.5);
2. Точность оценки риска опасного события на основе портфеля И-признаков должна быть не хуже заданного уровня (см. параграф 4.6).

*Приготовление оптимального портфеля И-признаков* — это решение эвентологической оптимизационной задачи, в которой оптимизируются экспертные затраты на проверку портфеля И-признаков, обеспечивающие заданную точность оценки риска опасного события (см. параграф 4.6); где под оптимизацией экспертных затрат понимается поиск оптимальных объема, структуры (см. параграф 5) и информационной емкости портфеля И-признаков (см. параграф 4.5).

##### 4.2 Экспертиза И-признаков

Экспертиза И-признаков из оптимально приготовленного портфеля заключается в экспертной оценке бинарных значений И-признаков: присутствует или отсутствует, по мнению эксперта, данный И-признак в характеристике текущего состояния безопасности предприятия.

### 4.3 Идентификация текущего состояния безопасности

Результаты экспертизы И-признаков из оптимально подготовленного портфеля служат исходными данными для идентификации текущего состояния безопасности предприятия. Одним и тем же стандартным алгоритмом [3, 4, 11, 12], основанным на гиббсовской событийной модели (4.4.1) для  $\mathfrak{A} = \mathfrak{M}, \mathfrak{B}$ , или  $\mathfrak{X}$ , проводится идентификация текущего *средневероятного состояния* каждого из трех фигурантов системы безопасности (см. рис. 5 в параграфе 3): *совокупных субъекта  $\mathfrak{M}$ , барьера  $\mathfrak{B}$  и объекта  $\mathfrak{X}$* , тем самым идентифицируется текущее *средневероятное состояние* системы безопасности (см. рис. 1 в параграфе 3). Идентификация текущего состояния системы проводится на основе той же гиббсовской событийной модели (4.4.1) для  $\mathfrak{A} = \mathfrak{M} \cup \mathfrak{B} \cup \mathfrak{X}$ .

### 4.4 Расчет риска опасного события

На основе гиббсовской событийной модели [3, 4, 11, 12] предлагается риск опасного события  $s$  на предприятии при условии текущего состояния его безопасности  $\text{ter}(A//\mathfrak{A})$  рассчитывать по формуле:

$$\mathbf{P}(s|\text{ter}(A//\mathfrak{A})) = \gamma(\emptyset) \exp \left( \frac{|A|_{v_s}}{|\mathfrak{A}|_{v_s}} \ln \frac{\gamma(\mathfrak{A})}{\gamma(\emptyset)} \right), \quad (4.4.1)$$

где  $\gamma(\emptyset)$  — наихудший риск,  $\gamma(\mathfrak{A})$  — наилучший риск,  $\gamma_0 = \mathbf{P}(s)$  — средний риск опасного события (см. рис. 6 в параграфе 3).

### 4.5 Оптимизация экспертных затрат на проверку портфеля И-признаков

Для решения задачи оптимизации экспертных затрат на проверку портфеля И-признаков требуются понятия *портфельного веса* и *информационной емкости И-признака*.

*Портфельный вес*  $w_a = \mathbf{V}(a)$  И-признака/И-события  $a \in \mathfrak{A}$  определяется на основе гиббсовской модели, которая связывает его с вероятностью  $p_a = \mathbf{P}(a)$  наступления И-события  $a \in \mathfrak{A}$  известной формулой [1, 9]

$$p_a = \frac{1}{Z} \exp\{\alpha w_a\}, \quad (4.5.1)$$

где  $\alpha$  и  $Z$  — параметры гиббсовской модели портфеля И-признаков.

Формула (4.5.1) позволяет решать две взаимно обратные вероятностные задачи:

- по известному портфельному весу  $w_a$  И-события  $a \in \mathfrak{A}$  искать его вероятность  $p_a$ ;
- по известной вероятности  $p_a$  И-события  $a \in \mathfrak{A}$  искать его портфельный вес

$$w_a = \frac{1}{\alpha} \ln(p_a Z), \quad (4.5.1')$$

а также задачи управления риском опасного события при различных ограничениях на экспертные затраты [17, 18, 19, 20].

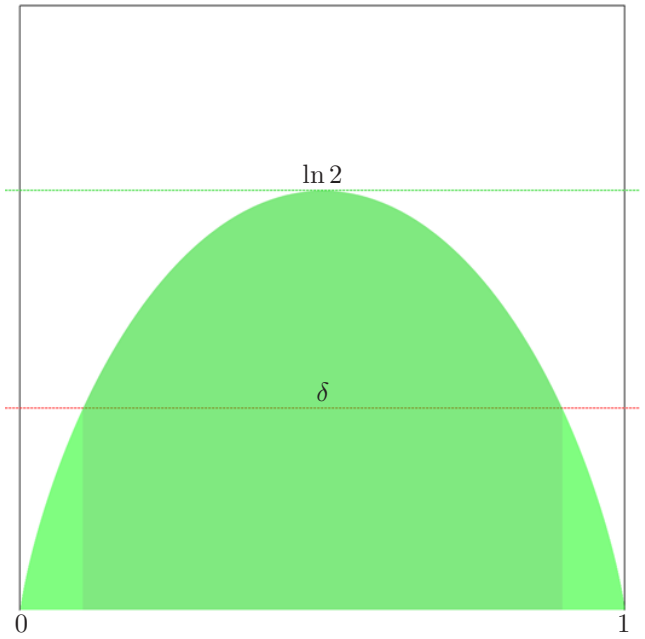


Рис. 8: На единичном квадрате  $[0, 1]^2$  показан подграфик (зеленый) функции информационной емкости  $\mathcal{I}_a$  И-признака  $a$  в зависимости от его вероятности  $p_a$  (горизонтальная ось) (4.5.2), используемой для отбора в оптимальный экспертный портфель тех И-признаков, вероятность которых не меньше порогового значения  $\delta$ .

*Информационная емкость И-признака/И-события* измеряется энтропией И-события  $a \in \mathfrak{A}$  по известной формуле [1, 9] (см. рис. 8)

$$\mathcal{I}_a = -p_a \ln p_a - (1 - p_a) \ln(1 - p_a), \quad (4.5.2)$$

где  $p_a = \mathbf{P}(a)$  — вероятность И-события  $a \in \mathfrak{A}$ .

Интерпретация этой информационной характеристики И-события заключается в том, что если вероятность И-события близка к 0 или к 1, то экспертная проверка его наступления дает мало дополнительной информации, и его включение в экспертный портфель не слишком оправдано. Критерием отбора И-признака  $a \in \mathfrak{A}$  из портфеля  $\mathfrak{A}$  в *оптимальный экспертный портфель*

$$\mathfrak{A}^{(e)}(\delta) = \{a \in \mathfrak{A} : \mathcal{I}_a \geq \delta\} \subseteq \mathfrak{A}, \quad (4.5.3)$$

является значительная величина его информационной емкости  $\mathcal{I}_a$ , превосходящая пороговое значение  $\delta \in [0, \ln 2]$ , которое определяется требуемой точностью оценки риска опасного события.

### 4.6 Оптимизация точности оценки риска опасного события

Погрешность  $\varepsilon$  оценки  $\hat{P}(s)$  риска  $\mathbf{P}(s|\text{ter}(A//\mathfrak{A}))$  опасного события  $s$  на основании выбранного экспертного портфеля  $\mathfrak{A}^{(e)}(\delta)$  может быть вычислена по



формуле:

$$\varepsilon = t_{(1-\alpha)/2, N-1} \frac{\sigma}{\sqrt{N}},$$

где  $t_{(1-\alpha)/2, N-1} = (1-\alpha)/2$ -квантиль  $t$ -распределения Стьюдента с  $N - 1$  степенью свободы,  $\sigma$  — стандартное отклонение,  $\alpha$  — требуемый уровень доверия, обычно принимаемый равным 0.95, а  $N = |\mathfrak{A}^{(e)}(\delta)|$  — мощность экспертного портфеля, т.е. количество И-признаков, включенных в  $\mathfrak{A}^{(e)}(\delta)$ . Таким образом погрешность  $\varepsilon$ , которая тем меньше, чем больше квадратный корень из его мощности  $N$ , определяет доверительный интервал

$$\hat{P}(s) - \varepsilon \leq \mathbf{P}(s|\text{ter}(A/\mathfrak{A})) \leq \hat{P}(s) + \varepsilon,$$

куда значение риска опасного события попадает с вероятностью, не меньшей чем  $\alpha$ .

Кроме мощности экспертного портфеля  $N$ , погрешность  $\varepsilon$  оценки риска существенно зависит еще и от того, насколько зависимость между И-признаками характеризуется данным портфелем отобранных И-признаков. Управлять взаимозависимостью И-признаков в портфеле можно, управляя его событийно-вероятностной структурой. Принципы управления событийно-вероятностной структурой экспертного портфеля, которые приводят к снижению погрешности, изложены в параграфе 5.

## 5 Справка: эвентологическая модель экспертного портфеля И-признаков безопасности

Ключом к применимости эвентологического подхода в области безопасности служит эвентологическая модель структуры ответов эксперта на совокупность нормативных вопросов о безопасности предприятия; иначе говоря, — структуры *экспертного портфеля идентификационных признаков/событий (И-признаков/И-событий) безопасности предприятия*<sup>5</sup>.

Такая событийно-вероятностная структура *И-признаков безопасности* должна быть организована оптимальным образом так, чтобы при минимальных экспертных затратах обеспечить требуемую точность оценки риска (вероятности) опасного события. Теоретически ясно, что большая совокупность нормативных вопросов ведет в результате к большей точности оценки риска. Однако с увеличением этой совокупности, во-первых, растут затраты на получение экспертных ответов, во-вторых, снижается информационная емкость ответа на каждый добавляемый вопрос. Предлагаемая оптимально

организованная структура *экспертного портфеля И-признаков безопасности* решает эту дилемму, обеспечивая требуемую точность минимальными затратами.

### 5.1 Эвентологическая модель портфеля И-признаков

Эвентологическая модель безопасности, предложенная в [2, 3, 4, 5, 6, 7, 8], — это событийно-вероятностная модель системы безопасности предприятия, которая образуется событийным взаимодействием трех системных совокупных фигурантов, каждый из которых отвечает за событийное поведение соответствующей совокупности:  $\mathfrak{M}$  — *субъектов*,  $\mathfrak{B}$  — *барьеров* и  $\mathfrak{X}$  — *объектов* предприятия, совместные событийные состояния которых формируют событийное состояние безопасности всего предприятия в целом.

Эвентологически риск (вероятность) опасного события на предприятии зависит от того, в каком событийном состоянии безопасности находится предприятие. Это событийное состояние характеризуется портфелем И-признаков безопасности данного предприятия, который обычно представляет из себя совокупность ответов экспертов на специально подобранные нормативные вопросы. Состав портфеля И-признаков и организация событийной структуры этих признаков в портфеле во многом определяют точность экспертной оценки риска и объем затрат на проведение экспертизы состояния безопасности предприятия.

Общая структура портфеля И-признаков безопасности определяется иерархической трехфигурантной структурой эвентологической модели безопасности, где каждый из трех совокупных фигурантов обладает собственной двух-множественной структурой. Эта иерархическая структура образуется тремя *фигурантными портфелями И-признаков*

$$\mathfrak{M}, \mathfrak{B}, \mathfrak{X},$$

каждый из которых образован объединением<sup>6</sup> двух *внутренних фигурантных подпортфелей И-признаков*<sup>7</sup>:

$$\mathfrak{M} = \mathfrak{M}_1 \cup \mathfrak{M}_2, \quad \mathfrak{B} = \mathfrak{B}_1 \cup \mathfrak{B}_2, \quad \mathfrak{X} = \mathfrak{X}_1 \cup \mathfrak{X}_2. \quad (5.1.1)$$

<sup>6</sup>Не обязательно непересекающимся.

<sup>7</sup>Каждая пара внутренних фигурантных подпортфелей И-признаков требуется, главным образом, для того, чтобы оценить, в каком состоянии из 4-х состояний цикла безопасности (норма, угроза, опасность, восстановление) находится каждый из трех фигурантов, а, следовательно, в каком состоянии из 64-х находится вся система безопасности данного предприятия. Кроме того, внутренние фигурантные подпортфели И-признаков могут быть использованы для оценки *частных рисков* опасного события для каждого фигуранта в отдельности, или для каждой из пары сторон<sup>8</sup> фигуранта в отдельности.

<sup>5</sup>Определения основных понятий и краткие формулировки задач (обычно выделенные в тексте курсивом), связанные с оптимальным выбором экспертного портфеля идентификационных признаков безопасности предприятия приведены в справочном параграфе 2 на стр. 106.

В итоге структура всего *портфеля И-признаков* безопасности предприятия характеризуется совокупностью

$$\mathfrak{A} = \mathfrak{M}_1 \cup \mathfrak{M}_2 \cup \mathfrak{B}_1 \cup \mathfrak{B}_2 \cup \mathfrak{X}_1 \cup \mathfrak{X}_2, \quad (5.1.2)$$

которая образована объединением<sup>6</sup> шести *фигурантных подпортфелей И-признаков*.

Задача выбора оптимальной структуры экспертного портфеля сводится к выбору шести *фигурантных экспертных подпортфелей И-признаков* (5.1.2), формирующих общий *экспертный портфель И-признаков безопасности* данного предприятия  $\mathfrak{A}^{(e)} \subseteq \mathfrak{A}$ , обеспечивающий требуемую точность оценки минимальными экспертными затратами.

## 5.2 Полный экспертный портфель И-признаков

Таблица 1 иллюстрирует минимальный *полный экспертный портфель И-признаков*

$$\mathfrak{A} = \mathfrak{M}_1 \cup \mathfrak{M}_2 \cup \mathfrak{B}_1 \cup \mathfrak{B}_2 \cup \mathfrak{X}_1 \cup \mathfrak{X}_2 \subset \mathcal{A}, \quad (5.2.1)$$

составленный объединением шести подпортфелей, каждый из которых определяется соответствующим *подмножеством И-событий*

$$\mathfrak{M}_1, \mathfrak{M}_2, \mathfrak{B}_1, \mathfrak{B}_2, \mathfrak{X}_1 \text{ и } \mathfrak{X}_2 \subset \mathcal{A}. \quad (5.2.2)$$

Этот полный экспертный портфель  $\mathfrak{A}$  имеет любопытную событийную интерпретацию, пожалуй, единственно корректную лишь при использовании понятия некой средней событийной характеристики *подмножеств И-событий*. Роль этой средней характеристики может с успехом исполнить, например, недавно введенное в эвентологии понятие *средневероятного события* [7, 15, 8, 16]. В этой интерпретации весь экспертный портфель аппроксимируется множеством всего из шести *средневероятных И-событий*

$$\widehat{\mathfrak{M}}_1, \widehat{\mathfrak{M}}_2, \widehat{\mathfrak{B}}_1, \widehat{\mathfrak{B}}_2, \widehat{\mathfrak{X}}_1 \text{ и } \widehat{\mathfrak{X}}_2 \in \mathcal{A}, \quad (5.2.3)$$

каждое из которых аппроксимирует одно из шести *подмножеств И-событий* (5.2.2), которое, например, для *подмножества И-событий*  $\mathfrak{M}_1 \subset \mathcal{A}$  толкуется так:

$$\widehat{\mathfrak{M}}_1 = \mathcal{E}(\mu/\mathfrak{M}_1) = \widehat{\mu}_{\mathfrak{M}_1} \in \mathcal{A}.$$

Обозначу

$$\widehat{\mathfrak{A}} = \{\widehat{\mathfrak{M}}_1, \widehat{\mathfrak{M}}_2, \widehat{\mathfrak{B}}_1, \widehat{\mathfrak{B}}_2, \widehat{\mathfrak{X}}_1, \widehat{\mathfrak{X}}_2\} \subset \mathcal{A} \quad (5.2.4)$$

— множество, составленное, как из элементов, из шести *средневероятных И-событий* (5.2.3), *аппроксимирующее в средневероятном* весь экспертный портфель  $\mathfrak{A}$  и для  $A \subseteq \mathfrak{A}$  порождающее  $64 = 2^6$  *террасных И-событий II-го рода*

$$\text{ter}_{A//\widehat{\mathfrak{A}}} = \bigcap_{a \in A} a \in \mathcal{A},$$

$\mathfrak{A} \stackrel{mp}{\sim} \widehat{\mathfrak{A}}$	$\widehat{\mathfrak{M}}_1$	$\widehat{\mathfrak{M}}_2$	$\widehat{\mathfrak{B}}_1$	$\widehat{\mathfrak{B}}_2$	$\widehat{\mathfrak{X}}_1$	$\widehat{\mathfrak{X}}_2$
ter--Y	o	o	o	o	o	•
ter--B	o	o	o	o	•	o
ter--H	o	o	o	o	•	•
ter-Y-	o	o	o	•	o	o
ter-YY	o	o	o	•	o	•
ter-YB	o	o	o	•	•	o
ter-YH	o	o	o	•	•	•
ter-B-	o	o	•	o	o	o
ter-BY	o	o	•	o	o	•
ter-BB	o	o	•	o	•	o
ter-BH	o	o	•	o	•	•
ter-H-	o	o	•	•	o	o
ter-HY	o	o	•	•	o	•
ter-HB	o	o	•	•	•	o
ter-HH	o	o	•	•	•	•
terY--	o	•	o	o	o	o
terY-Y	o	•	o	o	o	•
terY-B	o	•	o	o	•	o
terY-H	o	•	o	o	•	•
terYY-	o	•	o	•	o	o
terYYY	o	•	o	•	o	•
terYYB	o	•	o	•	•	o
terYYH	o	•	o	•	•	•
terYB-	o	•	•	o	o	o
terYBY	o	•	•	o	o	•
terYBB	o	•	•	o	•	o
terYBH	o	•	•	o	•	•
terYH-	o	•	•	•	o	o
terYHY	o	•	•	•	o	•
terYHB	o	•	•	•	•	o
terYHH	o	•	•	•	•	•
terB--	•	o	o	o	o	o
terB-Y	•	o	o	o	o	•
terB-B	•	o	o	o	•	o
terB-H	•	o	o	o	•	•
terBY-	•	o	o	•	o	o
terBYY	•	o	o	•	o	•
terBYB	•	o	o	•	•	o
terBYH	•	o	o	•	•	•
terBB-	•	o	•	o	o	o
terBBY	•	o	•	o	o	•
terBBB	•	o	•	o	•	o
terBBH	•	o	•	o	•	•
terBH-	•	o	•	•	o	o
terBHY	•	o	•	•	o	•
terBHB	•	o	•	•	•	o
terBHH	•	o	•	•	•	•
terH--	•	•	o	o	o	o
terH-Y	•	•	o	o	o	•
terH-B	•	•	o	o	•	o
terH-H	•	•	o	o	•	•
terHY-	•	•	o	•	o	o
terHYY	•	•	o	•	o	•
terHYB	•	•	o	•	•	o
terHYH	•	•	o	•	•	•
terHB-	•	•	•	o	o	o
terHBY	•	•	•	o	o	•
terHBB	•	•	•	o	•	o
terHBH	•	•	•	o	•	•
terHH-	•	•	•	•	o	o
terHHY	•	•	•	•	o	•
terHHB	•	•	•	•	•	o
terHHH	•	•	•	•	•	•

Таблица 1: Шесть подпортфелей И-событий:  $\mathfrak{M}_1, \mathfrak{M}_2, \mathfrak{B}_1, \mathfrak{B}_2$  и  $\mathfrak{X}_1, \mathfrak{X}_2$ , формирующие *минимальный полный портфель*  $\mathfrak{A} = \{\text{ter}_{A//\widehat{\mathfrak{A}}}, A \subseteq \widehat{\mathfrak{A}}\}$  из всех возможных (для портфеля  $\mathfrak{A}$ ) 63-х *террасных И-событий II-го рода*, порожденных шестью *средневероятными событиями* (5.2.3) их секстета  $\widehat{\mathfrak{A}}$ , аппроксимирующего  $\mathfrak{A}$  в *средневероятном*:  $\widehat{\mathfrak{A}} \stackrel{mp}{\sim} \mathfrak{A}$  (об аббревиатуре ННН, ..., YYY, ..., BBV для подмножеств  $A \subseteq \mathfrak{A}$  см. замечание 2 на стр. 112).

вероятности которых образуют *вероятностное распределение II-рода*

$$\{p_{A//\hat{\mathfrak{A}}}, A \subseteq \hat{\mathfrak{A}}\}$$

секстета И-событий  $\hat{\mathfrak{A}}$ , где

$$p_{A//\hat{\mathfrak{A}}} = \mathbf{P}(\text{ter}_{A//\hat{\mathfrak{A}}}).$$

**Замечание 1.** Все И-события из полного портфеля  $\mathfrak{A}$  совместны и порождаются шестью *средневероятными* И-событиями из секстета  $\hat{\mathfrak{A}}$  как различные результаты *террасной*<sup>9</sup> операции II-го рода над секстетом  $\hat{\mathfrak{A}}$ , занумерованные различными его подмножествами<sup>10</sup>  $A \subseteq \hat{\mathfrak{A}}$ .

**Замечание 2.** В таблице 1 использована аббревиатура для обозначения подмножеств секстета  $A \subseteq \hat{\mathfrak{A}}$ , отсылающая к состояниям Нормы (Н), Угрозы (У), Восстановления (В) и Опасности (О) каждого из трех совокупных фигурантов системы безопасности по следующему правилу. Для произвольного совокупного фигуранта  $\mathfrak{F} = \mathfrak{M}, \mathfrak{B}$ , или  $\mathfrak{X}$ , аппроксимируемого дуэтом средневероятных событий  $\{\hat{\mathfrak{F}}_1, \hat{\mathfrak{F}}_2\}$ , его пустое подмножество  $\emptyset \subseteq \{\hat{\mathfrak{F}}_1, \hat{\mathfrak{F}}_2\}$  соответствует *средневероятному состоянию* Опасности (О), подмножество  $\{\hat{\mathfrak{F}}_1\}$  — *средневероятному состоянию* Угрозы (У), подмножество  $\{\hat{\mathfrak{F}}_2\}$  — *средневероятному состоянию* Восстановления (В), а подмножество  $\{\hat{\mathfrak{F}}_1, \hat{\mathfrak{F}}_2\}$  — *средневероятному состоянию* Нормы (Н). Поэтому, например, аббревиатура “-УН” соответствует *средневероятному состоянию* системы безопасности, когда *совокупный субъект*  $\mathfrak{M}$  находится в произвольном *средневероятном состоянии*, *совокупный барьер*  $\mathfrak{B}$  — в *средневероятном состоянии* Угрозы (У), а *совокупный объект*  $\mathfrak{X}$  — в *средневероятном состоянии* Нормы (Н). Примеры использования аббревиатуры:

$$\text{ter}_{-} = \text{ter}_{\emptyset//\hat{\mathfrak{A}}},$$

$$\text{ter}_{\text{-Y}} = \text{ter}_{\{\hat{\mathfrak{B}}_2\}/\hat{\mathfrak{A}}},$$

$$\text{ter}_{\text{-YH}} = \text{ter}_{\{\hat{\mathfrak{B}}_2, \hat{\mathfrak{X}}_1, \hat{\mathfrak{X}}_2\}/\hat{\mathfrak{A}}},$$

$$\text{ter}_{\text{-BH}} = \text{ter}_{\{\hat{\mathfrak{B}}_1, \hat{\mathfrak{X}}_1, \hat{\mathfrak{X}}_2\}/\hat{\mathfrak{A}}},$$

$$\text{ter}_{\text{Y-Y}} = \text{ter}_{\{\hat{\mathfrak{M}}_2, \hat{\mathfrak{X}}_2\}/\hat{\mathfrak{A}}},$$

$$\text{ter}_{\text{HHH}} = \text{ter}_{\hat{\mathfrak{A}}//\hat{\mathfrak{A}}}.$$

<sup>9</sup>Террасной называется такая теоретико-множественная операция над множеством событий, результатом которой служит террасное событие одного из шести стандартных родов [1, 9], порожденное этим множеством событий.

<sup>10</sup>Результатами террасной операции II-го рода над событиями из  $\hat{\mathfrak{A}}$  служат террасные события II-го рода  $\text{ter}(A//\hat{\mathfrak{A}}) = \bigcap_{a \in A} a \subseteq \Omega$ , занумерованные подмножествами  $A \subseteq \hat{\mathfrak{A}}$ .

### 5.3 Минимальный экспертный портфель И-признаков

Минимальный экспертный портфель И-признаков (см. табл. 2) хотя и позволяет решить все задачи, стоящие перед эвентологической системой безопасности, но не гарантирует требуемую точность оценки риска.

И-признаки	★★★	Наполнение подпортфелей					
		$\hat{\mathfrak{M}}$		$\hat{\mathfrak{B}}$		$\hat{\mathfrak{X}}$	
		$\hat{\mathfrak{M}}_1$	$\hat{\mathfrak{M}}_2$	$\hat{\mathfrak{B}}_1$	$\hat{\mathfrak{B}}_2$	$\hat{\mathfrak{X}}_1$	$\hat{\mathfrak{X}}_2$
О-противодействие объекта	-Y	-	-	-	-	-	•
О-стойкость объекта?	-B	-	-	-	-	•	-
О-противодействие барьера?	-Y	-	-	-	•	-	-
О-стойкость барьера?	-B	-	-	•	-	-	-
О-противодействие субъекта?	Y	-	•	-	-	-	-
О-стойкость субъекта?	B	-	•	-	-	-	-

Таблица 2: Пример наполнения фигурантных подпортфелей И-признаками в *минимальном* экспертном портфеле, который состоит из 6-ти И-признаков с весами: 1,1,1,1,1,1 (сверху-вниз). В столбце ★★★ указана аббревиатура подмножеств (см. Замечание 2 на стр. 112), которыми занумерованы соответствующие И-события (террасные события II-рода).

### 5.4 Взвешенный минимальный экспертный портфель И-признаков

Взвешенный минимальный экспертный портфель И-признаков (см. табл. 3) также позволяет решить все задачи, стоящие перед системой безопасности, может обеспечить требуемую точность оценки риска при помощи варьирования весов И-признаков в портфеле, но не имеет возможности учета взаимосвязей И-признаков в портфеле.

И-признаки	★★★	Наполнение подпортфелей					
		$\hat{\mathfrak{M}}$		$\hat{\mathfrak{B}}$		$\hat{\mathfrak{X}}$	
		$\hat{\mathfrak{M}}_1$	$\hat{\mathfrak{M}}_2$	$\hat{\mathfrak{B}}_1$	$\hat{\mathfrak{B}}_2$	$\hat{\mathfrak{X}}_1$	$\hat{\mathfrak{X}}_2$
Кол-во въездов-проездов в НБ?	-Y	-	-	-	-	-	•
Уровень дежурной охраны в НБ?	-Y	-	-	-	-	-	•
Сроки энергоустановок в НБ?	-Y	-	-	-	-	-	•
Уровень опасных зон в НБ?	-Y	-	-	-	-	-	•
Системы внутр.Б в НБ?	-Y	-	-	-	-	-	•
Потенц.источн. О в НБ?	-Y	-	-	-	-	-	•
Состояние до СБО?	-Y	-	-	-	-	-	•
Есть дежурная охрана в НБ?	-B	-	-	-	-	-	•
ТО систем внутр.Б в НБ?	-Y	-	-	-	-	-	•
О-противодействие объекта	-Y	-	-	-	-	-	•
Авто-сигнализация об О в НБ?	-Y	-	-	-	-	-	•
Системы жизнеоб-ния в НБ?	-Y	-	-	-	-	-	•
О-стойкость объекта	-B	-	-	-	-	-	•
Есть дежурная охрана Б?	-B	-	-	-	-	-	•
Энергоустановки в НБ?	-B	-	-	-	-	-	•
Есть защита энерго-коммуникаций?	-B	-	-	-	-	-	•
Проведение О-работ в НБ?	-B	-	-	-	-	-	•
О-зон нет?	-B	-	-	-	-	-	•
Есть системы внутр.Б?	-B	-	-	-	-	-	•
Нет потенц. источников О?	-B	-	-	-	-	-	•
Служба борьбы с О (СБО)?	-B	-	-	-	-	-	•
Квал-ция дежурной охраны Б?	-B	-	-	-	-	-	•
О-вещ-ва на раб.местах в НБ?	-B	-	-	-	-	-	•
Класс систем жизнеоб-ния в НБ?	-B	-	-	-	-	-	•
Населенный пункт город?	-B	-	-	-	-	-	•
Время до О-прибора в НБ?	-B	-	-	-	-	-	•
Транспортировка О-веществ нет?	-B	-	-	-	-	-	•
Вещ-в. кот.могут стать О. нет?	-B	-	-	-	-	-	•
О-противодействие барьера?	-Y	-	-	-	-	•	-
Система защиты исправна?	-Y	-	-	-	-	•	-
Первичные средства защиты в НБ?	-Y	-	-	-	-	•	-
Первичные средства защиты?	-B	-	-	-	•	-	-
Система защиты присутствует?	-B	-	-	-	•	-	-
О-стойкость барьера?	-B	-	-	-	•	-	-
Защита энерго-коммуникаций в НБ?	-B	-	-	-	•	-	-
Количество въездов в НБ?	-B	-	-	-	•	-	-
О-противодействие субъекта?	Y	-	•	-	-	-	-
О-стойкость субъекта?	B	-	•	-	-	-	-

Таблица 3: Пример наполнения фигурантных подпортфелей И-признаками в *взвешенном минимальном* экспертном портфеле, который состоит из 38-ми И-признаков, соответствующих 6-ти террасным И-событиям II-го рода с весами: 12,16,3,5,1,1 (сверху-вниз). В столбце ★★★ указана аббревиатура подмножеств (см. Замечание 2 на стр. 112), которыми занумерованы соответствующие И-события II-рода. НБ — нормы безопасности; О — опасный; СБО — служба борьбы с опасностью; САБ — система автоматической безопасности; ТО — техническое обслуживание; ТБ — техника безопасности; АС — аварийные службы; ИТР — инженерно-технические работники.

## 5.5 Произвольный экспертный портфель И-признаков

Произвольный экспертный портфель И-признаков (см. табл. 2), содержащий взвешенный минимальный портфель, позволяет решить все задачи, стоящие перед эвентологической системой безопасности, может обеспечить требуемую точность оценки риска, и имеет возможность учета взаимосвязей И-признаков в портфеле при помощи И-признаков, содержащихся более чем в одном подпортфеле из шести.

И-признаки	★★★	Наполнение подпортфелей					
		М		В		Э	
		М <sub>1</sub>	М <sub>2</sub>	В <sub>1</sub>	В <sub>2</sub>	Э <sub>1</sub>	Э <sub>2</sub>
Кол-во въездов-проездов в НБ?	--Y	-	-	-	-	-	•
Уровень дежурной охраны в НБ?	--Y	-	-	-	-	-	•
Сроки энергоустановок в НБ?	--Y	-	-	-	-	-	•
Уровень опасных зон в НБ?	--Y	-	-	-	-	-	•
Системы внутр.Б в НБ?	--Y	-	-	-	-	-	•
Потенц.источн. О в НБ?	--Y	-	-	-	-	-	•
Расстояние до СБО?	--Y	-	-	-	-	-	•
Число перв. средств защиты в НБ?	--Y	-	-	-	-	-	•
ТО систем внутр.Б в НБ?	--Y	-	-	-	-	-	•
О-противодействие объекта	--Y	-	-	-	-	-	•
Авто-сигнализация об О в НБ?	--Y	-	-	-	-	-	•
Системы жизнеоб-ния в НБ?	--Y	-	-	-	-	-	•
О-стойкость объекта	--B	-	-	-	-	-	•
Есть дежурная охрана Б?	--B	-	-	-	-	-	•
Энергоустановки в НБ?	--B	-	-	-	-	-	•
Есть защита энерго-коммуникаций?	--B	-	-	-	-	-	•
Проведение О-работ в НБ?	--B	-	-	-	-	-	•
О-зон нет?	--B	-	-	-	-	-	•
Есть системы внутр.Б?	--B	-	-	-	-	-	•
Нет потенц. источников О?	--B	-	-	-	-	-	•
Служба борьбы с О (СБО)?	--B	-	-	-	-	-	•
Квал-ция дежурной охраны Б?	--B	-	-	-	-	-	•
О-вещ-ва на раб.местах в НБ?	--B	-	-	-	-	-	•
Класс систем жизнеоб-ния в НБ?	--B	-	-	-	-	-	•
Населенный пункт в НБ?	--B	-	-	-	-	-	•
Временные О-приборы в НБ?	--B	-	-	-	-	-	•
Транспортировки О-веществ нет?	--B	-	-	-	-	-	•
Вещ-я, кот.могут стать О, нет?	--B	-	-	-	-	-	•
О-противодействие барьера?	--Y	-	-	-	-	-	•
Система защиты исправна?	--Y	-	-	-	-	-	•
Первичные средства защиты в НБ?	--Y	-	-	-	-	-	•
Первичные средства защиты?	--B	-	-	-	-	-	•
Система защиты присутствует?	--B	-	-	-	-	-	•
О-стойкость барьера?	--B	-	-	-	-	-	•
Защита энерго-коммуникаций в НБ?	--B	-	-	-	-	-	•
Количество въездов в НБ?	--B	-	-	-	-	-	•
Противопожарные сигналы?	--B	-	-	-	-	-	•
Технические клетки?	--B	-	-	-	-	-	•
Материалы объекта?	--B	-	-	-	-	-	•
Аварийные выходы?	--B	-	-	-	-	-	•
О-противодействие субъекта?	--Y	-	-	-	-	-	•
Пещ.защита исправна?	--Y	-	-	-	-	-	•
Э-ионные пути исправны?	--Y	-	-	-	-	-	•
Аварийные выходы исправны?	--Y	-	-	-	-	-	•
Система эв-ции исправна?	--Y	-	-	-	-	-	•
Группы мал-моби-сти обучены?	--Y	-	-	-	-	-	•
Уровень обучения в НБ?	--Y	-	-	-	-	-	•
Уровень тренировок эв-ции в НБ?	--Y	-	-	-	-	-	•
Инструкции на раб.местах в НБ?	--Y	-	-	-	-	-	•
Работа с О-вещ-вами в НБ?	--Y	-	-	-	-	-	•
Пути эв-ции оборудованы по НБ?	--Y	-	-	-	-	-	•
Внезапная О возможна?	--Y	-	-	-	-	-	•
Тренировки ИТР по ТБ?	--Y	-	-	-	-	-	•
Тренировки ИТР по ТБ на местах?	--Y	-	-	-	-	-	•
О-стойкость субъекта?	--B	-	-	-	-	-	•
Перс.защита присутствует?	--B	-	-	-	-	-	•
Эв-ционные пути присутствуют?	--B	-	-	-	-	-	•
Аварийные выходы присутствуют?	--B	-	-	-	-	-	•
Система эв-ции присутствует?	--B	-	-	-	-	-	•
Групп мал-мобильности нет?	--B	-	-	-	-	-	•
Обучение ИТР по ТБ?	--B	-	-	-	-	-	•
Обучение ИТР по ТБ на местах?	--B	-	-	-	-	-	•
Проводятся тренировки эв-ции?	--B	-	-	-	-	-	•
Произв. дисциплина соблюдается?	--B	-	-	-	-	-	•
Есть инструкции на местах?	--B	-	-	-	-	-	•
План привлечения АС?	--B	-	-	-	-	-	•
САБ присутствует?	--B	-	-	-	-	-	•
Соц.обст-ка в насел.пункте в НБ?	--B	-	-	-	-	-	•
САБ исправна?	ННН	•	•	•	•	•	•

Таблица 4: Пример наполнения фигурантных подпортфелей И-признаками в *неполном* экспертном портфеле, который содержит минимальный портфель (см. табл. 2) и взвешенный минимальный портфель (см. табл. 3), состоят из 69-ти И-признаков, соответствующих 13-ти (вместо 63-х возможных) террасным И-событиям II-го рода с весами: 12,16,3,5,4,1,10,1,2,1,11,2,1 (сверху-вниз). В столбце ★★★ указана аббревиатура подмножеств (см. Замечание 2 на стр. 112), которыми занумерованы соответствующие И-события II-рода. НБ — нормы безопасности; О — опасный; СБО — служба борьбы с опасностью; САБ — система автоматической безопасности; ТО — техническое обслуживание; ТБ — техника безопасности; АС — аварийные службы; ИТР — инженерно-технические работники.

Учесть все виды взаимосвязей между И-признаками в портфеле можно при помощи экспертного портфеля, включающего *взвешенный полный портфель И-признаков безопасности*, который содержит И-

признаки, соответствующие всем возможным террасным событиям II-го рода, порожденным соответствующим *средневероятным экспертным портфелем*.

## 6 Итоги

Предлагаемое сжатое изложение итогов эвентологических исследований безопасности оказалось, хотя и далеко неисчерпывающим, но полезным предприятием, в процессе которого открылась очищенная от излишних подробностей общая конструкция эвентологической системы безопасности, породившая новые идеи, связанные с ключевой процедурой применимой эвентологии безопасности — приготовлением оптимального портфеля И-признаков. Подробное исследование и развитие этих эвентологических идей будет рассмотрено в дальнейших работах.

## Благодарности

Благодарю Аркадия Новосёлова за надежное сопротивление, которое позволяло оттачивать чуть ли не до аксиоматической строгости новые теоретические идеи в эвентологии безопасности, и Сергея Амелчугова за постоянное выдвигание новых проблем пожарной безопасности и их полезное обсуждение, которое неизменно приводило к новым эвентологическим подходам в моделировании и управлении безопасностью.

## Список литературы

- [1] О. Ю. Воробьев. *Эвентология*. Сибирский федеральный университет, Красноярск, 2007, 435с., <http://eventology-theory.ru/0-books/1-VorobyevOleg~2007~Eventology~435p.pdf>.
- [2] О. Ю. Воробьев. Эвентология безопасности. *Труды X Международ. ФАМЭБ конференции по финансово-актуарной математике и эвентологии безопасности*, Красноярск: НИИППБ, СФУ (под ред. Олега Воробьева):92–101, 2011.
- [3] О. Ю. Воробьев. Эвентологический системный анализ безопасности в картинках. *Труды XI Международ. ФАМЭБ конференции по финансово-актуарной математике и эвентологии безопасности*, Красноярск: НИИППБ, СФУ (под ред. Олега Воробьева):113–125, 2012.
- [4] О. Ю. Воробьев. Выразить и измерить: успехи применимой эвентологии. *Труды XVI Международ. ЭМ конференции по эвентологической математике и смежным вопросам*, Красноярск: СФУ (под ред. Олега Воробьева):44–49, 2012.
- [5] О. Ю. Воробьев. Эвентологический анализ систем: система событий во внесистемных обстоятельствах. *Труды XI Международ. ФАМЭБ конференции по финансово-актуарной математике и эвентологии безопасности*, Красноярск: НИИППБ, СФУ (под ред. Олега Воробьева):126–130, 2012.



- [6] О. Ю. Воробьев. Совокупная система и совокупность систем: эвентологическое сходство и различие. *Труды XI Международ. ФАМЭБ конференции по финансово-актуарной математике и эвентологии безопасности*, Красноярск: НИИППБ, СФУ (под ред. Олега Воробьева):131–138, 2012.
- [7] О. Ю. Воробьев. Средневероятное событие для множества событий. *Труды XI Международ. ФАМЭБ конференции по финансово-актуарной математике и эвентологии безопасности*, Красноярск: НИИППБ, СФУ (под ред. Олега Воробьева):139–147, 2012.
- [8] О. Ю. Воробьев. Событийные средние в эвентологии, их асимптотические свойства, интерпретация и визуализация. *Труды XVI Международ. ЭМ конференции по эвентологической математике и смежным вопросам*, Красноярск: СФУ (под ред. Олега Воробьева):50–56, 2012.
- [9] O. Yu. Vorobyev. *Eventology*. Siberian Federal University, Krasnoyarsk, Russia, 2007, 435p. (in Russian, abstract in English), <http://eventology-theory.com/0-books/1-VorobyevOleg~2007~Eventology~435p.pdf>.
- [10] O. Yu. Vorobyev. Eventology of safety. *Proc. of the X Intern. FAMES Conf. on Financial and Actuarial Mathematics and Eventology of Safety*, Krasnoyarsk, SFU (Oleg Vorobyev ed.):92–101, 2011 (in Russian, abstract in English).
- [11] O. Yu. Vorobyev. Eventological system analysis of safety in pictures. In. *Proc. of the XI Intern. FAMES Conf. on Financial and Actuarial Mathematics and Eventology of Safety*, Krasnoyarsk, SFU (Oleg Vorobyev ed.):113–125, 2012 (in Russian, abstract in English).
- [12] O. Yu. Vorobyev. To express and to measure: successes of applicable eventology. In. *Proc. of the XVI Intern. EM conference on eventological mathematics and related fields*, Krasnoyarsk: SFU (Oleg Vorobyev ed.):44–49, 2012 (in Russian, abstract in English).
- [13] O. Yu. Vorobyev. Eventological analysis of systems: an event system under the off-system circumstances. In. *Proc. of the XI Intern. FAMES Conf. on Financial and Actuarial Mathematics and Eventology of Safety*, Krasnoyarsk, SFU (Oleg Vorobyev ed.):126–130, 2012 (in Russian, abstract in English).
- [14] O. Yu. Vorobyev. A total system and a totality of systems: eventological similarity and distinction. In. *Proc. of the XI Intern. FAMES Conf. on Financial and Actuarial Mathematics and Eventology of Safety*, Krasnoyarsk, SFU (Oleg Vorobyev ed.):131–138, 2012 (in Russian, abstract in English).
- [15] O. Yu. Vorobyev. A mean probability event for a set of events. In. *Proc. of the XI Intern. FAMES Conf. on Financial and Actuarial Mathematics and Eventology of Safety*, Krasnoyarsk, SFU (Oleg Vorobyev ed.):139–147, 2012 (in Russian, abstract in English).
- [16] O. Yu. Vorobyev. Event means in eventology, its asymptotic properties, interpretations and visualization. In. *Proc. of the XVI Intern. EM conference on eventological mathematics and related fields*, Krasnoyarsk: SFU (Oleg Vorobyev ed.):50–56, 2012 (in Russian, abstract in English).
- [17] A. A. Novosyolov. Fire risk control and optimization under natural conditions and in technical objects. In. *Proc. of the XI Intern. FAMES Conf. on Financial and Actuarial Mathematics and Eventology of Safety*, Krasnoyarsk, SFU (Oleg Vorobyev ed.):285–287, 2012 (in Russian, abstract in English).
- [18] A. A. Novosyolov. Some problems of fire risk control. In. *Proc. of the XII Intern. FAMES Conf. on Financial and Actuarial Mathematics and Eventology of Safety*, Krasnoyarsk, SFU (Oleg Vorobyev ed.):259–265, 2013 (in Russian, abstract in English).
- [19] А. А. Новоселов. Управление пожарным риском в естественных условиях и на технических объектах. *Труды XI Международ. ФАМЭБ конференции по финансово-актуарной математике и эвентологии безопасности*, Красноярск: НИИППБ, СФУ (под ред. Олега Воробьева):285–287, 2012.
- [20] А. А. Новоселов. Некоторые задачи управления пожарным риском. *Труды XII Международ. ФАМЭБ конференции по финансово-актуарной математике и эвентологии безопасности*, Красноярск: НИИППБ, СФУ (под ред. Олега Воробьева):259–265, 2013.
- [21] С. П. Амелчугов, Ю. А. Андреев, С. Ю. Комаров. Оценка лесопожарных рисков. *Труды X Международ. ФАМЭБ конференции по финансово-актуарной математике и эвентологии безопасности*, Красноярск: НИИППБ, СФУ (под ред. Олега Воробьева):33–38, 2011.
- [22] Методика определения расчетных величин пожарного риска в зданиях, сооружениях и строениях различных классов функциональной пожарной опасности. *Приложение к приказу МЧС России от 30.06.2009, №382*, 2009, 14с.
- [23] Оценка международных стандартов для безопасной разведки, добычи и транспортировки нефти и газа в Баренцевом море. Гармонизация стандартов по охране труда, промышленной безопасности и охране окружающей среды для работ в Баренцевом море. *Баренц-2020. Российско-норвежский проект. Окончательный отчет*, 2009, 139с.



# Голубь и разум: эвентологическая трансдукция голубинового эксперимента Геррнштейна и проблема вероятностного выбора

Олег Юрьевич Воробьев

Сибирский федеральный университет  
Институт математики и фундаментальной информатики  
Торгово-экономический институт  
Красноярск  
oleg.yu.vorobyev@gmail.com

**Аннотация.** Материалы к лекциям по эвентологии. Рассматривается эвентологическая трансдукция известного голубинового эксперимента Геррнштейна [1] и его связь с задачами принятия вероятностных решений в условиях неопределенности и проблемой вероятностного выбора. Предлагается несколько вариантов событийно-точной постановки голубинового эксперимента в коробке Скиннера, наподобие того, который впервые был проведен Геррнштейном, чтобы, наконец, подтвердить или опровергнуть гипотезу Лефевра о решающей роли вероятностного выбора в возникновении разума в живых организмах. Формулируется необходимо строгая событийная постановка такого голубинового эксперимента, по результатам которого можно сделать окончательные выводы о судьбе гипотезы Лефевра. Построена новая модель соответствия реакции живых организмов подкреплению окружающей среды, которая названа гиббсовским соответствием и которая обобщает то, что в когнитивной психологии называлось законом соответствия. Прилагаются также справочные материалы об эвентологической Н-теореме, относительной энтропии изоморфных множеств событий, обращению Мебиуса вероятностных распределений множеств событий и когнитивной психологии.

**Ключевые слова.** Эвентология, теория вероятностей, элементарный исход, событие, вероятность, множество событий, алгебра событий, эвентологическая трансдукция, эксперимент Геррнштейна, принятие вероятностных решений, неопределенность, вероятностный выбор, закон соответствия, гиббсовское соответствие, эвентологическая Н-теорема, относительная энтропия, изоморфные множества событий, обращение Мебиуса, когнитивная психология, гипотеза Лефевра.

Все утро голубь ворковал...  
Борис Пастернак.  
“Еще более душный рассвет”, 1917.

## 1 Введение

В этой работе я попытаюсь связать воедино, применяя событийную трансдукцию, известный голубиный эксперимент [1, 1961], [2, 1970] Геррнштейна<sup>1</sup>, зада-

чи принятия решений в условиях неопределенности и проблему вероятностного выбора [3, 4].

Мой интерес к этой тройственной связи объясняется одним из основных постулатов, на которые опирается эвентология, как нарождающаяся наука о событиях. Это утверждение, принадлежащее Лефевру<sup>2</sup> [5, 6, 7], в моем парафразе выглядит так: “Разум возникает там и тогда, где и когда возникает способность делать вероятностный выбор”.

У меня это фундаментальное утверждение Лефевра сомнений не вызывает, и во мне крепнет устойчивое желание взглянуть на результаты точно поставленного психологами голубинового эксперимента в коробке Скиннера, наподобие того, который впервые был проведен Геррнштейном, чтобы, наконец, подтвердить или опровергнуть гипотезу о решающей роли вероятностного выбора в возникновении разума в живых организмах. Именно, это желание подвигло меня на данную работу, где формулируется необходимо строгая событийная постановка голубинового эксперимента Геррнштейна, по результатам которого можно сделать окончательные выводы о судьбе гипотезы Лефевра. Более того, в работе мне удалось построить новую модель соответствия реакции живых организмов подкреплению окружающей среды, которую я назвал гиббсовским соответствием и которая обобщает то, что в когнитивной психологии называлось законом соответствия<sup>3</sup>.

следователь обучения животных в скиннеровских традициях; работал со Скиннером в гарвардской голубиной лаборатории, где исследовал выбираемое поведение и поведенческую экономику; основатель количественного анализа поведения живых организмов; известен как первооткрыватель бихевиористского закона соответствия (*matching law*) — тенденции животных распределять свои выборы в прямой пропорции к подкреплениям, которые они обеспечивают, известной и в количественном анализе поведения, и в математической психологии.

<sup>2</sup>Лефевр, Владимир Александрович (р. 1936) — российский и американский психолог и математик; основатель теории рефлексии, находящейся на стыке многих наук: психологии, философии, математики, социологии, этики и др.; определяет разум, как способность делать вероятностный выбор.

<sup>3</sup>Закон соответствия (англ. *matching law*), см. справку на стр. 128.

© 2013 О.Ю.Воробьев

Олег Воробьев (ред.), Труды XII ФАМЭБ'2013, Красноярск: СФУ

<sup>1</sup>Геррнштейн, Ричард Дж. (англ. *Herrnstein, Richard J.*, 1930–1994) — американский экспериментальный психолог, ис-

## 2 Голубиный эксперимент Геррнштейна

Известный в бихевиоризме<sup>4</sup> голубиный эксперимент<sup>5</sup> Геррнштейна по описанию его автора заключается в следующем<sup>6</sup>:

*“Представьте себе коробку Скиннера<sup>7</sup>, в которую помещен голубь и размещены две кормушки, кляю в которые, голубь может получить еду, заложенную туда экспериментатором. В каждую кормушку экспериментатор бросается не более одного зерна с различной фиксированной частотой. Голубь может клевать кормушку, но “не знает”, получит ли он зерно наверняка. Если в кормушке находится зерно, голубь его получает. Если же зерна в кормушке нет, то голубь остается без еды. После каждого клевка кормушки очищаются от зерен, которые могли там остаться, и эксперимент повторяется”.*

В результате многократного повторения этого голубиногo эксперимента по словам Геррнштейна оказалось, что со временем голубь начинает клевать две кормушки с частотами, пропорциональными частотам<sup>8</sup>, с которыми туда закладываются зерна экспериментатором. Пропорционально-частотная связь поведения голубя с воздействием окружающей его среды — основной смысл бихевиористского закона соответствия (*matching law*)<sup>1,9</sup>

<sup>4</sup>Бихевиоризм (от англ. *behavior* — поведение) — наука о поведении; направление в психологии человека и животных; считает, что предметом психологии является поведение, а не сознание; категориями бихевиоризма являются стимул (любое воздействие на организм со стороны среды), реакция и подкрепление; позднее (во второй половине 20-го века) бихевиоризм был замещен когнитивной психологией.

<sup>5</sup>Опираясь на этот и другие подобные эксперименты бихевиористы пришли к формулировке того, что они называли законом соответствия (*matching law*, см. справку на стр. 128).

<sup>6</sup>Цитата не дословная.

<sup>7</sup>Скиннер, Беррес Фредерик (англ. *Skinner, Burrhus Frederic*; 1904–1990) — американский психолог, основатель радикального бихевиоризма, писатель, изобретатель и социальный философ; изобрел в конце 50-х оперантно обучающую камеру (*operant conditioning chamber*), известную как коробка Скиннера; его анализ поведения человека достиг пика в работе “Вербальное поведение (*Verbal Behaviour*)” [8], которая недавно пережила огромный рост интереса со стороны экспериментаторов и прикладных исследователей; ему принадлежит высказывание: “Образование — это то, что выживает, когда все, что было изучено, забыто (*Education is what survives when what has been learnt has been forgotten*)”; как-то заметил, что никогда не хотел быть эпонимом (*eponym*). Когнитивная революция 1960-ых выступила против взглядов радикального бихевиоризма [9], утверждая, что черный ящик разума имеет такое же большое влияние на поведение живого организма, как и воздействие окружающей среды (а возможно даже больше).

<sup>8</sup>или некоторым степеням частот.

<sup>9</sup>Короче говоря, закон соответствия полагает, что реакция живого организма на сценарий пропорциональна количеству/продолжительности положительного подкрепления, полученного им из окружающей среды. Например, вообразите, что есть два источника подкрепления, один из которых вдвое богаче, чем другой. Закон соответствия, который обнаружил Геррнштейн, утверждает, что в этой ситуации животные в два

Вероятностный характер этого закона допускает его событийную трансдукцию, в процессе которой, однако, приходится выдвигать несколько различных вариантов событийной модели голубиногo эксперимента Геррнштейна, так как в работах одного из самых известных бихевиористов модель эксперимента описана не достаточно эвентологически внятно. Поэтому я вынужден рассматривать ряд возможных вариантов модели эксперимента<sup>10</sup>, чтобы попытаться выбрать из них тот, который “наилучшим” образом отражает суть бихевиористского закона соответствия на событийном языке. Признаюсь, я уже настолько “свыкся” с этим законом, что мне хочется лишь выяснить возможности эвентологической теории “оправдать” его — воспользоваться шансом обнаружить такую эвентологическую модель голубиногo эксперимента, которая показывает его справедливость.

## 3 Варианты событийной модели голубиногo эксперимента Геррнштейна

Пусть  $(\Omega, \mathcal{A}, \mathbf{P})$  — всеобщее вероятностное пространство, в рамках которого я собираюсь провести эвентологическую трансдукцию возможных вариантов модели голубиногo эксперимента Геррнштейна: все рассматриваемые события будут взяты из алгебры всеобщих событий  $\mathcal{A}$ .

### 3.1 Поведение экспериментатора: событийно-вероятностное подкрепление

- 1) “Или-Или”-загрузка. Зерно бросается либо в одну, либо в другую кормушку со своей частотой для каждой кормушки; описывается событийным моноплетом

$$\mathfrak{X}_1 = \{x\},$$

где

- $x \in \mathcal{A}$  — “зерно в правую кормушку” с вероятностью  $p_x = \mathbf{P}(x)$ ;
- $x^c = \Omega - x \in \mathcal{A}$  — “зерно в левую кормушку” с вероятностью  $1 - p_x = \mathbf{P}(x^c)$ ;

раза чаще выбирают ту альтернативу, которая кажется им вдвое ценнее. У бихевиористов есть три основные идеи относительно того, как живые организмы управляют своими реакциями, чтобы оптимизировать подкрепление, получаемое из окружающей среды:

- I: “альтернативная” оптимизация: организмы всегда выбирают ту альтернативу реакции, которая наиболее вероятно будет тотчас подкреплена.
- II: “средняя” оптимизация: организмы распределяют свои ответы среди различных альтернатив, чтобы максимизировать количество подкрепления, которое они зарабатывают за длительный промежуток времени.
- III: “соответственная” оптимизация: организмы реагируют на подкрепления так, чтобы улучшить соответствие между пропорциями своих реакций и пропорциями подкреплений.

<sup>10</sup>которые только смог вообразить в этой событийно нестрогo описанной ситуации.

- 2) **“И”-загрузка.** Зерна бросаются в обе кормушки независимо или зависимым образом; описывается событийным дуплетом

$$\mathfrak{X} = \{x, y\},$$

где

- $x \in \mathcal{A}$  — “зерно в правую кормушку” с вероятностью  $p_x = \mathbf{P}(x)$ ;
- $y \in \mathcal{A}$  — “зерно в левую кормушку” с вероятностью  $p_y = \mathbf{P}(y)$ ;
- $x \cap y \in \mathcal{A}$  — “зерна в обе кормушки” с вероятностью  $p_{xy} = \mathbf{P}(x \cap y)$ ;
  - **“И”-независимая загрузка.** Зерна бросаются **независимо в каждую** кормушку со своей вероятностью (частотой) для каждой кормушки; поэтому
 
$$p_{xy} = \mathbf{P}(x \cap y) = p_x p_y;$$
  - **“И”-зависимая загрузка.** Зерна бросаются в обе кормушки **зависимым образом**, который определяется дуплетным распределением — набором вероятностей (частот) четырех вариантов бросания зерен экспериментатором в две кормушки; поэтому
 
$$p_{xy} = \mathbf{P}(x \cap y) \neq p_x p_y.$$

Например, **сначала в правую, потом в левую**<sup>11</sup>: “зерно в правую кормушку” с вероятностью  $p_x = \mathbf{P}(x)$ , а затем “зерно в левую кормушку” с условными вероятностями:  $p_{y|x} = p_{xy}/p_x$  или  $p_{y|x^c} = p_{x^c y}/p_{x^c} = \mathbf{P}(x^c \cap y)/\mathbf{P}(x^c)$  в зависимости от исхода бросания в правую кормушку:  $x$  или  $x^c$ .

### 3) Выдача

- **Обычная.** Если голубь клюет в кормушку, в которую загружено зерно, он получает его, иначе — не получает;
- **Бонусная.** Если голубь клюет в кормушку, в которую *не* загружено зерно, он ничего не получает, если загружено — получает, **но**, кроме того, может получить еще и “бонус” — оба зерна из обеих кормушек, **если**, оказывается (с вероятностью  $p_{xy}$ ), что клюет тогда, когда в обе кормушки загружены зерна.
- **Штрафная.** Если голубь клюет в любую кормушку, в тот момент, когда в обе кормушки загружены зерна (с вероятностью  $p_{xy}$ ), он подвергается “штрафу” — совсем не получает подкрепления. Иначе говоря, голубь получает подкрепление только если он клюет в загруженную кормушку в тот момент, когда другая кормушка пуста.

### 4) “Обнуление”

- После каждого голубиногo клевка и соответствующего подкрепления или его отсутствия обе кормушки “обнуляются” и в них вновь загружаются зерна по вероятностному правилу, которое контролирует экспериментатор.

## 3.2 Поведение голубя: событийно-вероятностная реакция

- 1) **“Или-Или”-реакция:** описывается событийным моноплетом

$$\mathfrak{B} = \{b_x\},$$

где

- $b_x \in \mathcal{A}$  — “клевок в правую кормушку” с вероятностью  $r_x = \mathbf{P}(b_x)$ ;
- $b_x^c = \Omega - b_x \in \mathcal{A}$  — “клевок в левую кормушку” с вероятностью  $r_y = 1 - r_x = \mathbf{P}(b_x^c)$ ;

\*\*\*

В итоге определены 7 альтернативных вариантов<sup>12</sup> (см. диаграммы Венна на рис. 1, 2, 4, 3, 5, 7 и 6 и таблицу на стр. 119) строгой событийной трансдукции голубиногo эксперимента Геррнштейна, которые я предлагаю рассмотреть в этой работе.

## 3.3 Реакция голубя и экспериментальное подкрепление на событийно-вероятностном языке

Голубю в своей *событийно-вероятностной* поведении в пределах коробки Скиннера приходится

- оценить *условные вероятностные распределения событийного* подкрепления (появления зерен в кормушках) на каждую его возможную *событийную* реакцию (клевок вправо, клевок влево);
- применить полученные оценки *условных вероятностных распределений* для генерирования своей *событийно-вероятностной* реакции на подкрепление.

**Экспериментатор**, используя коробку Скиннера, имитирует *событийно-вероятностное* воздействие окружающей среды, т.е. задает его *вероятностное* распределение и создает в соответствии с этим распределением *события*, внешние для голубя:

- загружает зерна в кормушки в соответствии с *вероятностным* распределением;
- выдает из кормушек загруженные зерна голубю по установленным правилам.

Итак, если экспериментатор, чтобы имитировать внешнее *событийно-вероятностное* подкрепление, должен лишь генерировать множество событий с заданным вероятностным распределением, то перед голубем стоит более сложная *событийно-вероятностная* задача, состоящая из двух подзадач:

- оценка условных вероятностных распределений внешнего подкрепления при каждой возможной собственной событийной реакции;
- выбор и реализация *оптимального* вероятностного распределения собственной событийной реакции;

<sup>11</sup>Или же в другом порядке, что эквивалентно: сначала в левую, потом в правую.

<sup>12</sup>Поскольку при одном варианте подкрепления — **“Или-Или”** загрузка — возможен только один **обычный** вариант выдачи; а при двух других вариантах подкрепления — **“И”-независимая**, **“И”-зависимая** загрузка — возможны три варианта выдачи: **обычный**, **бонусный** и **штрафной**.

которые он обязан решать *совместно*<sup>13</sup> и *постоянно*<sup>14</sup>.

### 3.4 Голубиная оптимальность

Возможно, вам трудно представить себя на месте голубя в коробке Скиннера. Позвольте мне проделать это вместо вас, чтобы поделиться с вами своими впечатлениями и событийно-вероятностными выводами о смысле голубиной оптимальности. А они таковы.

Опыт *событийно-вероятностной* борьбы за существование в той окружающей среде, в которой он оказался волею случая, дает голубю, по крайней мере, два важных урока:

- внешние подкрепления, которые голубь способен получить из окружающей среды в ответ на то или иное свое поведение, имеют *событийно-вероятностный* характер;
- *событийно-вероятностный* характер подкреплений на одно и то же его поведение со временем меняется.

Не выучив эти уроки, голубь, вряд ли, может рассчитывать на выживание в данных ему условиях.

**Первый урок** заставляет голубя научиться оценивать *событийно-вероятностный* характер подкрепления в ответ на каждый способ своего поведения, иначе говоря, оценивать условное вероятностное распределение множества внешних событий при условии того или иного способа голубинового поведения.

**Второй урок** постоянно напоминает голубю, что раз и навсегда оценить *событийно-вероятностный* характер подкрепления ему не удастся из-за того, что характер подкрепления может *событийно-вероятностно* измениться в любой момент. Это вынуждает голубя проводить *нескончаемые* текущие оценки условных вероятностных распределений множества внешних событий при условиях *всех возможных* способов своего голубинового *событийно-вероятностного* поведения.

В итоге, выбирая оптимальную стратегию своего поведения, голубь вынужден извлечь из этих уроков следующие критерии оптимальности:

- α) необходим определенный период времени, чтобы *событийно-вероятностно* приспособиться к окружающей среде оптимальным образом;
- β) оптимальная голубиная стратегия не может сводиться лишь к одному из возможных способов поведения, так как, чтобы следить за внешними изменениями и иметь возможность *событийно-вероятностного* выбора, голубь должен не только однажды провести испытания *всех* своих возможных способов поведения, но и нескончаемо продолжать испытывать *все* эти способы.

<sup>13</sup> Потому что результат решения одной из подзадач используется при решении другой.

<sup>14</sup> Потому что у голубя нет уверенности в том, что *событийно-вероятностное* поведение окружающей среды будет неизменным.

Кроме того, голубю нельзя забывать, что

- γ) оптимальная стратегия его *событийно-вероятностного* поведения должна обеспечивать ему подкрепление, достаточное для выживания.

### 3.5 Предполагаемые намерения

Для каждого из возможных вариантов эвентологической модели голубинового эксперимента в коробке Скиннера (см. параграф 3 на стр. 116) я предполагал сформулировать и решить три задачи<sup>15</sup> вероятностного выбора голубем оптимального событийного поведения.

- I: **Вероятностная оптимизация.** Оптимизация вероятности подкрепления;
- II: **Средняя оптимизация.** Оптимизация средней ценности (для голубя) подкрепления;
- III: **Соответственная оптимизация.** Оптимизация соответствия вероятностного распределения голубя вероятностному распределению окружающей среды при фиксированной средней ценности (для голубя) подкрепления.

Однако побывав на месте голубя в коробке Скиннера, я пришел к выводу, что постановки первой и второй задач *вероятностной и средней оптимизации* не соответствует критериям α), β) и γ) голубиной оптимальности на стр. 118. Дело в том, что в условиях данного голубинового эксперимента максимизация вероятности сиюминутного подкрепления, а также среднего подкрепления приводит к выбору одного из всех возможных способов, что противоречит критерию β), не позволяя голубю отслеживать событийно-вероятностные подкрепления при других способах его поведения<sup>16</sup>. Поэтому в работе я анализирую только третью задачу соответственной оптимизации.

Математическое решение этой задачи оптимизации, разумеется, не будет иметь решающего значения для когнитивной психологии. Эвентологическая теория может лишь предложить эталон для сравнения с результатами реальных голубиных экспериментов. Так что мне ничего не останется, как предложить психологам провести соответствующие и корректно эвентологически сформулированные голубиные эксперименты, чтобы, сравнив результаты этих экспериментов с полученными математическими решениями, иметь возможность делать эвентологически обоснованные выводы о справедливости или несправедливости так называемого *закона соответствия* в когнитивной психологии.

<sup>15</sup> В соответствии с тремя основными идеями психологов относительно того, как живые организмы управляют своими реакциями, чтобы оптимизировать подкрепление, получаемое из окружающей среды (см. сноску<sup>9</sup> на стр. 116).

<sup>16</sup> На мой взгляд, критерий β) следует из γ), так как, именно, стремление к выживанию заставляет голубя отслеживать событийно-вероятностные подкрепления при других способах его поведения, чтобы в каждый момент быть готовым к изменению законов подкрепления.



#### 4 Варианты голубиного эксперимента в диаграммах Венна

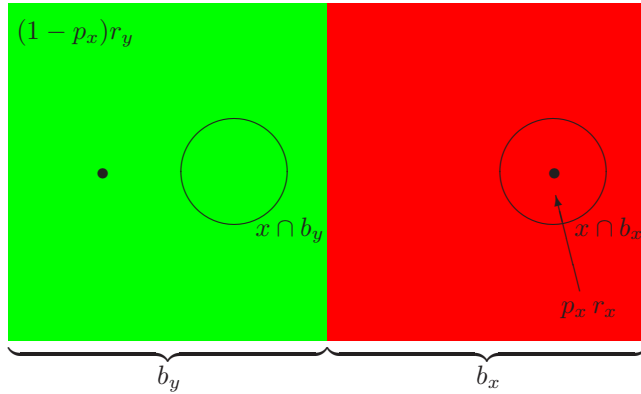


Рис. 1: I. “Или-Или”-загрузка и обычная выдача.

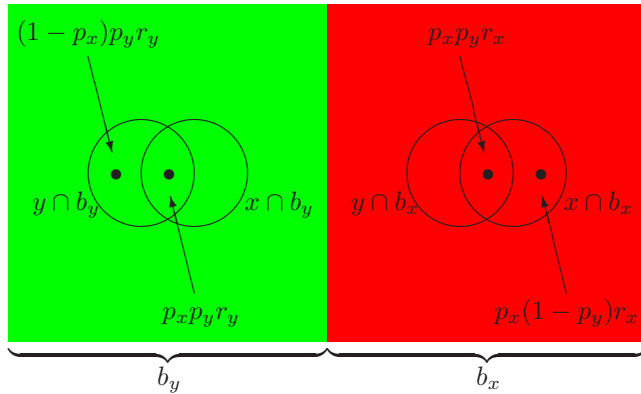


Рис. 2: II. “И”-независимая-загрузка, обычная выдача.

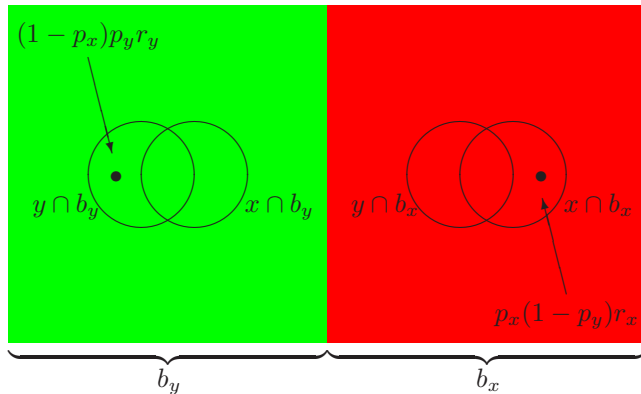


Рис. 3: III. “И”-независимая-загрузка и штрафная выдача.

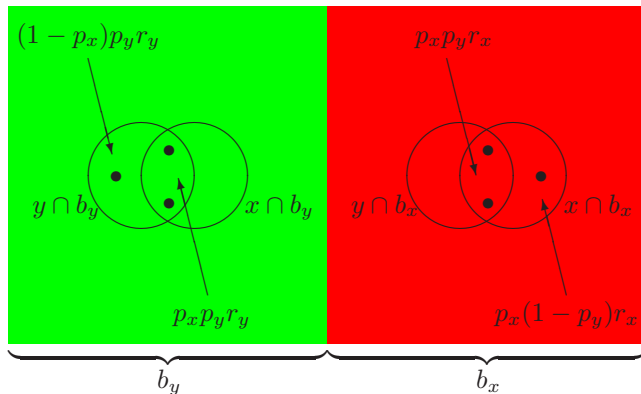


Рис. 4: IV. “И”-независимая-загрузка и бонусная выдача.

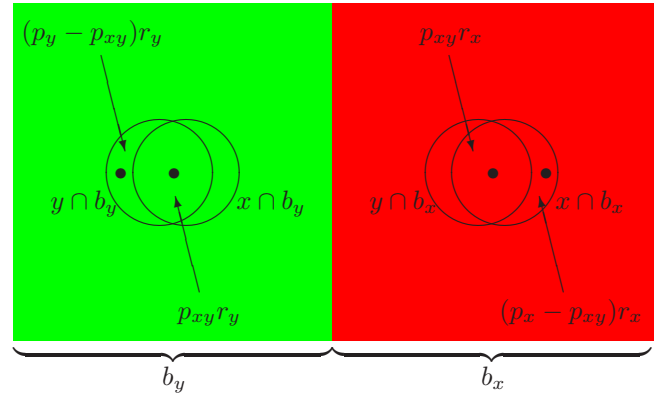


Рис. 5: V. “И”-зависимая-загрузка и обычная выдача.

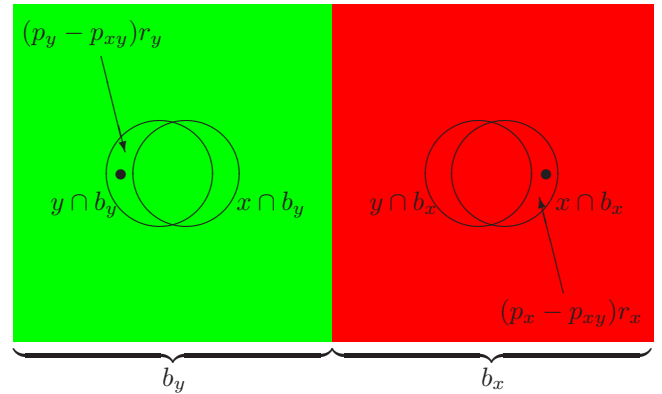


Рис. 6: VI. “И”-зависимая-загрузка и штрафная выдача.

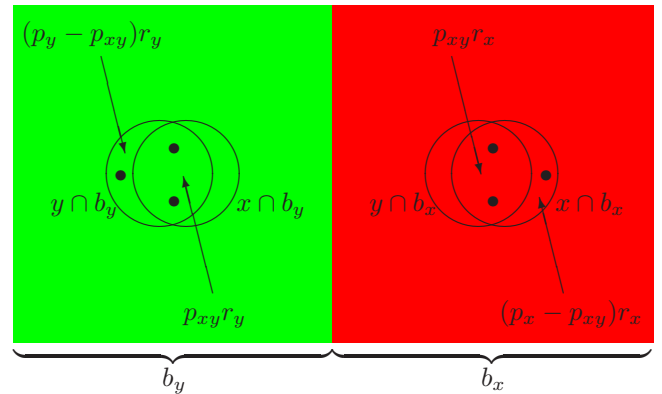


Рис. 7: VII. “И”-зависимая-загрузка и бонусная выдача.

	“Или-Или”-загрузка	“И”-независимая загрузка	“И”-зависимая загрузка
Обычная выдача	I вариант	II вариант	V вариант
Штрафная выдача	×	III вариант	VI вариант
Бонусная выдача	×	IV вариант	VII вариант

Итак предлагается 7 альтернативных вариантов (см. диаграммы Венна на рис. 1, 2, 4, 3, 5, 7 и 6 и таблицу) голубиного эксперимента Герриштейна. На всех семи диаграммах Венна кругами показаны события подкрепления:  $x$  — “зерно в правую кормушку”; цветом — события реакции:  $b_x$  — “клевок в правую кормушку” (красный);  $b_y$  — “клевок в левую кормушку” (зеленый). Террасные события, при наступлении которых голубь получает подкрепление, отмечены черными точками (их количество равно количеству зерен, полученных голубем при данном террасном событии); указаны вероятности только этих террасных событий.



## 5 Задача соответственной голубиной оптимизации

В голубином эксперименте Геррнштейна участвуют две стороны: окружающая среда и голубь. Но если окружающая среда характеризуется фиксированным вероятностным распределением своего событийного поведения, то голубь оптимизирует вероятностное распределение своего событийного поведения — приводит его в оптимальное соответствие с вероятностным распределением событийного поведения окружающей среды.

Я полагаю, что вероятностные распределения событийных поведений голубя и среды, не умаляя общности, можно считать вероятностно независимыми, поскольку голубь всегда находится в состоянии отсутствия какой бы-то ни было информации о том, в какие кормушки загружены зерна. Лишь по результатам серии своих клевков в правую или левую кормушки голубь может составить свое “голубиное” представление о величине условных вероятностей выдачи зерен из кормушек и об условных средних количествах выдаваемых зерен.

### 5.1 Модель подкрепления

В *событийно-вероятностном* описании голубиногo эксперимента (см. параграф 3 на стр. 116) участвуют три события, сгруппированные в два множества:

- моноплет событий  $\mathfrak{B} = \{b_x\}$ , описывающий событийно-вероятностную реакцию голубя;
- дуплет событий  $\mathfrak{X} = \{x, y\}$ , описывающий событийно-вероятностное подкрепление, имитируемое экспериментатором.

В этих терминах перед голубем стоит задача оптимизации своего моноплета событий  $\mathfrak{B}$  относительно экспериментаторского дуплета событий  $\mathfrak{X}$  при условии среднего подкрепления, зафиксированного голубем на уровне, достаточном для собственного выживания.

Чтобы продвинуться дальше в постановке задачи голубиной оптимизации, мне необходимо ввести математическую модель подкрепления, в качестве которой предлагается целочисленная функция

$$v : 2^{\mathfrak{X}} \rightarrow \mathbb{R},$$

определенная на  $2^{\mathfrak{X}}$  — множестве всех подмножеств дуплета событий  $\mathfrak{X}$  — и известная в эвентологии как функция ценности соответствующих террасных событий:

$$v(X//\mathfrak{X}) = \mathbf{V}(\text{ter}(X//\mathfrak{X}))$$

для  $X \subseteq \mathfrak{X}$ . Мне потребуется не сама функция ценности, а две ее условные “проекции”

$$v_{b_x}(X//\mathfrak{X}) = \mathbf{V}(\text{ter}(X//\mathfrak{X})|b_x),$$

$$v_{b_x^c}(X//\mathfrak{X}) = \mathbf{V}(\text{ter}(X//\mathfrak{X})|b_x^c),$$

которые должны быть изначально заданы в задаче голубиной оптимизации и значения которых имеют в смысл количества зерен получаемых голубем в каждой сложившейся в коробке Скиннера “X-ситуации”, описываемой “внешним” террасным событием  $\text{ter}(X//\mathfrak{X})$ , при условии одного ( $b_x$ ) или другого ( $b_x^c$ ) способа его “внутреннего” поведения.

### 5.2 Парафраз постановки простейшей задачи соответственной оптимизации для дуплетов

Перефразирую эквивалентно постановку задачи соответственной оптимизации для простейшего варианта событийной модели голубиногo эксперимента, используя для моделирования поведения голубя и среды два полных дуплета непересекающихся событий<sup>17</sup>.

Пусть  $\mathfrak{X} = \{x, y\}$  и  $\mathfrak{B} = \{b_x, b_y\}$  — два  $\varphi$ -изоморфных<sup>18</sup> полных дуплета непересекающихся событий, изоморфизм между которыми устанавливается “естественными” формулами:

$$b_x = \varphi(x), \quad b_y = \varphi(y).$$

Допустим еще, что каждый из дуплетов непересекающихся событий образует покрытие  $\Omega$ . Поэтому

$$p_x = \mathbf{P}(x), p_y = 1 - p_x = \mathbf{P}(y),$$

$$r_x = \mathbf{P}(b_x), r_y = 1 - r_x = \mathbf{P}(b_y).$$

В такой ситуации вероятностные распределения  $I$ -го рода этих дуплетов имеют вид:

$$\{p(\emptyset), p(\{x\})\} = \{p_y, p_x\},$$

$$\{r(\emptyset), r(\{x\})\} = \{r_y, r_x\}.$$

А относительные энтропии между этими  $\varphi$ -изоморфными полными дуплетами событий равны

$$\mathcal{H}_{\frac{x}{p}} = r_y \ln \frac{r_y}{p_y} + r_x \ln \frac{r_x}{p_x}, \quad (5.2.1)$$

$$\mathcal{H}_{\frac{y}{r}} = p_y \ln \frac{p_y}{r_y} + p_x \ln \frac{p_x}{r_x}. \quad (5.2.2)$$

### 5.3 Гиббсовское соответствие

Построю функцию ценности для голубя исходов эксперимента, моделью которого служат два независимых дуплета непересекающихся событий  $\mathfrak{X}$  и  $\mathfrak{B}$ , каждый из которых образует покрытие  $\Omega$ . Первый дуплет  $\mathfrak{X}$  — это событийная модель поведения окружающей среды, а второй  $\mathfrak{B}$  — событийная модель поведения голубя.

<sup>17</sup>Единственная цель такого парафраза — иметь затем возможность более удобного сравнения полученных здесь результатов с результатами для сложных вариантов событийной модели, основанных, как раз, на дуплетах произвольно пересекающихся событий.

<sup>18</sup>Об изоморфных множествах событий см. [10].

Террасное событие  $\text{ter}(\{b_x\} // \mathfrak{B})$  имеет для голубя среднюю (по дуплету  $\mathfrak{X}$ ) ценность  $p_x$ :

$$v(\{x\}) = p_x,$$

а террасное событие  $\text{ter}(\emptyset // \mathfrak{B})$  — среднюю (по дуплету  $\mathfrak{X}$ ) ценность  $p_y$ :

$$v(\emptyset) = p_y.$$

Голубю приходится учитывать, именно, среднюю по дуплету  $\mathfrak{X}$  ценность, поскольку при условии события  $b_x$  голубь получает зерно не с достоверностью: он получает зерно с вероятностью  $p_x$  и не получает — с дополнительной вероятностью  $p_y$ . Так как террасные события, порожденные голубиным дуплетом  $\mathfrak{B}$ , наступают с вероятностями  $r_x$  и  $r_y$  соответственно, то математическое ожидание (по дуплету  $\mathfrak{B}$ ) средней (по дуплету  $\mathfrak{X}$ ) ценности равно

$$v(\{x\})r_x + v(\emptyset)r_y = p_x r_x + p_y r_y = \langle v \rangle, \quad (5.3.1)$$

где  $\langle v \rangle$  — некоторое фиксированное значение этой (средней по обоим дуплетам) ценности исходов эксперимента для голубя, которую он устанавливает для себя, чтобы выжить в сложившихся обстоятельствах. Поскольку предполагается, что  $0 < p_x < 1$  и  $0 < r_x < 1$ , то

$$0 < \langle v \rangle < 1.$$

В силу эвентологической  $H$ -теоремы [11, 12, стр. 430–431] при фиксированной функции ценности и фиксированном распределении одного множества событий  $\mathfrak{X}$  минимум (по распределениям другого множества событий  $\mathfrak{B}$ ) относительной энтропии (5.2.1) между множествами  $\mathfrak{X}$  и  $\mathfrak{B}$  достигается на гиббсовском распределении множества  $\mathfrak{B}$ :

$$\frac{r(\{x\})}{r(\emptyset)} = \exp \left\{ \alpha [v(\{x\}) - v(\emptyset)] \right\} \frac{p(\{x\})}{p(\emptyset)},$$

в кратких обозначениях записываемого в виде соотношения

$$\frac{r_x}{r_y} = \exp \left\{ \alpha (p_x - p_y) \right\} \frac{p_x}{p_y} = \frac{p_x e^{\alpha p_x}}{p_y e^{\alpha p_y}} \quad (5.3.2)$$

и представляющего то, что можно назвать *гиббсовским соответствием* распределения голубя  $\mathfrak{B}$  распределению окружающей среды  $\mathfrak{X}$ , эвентологически корректно описывающим давно экспериментально отмеченную закономерность, которую психологи называют *законом соответствия* [1] поведения живых организмов поведению окружающей среды.

#### 5.4 Параметры простейшего гиббсовского соответствия

В соотношении (5.3.2) присутствует неизвестный параметр  $\alpha$ , который, так же, как и искомое вероятностное распределение  $\{1 - r_x, r_x\}$ , является решением задачи соответственной оптимизации при ограничении (5.3.1).

Перепишем соотношения (5.3.1) и (5.3.2), избавившись от  $r_y = 1 - r_x$  и  $p_y = 1 - p_x$ , чтобы получить систему из двух уравнений:

$$p_x r_x + (1 - p_x)(1 - r_x) = \langle v \rangle, \quad (5.4.1)$$

$$\frac{r_x}{1 - r_x} = \frac{p_x}{1 - p_x} e^{\alpha(2p_x - 1)} \quad (5.4.2)$$

относительно неизвестных  $r_x$  и  $\alpha$ .

Вероятность  $r_x$  находится из (5.4.1) по формуле

$$r_x = \frac{\langle v \rangle - (1 - p_x)}{2p_x - 1} = r_x = \frac{\langle v \rangle - (1 - p_x)}{p_x - (1 - p_x)}. \quad (5.4.3)$$

Отсюда

$$\frac{r_x}{1 - r_x} = \frac{\langle v \rangle - (1 - p_x)}{p_x - \langle v \rangle}. \quad (5.4.4)$$

Из (5.4.4) и (5.4.2) получаем формулу для  $\alpha$ :

$$\alpha = \frac{1}{p_x - (1 - p_x)} \ln \left( \frac{1 - p_x}{p_x} \cdot \frac{\langle v \rangle - (1 - p_x)}{p_x - \langle v \rangle} \right), \quad (5.4.5)$$

которую можно переписать, например, в следующем виде:

$$\alpha = \frac{1}{p_x - (1 - p_x)} \ln \frac{\frac{\langle v \rangle}{p_x} - \frac{1 - p_x}{p_x}}{\frac{p_x}{1 - p_x} - \frac{\langle v \rangle}{1 - p_x}}. \quad (5.4.5')$$

Решение задачи соответственной оптимизации дается формулами (5.4.3) и (5.4.5), или (5.4.5').

#### 5.5 Ограничения средней ценности

В примере от искомого вероятностного распределения  $\{1 - r_x, r_x\}$  требуется обеспечить ограничение

$$p_x r_x + (1 - p_x)(1 - r_x) = \langle v \rangle, \quad (5.4.1)$$

в котором  $\langle v \rangle$  формально может рассматриваться и как среднее значение величин  $1 - r_x$  и  $r_x$  по вероятностному распределению  $\{1 - p_x, p_x\}$ , и как среднее значение величин  $1 - p_x$  и  $p_x$  по вероятностному распределению  $\{1 - r_x, r_x\}$ . Отсюда, очевидно, следуют неравенства, которым удовлетворяет средняя ценность  $\langle v \rangle$ :

$$\min\{1 - r_x, r_x\} \leq \langle v \rangle \leq \max\{1 - r_x, r_x\}, \quad (5.4.6)$$

$$\min\{1 - p_x, p_x\} \leq \langle v \rangle \leq \max\{1 - p_x, p_x\}, \quad (5.4.7)$$

из которых следуют совместные неравенства:

$$\max \left\{ \min\{1 - r_x, r_x\}, \min\{1 - p_x, p_x\} \right\} \leq \langle v \rangle, \quad (5.4.8)$$

$$\langle v \rangle \leq \min \left\{ \max\{1 - r_x, r_x\}, \max\{1 - p_x, p_x\} \right\}. \quad (5.4.8')$$

### 5.6 Гиббсовское соответствие в общем варианте

В отличие от голубя, поведение которого событийно регламентировано в коробке Скиннера одним событием и его дополнением<sup>19</sup>, поведение экспериментатора может быть более разнообразным в той степени, в какой оно отражено в вариантах событийного описания эксперимента в параграфе 3 на стр. 116 и проиллюстрировано диаграммами Венна в параграфе 4 на стр. 119.

Однако это экспериментаторское событийное разнообразие голубь может воспринимать, только наблюдая за появлением зерен в кормушках после своих “событийно-моноплетных” клевков. Иными словами, событийное разнообразие поведения экспериментатора воспринимается им исключительно через событийные моноплеты, порождающие в общей сложности такие террасные события, которые голубь может отличать лишь по количеству зерен подкрепления и по левой или правой кормушке, где он обнаруживает зерно.

**Вариант условного распределения наблюдаемого подкрепления.** Переписав гиббсовское соответствие (5.3.2), полученное в параграфе 5.3 на основе эвентологической  $H$ -теоремы, для оставшихся общих вариантов<sup>20</sup>, в которых, вообще говоря,  $p_x + p_y \neq 1$ , предварительно нормировав эти вероятности специальным образом так, чтобы они образовали условное распределение наблюдаемого подкрепления по кормушкам при условии, что голубь вообще получает подкрепление:

$$\left\{ \frac{p_y r_y}{p_x r_x + p_y r_y}, \frac{p_x r_x}{p_x r_x + p_y r_y} \right\}, \quad (5.6.1)$$

получим гиббсовское соответствие в виде:

$$\frac{r_x}{r_y} = \frac{p_x r_x e^{\frac{\alpha}{p_x r_x + p_y r_y} p_x r_x}}{p_y r_y e^{\frac{\alpha}{p_x r_x + p_y r_y} p_y r_y}}. \quad (5.6.2)$$

Подчеркнем еще раз, что гиббсовское соотношение (5.6.2) можно получить из эвентологической  $H$ -теоремы на, вроде бы, законных<sup>21</sup> основаниях, если вероятностные распределения выдачи:

$$\{p_y, p_x\}$$

заменить на условные вероятностные распределения получения подкрепления по левой и правой кормушкам при условии получения подкрепления (5.6.1).

<sup>19</sup>Строго говоря, голубь всегда имеет третью альтернативу — “не клевать”, которая приводит к дуплету непересекающихся событий, сумма вероятностей которых меньше единицы. Но в работе упрощенно считается, что речь идет о поведении голубя при условии, что он клюет в кормушки, а весь вопрос лишь в том — в какую из двух кормушек.

<sup>20</sup>Например, для варианта II.

<sup>21</sup>Подробнее о “законности” этих оснований см. замечание чуть ниже.

По условию эвентологической  $H$ -теоремы

$$p_x r_x + p_y r_y = \langle v \rangle.$$

Поэтому из (5.6.1) получаем

$$1 = \frac{p_x e^{\frac{\alpha}{\langle v \rangle} p_x r_x}}{p_y e^{\frac{\alpha}{\langle v \rangle} p_y r_y}}. \quad (5.6.2)$$

Отсюда

$$\frac{r_x}{r_y} = \frac{p_y \ln p_y}{p_x \ln p_x}. \quad (5.6.3)$$

**Замечание.** В рассмотренном варианте вероятностное распределение голубя энтропийно сравнивается с условным вероятностным распределением наблюдаемого им подкрепления, в то время как в эвентологической  $H$ -теореме не идет речь об условных распределениях. Строго говоря, для этого требуется построить обобщение формулировки Э- $H$ -теоремы, включающее условные распределения и, тем самым, поглощающее рассмотренный вариант.

## 6 Голубиная оптимизация как принятие вероятностных решений

По сути дела, голубь в коробке Скиннера вынужден решать известную задачу принятия решений в условиях вероятностной неопределенности. Причем критерием его собственного *вероятностного выбора*<sup>22</sup> одного из двух предлагаемых ему средой условных вероятностных распределений подкрепления служит стремление выжить в сложившихся обстоятельствах. Выбор в этой событийной модели без сомнения *вероятностный* потому, что его результатом служит *вероятностное распределение* поведения голубя, а не набор детерминированных решений, которые он принимает при различных стечениях событийных обстоятельств<sup>23</sup>. Следовательно, в рамках предлагаемой событийной модели голубиного эксперимента Геррнштейна перед голубем поставлена не классическая задача принятия решений, а задача принятия *вероятностных* решений, иными словами, задача выбора *вероятностного* распределения *собственного* поведения, давно именуемая не только в эвентологии *задачей вероятностного выбора* [3, 4].

## 7 Голубиная оптимизация как вероятностный выбор

Уже в простейшем (из семи предложенных) варианте событийной модели подкрепления опора на эвен-

<sup>22</sup>О голубиной оптимизации как вероятностном выборе см. в параграфе 7 на стр. 122.

<sup>23</sup>Совсем другой вопрос, описывает ли данная событийная модель поведение реального голубя в коробке Скиннера при вероятностных распределениях “окружающей среды”, *контролируемых* экспериментатором. Ответ на этот вопрос может дать лишь голубиный эксперимент, описание которого я и предлагаю специалистам по когнитивной психологии.

тологическую *H*-теорему дает решение *задачи вероятностного выбора* в виде гиббсовского соответствия вероятностного поведения голубя вероятностному поведению окружающей среды, которое получено в параграфе 5.3 на стр. 120 с оптимальными параметрами соответствия из параграфа 5.4 на стр. 121, где оптимальность понимается как минимум *энтропийного отклонения*<sup>24</sup> между вероятностными распределениями поведений голубя и среды при ограничении на величину *среднего подкрепления*<sup>25</sup>.

Однако метод решения задачи вероятностного выбора, использованный в простейшем варианте, оказался недостаточно универсальным, чтобы с его помощью можно было решать эту же задачу в более сложных постановках (варианты II – VII). Объяснение обычное: в постановке простейшего варианта задачи не потребовалась одна ее особенность, важная для проблемы вероятностного выбора в целом.

Дело в том, что голубь не имеет прямого доступа к вероятностному распределению поведения окружающей среды (в эксперименте — *вероятностное распределение загрузки* зерен экспериментатором). *Вероятностное распределение загрузки* для него не только неизвестно, но и находится за пределами его прямых наблюдений. Однако он может воспринимать вероятностное следствие этой вероятностной причины — *вероятностное распределение подкрепления*, порождаемое *вероятностным распределением загрузки*.

Так как голубю вполне доступны наблюдения того, что происходит после его клевков, т.е. наблюдения подкрепления — сколько зерен появляется или не появляется в одной и/или в другой кормушках. Вместе с тем, правила загрузки и выдачи в разных вариантах таковы, что по результатам наблюдений за подкреплением невозможно отличить одно от другого *все* террасные события, порожденные событиями загрузки и правилами выдачи. Некоторые из этих террасных событий “сливаются” для голубя в одно и то же неразличимое подкрепление, а сами правила “слияния” определяются вариантами загрузки и выдачи подкрепления. В этом и кроется причина того, что в более сложных вариантах эксперимента взамен прямых наблюдений террасных событий загрузки голубю даны лишь наблюдения террасных событий подкрепления, порождаемых событиями загрузки и правилами выдачи.

Следовательно, если голубиный эксперимент рассматривать как проверку характера соответствия вероятностного поведения голубя вероятностному поведению окружающей его среды — соответствия, ко-

торое проявляется как результат решения голубем задачи вероятностного выбора, то требуется внести в постановку этой задачи одно простое, но важное добавление. Так как голубь не может искать соответствия с тем, что не доступно его наблюдениям, то задача вероятностного выбора — это задача поиска вида соответствия вероятностного поведения голубя *наблюдаемому* вероятностному поведению окружающей среды. В голубином эксперименте Геррнштейна *наблюдаемым* вероятностным поведением окружающей среды служит *вероятностное распределение подкрепления*, а не *вероятностное поведение экспериментатора*<sup>26</sup>.

## 8 Голубиная интерпретация эвентологической *H*-теоремы

Идея привлечь к событийной трансдукции голубино-го эксперимента Геррнштейна эвентологическую *H*-теорему возникла не только ради построения строгой событийной интерпретации этого давно известного эксперимента. Я чувствовал безотчетное намерение использовать результаты событийной трансдукции для формулировки более общего утверждения, пригодного для описания конфликтного поведения двух сторон, каждая из которых способна делать событийно-вероятностный выбор на основе собственных представлений о ценности происходящего<sup>27</sup>.

В обычной формулировке эвентологической *H*-теоремы<sup>28</sup>, которая по сути является событийной калькой с основной модификации классической *H*-теоремы из статистической физики, фигурирует две стороны. Из них одна (окружающая среда) совершенно “равнодушна” к происходящему и не оценивает его, в то время, как другая (система) способна выбирать собственное вероятностное распределение на основе своих представлений о ценности для нее всего, что происходит вокруг и дано ей в виде *наблюдаемого распределения*.

Ценность происходящего для “равнодушной” среды — постоянная функция на  $2^{\mathbb{X}}$ . Общая же ценность происходящего террасного события для обеих сторон конфликта равна сумме двух его ценностей для каждой стороны в отдельности. Причем в силу противостояния сторон их относительные ценности имеют (с точки зрения одной из сторон) противоположные знаки — то, что ценится одной стороной, отбрасывается у другой и потому имеет для той, другой, отрицательную ценность.

На этом пути можно построить обобщенную *H*-модель событийного конфликта, в которой обе сто-

<sup>24</sup> Я имею в виду *относительную энтропию*, известную также как *расстояние Кульбака-Лейблера* [13] между вероятностными распределениями.

<sup>25</sup> Величина *среднего подкрепления* определяется голубем из “соображений” выживания.

<sup>26</sup> Т.е., вероятностное распределение загрузки и правила выдачи, контролируемые экспериментатором.

<sup>27</sup> В эвентологии постулируется, что проявление этой способности характерно для разумного поведения.

<sup>28</sup> См. справочный параграф 9 на стр. 125.



роны равноправны в способности делать событийно-вероятностный выбор, а их событийное поведение описывается совместным *вероятностным распределением обобщенного гиббсовского типа*.

Напомню, что маргинальное вероятностное распределение гиббсовского типа одной стороны при наблюдаемом маргинальном вероятностном распределении другой стороны определяется на основе эвентологической  $\mathcal{H}$ -теоремы следующим образом.

**Определение.** Вероятностное распределение  $r_\varphi$  гиббсовского типа относительно наблюдаемого вероятностного распределения  $p$  и функции ценности  $v$  на  $2^{\mathfrak{X}}$  определяется для  $X \subseteq \mathfrak{X}$  формулами:

$$\frac{r_\varphi(X)}{r_\varphi(\emptyset)} = \exp \left\{ \alpha [v(X) - v(\emptyset)] \right\} \frac{p(X)}{p(\emptyset)}, \quad (8.1)$$

где параметр  $\alpha$  удовлетворяет соотношению

$$\langle v \rangle = \sum_{X \subseteq \mathfrak{X}} r_\varphi(X) v(X). \quad (8.2)$$

Формулы (8.1) можно сначала переписать в эквивалентном виде

$$\frac{r_\varphi(X) e^{\alpha_\varphi v_\varphi(X)}}{r_\varphi(\emptyset) e^{\alpha_\varphi v_\varphi(\emptyset)}} = \frac{p(X)}{p(\emptyset)}, \quad (8.3)$$

где под  $v = v_\varphi$  понимается функция ценности для стороны с распределением  $r_\varphi$ , а под  $\alpha = \alpha_\varphi$  — параметр, удовлетворяющий соотношению

$$\langle v_\varphi \rangle = \sum_{X \subseteq \mathfrak{X}} r_\varphi(X) v_\varphi(X). \quad (8.3)$$

Затем — неэквивалентно “расширить” до общих симметричных<sup>29</sup> формул

$$\frac{r_\varphi(X) e^{\alpha_\varphi v_\varphi(X)}}{r_\varphi(\emptyset) e^{\alpha_\varphi v_\varphi(\emptyset)}} = \frac{p(X) e^{\alpha_p v_p(X)}}{p(\emptyset) e^{\alpha_p v_p(\emptyset)}}, \quad (8.4)$$

где под  $v_p$  понимается функция ценности для стороны с распределением  $p$ , а под  $\alpha = \alpha_p$  — параметр, удовлетворяющий соотношению

$$\langle v_p \rangle = \sum_{X \subseteq \mathfrak{X}} p(X) v_p(X). \quad (8.5)$$

В итоге такого симметричного “расширения” формулы (8.3) оказываются ожидаемым частным случаем формул (8.4), если функция ценности  $v_p$  постоянна<sup>30</sup> на  $2^{\mathfrak{X}}$ , а  $\langle v_p \rangle$  оказывается константой, которая не зависит от вероятностного распределения  $p$ .

Общие формулы (8.4) я интерпретирую, как связь между гиббсовскими распределениями двух сторон

событийного конфликта, находящихся в конфликтном равновесии и использующих для принятия решений наблюдаемые распределения о событийном поведении противостоящей стороны. Иными словами, — в таких равновесных событийно-вероятностных распределениях в которых любое изменение поведения одной из сторон приводит к ухудшению ее положения в смысле ее же функции ценности.

Интерпретация общей эвентологической  $\mathcal{H}$ -теоремы видятся мне *теперь* как утверждение о стремлении обеих сторон событийного конфликта, способных делать вероятностный выбор, к равновесным гиббсовским распределениям собственного событийного поведения. Это утверждение должно выполняться для любого изначального совместного вероятностного распределения сторон, в котором зафиксированы лишь две их функции ценности и также две средние ценности происходящего, обеспечивающие выживание каждой стороне.

Если на событийную систему, составленную из двух событийно-конфликтующих сторон, взглянуть, как на единую систему в “равнодушном” и равновесном событийном окружении, то к ней применима обычная эвентологическая  $\mathcal{H}$ -теорема. Из этой теоремы следует, что такая событийная система, будучи в равновесии с “равнодушной” и равновесной окружающей средой, имеет вероятностное распределение гиббсовского типа. Другими словами, совместное вероятностное распределение двух сторон внутрисистемного событийного конфликта — это гиббсовское распределение, определяемое некоторой фиксированной совместной функцией ценности единой системы и фиксированным средним уровнем совместной ценности, обеспечивающей выживание этой составной системы.

Итак, множества событий  $\mathfrak{X}$  и  $\varphi(\mathfrak{X})$  имеют маргинальные вероятностные распределения  $p$  и  $r$  соответственно. Обозначу  $s$  их совместное распределение, т.е. вероятностное распределение множества событий  $\mathfrak{X} + \varphi(\mathfrak{X})$ , которое состоит из вероятностей террасных событий следующего вида для  $X \subseteq \mathfrak{X}$  и  $Y \subseteq \varphi(\mathfrak{X})$ :

$$s(X + Y // \mathfrak{X} + \varphi(\mathfrak{X})) = \mathbf{P}(\text{ter}(X + Y // \mathfrak{X} + \varphi(\mathfrak{X}))),$$

и обеспечивает заданные проекции на  $\mathfrak{X}$  и на  $\varphi(\mathfrak{X})$  соответственно:

$$p(X // \mathfrak{X}) = \sum_{Y \subseteq \varphi(\mathfrak{X})} s(X + Y // \mathfrak{X} + \varphi(\mathfrak{X})),$$

$$r_\varphi(Y // \varphi(\mathfrak{X})) = \sum_{X \subseteq \mathfrak{X}} s(X + Y // \mathfrak{X} + \varphi(\mathfrak{X})).$$

Из обычного предположения о “равнодушии” окружающей среды к происходящему и дополнительного предположения о ее равновесном состоянии следует,

<sup>29</sup>Это требование следует из симметрии участия двух сторон в событийном конфликте.

<sup>30</sup>Иначе говоря, если сторона с распределением  $p$  “равнодушно взирает” на происходящее.



что вероятностное распределение среды равновероятно. Это ведет к тому, что для совместного событийного описания составной системы и окружающей ее среды справедлив частный вариант эвентологической  $\mathcal{H}$ -теоремы, в котором утверждается, что для равновесия с равновесной окружающей средой совместное вероятностное распределение составной системы должно быть следующего гиббсовского типа:

$$\frac{s(X + Y // \mathfrak{X} + \varphi(\mathfrak{X}))}{s(\emptyset // \mathfrak{X} + \varphi(\mathfrak{X}))} = \frac{\exp \left\{ \alpha_s v_s(X + Y // \mathfrak{X} + \varphi(\mathfrak{X})) \right\}}{\exp \left\{ \alpha_s v_s(\emptyset // \mathfrak{X} + \varphi(\mathfrak{X})) \right\}}, \quad (8.1)$$

где  $v_s$  — функция совместной ценности происходящего в окружающей среде для составной системы, а параметр  $\alpha_s$  обеспечивает фиксированный средний уровень совместной ценности:

$$\langle v_s \rangle = \sum_{X \subseteq \mathfrak{X}} \sum_{Y \subseteq \varphi(\mathfrak{X})} s(X + Y // \mathfrak{X} + \varphi(\mathfrak{X})) v_s(X + Y // \mathfrak{X} + \varphi(\mathfrak{X})). \quad (8.2)$$

Далее естественно предположить, что распределения  $r_\varphi$  и  $p$  независимы. При таком предположении совместное распределение определяется маргинальностями:

$$s(X + Y // \mathfrak{X} + \varphi(\mathfrak{X})) = p(X // \mathfrak{X}) r_\varphi(Y // \varphi(\mathfrak{X})),$$

функция совместной ценности равна сумме функций маргинальных ценностей:

$$v_s(X + Y // \mathfrak{X} + \varphi(\mathfrak{X})) = v_p(X // \mathfrak{X}) + v_\varphi(Y // \varphi(\mathfrak{X})),$$

а средняя совместная ценность равна сумме средних маргинальных ценностей:

$$\langle v_s \rangle = \langle v_p \rangle + \langle v_\varphi \rangle.$$

Дальнейшая интерпретация двустороннего гиббсовского событийного конфликта двух сторон, основанного на эвентологической  $\mathcal{H}$ -теореме, думаю, не встретит особых трудностей и, скорее всего, будет рассмотрена в последующих работах.

\*\*\*

## 9 Справка: эвентологическая $\mathcal{H}$ -теорема [14, 15]

Рассматривая в статистической физике вопрос о приближении термодинамической системы к равновесию, Больцман<sup>31</sup> сформулировал [16, 1866], [17, стр.

<sup>31</sup> Больцман, Людвиг Эдуард (нем. Boltzmann, Ludwig Eduard, 1844–1906) — австрийский физик, один из основоположников статистической механики и статистической термодинамики; первопроходец и герой в науке, часто бывал не понят, иногда преследуем, но в конце концов был полностью

191] свою знаменитую  $\mathcal{H}$ -теорему. Она вызвала споры и, как следствие, подверглась некоторым видоизменениям. В настоящее время имеется несколько основных ее модификаций, отличающихся способом рассмотрения. Несмотря на разные способы рассмотрения, все эти модификации тождественны в главном: они указывают на возрастание энтропии при приближении к статистическому равновесию. Необратимость, вытекающая из  $\mathcal{H}$ -теоремы, вызвала много вопросов и возражений. Отвечая на них, Больцман интерпретировал локальное убывание энтропии как статистическое, а не динамическое изменение. Стоит подчеркнуть, что исходная  $\mathcal{H}$ -теорема Больцмана определяет равновесие только для одночастичной системы. В модификации Гиббса<sup>32</sup>  $\mathcal{H}$ -теорема определяет равновесную многочастичную функцию распределения системы. Поэтому я рассматриваю модификацию Гиббса<sup>33</sup> в качестве образца для аналогичной эвентологической  $\mathcal{H}$ -теоремы.

Введя величину

$$\mathcal{H}_{\frac{r_\varphi}{p}} = \sum_{X \subseteq \mathfrak{X}} r_\varphi(X) \ln \frac{r_\varphi(X)}{p(X)}, \quad (9.★)$$

которая называется *относительной энтропией распределения  $r_\varphi$  к распределению  $p$* <sup>34</sup>, замечу, что  $\mathcal{H}$ -функция (9.★) равна среднему по  $r_\varphi$  от  $\ln(r_\varphi/p)$  и не зависит от “террасных координат”  $X \subseteq \mathfrak{X}$  всего вероятностного пространства  $(\Omega, \mathcal{A}, \mathbf{P})$ , порожденных  $\mathfrak{X} \subset \mathcal{A}$ . Однако, так как  $\mathcal{H}$ -функция содержит суммирование по этим “террасным координатам”, она может зависеть от распределений  $r_\varphi$  и  $p$  в целом. Кроме того,  $\mathcal{H}_{\frac{r_\varphi}{p}}$  является функцией от наложенных на множество событий  $\mathfrak{X}$  внешних ограничений.

Если определять распределение  $r_\varphi$  при фиксированной средней ценности, то имеет место следующая

### Теорема 2 (эвентологическая $\mathcal{H}$ -теорема об

оправдан позднейшим развитием физики; научные интересы охватывали почти все области физики и ряд областей математики; важнейшие заслуги — кинетическая теория газов, исследование необратимых процессов и статистическая трактовка второго начала термодинамики, основанная на выведенной им связи энтропии и вероятности, которая была записана позже (1906) Планком в виде больцмановского соотношения “энтропия ~ вероятность”:  $S = k_B \cdot \ln W$  его имя носят физическая постоянная  $k_B$ , кинетическое уравнение, фактор, распределение частиц по дискретным уровням энергии в состоянии равновесия.

<sup>32</sup> Гиббс, Джозайя Уиллард (англ. Gibbs, Josiah Willard, 1839–1903) — американский математик, физик, химик и механик; один из основоположников, статистической физики, математической теории термодинамики, векторного анализа; установил фундаментальный закон статистической механики — распределение Гиббса, обобщил принцип энтропии, применяя второе начало термодинамики к широкому кругу физических процессов; во многом предопределил развитие современных точных наук и естествознания в целом.

<sup>33</sup> которую с небольшими обобщениями излагаю в версии, взятой из [18, 19].

<sup>34</sup> Для  $\varphi$ -изоморфных полных множеств событий  $\mathfrak{X}_\varphi = \varphi(\mathfrak{X})$  и  $\mathfrak{X}$  соответственно.

экстремальных свойствах событийных распределений гиббсовского типа). Пусть  $(\Omega, \mathcal{A}, \mathbf{P})$  — всеобщее вероятностное пространство<sup>35</sup>,  $\mathfrak{X} \subset \mathcal{A}$  и  $\mathfrak{X}_\varphi = \varphi(\mathfrak{X}) \subset \mathcal{A}$  — два конечных  $\varphi$ -изоморфных полных множества событий,  $v(X)$  — ограниченная действительная функция на  $2^{\mathfrak{X}}$ ,  $p(X)$  — некоторое фиксированное событийное распределение на  $2^{\mathfrak{X}}$ , а событийные распределения  $r_\varphi(X) = \mathbf{P}(\text{ter}(\varphi(X)))$  обеспечивают на  $2^{\mathfrak{X}}$  среднее значение данной функции  $v(X)$  на фиксированном уровне

$$\langle v \rangle = \sum_{X \subseteq \mathfrak{X}} r_\varphi(X) v(X). \quad (9.1)$$

Тогда минимум по  $r_\varphi$  относительной энтропии распределения  $r_\varphi$  к распределению  $p$ :

$$\mathcal{H}_{\frac{r_\varphi}{p}} = \sum_{X \subseteq \mathfrak{X}} r_\varphi(X) \ln \frac{r_\varphi(X)}{p(X)} \rightarrow \min_{r_\varphi} \quad (9.2)$$

среди всех событийных распределений  $r_\varphi$ , удовлетворяющих ограничению (9.1), достигается на событийных распределениях гиббсовского типа ( $\alpha < 0, \alpha > 0$ ):

$$r_\varphi(X) = \frac{1}{\mathcal{Z}_p} \exp \left\{ \alpha v(X) \right\} p(X), \quad (9.3)$$

которые без нормирующего множителя  $\mathcal{Z}_p = \sum_{X \subseteq \mathfrak{X}} \exp \left\{ \alpha v(X) \right\} p(X)$  записываются в эквивалентном виде

$$\frac{r_\varphi(X)}{r_\varphi(\emptyset)} = \exp \left\{ \alpha [v(X) - v(\emptyset)] \right\} \frac{p(X)}{p(\emptyset)}, \quad (9.4)$$

где  $\alpha$  — параметр, при отрицательных значениях которого событийные распределения (9.3) относятся к классу гиббсовских, а при положительных — к классу противоположных гиббсовских событийных распределений.

**Доказательство**<sup>36</sup>. Введем обозначение

$$h(X) = \exp \left\{ \alpha v(X) \right\} p(X) \quad (9.6)$$

и сравним относительные энтропии (9.2) для

$$r_\varphi(X) = \frac{1}{\mathcal{Z}_p} h(X) \text{ и } r_f(X) = \frac{1}{\mathcal{Z}_p} f(X),$$

<sup>35</sup>Напомним, что алгебра всеобщих событий  $\mathcal{A}$  в силу эвентологической аксиомы VII [20, 21, 22] обладает свойством симплициальной замкнутости относительно вероятности  $\mathbf{P}$ , которое означает, что вместе с каждым конечным множеством событий  $\mathfrak{X} \subset \mathcal{A}$ , алгебра содержит весь событийный симплекс, все элементы (конечные множества событий) которого состоят из событий, имеющих те же имена [20, 21, 22], что и события из данного конечного множества событий  $\mathfrak{X}$ .

<sup>36</sup>Взято в [14, 15, стр. 51-58], где используется идея доказательства гиббсовской модификации  $H$ -теоремы Больцмана в формулировке Исихары [18, 19, стр. 41].

где

$$f(X) = h(X) \cdot g(X) \quad (9.7)$$

— функция, которая после нормировки на  $\mathcal{Z}_p$  определяет некоторое вероятностное распределение  $r_f$  на  $2^{\mathfrak{X}}$ ; а  $g$  — еще одна функция на  $2^{\mathfrak{X}}$ , определяемая функциями  $f$  и  $h$ .

Находим

$$\begin{aligned} \mathcal{H}_{\frac{r_f}{p}} - \mathcal{H}_{\frac{r_\varphi}{p}} &= \\ &= \frac{1}{\mathcal{Z}_p} \sum_{X \subseteq \mathfrak{X}} \left[ f(X) \ln \frac{f(X)}{p(X)} - h(X) \ln \frac{h(X)}{p(X)} \right] = \\ &= \frac{1}{\mathcal{Z}_p} \sum_{X \subseteq \mathfrak{X}} h(X) \left[ g(X) \ln \frac{h(X)g(X)}{p(X)} - \ln \frac{h(X)}{p(X)} \right]. \end{aligned} \quad (9.8)$$

Условие вероятностной нормировки дает

$$\sum_{X \subseteq \mathfrak{X}} [f(X) - h(X)] = \sum_{X \subseteq \mathfrak{X}} h(X) [g(X) - 1] = 0, \quad (9.9)$$

а условие теоремы (9.1) —

$$\sum_{X \subseteq \mathfrak{X}} [f(X) - h(X)] v(X) = 0. \quad (9.10)$$

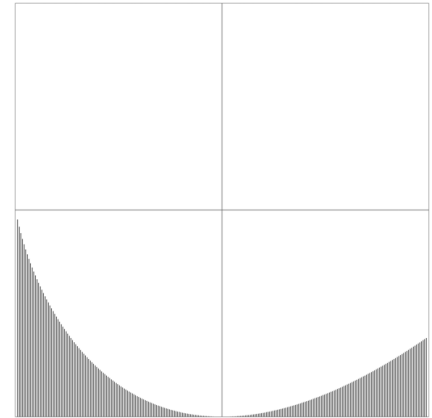


Рис. 8: Подграфик функции  $g \ln g - g + 1$  на квадрате  $[0, 2]^2$ .

Учитывая определение (9.6) в виде

$$\alpha v(X) = \ln \frac{h(X)}{p(X)},$$

получаем из (9.9) и (9.10)

$$\begin{aligned} &\frac{1}{\mathcal{Z}_p} \sum_{X \subseteq \mathfrak{X}} [f(X) - h(X)] \ln \frac{h(X)}{p(X)} = \\ &= \frac{1}{\mathcal{Z}_p} \sum_{X \subseteq \mathfrak{X}} h(X) [g(X) - 1] \ln \frac{h(X)}{p(X)} = 0. \end{aligned} \quad (9.11)$$

Вычтем (9.9) и (9.11) из (9.8), тогда

$$\mathcal{H}_{\frac{r_f}{p}} - \mathcal{H}_{\frac{r_\varphi}{p}} =$$

$$= \frac{1}{z_p} \sum_{X \subseteq \mathfrak{X}} h(X) [g(X) \ln g(X) - g(X) + 1].$$

По определению функция  $h(X)$  положительна, а величина

$$g \ln g - g + 1 = \int_1^g \ln g \, dg$$

неотрицательна для произвольных положительных  $g$  (см. рис. 8). Следовательно,

$$\mathcal{H}_{\frac{r_f}{p}} - \mathcal{H}_{\frac{r_\varphi}{p}} \geq 0$$

— функция  $\mathcal{H}_{\frac{r_\varphi}{p}}$  всегда не больше  $\mathcal{H}_{\frac{r_f}{p}}$ . Иными словами, распределения гиббсовского типа (9.3) обеспечивают минимум относительной энтропии (9.2). **Теорема доказана.**

**Замечание.** Поскольку вместе с конечным множеством событий  $\varphi(\mathfrak{X}) \subset \mathcal{A}$  алгебра всеобщих событий  $\mathcal{A}$  содержит весь событийный симплекс, порожденный  $\varphi(\mathfrak{X})$  (см. сноску<sup>35</sup> на стр. 126), то все конечные множества событий, на гиббсовских распределениях которых достигается минимум *энтропийного отклонения*, также содержатся в том же событийном симплексе.

## 10 Справка: изоморфные множества событий и относительная энтропия

### 10.1 Изоморфные множества событий

Пусть  $(\Omega, \mathcal{A}, \mathbf{P})$  — всеобщее вероятностное пространство<sup>37</sup>, а  $\mathfrak{X}$  и  $\mathfrak{Y} \subset \mathcal{A}$  — конечные множества событий.

Любые равномошные конечные множества событий изоморфны по отношению к операциям объединения, пересечения и взятия дополнения. Здесь мне интересны конечные множества событий, между которыми уже задан конкретный изоморфизм, порожденный некоторым автоморфизмом алгебры  $\mathcal{A}$ :

$$\varphi : \mathcal{A} \rightarrow \mathcal{A},$$

и такие множества я стану называть  *$\varphi$ -изоморфными конечными множествами событий*. Например, если для  $\mathfrak{X}, \mathfrak{Y} \subset \mathcal{A}$  либо  $\varphi(\mathfrak{X}) = \mathfrak{Y}$ , либо  $\varphi(\mathfrak{Y}) = \mathfrak{X}$ , то множества событий  $\mathfrak{X}$  и  $\mathfrak{Y}$  называются  $\varphi$ -изоморфными.

Ясно, что  $\varphi$ -изоморфные образы  $\varphi(X)$  подмножеств событий  $X \subseteq \mathfrak{X}$ , определяются формулами<sup>38</sup>

$$\varphi(X) = \{\varphi(x) : x \in X\} \subseteq \{\varphi(x) : x \in \mathfrak{X}\} = \varphi(\mathfrak{X}).$$

<sup>37</sup>Бывает полезно представлять, что  $(\Omega, \mathcal{A}, \mathbf{P})$  порождает эвентологическое пространство  $(\Omega, \mathcal{A}, \tilde{\mathcal{A}}, \mathbf{P})$  добавлением алгебры имен  $\tilde{\mathcal{A}}$  [23], погружение в которое требуют эвентологические аксиомы имени события, симплицальности и достаточности [20, 21, 22]. В эвентологическом пространстве каждому конечному множеству событий  $\mathfrak{X} \subset \mathcal{A}$  соответствует множество имен  $\tilde{\mathfrak{X}} \subset \tilde{\mathcal{A}}$ .

<sup>38</sup>Иногда, если изоморфизм  $\varphi$  фиксирован, произвольное событие  $y \in \mathfrak{Y} = \varphi(\mathfrak{X})$ , такое, что  $y = \varphi(x)$ , я буду записывать

Более того, поскольку изоморфизм  $\varphi$  сохраняет операции объединения, пересечения и взятия дополнения, то  $\varphi$ -образ произвольного террасного события  $\text{ter}(X//\mathfrak{X})$ , порожденного  $\mathfrak{X}$ , имеет вид

$$\varphi(\text{ter}(X//\mathfrak{X})) = \text{ter}(\varphi(X)//\varphi(\mathfrak{X})).$$

Вероятностное распределение  $I$ -го рода  $\varphi$ -образа  $\varphi(\mathfrak{X})$  множества событий  $\mathfrak{X}$  я буду определять как набор вероятностей

$$\{p(\varphi(X)//\varphi(\mathfrak{X})) : X \subseteq \mathfrak{X}\},$$

где

$$p(\varphi(X)//\varphi(\mathfrak{X})) = \mathbf{P}(\text{ter}(\varphi(X)//\varphi(\mathfrak{X})))$$

— вероятность  $\varphi$ -образа

$$\text{ter}(\varphi(X)//\varphi(\mathfrak{X})) = \bigcap_{x \in X} \varphi(x) \bigcap_{x \in \mathfrak{X} - X} (\varphi(x))^c$$

террасного события

$$\text{ter}(X//\mathfrak{X}) = \bigcap_{x \in X} x \bigcap_{x \in \mathfrak{X} - X} x^c.$$

### 10.2 Относительная энтропия изоморфных множеств событий

Проще всего начать с полных изоморфных множеств событий, т.е. с множеств событий  $\mathfrak{X}$ , порождающих лишь такие террасные события, которые отличны от невозможных событий:

$$\text{ter}(X//\mathfrak{X}) \neq \emptyset$$

для всех  $X \subseteq \mathfrak{X}$ . Поскольку вероятностная мера  $\mathbf{P}$  считается регулярной на  $\mathcal{A}$ , вероятностные распределения  $I$ -го рода полных множеств событий не содержат нулевых вероятностей:

$$p(X//\mathfrak{X}) = \mathbf{P}(\text{ter}(X//\mathfrak{X})) \neq 0.$$

**Определение 1<sup>39</sup>.** *Относительная энтропия между полным множеством событий  $\mathfrak{X}$  и  $\varphi$ -изоморфным ему полным множеством событий  $\varphi(\mathfrak{X})$  определяется двумя известными способами:*

$$\mathcal{H}_{\frac{\varphi(\mathfrak{X})}{\mathfrak{X}}} = \sum_{X \subseteq \mathfrak{X}} p(\varphi(X)//\varphi(\mathfrak{X})) \ln \frac{p(\varphi(X)//\varphi(\mathfrak{X}))}{p(X//\mathfrak{X})}, \quad (10.2.1)$$

$$\mathcal{H}_{\frac{\mathfrak{X}}{\varphi(\mathfrak{X})}} = \sum_{X \subseteq \mathfrak{X}} p(X//\mathfrak{X}) \ln \frac{p(X//\mathfrak{X})}{p(\varphi(X)//\varphi(\mathfrak{X}))}. \quad (10.2.2)$$

короче:

$$y_x = \varphi(x),$$

а  $\varphi$ -образ  $Y_X = \varphi(X)$  каждого подмножества событий  $X \subseteq \mathfrak{X}$  определять формулой

$$Y_X = \{y_x : x \in X\}.$$

<sup>39</sup>Простейший пример относительной энтропии изоморфных множеств событий рассмотрен в параграфе 5.2 на стр. 120.

Если же множества событий  $\mathfrak{X}$  или  $\varphi(\mathfrak{X})$  не являются полными множествами событий, т.е. среди определяющих их вероятностей  $I$ -го рода есть нулевые, то под знаком суммы могут возникать три вида неопределенностей.

Два вида этих неопределенностей устраняются принятием следующих обычных соглашений:

$$0 \ln \frac{0}{0} = 0, \quad 0 \ln \frac{0}{p} = 0, \quad (10.2.3)$$

а для устранения третьего возможного вида

$$p \ln \frac{p}{0} \quad (10.2.4)$$

ничего не остается делать, как модифицировать определение относительной энтропии таким образом, чтобы она была определена и в этой ситуации.

### 10.3 Модификация определения относительной энтропии

Начну с рассмотрения крайне “неблагоприятной” для определения 1 ситуации неполных множеств событий, чтобы потом перейти к общей.

#### 10.3.1 Крайняя ситуация

Пусть множество событий  $\mathfrak{X}$  и его  $\varphi$ -образ  $\varphi(\mathfrak{X})$  порождают террасные события, которые *отличны от невозможного* на непересекающихся подмножествах  $2^{\mathfrak{X}}$ . Иначе говоря,

$$\text{ter}(X//\mathfrak{X}) \neq \emptyset \text{ для } X \in \mathcal{O} \subset 2^{\mathfrak{X}},$$

$$\text{ter}(\varphi(X)//\varphi(\mathfrak{X})) \neq \emptyset \text{ для } X \in \mathcal{O}_{\varphi} \subset 2^{\mathfrak{X}},$$

где  $\mathcal{O} \cap \mathcal{O}_{\varphi} = \emptyset$ . Тогда суммы в определении 1 будут содержать только слагаемые всех трех видов неопределенности, из которых в первых двух видах (10.2.3) слагаемые обычно полагаются равными нулю, а в третьем (10.2.4) их можно попробовать положить равными единице, чтобы не разрушить смысл относительной энтропии, как средней меры расхождения распределений. При этой модификации и в такой “неблагоприятной” ситуации относительные энтропии оказываются равными:

$$\mathcal{H}_{\frac{\varphi(\mathfrak{X})}{\mathfrak{X}}} = |\mathcal{O}_{\varphi}|, \quad \mathcal{H}_{\frac{\mathfrak{X}}{\varphi(\mathfrak{X})}} = |\mathcal{O}|.$$

#### 10.3.2 Общая ситуация

В общей ситуации обозначим

$$\mathcal{O}_{+} = \mathcal{O} \cap \mathcal{O}_{\varphi} \neq \emptyset$$

непустое подмножество  $2^{\mathfrak{X}}$ , относительно которого и множество событий  $\mathfrak{X}$ , и его  $\varphi$ -образ  $\varphi(\mathfrak{X})$  являются полными.

**Определение 2.** Три соглашения

$$0 \ln \frac{0}{0} = 0, \quad 0 \ln \frac{0}{p} = 0, \quad p \ln \frac{p}{0} = 1 \quad (10.3.1)$$

и определение 1 приводят к формулам для относительных энтропий произвольного множества событий  $\mathfrak{X}$ , и его  $\varphi$ -образа  $\varphi(\mathfrak{X})$  вида:

$$\mathcal{H}_{\frac{\varphi(\mathfrak{X})}{\mathfrak{X}}} = \mathcal{H}_{\frac{\varphi(\mathfrak{X})}{\mathfrak{X}}}((\mathcal{O}_{\varphi} - \mathcal{O}_{+})^c) + |\mathcal{O}_{\varphi} - \mathcal{O}_{+}|, \quad (10.3.2)$$

$$\mathcal{H}_{\frac{\mathfrak{X}}{\varphi(\mathfrak{X})}} = \mathcal{H}_{\frac{\mathfrak{X}}{\varphi(\mathfrak{X})}}((\mathcal{O} - \mathcal{O}_{+})^c) + |\mathcal{O} - \mathcal{O}_{+}|, \quad (10.3.3)$$

где использованы сокращенные обозначения:

$$\begin{aligned} \mathcal{H}_{\frac{\varphi(\mathfrak{X})}{\mathfrak{X}}}((\mathcal{O}_{\varphi} - \mathcal{O}_{+})^c) &= \\ &= \sum_{X \notin \mathcal{O}_{\varphi} - \mathcal{O}_{+}} p(\varphi(X)//\varphi(\mathfrak{X})) \ln \frac{p(\varphi(X)//\varphi(\mathfrak{X}))}{p(X//\mathfrak{X})}, \\ \mathcal{H}_{\frac{\mathfrak{X}}{\varphi(\mathfrak{X})}}((\mathcal{O} - \mathcal{O}_{+})^c) &= \\ &= \sum_{X \notin \mathcal{O} - \mathcal{O}_{+}} p(X//\mathfrak{X}) \ln \frac{p(X//\mathfrak{X})}{p(\varphi(X)//\varphi(\mathfrak{X}))}. \end{aligned}$$

## 11 Справка: закон соответствия<sup>40</sup>

В бихевиоризме, а впоследствии и в когнитивной психологии *закон соответствия* — количественное соотношение, которое поддерживается между частотой реакции и частотой подкрепления в синхронных сценариях; попросту заявляет, что существует корреляция между поведением живого организма и окружающей средой; реально используется, когда животные субъекты (“нелюди”) подвергаются согласованным сценариям подкрепления; применимость закона в других ситуациях менее ясна и находится в зависимости от сделанных предположений и деталей эксперимента; оказал существенную помощь анализам поведения живого организма в установлении законов, связывающих это поведения с окружающей средой, а также в написании соответствующих уравнений, которые ясно характеризуют эту корреляцию.

Короче говоря, закон соответствия полагает, что реакция живого организма на сценарий подкрепления пропорциональна количеству/продолжительности полученного им подкрепления.

У бихевиористов есть три идеи относительно того, как люди и животные максимизируют подкрепление (сравни со сноской<sup>9</sup> на стр. 116):

- *молекулярная максимизация* (“альтернативная” оптимизация): организмы всегда выбирают ту альтернативу реакции, которая наиболее вероятно будет тотчас подкреплена.

<sup>40</sup>авторизованный выборочный перевод из Wikipedia ([http://en.wikipedia.org/wiki/Matching\\_law](http://en.wikipedia.org/wiki/Matching_law)).

- *молярная максимизация* (“средняя” оптимизация): организмы распределяют свои ответы среди различных альтернатив, чтобы максимизировать количество подкрепления, которое они зарабатывают за длительный промежуток времени.
- *усовершенствование* (“соответственная” оптимизация): организмы реагируют на подкрепления так, чтобы улучшить соответствие между пропорциями своих реакций и пропорциями подкреплений.

Закон соответствия впервые был сформулирован Геррнштейном<sup>1</sup> как следствие его эксперимента с голубями на синхронных сценариях подкреплений и реакций [1, 1961]. Голубям предоставляли две кнопки в коробке Скиннера, каждая из которых приводила к различным порциям пищи. Голуби имели тенденцию клевать кнопку, которая приводила к большему подкреплению пищей, чаще, чем другую кнопку; причем они делали это в той пропорции, которая соответствовала пропорции подкрепления.

Если  $R_1$  и  $R_2$  — нормы реакций в двух сценариях, которые позволяют получить размер подкрепления  $R_{v1}$  и  $R_{v2}$ , строгий закон соответствия предписывает, что

$$\frac{R_1}{R_2} = \frac{R_{v1}}{R_{v2}}.$$

Последующие исследования показали, что экспериментальные данные обычно отступают от строгого соответствия, но в очень хорошем приближении отвечают обобщенному закону соответствия [24]

$$\frac{R_1}{R_2} = b \left( \frac{R_{v1}}{R_{v2}} \right)^s,$$

что более удобно записывается в логарифмической форме

$$\log \left( \frac{R_1}{R_2} \right) = \log(b) + s \cdot \log \left( \frac{R_{v1}}{R_{v2}} \right).$$

Константы  $b$  и  $s$  обычно упоминаются как “склонность” и “чувствительность” соответственно.

Этот “обобщенный закон соответствия” объясняет высокую долю расхождений в большинстве экспериментов на синхронных сценариях для “нелюдей”. Значение  $b$  зависит от деталей проведения эксперимента, а значение  $s$  чаще всего приблизительно равно 0.8, тогда как для строгого соответствия значение  $s$  должно быть равно 1.0 [24, 25].

Закон соответствия теоретически важен по двум причинам. Во-первых, он предлагает простое определение количества поведения, которое можно распространить на множество других ситуаций. Во-вторых, кажется, предлагает законный и прогнозируемый расчет выбора; как Геррнштейн выражал это (1970)

в оперантном анализе, выбор — это всего лишь поведенческий набор в контексте другого поведения [2]. Таким образом закон бросает вызов любой идее относительно доброй воли, как утверждал Скиннер<sup>7</sup> об экспериментальном анализе поведения в своей книге [26, “Вне Свободы и Достоинства”, 1971]. Претензии закона соответствия достаточно серьезны, так как он может быть распространен с голубей на людей. Когда в экспериментах участвовали люди, которые выбирали свои реакции под воздействием синхронных подкреплений, в одних экспериментах соответствие соблюдалось [27], но в других [28] были обнаружены значительные отклонения от него.

Хотя Геррнштейн и был пионером в этой области, последние исследования показывают, что в отличие от обобщенного уравнения соответствия, первоначальное уравнение Геррнштейна не в состоянии точно описать экспериментальные данные синхронных сценариев при существенном диапазоне условий. Поэтому аналитиками поведения широко используется обобщенное уравнение соответствия — много более мощный описательный инструмент. Закон соответствия весьма расширяет понимание психологов и аналитиков таких естественных явлений поведения, как развитие и различные психопатологии, связанные с развитием. Обнаружено, что ответное соответствие предсказывает, какую тактику конфликта использовать в конфликте между детьми и родителями. Пропорция соответствия предсказывает, например, будущие аресты. Даже использование детьми ненормативной лексики, оказывается, следует правилу соответствия.

## 12 Справка: Скиннер об ощущении (*tacting*) частных и публичных событий<sup>41</sup>

*Частные события* — события, доступные только для говорящего. *Публичные события* — события, которые происходят за пределами организма, которые наблюдаются более, чем одним человеком. Головная боль — пример частного события, а автомобильная катастрофа — пример публичного события. Частные события были впервые описаны и признаны Скиннером<sup>7</sup>.

*Ощущение (*tacting*) частных событий* организмом формируется вербальным сообществом, которое по-разному подкрепляет разнообразие поведения и ответов на происходящие частные события. Например, если ребенок вербально заявляет, “круг”, когда круг находится в непосредственной окружающей среде, это может быть ощущением частного события. Если ребенок вербально заявляет, “у меня болит зуб”, она/он может ощущать частное событие, и стимул принадлежит ему — говорящему, а не остальной ча-

<sup>41</sup> авторизованный выборочный перевод из Wikipedia ([http://en.wikipedia.org/wiki/Verbal\\_Behavior](http://en.wikipedia.org/wiki/Verbal_Behavior)).



сти вербального сообщества.

Вербальное сообщество формирует первоначальное развитие и поддержание или прекращение ощущений частных событий. Организм подобным образом отвечает и на частные стимулы и на публичные стимулы [8, р. 130]. Однако, более трудно для вербального сообщества сформировать вербальное поведение, связанное с частными событиями. Может быть более трудно сформировать частные события, но есть критические вещи, которые происходят в пределах организма, которые не должны быть исключены из нашего понимания вербального поведения.

Несколько проблем связаны с ощущением частных событий. Скиннер (1957) признал две главных дилеммы. Сначала, он признал трудность с предсказанием и управлением стимулами, связанными с ощущением частных событий [8, р. 130]. Это описано как отсутствие стимула для членов вербального сообщества. Вторая проблема, которую описывает Скиннер (1957), — наша текущая неспособность понять, как развивается вербальное поведение, связанное с частными событиями [8, р. 131].

Скиннер (1957) описывает четыре потенциальных способа, которыми вербальное сообщество может опознать вербальное поведение без доступа к стимулам говорящего. Он предполагает, что самый частый метод — через “общее публичное сопровождение”. Примером может быть то, когда ребенок падает и разбивает коленку, а попечитель заявляет что-то вроде, “ты был травмирован”. Другой метод — “сопутствующий ответ”, связанный с частным стимулом. Примером служит, когда ребенок прибегает, кричит и держится руками за коленку, а попечитель и тут может заявить “ты был травмирован”. Третий путь состоит в том, когда вербальное сообщество обеспечивает подкрепление, зависящее от конкретного поведения, а организм обобщает это к частному событию, которое происходит. Скиннер именуется это как “метафорическое расширение”. Заключительный метод, который предлагает Скиннер, может помочь формироваться нашему вербальному поведению и способствовать тому, чтобы поведение, происходящее первоначально на нижнем уровне, затем превращалось в частное событие [8, р. 134]. Такое понятие может быть получено в итоге из понимания, что вербальное поведение частных событий может быть сформировано с помощью вербального сообщества через расширение языка ощущений.

Частные события ограничены и не должны служить “объяснениями поведения” [8, р. 254]. Скиннер (1957) предостерегает, что, “язык частных событий может легко отвлечь нас от публичных причин поведения”.

### 13 Справка: Лефевр о зарождении разума [29, 30]

Вы не найдете в работах Лефевра буквально такое же высказывание, как на стр. 115, но познакомившись поближе с его трудами [31, 5, 32, 33, 34, 35, 6, 7, 36], придете вместе с ним, именно, к такому выводу:

*“разум зарождается там и тогда, где и когда зарождается способность делать вероятностный выбор”.*

Эта событийно-вероятностная<sup>42</sup> гипотеза о зарождении разума, которую Лефевр выдвинул, опираясь на результаты экспериментальной психологии [37], как оказалось, сразу же следует из фундаментального бахтинского определения события как со~бытия [38] и одной основополагающей эвентологической идеи [3, 4]:

*“со~бытие и вероятность — два взаимно обусловленных понятия, теряющие смысл в отрыве друг от друга”.*

\*\*\*

### 14 Голубь и разум: итоги эвентологической трансдукции

Считаю, что эвентологическая цель, поставленная в этой работе, достигнута.

Во-первых, мне удалось сформулировать событийно-вероятностное правило *гиббсовского соответствия*, основанное на эвентологической *H*-теореме и заменяющее *закон соответствия*, предложенный психологами для объяснения соответствия реакций живых организмов подкреплениям из окружающей среды.

Во-вторых, “глубокое погружение” в коробку Скиннера открыло очевидную теперь для меня цепочку логических суждений, которые без сомнения доказывают обоснованность предположений, сделанных для вывода гиббсовского соответствия.

В-третьих, предлагаемое в работе ясное событийно-вероятностное описание возможных вариантов голубинового эксперимента позволяет провести проверку правила гиббсовского соответствия на практике.

В четвертых, если правила гиббсовского соответствия подтвердятся результатами экспериментов, то это можно будет рассматривать убедительным подтверждением гипотезы Лефевра о зарождении разума в живых организмах, как зарождении в них способности делать вероятностный выбор.

<sup>42</sup>Нельзя не признать, что Лефевр в своих работах, оперирует только с вероятностями, не используя явно понятие события, хотя всегда молчаливо его предполагая.

В итоге эвентология, как наука о событиях, получит весомое подкрепление из среды когнитивных психологов, так как гипотеза Лефевра — это один из ее фундаментальных постулатов.

\*\*\*

**Благодарности.** Автор благодарит всех участников семинара по эвентологической математике, а также студентов второго курса Института математики и фундаментальной информатики СФУ за полезные обсуждения проблемы вероятностного выбора на протяжении всего весеннего семестра 2013 года. Моя специальная благодарность Аркадию Арсеньевичу Новоселову за неперенные вопросы и активное сопротивление моей точке зрения на эту проблему, одним из результатов преодоления которого явилась предлагаемая работа.

## Список литературы

- [1] R. J. Herrnstein. Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of Experimental Analysis of Behavior*, 4:267–272, 1961.
- [2] R. J. Herrnstein. On the law of effect. *Journal of the Experimental Analysis of Behavior*, 13:243–266, 1970.
- [3] О. Ю. Воробьев. *Эвентология*. Сибирский федеральный университет, Красноярск, 2007, 435с., <http://eventology-theory.ru/0-books/1-VorobyevOleg~2007~Eventology~435p.pdf>.
- [4] O. Yu. Vorobyev. *Eventology*. Siberian Federal University, Krasnoyarsk, Russia, 2007, 435p. (in Russian, abstract in English), <http://eventology-theory.com/0-books/1-VorobyevOleg~2007~Eventology~435p.pdf>.
- [5] V. A. Lefebvre. *Algebra of Conscience*. Kluwer Academic Publishers, Boston, 2001.
- [6] В. А. Лефевр. Ментализм и бихевиоризм: слияние? *Рефлексивные процессы и управление*, 4(1):75–98, 2004.
- [7] В. А. Лефевр. Высшие ценности и формальная теория выбора. *Рефлексивные процессы и управление*, 1-2(9):51–56, 2009.
- [8] B. F. Skinner. *Verbal Behavior*. Copley Publishing Group, Acton, MA, 1957, 478p.
- [9] N. Chomsky. Review of Skinner's Verbal Behavior. *Language*, 35:26–58, 1959.
- [10] О. Ю. Воробьев. Относительная энтропия изоморфных множеств событий. *Заметки ЭМ-семинара*, 12:1–... — Красноярск: СФУ. 21 апреля, 2013.
- [11] О. Ю. Воробьев и Е. Е. Голденок. Многособытийные модели рынка, следующие из эвентологической H-теоремы. *Журнал Сибирского федерального университета. Математика & Физика*, 4(4):419–433, 2011 (на англ.).
- [12] O. Yu. Vorobyev and E. E. Goldenok. Many-event-based models of market following from the eventological H-theorem. *Journal of Siberian Federal University. Mathematics & Physics*, 4(4):419–433, 2011.
- [13] Leibler R.A. Kullback S. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951.
- [14] О. Ю. Воробьев. Эвентологическая H-теорема. *Труды VII Всерос. ФАМ конференции по финансово-актуарной математике и смежным вопросам*, Красноярск: СФУ (под ред. Олега Воробьева), 1:51–58, 2008.
- [15] O. Yu. Vorobyev. Eventological H-theorem. In. *Proc. of the VII All-Russian FAM Conf. on Financial and Actuarial Mathematics and Related Fields*, Krasnoyarsk, SFU (Oleg Vorobyev ed.), 1:51–58, 2008 (in Russian, abstract in English).
- [16] L. Boltzmann. Über die mechanische bedeutung des zweiten hauptsatzes der wärmetheorie. *Wiener Berichte*, 53:195–220, 1866.
- [17] Л. Больцман. *Избранные труды*. Наука, Москва, 1984, 593с.
- [18] A. Isihara. *Statistical Physics*. Academic Press, New York, London, 1971.
- [19] А. Исихара. *Статистическая физика*. Мир, М., 1973, 471с.
- [20] О. Ю. Воробьев. Начала аксиоматизации эвентологии. *Труды VIII Международ. ФАМ конференции по финансово-актуарной математике и смежным вопросам*, Красноярск: СФУ (под ред. Олега Воробьева), 1:77–82, 2009.
- [21] O. Yu. Vorobyev. Elements of axiomatizing the eventology. In. *Proc. of the VIII Intern. FAM Conf. on Financial and Actuarial Mathematics and Related Fields*, Krasnoyarsk, SFU (Oleg Vorobyev ed.), 1:77–82, 2009 (in Russian, abstract in English).
- [22] O. Yu. Vorobyev. On elements of axiomatizing eventology. *Journal of Siberian Federal University. Mathematics & Physics*, 3(2):157–164, 2010.
- [23] О. Ю. Воробьев. Невербальность события и его имя. *Заметки ЭМ-семинара*, 8:1–... — Красноярск: СФУ. 6 апреля, 2013.
- [24] W. M. Baum. On two types of deviation from the matching law: Bias and undermatching. *Journal of the Experimental Analysis of Behavior*, 22:231–242, 1974.
- [25] M. Davison and D. McCarthy. *The matching law: A research review*. Hillsdale, NJ: Erlbaum, 1988.
- [26] B. F. Skinner. *Beyond Freedom and Dignity*. Alfred A. Knopf, New York, 1971.
- [27] C. M. Bradshaw, E. Szabadi, and P. Bevan. Behavior of humans in variable-interval schedules of reinforcement. *Journal of the Experimental Analysis of Behavior*, 26:135–141, 1976.
- [28] P.J. Horne and C.F. Lowe. Determinants of human performance on concurrent schedules. *Journal of the Experimental Analysis of Behavior*, 59:29–60, 1993.

- [29] О. Ю. Воробьев. Трансформация по Эпштейну и информация: эвентологический взгляд. *Труды XVI Международ. ЭМ конференции по эвентологической математике и смежным вопросам*, Красноярск: СФУ (под ред. Олега Воробьева):57–65, Декабрь 2012. ISBN 978-5-9903358-2-0.
- [30] O. Yu. Vorobyev. Transformation by Epstein and information: an eventological view. In. *Proc. of the XVI Intern. EM conference on eventological mathematics and related fields*, Krasnoyarsk: SFU (Oleg Vorobyev ed.):57–65, December 2012. ISBN 978-5-9903358-2-0.
- [31] В.А. Лефевр. *Космический субъект*. Инкварто, М., 1996, 184с.
- [32] В. А. Лефевр. *Алгебра совести*. Когито-Центр, Москва, 2003, 412с.
- [33] В. А. Лефевр. *Рефлексия*. Когито-Центр, Москва, 2003, 496с.
- [34] В. А. Лефевр. *Лекции по теории рефлексивных игр*. Когито-Центр, Москва, 2009.
- [35] В. А. Лефевр. Закон само-рефлексии: возможное общее пояснение трех различных психологических феноменов. *Рефлексивные процессы и управление*, 3(1):64–73, 2003.
- [36] В. А. Лефевр. А что если Платон прав? *Рефлексивные процессы и управление*, 1-2(10):23–32, 2010.
- [37] R. J. Herrnstein. Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of Experimental Analysis of Behavior*, 4:267–272, 1961.
- [38] M. M. Bakhtin. *Toward a Philosophy of the Act*. University of Texas Press, Austin (1993), St.Petersburg, 1920.

# Эвентологические особенности выбора точки из Фреше-многогранника в симплексе распределений

Олег Юрьевич Воробьев

Сибирский федеральный университет  
Институт математики и фундаментальной информатики  
Торгово-экономический институт  
Красноярск  
oleg.yu.vorobyev@gmail.com

**Аннотация.** Рассматриваются эвентологические особенности одного алгоритма выбора вероятностного распределения множества событий из Фреше-многогранника в симплексе распределений.

**Ключевые слова.** Эвентология, теория вероятностей, элементарный исход, событие, вероятность, множество событий, алгебра событий, равномерный выбор, симплекс распределений, Фреше-многогранник.

Состава и события преступления не было.

Россия, 2013.

Считаю известным общий алгоритм<sup>1</sup> равномерного выбора  $m$ -мерной точки из стандартного единичного  $m$ -вершинного симплекса

$$\mathbb{S}^m = \left\{ (r_1, \dots, r_m) : \sum_{i=1}^m r_i = 1; r_i \geq 0, i = 1, \dots, m \right\}.$$

## 1 Равномерный выбор вероятностного распределения из симплекса всех распределений

Поставим задачу равномерного выбора вероятностных распределений  $N$ -множеств событий  $\mathfrak{X}$ , которые рассматриваемые как  $2^N$ -мерные вектора

$$(p(X//\mathfrak{X}), X \subseteq \mathfrak{X}) \in \mathbb{R}^{(2^N)}.$$

© 2013 О.Ю.Воробьев

Олег Воробьев (ред.), Труды ХII ФАМЭБ'2013, Красноярск: СФУ

<sup>1</sup>Пусть

$$\alpha_1, \dots, \alpha_{m-1}$$

— значения случайных величин, каждая из которых равномерно распределена на отрезке  $[0, 1]$ , а

$$\alpha_1^* \leq \dots \leq \alpha_{m-1}^*$$

— те же значения, расположенные в порядке возрастания. Если положить  $\alpha_0^* = 0, \alpha_m^* = 1$ , то  $m$ -мерная точка

$$(r_i = \alpha_i^* - \alpha_{i-1}^*, i = 1, \dots, m) \in \mathbb{R}^m$$

— это значение случайного  $m$ -мерного вектора, равномерно распределенного в симплексе  $\mathbb{S}^m \subset \mathbb{R}^m$ .

равномерно выбираются из  $2^N$ -вершинного симплекса  $\mathbb{S}^{(2^N)}$  таких распределений.

С поставленной задачей принципиально справляется упомянутый выше общий алгоритм. Однако меня здесь интересуют особенности подробной реализации этого общего алгоритма в той частной ситуации, когда компоненты выбираемого вектора занумерованы подмножествами  $X \subseteq \mathfrak{X}$ . Кроме того, меня интересует еще одна частная особенность, которая приобретает значение из-за моих ограниченных программистских возможностей. Я буду считать, что алгоритм не имеет прямого доступа к разрядам чисел, записанных в двоичной системе счисления и что для обеспечения такого доступа требуются специальные процедуры, о которых коротко пойдет речь ниже.

Для описания частных особенностей такого рода мне придется определить взаимно однозначное отображение

$$g : 2^{\mathfrak{X}} \rightarrow \{0, \dots, 2^N - 1\} \subset \mathbb{Z}. \quad (1.1)$$

при некотором фиксированном порядке событий в

$$\mathfrak{X} = \{x_N, \dots, x_1\} \quad (1.2)$$

формулой

$$g(X//\mathfrak{X}) = \sum_{i=1}^N 1_X(x_i) 2^i, \quad (1.3)$$

в силу которой

$$0 = g(\emptyset) < \dots < g(X) < \dots < g(\mathfrak{X}) = 2^N - 1.$$

Взаимно однозначное отображение  $g$  называется *координатным отображением* и определяется порядком событий (1.2).

Замечу также, что значение  $g(X)$  координатного отображения на подмножестве  $X \subseteq \mathfrak{X}$  это число, двоичная запись которого составлена из значений индикатора  $1_X(x)$  на упорядоченных событиях  $x \in \mathfrak{X}$ :

$$g(X//\mathfrak{X}) = 1_X(x_N) \dots 1_X(x_1). \quad (1.4)$$

Например,

$$g(\emptyset//\mathfrak{X}) = \underbrace{0 \dots 0}_N, \quad (1.4')$$

$$g(\mathfrak{X}//\mathfrak{X}) = \underbrace{1 \dots 1}_N. \quad (1.4'')$$

Обозначу еще

$$g(X//\mathfrak{X}) = g_N(X) \dots g_1(X) \quad (1.5)$$

— запись числа  $g(X//\mathfrak{X})$  в двоичной системе счисления, а  $g_i(X)$  — значение  $i$ -го двоичного разряда в этой записи.

Обратное к  $g$  отображение

$$g^{-1} : \{0, \dots, 2^N - 1\} \rightarrow 2^{\mathfrak{X}} \quad (1.6)$$

определяется для  $\gamma \in \{0, \dots, 2^N - 1\}$  очевидной формулой

$$g^{-1}(\gamma) = \{x_n \in \mathfrak{X} : \gamma_n = 1\} \subseteq \mathfrak{X}, \quad (1.7)$$

где  $\gamma_n$  —  $n$ -й двоичный разряд числа  $\gamma$ , двоичная запись которого имеет вид

$$\gamma = \gamma_N \dots \gamma_1.$$

## 2 Процедуры “сет-инт” и “инт-сет”

Для работы алгоритма равномерного выбора необходимы две взаимно обратные процедуры: *сет-инт* и *инт-сет*.

1. Процедура *сет-инт* (от множества к целому числу)

$$g : X \rightarrow g(X)$$

определяется формулой (1.2) и позволяет по известному подмножеству  $X \subseteq \mathfrak{X}$ , т.е. по значениям индикатора  $X$  на событиях  $x_i \in \mathfrak{X}$ , вычислять  $i$ -тые двоичные разряды

$$g_i(X) = 1_X(x_i)$$

координатного отображения  $g(X)$ , а, значит, и его значение в другой системе счисления.

2. Обратная процедура *инт-сет* (от целого числа к множеству) по значению координатного отображения вычисляет подмножество, на котором координатное отображение принимает данное значение:

$$g^{-1} : g(X) \rightarrow X.$$

Обратная процедура *инт-сет* сводится к последовательному вычислению двоичных разрядов  $g_i(X)$ ,  $i = 1, \dots, N$ , известного значения  $\gamma(X)$ , записанного в другой, например, десятичной системе счисления. Значения двоичных разрядов позволяет определить подмножество формулой (1.7):

$$X = \{x_i \in \mathfrak{X} : g_i(X) = 1\} \subseteq \mathfrak{X},$$

так как единичное значение  $i$ -того двоичного разряда указывает на то, что  $x_i \in X$ , а нулевое значение — на то, что  $x_i \in \mathfrak{X} - X$ .

Процедуры *сет-инт* и *инт-сет* дают возможность переходить по мере надобности от нумерации вероятностей в распределении первого рода подмножествами событий  $X \subseteq \mathfrak{X}$ :

$$\{p(X//\mathfrak{X}), X//\mathfrak{X}\}$$

к их взаимно однозначной нумерации числами из  $\{0, 1, \dots, 2^N - 1\}$ :

$$\{p^*(\gamma), \gamma \in \{0, 1, \dots, 2^N - 1\}\}$$

и обратно. При этом в одну сторону для  $X \subseteq \mathfrak{X}$

$$p(X//\mathfrak{X}) = p^*(g(X//\mathfrak{X})),$$

а в другую сторону для  $\gamma \in \{0, 1, \dots, 2^N - 1\}$

$$p^*(\gamma) = \begin{cases} p(\emptyset//\mathfrak{X}), & \gamma = g(\emptyset//\mathfrak{X}); \\ \dots & \dots \\ p(X//\mathfrak{X}), & \gamma = g(X//\mathfrak{X}); \\ \dots & \dots \\ p(\mathfrak{X}//\mathfrak{X}), & \gamma = g(\mathfrak{X}//\mathfrak{X}). \end{cases}$$

### 2.1 Справка: коды на Паскале для процедур “сет-инт” и “инт-сет”

```
function Set32ToInt(Value: string): Cardinal;
var
  i: integer;
begin
  Result := 0;
  for i := 1 to Length(Value) do
    if Value[i] = '1' then
      Result := Result + Round(exp((i - 1) * ln(2)));
  end;

function IntToSet32(Value: Cardinal): string;
var
  i: integer;
begin
  SetLength(Result, 32);
  for i := 1 to 32 do
    if (Value shr (i - 1)) shr 31 = 0 then
      Result[33 - i] := '0'
    else
      Result[33 - i] := '1';
  end;
```

## 3 Алгоритм выбора распределения множества событий из Фреше-многогранника

Предположим, что требуется выбрать из симплекса всех распределений множества событий  $\mathfrak{X}$  из  $N = |\mathfrak{X}|$  событий такое распределение, которое попадает в ту часть симплекса, которая заполнена распределениями с зафиксированными вероятностями событий. А поскольку такие части симплекса вероятностных распределений называются *Фреше-многогранниками* [1, 2], то речь идет о выборе вероятностного распределения  $\mathfrak{X}$  из Фреше-многогранника, определяемого данными вероятностями событий из  $\mathfrak{X}$ .

Допустите, что вам уже удалось решить эту проблему для некоторого  $n < N$ , т.е. выбрать распределение  $(n - 1)$ -подмножества  $\mathfrak{X}_{n-1} \subset \mathfrak{X}$ .

Тогда на очередном  $n$ -м этапе вам остается лишь “дополнить” его еще одним  $n$ -м событием, которое



обозначьте  $x \in \mathfrak{X} - \mathfrak{X}_{n-1}$ , и которое превратит это  $2^{n-1}$ -распределение в  $2^n$ -распределение, разбив каждое его террасное событие на два, одно из которых попадет в событие  $x$ , а другое в его дополнение  $x^c$ . И сделать эти разбиения надо так, чтобы вероятность добавляемого события  $x$  оказалась равной заранее фиксированному числу  $\mathbf{P}(x)$ .

### 3.1 Алгоритм в числовых обозначениях

Итак, на очередном этапе требуется “разбить (распределить) вероятностную массу”  $\mathbf{P}(x)$  на  $2^{n-1}$  частей так, чтобы каждая “уместилась” в соответствующую вероятность террасного события из  $2^{n-1}$  вероятностей, полученных на предыдущем этапе. Решение этой задачи всегда существует, так как сумма всех вероятностей террасных событий равна единице. Допустите, что вероятности таких террасных событий образуют набор

$$\{p_{n-1}^*(\gamma), \gamma \in \{0, \dots, 2^{n-1} - 1\}\} \quad (3.1.1)$$

— вероятностное  $2^{n-1}$ -распределение, полученное на  $(n-1)$ -м этапе.

Затем представьте, что в пределах каждого очередного этапа вы реализуете распределение “вероятностной массы”  $\mathbf{P}(x)$  на  $2^{n-1}$  частей *шаг за шагом* по определенному правилу. И представьте, что на текущем  $\gamma$ -шаге ( $\gamma = 0, \dots, 2^{n-1} - 2$ )<sup>2</sup> перед вами стоит задача распределить оставшуюся нераспределенной “массу”  $\pi^{(\gamma)} \in (0, \mathbf{P}(x)]$ , иначе говоря,

$$0 < \pi^{(\gamma)} \leq \mathbf{P}(x).$$

Введите следующие обозначения<sup>3</sup>

$$\sum_{j=0}^{\gamma-1} p_{n-1}^*(j) = S_{\gamma-1}, \quad \sum_{j=\gamma+1}^{2^{n-1}-1} p_{n-1}^*(j) = D_{\gamma+1},$$

в которых вероятностная нормировка распределения (3.1.1) принимает вид

$$S_{\gamma-1} + p_{n-1}^*(\gamma) + D_{\gamma+1} = 1,$$

и ясно, что

$$\pi^{(\gamma)} \leq p_{n-1}^*(\gamma) + D_{\gamma+1}.$$

Процедура распределения вероятностной массы  $\mathbf{P}(x)$  начинается с нулевого шага для  $\gamma = 0$  и  $\pi(0) = \mathbf{P}(x)$ . На текущем  $\gamma$ -шаге очередного этапа вам надо решить задачу распределения оставшейся вероятностной массы  $\pi^{(\gamma)}$  на две половины:

$$\pi^{(\gamma)} = \rho_\gamma + \delta_\gamma, \quad (3.1.2)$$

<sup>2</sup>Чтобы разбить “вероятностную массу”  $\mathbf{P}(x)$  на  $2^{n-1}$  частей требуется  $2^{n-1} - 1$  шаг.

<sup>3</sup>Здесь надо договориться, что  $S_{-1} = 0$  при начальном  $\gamma = 0$ , так как  $S_{-1}$  — это сумма пустого множества слагаемых.

которые “вмещаются в дваместилища”

$$R^{(\gamma)} = p_{n-1}^*(\gamma) \quad \text{и} \quad \Delta^{(\gamma)} = D_{\gamma+1} \quad (3.1.3)$$

соответственно. Вам стоит, конечно, понять, что выбор здесь определяется одним случайным числом  $r$ , например, определяющим величину  $\rho_\gamma$ , а вторая величина  $\delta_\gamma$  определится из ограничения на их сумму (3.1.2). Не трудно догадаться<sup>4</sup>, что решение задачи  $\rho_\gamma$  на  $\gamma$ -шаге дается случайным числом  $r_{n-1}(\gamma)$ , равномерным в интервале:

$$\max\{0, \pi^{(\gamma)} - \Delta^{(\gamma)}\} \leq r_{n-1}(\gamma) \leq \min\{\pi^{(\gamma)}, R^{(\gamma)}\}, \quad (3.1.4)$$

границы которого определяются соответствующими границами Фреше. При этом вероятность  $p_{n-1}^*(\gamma)$  с помощью числа  $r_{n-1}(\gamma)$  разбивается на две вероятности:

$$p_{n-1}^*(\gamma) = p_n(\gamma) + p_n(\gamma + 2^{n-1}), \quad (3.1.5)$$

где

$$\begin{cases} p_n(\gamma + 2^{n-1}) = p_{n-1}^*(\gamma) - r_{n-1}(\gamma), \\ p_n(\gamma) = r_{n-1}(\gamma). \end{cases} \quad (3.1.6)$$

На следующем  $(\gamma + 1)$ -м шаге очередного этапа решение  $\rho_{\gamma+1}$ , записанное в виде (3.1.4), также годится после соответствующего пересчета величин

$$\pi^{(\gamma+1)} = \pi^{(\gamma)} - \rho_\gamma,$$

$$\Delta^{(\gamma+1)} = \Delta^{(\gamma)} - p_{n-1}^*(\gamma + 1),$$

$$R^{(\gamma+1)} = p_{n-1}^*(\gamma + 1).$$

Завершив этот *пошаговый* процесс на  $n$ -м этапе, используя на каждом шаге формулы (3.1.5) и (3.1.6), вы получите  $2^n$ -распределение, чтобы затем этап за этапом дойти на  $N$ -м этапе до выбора  $2^N$ -распределения.

### 3.2 Алгоритм в сет-обозначениях

А теперь рассмотрите тот же алгоритм, но представленный в сет-обозначениях, чтобы убедиться в том, что, хотя, на первый взгляд, структура алгоритма не претерпевает существенных изменений, но сет-обозначения дают решающее эвентологическое преимущество — они позволяют событийно осмысленно интерпретировать каждый шаг алгоритма и получаемые результаты.

Итак, на очередном этапе требуется “разбить (распределить) вероятностную массу”  $\mathbf{P}(x)$  на  $2^{n-1}$  частей так, чтобы каждая “уместилась” в соответствующую вероятность террасного события  $\text{ter}(X/\mathfrak{X}_{n-1})$

<sup>4</sup>Изобразите на плоскости прямоугольник  $R^{(\gamma)} \times \Delta^{(\gamma)}$  и прямую, которая задается уравнением (3.1.2). Решение (3.1.3) станет сразу наглядным.

из полученных на предыдущем этапе. Допустите, что вероятности таких террасных событий, занумерованные подмножествами  $X \subseteq \mathfrak{X}_{n-1} \subseteq \mathfrak{X}$ , образуют набор

$$\left\{ \underbrace{p(\emptyset // \mathfrak{X}_{n-1}), \dots, p(X // \mathfrak{X}_{n-1}), \dots, p(\mathfrak{X}_{n-1} // \mathfrak{X}_{n-1})}_{2^{n-1}} \right\} = \\ = \left\{ p(X // \mathfrak{X}_{n-1}) : X \subseteq \mathfrak{X}_{n-1} \right\},$$

— вероятностное  $2^{n-1}$ -распределение множества событий  $\mathfrak{X}_{n-1}$ , полученное на  $(n-1)$ -м этапе. Вероятности из этого распределения с помощью “координатного” отображения, определенного в параграфе 1 на стр. 133, могут быть занумерованы числами  $\gamma \in \{0, \dots, 2^{n-1} - 1\}$  и записаны в виде набора

$$\left\{ p_{n-1}^*(\gamma(X)) : X \subseteq \mathfrak{X}_{n-1} \right\} = \\ = \left\{ p_{n-1}^*(\gamma) : \gamma \in \{0, \dots, 2^{n-1} - 1\} \right\}, \quad (3.2.1)$$

который совпадает с (3.1.1).

Затем представьте, что в пределах каждого очередного этапа вы реализуете распределение “вероятностной массы”  $\mathbf{P}(x)$  на  $2^{n-1}$  частей *шаг за шагом* по определенному правилу. И представьте, что на текущем  $\gamma$ -шаге ( $\gamma = 0, \dots, 2^{n-1} - 2$ )<sup>2</sup> перед вами стоит задача распределить оставшуюся нераспределенной “массу”  $\pi^{(\gamma)} \in (0, \mathbf{P}(x)]$ , иначе говоря,

$$0 < \pi^{(\gamma)} \leq \mathbf{P}(x).$$

Наступил момент, когда вы можете заняться событийным осмыслением происходящего на данном  $\gamma$ -шаге алгоритма. Обратное “координатное” отображение  $g^{-1}$  позволяет определить, что на  $\gamma$ -шаге стоит задача разбиения событием  $x_n \in \mathfrak{X} - \mathfrak{X}_{n-1}$  террасного события  $\text{ter}(g^{-1}(\gamma) // \mathfrak{X}_{n-1})$ , порожденного  $\mathfrak{X}_{n-1}$ , на два террасных события

$$\text{ter}(g^{-1}(\gamma) // \mathfrak{X}_{n-1}) = \\ = x_n \cap \text{ter}(g^{-1}(\gamma) // \mathfrak{X}_{n-1}) + x_n^c \cap \text{ter}(g^{-1}(\gamma) // \mathfrak{X}_{n-1}) = \\ = \text{ter}(g^{-1}(\gamma) + \{x_n\} // \mathfrak{X}_n) + \text{ter}(g^{-1}(\gamma) // \mathfrak{X}_n), \quad (3.2.2)$$

которые порождены  $\mathfrak{X}_n = \mathfrak{X}_{n-1} + \{x_n\}$ . Данное разбиение, естественно, определяет и соответствующее разбиение вероятности из распределения  $\mathfrak{X}_{n-1}$

$$p_{n-1}^*(\gamma) = p_{n-1}(g^{-1}(\gamma) // \mathfrak{X}_{n-1}) \\ \text{на две вероятности из распределения } \mathfrak{X}_n: \\ p_{n-1}(g^{-1}(\gamma) // \mathfrak{X}_{n-1}) = \\ = p_n(g^{-1}(\gamma) + \{x_n\} // \mathfrak{X}_n) + p_n(g^{-1}(\gamma) // \mathfrak{X}_n). \quad (3.2.3)$$

На решение этой задачи событийного разбиения наложены ограничения, аналогичные ограничениям из

числового алгоритма. Для их формулировки так же, как и в параграфе 3.1, введите похожие обозначения

$$\sum_{j=0}^{\gamma-1} p_{n-1}^*(j) = S_{\gamma-1}, \quad \sum_{j=\gamma+1}^{2^{n-1}-1} p_{n-1}^*(j) = D_{\gamma+1},$$

в которых вероятностная нормировка распределения (3.2.1) принимает вид

$$S_{\gamma-1} + p_{n-1}^*(\gamma) + D_{\gamma+1} = 1,$$

и ясно, что

$$\pi^{(\gamma)} \leq p_{n-1}^*(\gamma) + D_{\gamma+1}.$$

Тогда на текущем  $\gamma$ -шаге очередного этапа вам надо решить задачу распределения оставшейся вероятностной массы  $\pi^{(\gamma)}$  на две половины:

$$\pi^{(\gamma)} = \rho_\gamma + \delta_\gamma, \quad (3.2.4)$$

которые “вмещаются в два вместилища” вероятностей

$$R^{(\gamma)} = p_{n-1}^*(\gamma) \quad \text{и} \quad \Delta^{(\gamma)} = D_{\gamma+1} \quad (3.2.5)$$

соответственно.

Наступил второй момент, удобный для событийного осмысления.

Во-первых, двум “вместилищам” вероятностей (3.2.5) соответствуют два событийных “вместилища”:

- происходящее с вероятностью  $p_{n-1}^*(\gamma)$  террасное событие

$$\text{ter}(g^{-1}(\gamma) // \mathfrak{X}), \quad (3.2.6)$$

которое разбивается на  $\gamma$ -шаге, и

- происходящее с вероятностью  $D_{\gamma+1}$  объединение террасных событий

$$\sum_{\gamma'=\gamma+1}^{2^{n-1}-1} \text{ter}(g^{-1}(\gamma') // \mathfrak{X}), \quad (3.2.6')$$

до которых еще не дошло разбиение на две половины.

Во-вторых, надеясь, что вы не перепутали разбиение террасного события (3.2.2) на два террасных события с разбиением оставшейся части события  $x_n$  по двум событийным “вместилищам” (3.2.6) и (3.2.6'), предлагаю заметить тот факт, что

$$p_n(g^{-1}(\gamma) + \{x_n\} // \mathfrak{X}_n) = \rho_\gamma, \quad (3.2.7)$$

$$p_n(g^{-1}(\gamma) // \mathfrak{X}_n) = p_{n-1}^*(\gamma) - \rho_\gamma \quad (3.2.8)$$

— искомое решение  $\rho_\gamma$  определяет вероятности обоих террасных событий, порожденных  $\mathfrak{X}_n$ , на которые разбивается текущее террасное событие  $\text{ter}(g^{-1}(\gamma) // \mathfrak{X}_{n-1})$ , порожденное  $\mathfrak{X}_{n-1}$  (3.2.2).

Как и в числовом алгоритме, вам стоит еще раз понять, что выбор здесь определяется одним случайным числом  $r$ , например, определяющим искомую

величину  $\rho_\gamma$ , а вторая величина  $\delta_\gamma$  определится из ограничения на их сумму (3.2.2). Всё также не трудно догадаться, что решение задачи  $\rho_\gamma$  на  $\gamma$ -шаге дается случайным числом  $r_{n-1}(\gamma)$ , равномерным в интервале:

$$\max\{0, \pi^{(\gamma)} - \Delta^{(\gamma)}\} \leq r_{n-1}(\gamma) \leq \min\{\pi^{(\gamma)}, R^{(\gamma)}\}, \quad (3.2.9)$$

границы которого определяются соответствующими границами Фреше.

В третий и последний раз весьма кстати заняться событийным осмыслением.

Во-первых, случайное число  $r$ , равномерно выбираемое в ограничениях (3.2.9), определяет искомую величину  $\rho_\gamma$ , а вместе с ней вероятности двух террасных событий (3.2.7) и (3.2.8), порожденных  $\mathfrak{X}_n$ , на которые разбивается текущее террасное событие (3.2.2), порожденное  $\mathfrak{X}_n$ .

Во-вторых, границы Фреше<sup>5</sup>, которые были упомянуты выше, — это известные границы Фреше для вероятностей двух пересечений:

$$x_n \cap \text{ter}(g^{-1}(\gamma) // \mathfrak{X}_{n-1}) = \text{ter}(g^{-1}(\gamma) + \{x_n\} // \mathfrak{X}_n),$$

$$x_n \cap \sum_{\gamma'=\gamma+1}^{2^{n-1}} \text{ter}(g^{-1}(\gamma') // \mathfrak{X}),$$

совместное выполнение которых приводит к ограничениям (3.2.9).

Наконец, завершающее замечание сет-алгоритма, попросту повторяет замечание из его числового варианта. На следующем  $(\gamma+1)$ -м шаге очередного этапа решение  $\rho_{\gamma+1}$ , записанное в виде (3.2.9), также годится после соответствующего пересчета величин

$$\pi^{(\gamma+1)} = \pi^{(\gamma)} - \rho_\gamma,$$

$$\Delta^{(\gamma+1)} = \Delta^{(\gamma)} - p_{n-1}^*(\gamma+1),$$

$$R^{(\gamma+1)} = p_{n-1}^*(\gamma+1).$$

Завершив этот *пошаговый* процесс на  $n$ -м этапе, вы получите  $2^n$ -распределение, чтобы затем этап за этапом дойти на  $N$ -м этапе до выбора  $2^N$ -распределения.

В итоге, сет-алгоритм в отличие от числового алгоритма строит выбор не просто точки из единичного  $2^N$ -вершинного симплекса, но — точки, координаты которой нумеруются подмножествами событий  $X \subseteq \mathfrak{X}$  и тем самым могут быть интерпретированы как вероятности соответствующих террасных событий  $\text{ter}(X // \mathfrak{X})$ . Это решающее эвентологическое преимущество достигается за счет введения некоторого предварительного порядка на  $\mathfrak{X}$  и использования так

называемого “координатного” отображения, определяемого этим порядком. Разумеется, это преимущество критическим образом проявляется лишь в ситуации множеств, составленных из большого числа событий и требующих универсальных алгоритмов обработки. Для дуплетов и триплетов событий и даже немного более того можно с успехом обойтись и без помощи “координатного” отображения, что будет вам крайне полезно продемонстрировать ниже.

### 3.3 Пример без “координатного” отображения

**Дуплет.** Рассмотрите, не прибегая к услугам “координатного” отображения, пример выбора вероятностного распределения *сначала* дуплета событий  $\mathfrak{X}_2 = \{x, y\}$  из симплекса всех распределений дуплета

$$\mathbb{S}_{\mathfrak{X}} = \left\{ \{p(X) : X \subseteq \mathfrak{X}\} : p(X) \geq 0, \sum_{X \subseteq \mathfrak{X}_2} p(X) = 1 \right\},$$

если заданы вероятности двух событий из дуплета:

$$p(\{x\}) + p(\{x, y\}) = \mathbf{P}(x) = p_x,$$

$$p(\{y\}) + p(\{x, y\}) = \mathbf{P}(y) = p_y.$$

Изобразите  $\Omega$  в виде единичного отрезка  $[0, 1]$ , а вероятность  $\mathbf{P}$  представьте в виде лебеговой меры.

**1-й этап алгоритма** выбора не требует случайных чисел, так как он состоит в разбиении  $\Omega$  — единичного отрезка  $[0, 1]$  — на два террасных события:

$$[0, 1] = \text{ter}(\{x\} // \{x\}) + \text{ter}(\emptyset // \{x\}) = [0, p_x] + (p_x, 1],$$

вероятности которых равны соответственно

$$p(\{x\} // \{x\}) = p_x, \quad p(\emptyset // \{x\}) = 1 - p_x.$$

**2-й этап алгоритма** состоит из одного 1-го шага для которого параметры (3.2.3) имеют вид:

$$\pi^{(1)} = p_y, \quad p_1 = p_x, \quad R^{(1)} = p_x, \quad \Delta^{(1)} = 1 - p_x.$$

Выбор осуществляется выбором случайного числа  $r_1$ , равномерного на отрезке (3.2.9):

$$\max\{0, p_x + p_y - 1\} \leq r_1 \leq \min\{p_y, p_x\}, \quad (3.3.1)$$

границами которого служат границы Фреше для вероятности пересечения событий  $x, y \in \mathfrak{X}$ . Таким образом, выбор  $r_1$  в границах (3.3.1) равносителен выбору вероятности парного пересечения

$$p_{xy} = \mathbf{P}(x \cap y) = r_1.$$

После выбора вероятности парного пересечения событий дуплета  $\mathfrak{X}$  его распределение определено, и его

<sup>5</sup>Краткую справку о границах Фреше см. в параграфе 4.

можно изобразить разбиением отрезка  $[0, 1]$  на четыре части, каждая из которых “изображает” одно из четырех террасных событий:

$$\begin{aligned} [0, 1] &= [0, p_{xy}] + (p_{xy}, p_x] + \\ &+ (p_x, p_x + p_y - p_{xy}] + (p_x + p_y - p_{xy}, 1]. \end{aligned} \quad (3.3.2)$$

Эти части “изображают” на единичном отрезке террасные события:

$$\begin{aligned} [0, p_{xy}] &\sim \text{ter}(\{x, y\}) = x \cap y, \\ (p_{xy}, p_x] &\sim \text{ter}(\{x\}) = x \cap y^c, \\ (p_x, p_x + p_y - p_{xy}] &\sim \text{ter}(\{y\}) = x^c \cap y, \\ (p_x + p_y - p_{xy}, 1] &\sim \text{ter}(\emptyset) = x^c \cap y^c, \end{aligned}$$

образующие разбиение пространства всеобщих элементарных исходов

$$\Omega = \text{ter}(\{x, y\}) + \text{ter}(\{y\}) + \text{ter}(\{x\}) + \text{ter}(\emptyset), \quad (3.3.3)$$

которое в полных обозначениях имеет вид:

$$\begin{aligned} \Omega &= \text{ter}(\emptyset // \mathfrak{X}_2) + \text{ter}(\{x\} // \mathfrak{X}_2) + \\ &+ \text{ter}(\{y\} // \mathfrak{X}_2) + \text{ter}(\{x, y\} // \mathfrak{X}_2), \end{aligned} \quad (3.4.3')$$

и которому соответствует разбиение единичной “вероятностной массы”:

$$\begin{aligned} 1 &= p(\emptyset // \mathfrak{X}_2) + p(\{x\} // \mathfrak{X}_2) + \\ &+ p(\{y\} // \mathfrak{X}_2) + p(\{x, y\} // \mathfrak{X}_2), \end{aligned} \quad (3.4.3'')$$

или в сокращенных обозначениях:

$$1 = p_2(\emptyset) + p_2(x) + p_2(y) + p_2(xy). \quad (3.4.3''')$$

**От дуплета к триплету.** *Теперь*, считая, что точка в симплексе для дуплета  $\mathfrak{X}_2$  выбрана, перейдите к следующему этапу выбора точки в симплексе для триплета событий  $\mathfrak{X}_3 = \{x, y, z\}$ . Для этого можно воспользоваться полученным дуплетным результатом и при помощи того же алгоритма разбить вероятность  $p_z = \mathbf{P}(z)$  третьего события  $z \in \mathfrak{X}$  по четырем террасным событиям (3.3.3), которые порождены дуплетом  $\mathfrak{X}_2$ .

**3-й этап алгоритма** состоит из 3-х шагов, для первого из которых параметры (3.2.3) имеют вид:

$$\begin{aligned} \pi^{(1)} &= p_z, \quad p_1 = p_2(xy), \\ R^{(1)} &= p_2(xy), \quad \Delta^{(1)} = 1 - p_2(xy). \end{aligned}$$

**1-й шаг.** Выбор осуществляется выбором случайного числа  $r_1$ , равномерного на отрезке (3.2.9):

$$\begin{aligned} \max\{0, p_z + p_2(xy) - 1\} &\leq r_1, \\ r_1 &\leq \min\{p_z, p_2(xy)\}, \end{aligned} \quad (3.3.4)$$

границами которого служат границы Фреше для вероятности пересечения событий  $z$  и  $x \cap y$ . Таким образом, выбор  $r_1$  в границах (3.3.4) равносильен выбору вероятности тройного пересечения

$$p_3(xyz) = \mathbf{P}(x \cap y \cap z) = r_1. \quad (3.3.5)$$

**2-й шаг.** Пересчитайте параметры (3.2.3)

$$\begin{aligned} \pi^{(2)} &= p_z - p_3(xyz), \quad p_2 = p_2(x), \\ R^{(2)} &= p_2(x), \quad \Delta^{(2)} = 1 - p_2(xy) - p_2(x). \end{aligned}$$

Выбор осуществляется выбором случайного числа  $r_2$ , равномерного на отрезке (3.2.9):

$$\begin{aligned} \max\{0, p_z - p_3(xyz) + p_2(xy) - p_2(x) - 1\} &\leq r_2, \\ r_2 &\leq \min\{p_z - p_3(xyz), p_2(x)\}, \end{aligned} \quad (3.3.6)$$

границами которого служат границы Фреше для вероятности пересечения событий  $z - x \cap y \cap z$  и  $x \cap y^c$ . Таким образом, выбор  $r_2$  в границах (3.3.6) равносильен выбору вероятности тройного пересечения

$$p_3(xy^c z) = \mathbf{P}(x \cap y^c \cap z) = r_2. \quad (3.3.7)$$

**3-й шаг.** Пересчитайте параметры (3.2.3)

$$\begin{aligned} \pi^{(2)} &= p_z - p_3(xyz) - p_3(xy^c z), \quad p_2 = p_2(y), \\ R^{(2)} &= p_2(y), \quad \Delta^{(2)} = 1 - p_2(xy) - p_2(x) - p_2(y). \end{aligned}$$

Выбор осуществляется выбором случайного числа  $r_3$ , равномерного на отрезке (3.2.9):

$$\begin{aligned} \max\{0, p_z - p_3(xyz) - p_3(xy^c z) - p_3(x^c y^c)\} &\leq r_3, \\ r_3 &\leq \min\{p_z - p_3(xyz) - p_3(xy^c z), p_2(y)\}, \end{aligned} \quad (3.3.8)$$

границами которого служат границы Фреше для вероятности пересечения событий  $z \cap x^c$  и  $x^c \cap y$ . Таким образом, выбор  $r_3$  в границах (3.3.8) равносильен выбору вероятности тройного пересечения

$$p_3(xy^c z) = \mathbf{P}(x^c \cap y \cap z) = r_3. \quad (3.3.9)$$

Оставшаяся четвертая часть события  $z$ , т.е. событие  $x^c \cap y^c \cap z$  получает оставшуюся часть вероятностной массы

$$p_3(x^c y^c z) = p_z - r_1 - r_2 - r_3. \quad (3.3.10)$$

При этом вероятности остальных четырех террасных событий триплета  $\mathfrak{X}_3 = \{x, y, z\}$  также определены очевидными “дополнительными” формулами:

$$p_3(xyz^c) = p_2(xy) - r_1, \quad (3.3.11)$$

$$p_3(xy^c z^c) = p_2(x) - r_2, \quad (3.3.12)$$

$$p_3(xy^c z^c) = p_2(y) - r_3. \quad (3.3.13)$$

$$p_3(x^c y^c z^c) = p_2(\emptyset) - (\mathbf{P}(z) - r_1 - r_2 - r_3). \quad (3.3.14)$$

В итоге на 3-м этапе из многогранника Фреше выбрано распределение триплета событий  $\mathfrak{X}_3 = \{x, y, z\}$ , восемь вероятностей которого определяются формулами (3.3.5, 7, 9, 10–14).

Этот рекурсивный процесс может быть продолжен без услуг “координатного” отображения до  $N$ -множества событий  $\mathfrak{X}_N$ , у всех событий  $x \in \mathfrak{X}_N$  которого заданы вероятности  $\mathbf{P}(x)$  и мощность которого приемлема для уникальной *ручной* обработки.

### 3.4 Пример с “координатным” отображением

**Дуплет.** Рассмотрим пример использования “координатного” отображения  $g$  для выбора вероятностного распределения *сначала* дуплета событий  $\mathfrak{X}_2 = \{x_2, x_1\}$  из симплекса всех распределений дуплета

$$\mathbb{S}_{\mathfrak{X}} = \left\{ \{p(X) : X \subseteq \mathfrak{X}\} : p(X) \geq 0, \sum_{X \subseteq \mathfrak{X}_2} p(X) = 1 \right\},$$

если заданы вероятности двух событий из дуплета:

$$p(\{x_1\}) + p(\{x_2, x_1\}) = \mathbf{P}(x_1) = p_{x_1},$$

$$p(\{x_2\}) + p(\{x_2, x_1\}) = \mathbf{P}(x_2) = p_{x_2}.$$

Изобразите  $\Omega$  в виде единичного отрезка  $[0, 1]$ , а вероятность  $\mathbf{P}$  представьте в виде лебеговой меры.

**1-й этап алгоритма** выбора не требует случайных чисел, так как он состоит в разбиении  $\Omega$  — единичного отрезка  $[0, 1]$  — на два террасных события:

$$\begin{aligned} [0, 1] &= \text{ter}(\emptyset // \{x_1\}) + \text{ter}(\{x_1\} // \{x_1\}) = \\ &= [0, 1 - p_{x_1}] + (1 - p_{x_1}, 1], \end{aligned}$$

которому соответствует вероятностное распределение

$$\{p(\emptyset // \{x_1\}), p(\{x_1\} // \{x_1\})\} = \{1 - p_{x_1}, p_{x_1}\},$$

записываемое в “координатной” форме

$$\{p^*(1), p^*(2)\} = \{1 - p_{x_1}, p_{x_1}\}.$$

**2-й этап алгоритма** состоит из одного 1-го шага для которого параметры (3.2.3) имеют вид:

$$\pi^{(1)} = p_{x_2}, \quad R^{(1)} = p^*(1), \quad \Delta^{(1)} = 1 - p^*(1).$$

Выбор осуществляется выбором случайного числа  $r_1$ , равномерного на отрезке (3.2.9):

$$\max\{0, p_{x_2} + p^*(1) - 1\} \leq r_1 \leq \min\{p_{x_2}, p^*(1)\}, \quad (3.4.1)$$

границами которого служат границы Фреше для вероятности пересечения событий  $x_1^c$  и  $x_2$ . Таким образом, выбор  $r_1$  в границах (3.4.1) равносильно выбору вероятности парного пересечения

$$p(x_2 x_1^c) = \mathbf{P}(x_2 \cap x_1^c) = r_1.$$

После выбора вероятности парного пересечения событий дуплета  $\mathfrak{X}$  его распределение определено, и его можно изобразить разбиением отрезка  $[0, 1]$  на четыре части, каждая из которых “изображает” одно из четырех террасных событий:

$$\begin{aligned} [0, 1] &= [0, p(x_2 x_1^c)] + (p(x_2 x_1^c), 1 - p_{x_1}] + \\ &+ (1 - p_{x_1}, p_{x_1} + p_{x_2} - p(x_2 x_1^c)] + \\ &+ (1 - p_{x_1} - p_{x_2} - p(x_2 x_1^c), 1]. \end{aligned} \quad (3.4.2)$$

Эти части “изображают” на единичном отрезке террасные события:

$$\begin{aligned} [0, p(x_2 x_1^c)] &\sim \text{ter}(\{x_2\}) = x_2 \cap x_1^c, \\ (p(x_2 x_1^c), 1 - p_{x_1}] &\sim \text{ter}(\emptyset) = x_2^c \cap x_1^c, \\ (1 - p_{x_1}, p_{x_1} + p_{x_2} - p(x_2 x_1^c)] &\sim \text{ter}(\{x_2, x_1\}) = x_2 \cap x_1, \\ (1 - p_{x_1} - p_{x_2} + p(x_2 x_1^c), 1] &\sim \text{ter}(\{x_1\}) = x_2^c \cap x_1, \end{aligned}$$

образующие разбиение пространства всеобщих элементарных исходов

$$\begin{aligned} \Omega &= \text{ter}(\emptyset) + \text{ter}(\{x_2\}) + \\ &+ \text{ter}(\{x_1\}) + \text{ter}(\{x_2, x_1\}), \end{aligned} \quad (3.4.3)$$

которое в полных обозначениях имеет вид:

$$\begin{aligned} \Omega &= \text{ter}(\emptyset // \mathfrak{X}_2) + \text{ter}(\{x_2\} // \mathfrak{X}_2) + \\ &+ \text{ter}(\{x_1\} // \mathfrak{X}_2) + \text{ter}(\{x_2, x_1\} // \mathfrak{X}_2), \end{aligned} \quad (3.4.3')$$

и которому соответствует разбиение единичной “вероятностной массы”:

$$\begin{aligned} 1 &= p(\emptyset // \mathfrak{X}_2) + p(\{x_2\} // \mathfrak{X}_2) + \\ &+ p(\{x_1\} // \mathfrak{X}_2) + p(\{x_2, x_1\} // \mathfrak{X}_2), \end{aligned} \quad (3.4.3'')$$

или в сокращенных обозначениях:

$$1 = p_2(\emptyset) + p_2(x_2) + p_2(x_1) + p_2(x_2 x_1) \quad (3.4.3''')$$

и, наконец, в “координатной” форме:

$$1 = p_2^*(1) + p_2^*(2) + p_2^*(3) + p_2^*(4). \quad (3.4.4)$$

**От дуплета к триплету.** Теперь, считая, что точка в симплексе для дуплета  $\mathfrak{X}_2$  выбрана с координатами

$$(p_2^*(1), p_2^*(2), p_2^*(3), p_2^*(4)), \quad (3.4.5)$$

перейдите к следующему этапу выбора точки в симплексе для триплета событий

$$\mathfrak{X}_3 = \{x_3\} + \mathfrak{X}_2 = \{x_3, x_2, x_1\}$$



при помощи “координатного” отображения\*. Для этого можно воспользоваться полученным дуплетным результатом (3.4.5) и при помощи стандартного алгоритма разбить вероятность  $p_{x_3} = \mathbf{P}(x_3)$  третьего события  $x_3 \in \mathfrak{X}_3$  по четырем террасным событиям (3.4.3), которые порождены дуплетом  $\mathfrak{X}_2$ .

**3-й этап алгоритма** состоит из 3-х шагов, для первого из которых параметры (3.2.3) имеют вид:

$$\pi^{(1)} = p_{x_3}, \quad R^{(1)} = p_2^*(1), \quad \Delta^{(1)} = 1 - p_2^*(1).$$

**1-й шаг.** Выбор осуществляется выбором случайного числа  $r_1$ , равномерного на отрезке (3.2.9):

$$\max\{0, p_{x_3} + p_2^*(1) - 1\} \leq r_1, \\ r_1 \leq \min\{p_{x_3}, p_2^*(1)\}, \quad (3.4.4)$$

границами которого служат границы Фреше для вероятности пересечения событий  $x_3$  и  $x_2^c \cap x_1^c$ , т.е. вероятности террасного события

$$\text{ter}(\{x_3\}) = x_3 \cap x_2^c \cap x_1^c,$$

которое нумеруется соответствующим значением координатного отображения:

$$g(\{x_3\}) = 4.$$

Таким образом, выбор  $r_1$  в границах (3.4.4) равносильно выбору вероятности тройного пересечения

$$p_3^*(5) = p_3^*(g\{x_3\} + 1) = \\ = \mathbf{P}(x_3 \cap x_2^c \cap x_1^c) = r_1, \quad (3.4.5)$$

которая определяет еще одну, “дополнительную”, вероятность тройного пересечения

$$p_3^*(1) = p_3^*(g\{\emptyset\} + 1) = \\ = \mathbf{P}(x_3^c \cap x_2^c \cap x_1^c) = p_2^*(1) - r_1, \quad (3.4.5')$$

где  $g\{\emptyset\} = 0$  — номер террасного события

$$\text{ter}(\emptyset) = x_3^c \cap x_2^c \cap x_1^c.$$

**2-й шаг.** Пересчитайте параметры (3.2.3)

$$\pi^{(2)} = p_{x_3} - p_3^*(5), \\ R^{(2)} = p_2^*(2), \quad \Delta^{(2)} = 1 - p_2^*(1) - p_2^*(2).$$

Выбор осуществляется выбором случайного числа  $r_2$ , равномерного на отрезке (3.2.9):

$$\max\{0, p_{x_3} - p_3^*(5) + p_2^*(1) + p_2^*(2) - 1\} \leq r_2, \\ r_2 \leq \min\{p_{x_3} - p_3^*(5), p_2^*(2)\}, \quad (3.4.6)$$

границами которого служат границы Фреше для вероятности пересечения событий  $x_3 - x_3 \cap x_2^c \cap x_1^c$  и  $x_2^c \cap x_1^c$ . Таким образом, выбор  $r_2$  в границах (3.4.6)

равносильно выбору вероятности тройного пересечения

$$p_3^*(6) = p_3^*(g\{x_3, x_1\} + 1) = \\ = \mathbf{P}(x_3 \cap x_2^c \cap x_1) = r_2. \quad (3.4.7)$$

которая определяет еще одну, “дополнительную”, вероятность тройного пересечения

$$p_3^*(2) = p_3^*(g\{x_1\} + 1) = \\ = \mathbf{P}(x_3^c \cap x_2^c \cap x_1) = p_2^*(2) - r_2, \quad (3.4.7')$$

где  $g\{x_1\} = 0$  — номер террасного события

$$\text{ter}(\{x_1\}) = x_3^c \cap x_2^c \cap x_1.$$

**3-й шаг.** Пересчитайте параметры (3.2.3)

$$\pi^{(2)} = p_z - p_3(xyz) - p_3(xy^c z), \quad p_2 = p_2(y), \\ R^{(2)} = p_2(y), \quad \Delta^{(2)} = 1 - p_2(xy) - p_2(x) - p_2(y).$$

Выбор осуществляется выбором случайного числа  $r_3$ , равномерного на отрезке (3.2.9):

$$\max\{0, p_z - p_3(xyz) - p_3(xy^c z) - p_3(x^c y^c)\} \leq r_3, \\ r_3 \leq \min\{p_z - p_3(xyz) - p_3(xy^c z), p_2(y)\}, \quad (3.4.8)$$

границами которого служат границы Фреше для вероятности пересечения событий  $z \cap x^c$  и  $x^c \cap y$ . Таким образом, выбор  $r_3$  в границах (3.4.8) равносильно выбору вероятности тройного пересечения

$$p_3(xy^c z) = \mathbf{P}(x^c \cap y \cap z) = r_3. \quad (3.4.9)$$

Оставшаяся четвертая часть события  $z$ , т.е. событие  $x^c \cap y^c \cap z$  получает оставшуюся часть вероятностной массы

$$p_3(x^c y^c z) = p_z - r_1 - r_2 - r_3. \quad (3.4.10)$$

При этом вероятности остальных четырех террасных события триплета  $\mathfrak{X}_3 = \{x, y, z\}$  также определены очевидными “дополнительными” формулами:

$$p_3(xyz^c) = p_2(xy) - r_1, \quad (3.4.11)$$

$$p_3(xy^c z^c) = p_2(x) - r_2, \quad (3.4.12)$$

$$p_3(xy^c z^c) = p_2(y) - r_3. \quad (3.4.13)$$

$$p_3(x^c y^c z^c) = p_2(\emptyset) - (\mathbf{P}(z) - r_1 - r_2 - r_3). \quad (3.4.14)$$

В итоге на 3-м этапе из многогранника Фреше выбрано распределение триплета событий  $\mathfrak{X}_3 = \{x, y, z\}$ , восемь вероятностей которого определяются формулами (3.4.5, 7, 9, 10–14).

Этот рекурсивный процесс может быть продолжен до  $N$ -множества событий  $\mathfrak{X}_N$ , у всех событий  $x \in \mathfrak{X}_N$  которого заданы вероятности  $\mathbf{P}(x)$  и мощность которого приемлема для универсальной компьютерной обработки.

#### 4 Справка: границы и неравенства Фреше

В теории вероятности и логике *неравенства Фреше*, также известные как неравенства Буля-Фреше, неявно использованные еще в работе Джорджа Буля [3] и явно полученные Морисом Фреше [4, 5], — это неравенства, которые ограничивают вероятности комбинаций логических высказываний или событий, связанных между собой в пересечениях (И-операциях) или объединениях (ИЛИ-операциях); используются в теории вероятностей, математической статистике и вероятностной логике; распространены в оценке рисков, инженерном проектировании и искусственном интеллекте. Эти неравенства можно рассматривать как правила, ограничивающие расчеты с вероятностями без предположения независимости событий или, по сути, без каких бы-то ни было предположений зависимости событий. Фреше-неравенства тесно связаны с *неравенствами Буля-Бонферрони-Фреше*, и *границами Фреше*.

В эвентологии *неравенства и границы Фреше* используются в определении *Фреше-корреляции* множества событий и *Фреше-многогранника* вероятностных распределений множеств событий [1, 2] с заданными вероятностями событий, а также в алгоритмах выбора из *Фреше-многогранника* распределений множества событий с заданными вероятностями событий.

#### Список литературы

- [1] О. Ю. Воробьев. Фреше-граничные эвентологические распределения и их применения. *Труды XIV Международ. ЭМ конференции по эвентологической математике и смежным вопросам*, Красноярск: КГТЭИ, СФУ (под ред. Олега Воробьева):57–69, 2010.
- [2] O. Yu. Vorobyev. Fréchet boundary eventological distributions and its applications. In. *Proc. of the XV Intern. EM Conf. on Eventological Mathematics*, Krasnoyarsk, SFU (Oleg Vorobyev ed.):57–69, 2010 (in Russian, abstract in English).
- [3] G. Boole. *An Investigation of the Laws of Thought, On Which Are Founded the Mathematical Theories of Logic and Probability*. Walton and Maberly, London, 1854, p.299.
- [4] M. Fréchet. Généralisations du théorème des probabilités totales. *Fundamenta Mathematica*, 25:379–387, 1935.
- [5] M. Fréchet. Sur les tableaux de corrélation dont les marges sont données. *Annales de l'Université de Lyon. Section A: Sciences mathématiques et astronomie*, 9:53–77, 1951.

# Измерение многомерных событийных зависимостей в эвентологическом портфельном анализе

*финансово-актуарных, гуманитарных и социо-экономических рынков*

**Олег Юрьевич Воробьев**

Сибирский федеральный университет  
Институт математики и фундаментальной информатики  
Торгово-экономический институт  
Красноярск  
oleg.yu.vorobyev@gmail.com

**Аннотация.** Рассматриваются новые эвентологические методы измерения многомерной событийной зависимости в портфельном анализе финансово-актуарных, гуманитарных и социо-экономических рынков. Вводится новое для портфельного анализа понятие ковариационной пули, которая характеризует структуру многомерных зависимостей портфеля рыночных событий.

**Ключевые слова.** Эвентология, теория вероятностей, элементарный исход, событие, вероятность, множество событий, алгебра событий, многомерная зависимость событий, ковариационная мера многомерной зависимости, сет-рассеяние, эвентологический портфельный анализ, портфельное событие, структура многомерных зависимостей портфеля событий.

“Портфель. Портфель”,  
— два-три раза прочел он.  
Борис Пастернак,  
“Доктор Живаго”, 1957.

Не только на фондовых рынках хорошо известен факт снижения риска портфеля ценных бумаг за счет отрицательных ковариационных связей между портфельными компонентами. Это происходит из-за того, что отрицательные ковариационные связи между компонентами способствуют взаимной компенсации потерь и доходов. Однако при росте числа портфельных компонент, сложность структуры статистических взаимозависимостей в портфеле возрастает экспоненциально. Поэтому анализ структуры многомерных взаимозависимостей компонент портфеля в целом, вряд ли, кем-то ставился в качестве практической задачи. Классическая постановка прямой задачи Марковица ограничивается, к примеру, только парными статистическими взаимодействиями.

Эвентологическая теория событийных взаимодействий в ее нынешнем виде располагает изрядным арсеналом инструментов измерения зависимостей событий без ограничений на размерность статистиче-

ского взаимодействия. Разумеется, собственными измерителями зависимостей любой размерности располагает любая классическая теория, но эвентологические измерители многомерных зависимостей событий, по-видимому, наиболее просты и прозрачны из всех имеющихся в распоряжении современных статистиков.

## 1 Меры рассеяния множества событий

В нашей работе [1, 2, 2011] введены меры рассеяния множества событий  $\mathfrak{X} \subset \mathcal{A}$ , которые характеризуют структуру многомерных взаимозависимостей этого множества событий в целом, и показано, что эти меры обладают некоторыми характерными аксиоматическими свойствами расстояния.

Одна из этих мер, сет-рассеяние  $\mathfrak{X}$ , определяемая как полная ковариация  $\mathfrak{X}$  [1, 2]:

$$S_{\mathfrak{X}} = \sum_{X \subseteq \mathfrak{X}} |X| \text{Kov}_X, \quad (1.1)$$

по сути, является ковариационной мерой структуры многомерных зависимостей всего множества событий, так как составлена из ковариаций (II-го рода) [3, 4] всех подмножеств событий  $X \subseteq \mathfrak{X}$ .

Сет-рассеяние  $\mathfrak{X}$  меньше нуля, если события из  $\mathfrak{X}$  в целом “статистически отталкиваются”; больше нуля, если — “притягиваются”; и равна нулю, если статистические “отталкивание” и “притягивание” событий находятся в “ковариационном равновесии”.

Помимо того, что в эвентологической теории так определенное сет-рассеяние множества событий характеризует несколько события “рассыпаны” по пространству всеобщих элементарных исходов  $\Omega$  в сравнении с тотально независимыми событиями, происходящими с теми же вероятностями [1, 2], в портфельном анализе событий оно обобщает “ковариационный довесок” в дисперсии доходности портфеля событий<sup>1</sup>, который возникает для зависимых порт-

<sup>1</sup>В простейшем варианте доходность событийного портфе-

фельных событий:

$$\begin{aligned} D1_{\mathfrak{X}} &= D \left( \sum_{x \in \mathfrak{X}} 1_x \right) = \\ &= \sum_{x \in \mathfrak{X}} D1_x + 2 \sum_{\{x, y\} \subseteq \mathfrak{X}} \text{Kov}_{xy}, \end{aligned} \quad (1.2)$$

дополняя его ковариациями всех порядков без исключения, а не ограничиваясь только вторым порядком зависимости, который имеют дисперсия и парная ковариация.

Сет-рассеяние (1.1) обобщается до *взвешенного сет-рассеяния*

$$S_{\langle \mathfrak{X}, \mathbf{a} \rangle} = \sum_{X \subseteq \mathfrak{X}} a_X |X| \text{Kov}_X, \quad (1.3)$$

введением весов подмножеств событий  $X \subseteq \mathfrak{X}$ , определяемых набором неотрицательных весов событий  $\mathbf{a} = \{a_x, x \in \mathfrak{X}\}$ , по формуле

$$a_X = \prod_{x \in X} a_x, \quad (1.4)$$

которые соответствуют долям портфельных событий в эвентологическом портфельном анализе<sup>2</sup>.

## 2 Справка: свойства ковариаций и ковариационный вес события

Как известно, вероятностные распределения первого и второго рода множества событий  $\mathfrak{X}$  связаны между собой формулами обращения Мёбиуса:

$$p(X//\mathfrak{X}) = \sum_{Y \subseteq X} (-1)^{|X-Y|} p_{Y//\mathfrak{X}}, \quad (2.1)$$

$$p_{X//\mathfrak{X}} = \sum_{Y \subseteq X} p(Y//\mathfrak{X}). \quad (2.2)$$

Ковариации первого и второго рода определяются формулами для  $X \subseteq \mathfrak{X}$

$$\text{Kov}(X//\mathfrak{X}) = p(X//\mathfrak{X}) - p^i(X//\mathfrak{X}), \quad (2.3)$$

$$\text{Kov}_{X//\mathfrak{X}} = p_{X//\mathfrak{X}} - p_{X//\mathfrak{X}}^i, \quad (2.4)$$

где

$$p^i(X//\mathfrak{X}) = \prod_{x \in X} p_x \prod_{x \in \mathfrak{X}-X} (1 - p_x), \quad (2.5)$$

$$p_{X//\mathfrak{X}}^i = \prod_{x \in X} p_x \quad (2.6)$$

<sup>2</sup>Если потребовать стандартную нормировку:  $\sum_{x \in \mathfrak{X}} a_x = 1$ , что, впрочем, не принципиально.

— соответствующие вероятностные распределения множества событий  $\mathfrak{X}$ , в предположении, что все его события наступают с прежними вероятностями, но тотально независимы.

Из формул обращения Мёбиуса для распределений первого и второго рода следуют аналогичные формулы обращения для ковариаций соответствующего рода:

$$\text{Kov}_{X//\mathfrak{X}} = \sum_{Y \subseteq X} \text{Kov}(Y//\mathfrak{X}), \quad (2.7)$$

$$\text{Kov}(X//\mathfrak{X}) = \sum_{Y \subseteq X} (-1)^{|X-Y|} \text{Kov}_{Y//\mathfrak{X}}. \quad (2.8)$$

**Определение (ковариационный вес события).** Подставив (2.7) в (1.2), получим

$$\begin{aligned} S_{\mathfrak{X}} &= \sum_{X \subseteq \mathfrak{X}} a_X |X| \sum_{Y \subseteq X} \text{Kov}(Y) = \\ &= \sum_{Y \subseteq \mathfrak{X}} \text{Kov}(Y) \sum_{X \subseteq Y} a_X |X| = \\ &= \sum_{x \in \mathfrak{X}} \frac{a_x}{1 + a_x} \sum_{x \in Y \subseteq \mathfrak{X}} \text{Kov}(Y) (1 + a)_Y = \\ &= \sum_{x \in \mathfrak{X}} \frac{a_x}{1 + a_x} K_a(x), \end{aligned} \quad (2.9)$$

— формулу, выражающую взвешенное сет-рассеяние  $\mathfrak{X}$  в виде линейной комбинации *ковариационных а-весов событий* из  $\mathfrak{X}$ , где

$$K_a(x) = \sum_{x \in Y \subseteq \mathfrak{X}} \text{Kov}(Y) (1 + a)_Y \quad (2.10)$$

— *ковариационный а-вес события*  $x \in \mathfrak{X}$ , а

$$(1 + a)_Y = \prod_{x \in Y} (1 + a_x). \quad (2.11)$$

**Замечание 1.** Взвешенное сет-рассеяние (1.2) превращается в сет-рассеяние (1.1), если веса всех событий равны единице:  $a_x = 1, x \in \mathfrak{X}$ . В этой ситуации ковариационный вес каждого события  $x \in \mathfrak{X}$  (2.10) равен

$$K(x) = \sum_{x \in Y \subseteq \mathfrak{X}} \text{Kov}(Y) 2^{|Y|}, \quad (2.12)$$

а сет-рассеяние равно

$$S_{\mathfrak{X}} = \frac{1}{2} \sum_{x \in \mathfrak{X}} K(x). \quad (2.13)$$

**Замечание 2.** Отмечу еще, что по определению ковариации второго рода подмножеств событий  $X \subseteq \mathfrak{X}$  ковариации пустого подмножества  $\emptyset \subseteq \mathfrak{X}$  и каждого событийного моноплета  $\{x\} \subseteq \mathfrak{X}$  для  $x \in \mathfrak{X}$  равны нулю:

$$\text{Kov}_{\emptyset} = p_{\emptyset} - p_{\emptyset}^i = 0,$$

$$\text{Kov}_{\{x\}} = p_x - p_x^i = 0.$$

Поэтому ненулевыми могут быть только те слагаемые, например, в суммах (2.1), (2.2), содержащих такие ковариации, которые “занумерованы” дуплетами событий и выше.

**Замечание 3.** Для тотально независимого множества событий  $\mathfrak{X}$  ковариационные веса (2.10) и (2.12) всех событий  $x \in \mathfrak{X}$ , разумеется, равны нулю:

$$K_a(x) = K(x) = 0,$$

так как для всех  $X \subseteq \mathfrak{X}$

$$\text{Kov}(X) = 0$$

— равны нулю ковариации первого<sup>3</sup> рода.

## 2.1 Примеры измерения структур зависимостей сет-рассеянием

Рассмотрим триплеты событий  $\mathfrak{X} = \{x, y, z\}$  с различными структурами взаимозависимостей и значения их сет-рассеяния.

**Крайне отрицательная ковариационная зависимость.** Пусть сначала все события из  $\mathfrak{X}$  происходят с одинаковой вероятностью  $p \leq 1/3$  и не пересекаются, т.е. статистически “отталкиваются”. Это крайняя отрицательная ковариационная зависимость. Вычислим ковариационные веса событий в этой ситуации. Сразу ясно, что они в силу симметрии одинаковы. Вычислим следующие<sup>4</sup> ковариации первого рода<sup>5</sup>:

$$\text{Kov}(x) = \text{Kov}(y) = \text{Kov}(z) = p - p(1 - p)^2;$$

$$\text{Kov}(x, y) = \text{Kov}(x, z) = \text{Kov}(y, z) = -p^2(1 - p);$$

$$\text{Kov}(x, y, z) = -p^3;$$

Отсюда и из (2.12) получаем для  $p \leq 1/3$

$$\begin{aligned} K(x) &= [p - p(1 - p)^2]2 - p^2(1 - p)2^2 - p^32^3 = \\ &= 2p[1 - (1 - p)^2] - 4p^2[(1 - p)^2 + 2p] = \\ &= 2p^2[2 - p] - 4p^2[1 + p^2] = -2p^3[1 + 2p] =^6 \\ &= -2/27[1 + 2/3] = -10/81. \end{aligned}$$

Сет-рассеяние всего этого триплета по формуле (2.13) равно

$$S_{\mathfrak{X}} = -\frac{5}{27}.$$

**Крайне положительная ковариационная зависимость.** Пусть теперь все события из  $\mathfrak{X}$  совпадают

<sup>3</sup>Впрочем и второго рода также.

<sup>4</sup>Кроме  $\text{Kov}(\emptyset)$ , которая нам не нужна, так как в (2.10) и (2.12) суммирование начинается с событийных моноплетов.

<sup>5</sup>В сокращенных обозначениях:  $\text{Kov}(x) = \text{Kov}(\{x\})$ ,  $\text{Kov}(x, y) = \text{Kov}(\{x, y\})$ ,  $\text{Kov}(x, y, z) = \text{Kov}(\{x, y, z\})$ .

<sup>6</sup>При  $p = 1/3$ .

и происходят с вероятностью  $p$ . Вычислим те же ковариации первого рода:

$$\text{Kov}(x) = \text{Kov}(y) = \text{Kov}(z) = -p(1 - p)^2;$$

$$\text{Kov}(x, y) = \text{Kov}(x, z) = \text{Kov}(y, z) = -p^2(1 - p);$$

$$\text{Kov}(x, y, z) = p - p^3.$$

Отсюда и из (2.12) получаем

$$\begin{aligned} K(x) &= -2p(1 - p)^2 - 4p^2(1 - p) + 8p[1 - p^2] = \\ &= 6p(1 - p^2) =^7 = 16/9. \end{aligned}$$

Сет-рассеяние всего этого триплета по формуле (2.13) равно

$$S_{\mathfrak{X}} = \frac{8}{3}.$$

**Более отрицательная ковариационная зависимость.** Представьте два вложенных события  $x, y \in \mathfrak{X}$ , происходящих с вероятностями  $1/2$  и  $1/3$  соответственно, которые оба не пересекаются с третьим событием  $z \in \mathfrak{X}$ , происходящим с вероятностью  $1/3$ . Вычислим ковариации первого рода:

$$\text{Kov}(x) = 1/6 - 1/2 \cdot 2/3 \cdot 2/3 = -1/18;$$

$$\text{Kov}(y) = 0 - 1/3 \cdot 1/2 \cdot 2/3 = -2/18;$$

$$\text{Kov}(z) = 1/3 - 1/3 \cdot 1/2 \cdot 2/3 = 4/18;$$

$$\text{Kov}(x, y) = 1/3 - 1/3 \cdot 1/2 \cdot 2/3 = 4/18;$$

$$\text{Kov}(x, z) = -1/2 \cdot 1/3 \cdot 2/3 = -2/18;$$

$$\text{Kov}(y, z) = -1/3 \cdot 1/3 \cdot 1/2 = -1/18;$$

$$\text{Kov}(x, y, z) = -1/3 \cdot 1/3 \cdot 1/2 = -1/18;$$

Отсюда и из (2.12) получаем

$$K(x) = -2/18 + 16/18 - 8/18 - 8/18 = -2/18;$$

$$K(y) = -4/18 + 16/18 - 4/18 - 8/18 = 0;$$

$$K(z) = 8/18 - 8/18 - 4/18 - 8/18 = -12/18;$$

Сет-рассеяние всего этого триплета по формуле (2.13) равно

$$S_{\mathfrak{X}} = -7/18.$$

**Ковариационно-равновесная зависимость.** Представьте два непересекающихся события  $y, z \in \mathfrak{X}$ , происходящих с одинаковой вероятностью  $1/3$ , которые оба вложены в третье событие  $x \in \mathfrak{X}$ , происходящее с вероятностью  $2/3$ . Иначе говоря, события  $y, z$  разбивают событие  $x$  на два равновероятных события. Вычислим ковариации первого рода:

$$\text{Kov}(x) = -2/3 \cdot 2/3 \cdot 2/3 = -8/27;$$

$$\text{Kov}(y) = -1/3 \cdot 2/3 \cdot 1/3 = -2/27;$$

<sup>7</sup>При той же вероятности  $p = 1/3$ .



$$\text{Kov}(z) = -1/3 \cdot 2/3 \cdot 1/3 = -2/27;$$

$$\text{Kov}(x, y) = 1/3 - 1/3 \cdot 2/3 \cdot 2/3 = 5/27;$$

$$\text{Kov}(x, z) = 1/3 - 1/3 \cdot 2/3 \cdot 2/3 = 5/27;$$

$$\text{Kov}(y, z) = -1/3 \cdot 1/3 \cdot 1/3 = -1/27;$$

$$\text{Kov}(x, y, z) = -1/3 \cdot 1/3 \cdot 2/3 = -2/27;$$

Отсюда и из (2.12) получаем

$$K(x) = -16/27 + 20/27 + 20/27 - 16/27 = 8/27;$$

$$K(y) = -4/27 + 20/27 - 4/27 - 16/27 = -4/27;$$

$$K(z) = -4/27 + 20/27 - 4/27 - 16/27 = -4/27;$$

Сет-рассеяние всего этого триплета по формуле (2.13) равно

$$\mathbf{S}_{\mathfrak{X}} = 0.$$

**Более положительная ковариационная зависимость.** Представьте три пересекающихся события  $x, y, z \in \mathfrak{X}$ , происходящих с одинаковой вероятностью  $1/2$ , которые никогда не происходят в одиночку,  $\mathfrak{X}$ -строго по два<sup>8</sup> наступают с вероятностью  $1/5$ , а втроем происходят с вероятностью  $1/10$ . Вычислим ковариации первого рода :

$$\text{Kov}(x) = -1/2 \cdot 1/2 \cdot 1/2 = -1/8;$$

$$\text{Kov}(y) = -1/2 \cdot 1/2 \cdot 1/2 = -1/8;$$

$$\text{Kov}(z) = -1/2 \cdot 1/2 \cdot 1/2 = -1/8;$$

$$\text{Kov}(x, y) = 1/5 - 1/2 \cdot 1/2 \cdot 1/2 = 3/40;$$

$$\text{Kov}(x, z) = 1/5 - 1/2 \cdot 1/2 \cdot 1/2 = 3/40;$$

$$\text{Kov}(y, z) = 1/5 - 1/2 \cdot 1/2 \cdot 1/2 = 3/40;$$

$$\text{Kov}(x, y, z) = 1/10 - 1/2 \cdot 1/2 \cdot 1/2 = -1/40;$$

Отсюда и из (2.12) получаем

$$K(x) = -5/20 + 6/20 + 6/20 - 4/20 = 3/20;$$

$$K(y) = -5/20 + 6/20 + 6/20 - 4/20 = 3/20;$$

$$K(z) = -5/20 + 6/20 + 6/20 - 4/20 = 3/20;$$

Сет-рассеяние всего этого триплета по формуле (2.13) равно

$$\mathbf{S}_{\mathfrak{X}} = 9/40.$$

<sup>8</sup> $\mathfrak{X}$ -строго-парными называются парные пересечения  $I$ -го рода в рамках множества событий  $\mathfrak{X}$ , т.е. пересечения вида  $\text{ter}(\{x, y\} // \mathfrak{X}) = x \cap y \cap \bigcap_{z \in \mathfrak{X} - \{x, y\}} z^c$ .

### 3 Справка: прямая эвентологическая портфельная задача

Пусть  $\mathfrak{X} \subset \mathcal{A}$  — конечное множество из  $N = |\mathfrak{X}|$  рыночных событий, наступление каждого из которых интерпретируется как доходность некоторого рыночного финансового инструмента, а ненаступление — как его недоходность на рассматриваемом рынке при данном стечении рыночных обстоятельств. Образует из этих событий портфель  $\langle \mathfrak{X}, \mathbf{a} \rangle$ , в который каждое рыночное событие  $x \in \mathfrak{X}$  входит в неотрицательной доле  $a_x \geq 0$  так, что сумма всех долей равна единице:

$$\sum_{x \in \mathfrak{X}} a_x = 1, \quad (3.1)$$

где  $\mathbf{a} = \{a_x, x \in \mathfrak{X}\}$  — набор долей портфельных событий, которому разрешено быть произвольной  $N$ -мерной точкой из стандартного единичного  $N$ -вершинного симплекса

$$\mathbb{S}_N = \left\{ \mathbf{a} : \sum_{x \in \mathfrak{X}} a_x = 1; a_x \geq 0 \right\} \subset \mathbb{R}^{N-1}.$$

Для такого простейшего портфеля событий  $\langle \mathfrak{X}, \mathbf{a} \rangle$  классическая постановка *прямой эвентологической портфельной задачи* — экстремальной задачи второго порядка с линейным ограничением — записывается так:

$$\begin{cases} \mathbf{E} \mathbf{1}_{\langle \mathfrak{X}, \mathbf{a} \rangle} = \sum_{x \in \mathfrak{X}} a_x p_x = \langle \mathbf{a} \rangle, \\ \mathbf{D} \mathbf{1}_{\langle \mathfrak{X}, \mathbf{a} \rangle} = \sum_{x \in \mathfrak{X}} \mathbf{D} a_x^2 \mathbf{1}_x + 2 \sum_{\{x, y\} \subseteq \mathfrak{X}} a_x a_y \text{Kov}_{xy} \rightarrow \min_{\mathbf{a} \in \mathbb{S}_N}, \end{cases} \quad (3.2)$$

где

$$\mathbf{1}_{\langle \mathfrak{X}, \mathbf{a} \rangle} = \sum_{x \in \mathfrak{X}} a_x \mathbf{1}_x \quad (3.3)$$

— случайная доходность портфеля событий  $\langle \mathfrak{X}, \mathbf{a} \rangle$ ;  $\mathbf{E}$  и  $\mathbf{D}$  — операторы *ожидания* и *дисперсии*, измерителей *средней доходности* и *риска* портфеля, соответственно;  $\mathbf{a}$

$$\mathbf{p} = \{p(X // \mathfrak{X}), X \subseteq \mathfrak{X}\} \quad (3.4)$$

— вероятностное распределение  $I$ -рода множества портфельных событий  $\mathfrak{X}$ , которое определяет все вероятностные характеристики портфеля, в том числе

$$p_x = \mathbf{P}(x) = \sum_{x \in X} p(X // \mathfrak{X}), \quad (3.5)$$

— вероятность портфельного события  $x \in \mathfrak{X}$ ,

$$\text{Kov}_{xy} = \mathbf{P}(x \cap y) - p_x p_y =$$

$$= \sum_{\{x, y\} \subseteq X} p(X // \mathfrak{X}) - \sum_{x \in X} p(X // \mathfrak{X}) \sum_{y \in Y} p(Y // \mathfrak{X}). \quad (3.6)$$

— парную ковариацию двух портфельных событий  $x \in \mathfrak{X}$ ,  $y \in \mathfrak{X}$ .

В этой экстремальной задаче для известного вероятностного распределения  $\mathbf{p}$  (3.4) множества портфельных событий  $\mathcal{X}$  ищется такой набор долей  $\mathbf{a} \in \mathbb{S}_N$  событий в портфеле  $\langle \mathcal{X}, \mathbf{a} \rangle$ , который, обеспечивая данный уровень его средней доходности:  $\mathbf{E}\mathbf{1}_{\langle \mathcal{X}, \mathbf{a} \rangle} = \langle \mathbf{a} \rangle$ , минимизирует его риск:  $\mathbf{D}\mathbf{1}_{\langle \mathcal{X}, \mathbf{a} \rangle} \rightarrow \min_{\mathbf{a} \in \mathbb{S}_N}$ .

**Пуля Марковица.** В портфельном анализе давно вошло в обиход понятие “пули Марковица” — множества точек плоскости

$$\mathfrak{B}_{\mathcal{X}}(\mathbf{E}, \mathbf{D}) = \left\{ \left( \mathbf{E}\mathbf{1}_{\langle \mathcal{X}, \mathbf{a} \rangle}, \mathbf{D}\mathbf{1}_{\langle \mathcal{X}, \mathbf{a} \rangle} \right) : \mathbf{a} \in \mathbb{S}_N \right\} \subset \mathbb{R}^2, \quad (3.7)$$

служащего образом стандартного единичного симплекса  $\mathbb{S}_N$  при отображении

$$(\mathbf{E}, \mathbf{D}) : \mathbb{S}_N \rightarrow \mathfrak{B}_{\mathcal{X}}(\mathbf{E}, \mathbf{D}), \quad (3.8)$$

которое для каждой  $N$ -мерной точки  $\mathbf{a} \in \mathbb{S}_N$  определяется как точка плоскости

$$\left( \mathbf{E}\mathbf{1}_{\langle \mathcal{X}, \mathbf{a} \rangle}, \mathbf{D}\mathbf{1}_{\langle \mathcal{X}, \mathbf{a} \rangle} \right) \in \mathbb{R}^2 \quad (3.9)$$

с координатами “средняя доходность” и “риск”, вычисляемыми по формулам (3.2).

Пуля Марковица  $\mathfrak{B}_{\mathcal{X}}(\mathbf{E}, \mathbf{D})$  (3.7) характеризует связь между *средней доходностью* (определяемой оператором ожидания  $\mathbf{E}$ ) и *риском* (определяемым оператором дисперсии  $\mathbf{D}$ ) портфеля событий  $\langle \mathcal{X}, \mathbf{a} \rangle$ , навязываемую структурой не более чем парных зависимостей портфельных событий в  $\mathcal{X}$  и определяет для такой структуры зависимостей решение  $\mathbf{a}$  прямых эвентологических портфельных задач (3.2) для всех допустимых уровней средней доходности  $\langle \mathbf{a} \rangle$  данного портфеля событий.

#### 4 Ковариационная пуля эвентологического портфеля событий

Чтобы определить новое понятие *ковариационной пули* эвентологического портфеля событий, мне потребуется структурировать обозначения дисперсионных и ковариационных мер зависимостей множества событий, разложив их на суммы мер, каждая из которых отсылает к зависимости той или иной размерности.

Взвешенное сет-рассеяние (1.3) портфеля событий  $\langle \mathcal{X}, \mathbf{a} \rangle$ , или его полная ковариация, представима в виде

$$\mathbf{S}_{\langle \mathcal{X}, \mathbf{a} \rangle} = \sum_{m=2}^N \mathbf{S}_{\langle \mathcal{X}, \mathbf{a} \rangle}^{(m)}, \quad (4.1)$$

где

$$\mathbf{S}_{\langle \mathcal{X}, \mathbf{a} \rangle}^{(m)} = \sum_{X \in C_{\mathcal{X}}^m} a_X |X| \text{Kov}_X \quad (4.2)$$

— *полная  $m$ -слойная ковариация* портфеля  $\langle \mathcal{X}, \mathbf{a} \rangle$ , а

$$C_{\mathcal{X}}^m = \{X \subseteq \mathcal{X} : |X| = m\} \subset 2^{\mathcal{X}}$$

—  *$m$ -слой подмножество*  $\mathcal{X}$ .

*Полная  $m$ -слойная ковариация* измеряет зависимость  $m$ -го порядка между портфельными событиями. Например, дисперсию (1.2) портфеля событий  $\langle \mathcal{X}, \mathbf{a} \rangle$  можно разложить на сумму двух измерителей:

$$\mathbf{D}\mathbf{1}_{\langle \mathcal{X}, \mathbf{a} \rangle} = \mathbf{D}^i \mathbf{1}_{\langle \mathcal{X}, \mathbf{a} \rangle} + \mathbf{S}_{\langle \mathcal{X}, \mathbf{a} \rangle}^{(2)}, \quad (4.3)$$

где первый измеритель<sup>9</sup>

$$\mathbf{D}^i \mathbf{1}_{\langle \mathcal{X}, \mathbf{a} \rangle} = \sum_{x \in \mathcal{X}} a_x^2 \mathbf{D}\mathbf{1}_x \quad (4.4)$$

— это взвешенная сумма *дисперсий всех портфельных событий*<sup>10</sup>  $x \in \mathcal{X}$ , т.е., — *дисперсия множества взвешенных портфельных событий*<sup>10</sup>  $\mathcal{X}$ , которые происходят с теми же вероятностями, но *тотально независимы*; эта дисперсия измеряет неопределенность, вносимую в портфель каждым взвешенным событием в отдельности; а второй измеритель

$$\mathbf{S}_{\langle \mathcal{X}, \mathbf{a} \rangle}^{(2)} = 2 \sum_{\{x, y\} \subseteq \mathcal{X}} a_x a_y \text{Kov}_{xy} \quad (4.5)$$

— это *полная 2-слойная ковариация* портфеля, характеризующая попарную зависимость между всеми взвешенными портфельными событиями.

Введем также техническое обозначение

$$\mathbf{S}_{\langle \mathcal{X}, \mathbf{a} \rangle}^{(\leq m)} = \sum_{n \leq m} \mathbf{S}_{\langle \mathcal{X}, \mathbf{a} \rangle}^{(n)} \quad (4.6)$$

для сумм ковариационных измерителей зависимостей между взвешенными портфельными событиями до  $m$ -го порядка включительно и будем называть этот измеритель  *$m$ -усеченной ковариацией* портфеля событий  $\langle \mathcal{X}, \mathbf{a} \rangle$ .

**Ковариационная пуля** в эвентологическом портфельном анализе определяется как множество точек плоскости

$$\mathfrak{B}_{\mathcal{X}}(\mathbf{E}, \mathcal{K}) = \left\{ \left( \mathbf{E}\mathbf{1}_{\langle \mathcal{X}, \mathbf{a} \rangle}, \mathcal{K}_{\langle \mathcal{X}, \mathbf{a} \rangle} \right) : \mathbf{a} \in \mathbb{S}_N \right\} \subset \mathbb{R}^2, \quad (4.6)$$

служащее образом стандартного единичного симплекса  $\mathbb{S}_N$  при отображении

$$(\mathbf{E}, \mathcal{K}) : \mathbb{S}_N \rightarrow \mathfrak{B}_{\mathcal{X}}(\mathbf{E}, \mathcal{K}), \quad (4.7)$$

которое для каждой  $N$ -мерной точки  $\mathbf{a} \in \mathbb{S}_N$  определяется как точка плоскости

$$\left( \mathbf{E}\mathbf{1}_{\langle \mathcal{X}, \mathbf{a} \rangle}, \mathcal{K}_{\langle \mathcal{X}, \mathbf{a} \rangle} \right) \in \mathbb{R}^2 \quad (4.8)$$

<sup>9</sup>“ $i$ ” от англ. *independent* — независимый.

<sup>10</sup>Здесь используется специальная терминология, принятая в эвентологии: *дисперсия события* определяется как дисперсия его индикатора, а *дисперсия множества взвешенных событий* — как дисперсия взвешенной суммы их индикаторов.

с координатами “средняя доходность” и “ковариационная зависимость”, где “средняя доходность” вычисляется по формуле (3.2), а “ковариационная зависимость” определяется ковариационным оператором  $\mathcal{K}$  одного из следующих видов (4.1), (4.2) и (4.6):

$$\mathcal{K}_{\langle \mathfrak{X}, \mathbf{a} \rangle} = \begin{cases} \mathbf{S}_{\langle \mathfrak{X}, \mathbf{a} \rangle} & \text{— полная ковариация } \langle \mathfrak{X}, \mathbf{a} \rangle; \\ \mathbf{S}_{\langle \mathfrak{X}, \mathbf{a} \rangle}^{(m)} & \text{— } m\text{-слойная ковариация } \langle \mathfrak{X}, \mathbf{a} \rangle; \\ \mathbf{S}_{\langle \mathfrak{X}, \mathbf{a} \rangle}^{(\leq m)} & \text{— } m\text{-усеченная ковариация } \langle \mathfrak{X}, \mathbf{a} \rangle. \end{cases} \quad (4.9)$$

Ковариационная пуля  $\mathfrak{B}_{\mathfrak{X}}(\mathbf{E}, \mathcal{K})$  (4.6) характеризует связь между *средней доходностью* (определяемой оператором ожидания  $\mathbf{E}$ ) и *ковариационной зависимостью* (определяемой ковариационным оператором  $\mathcal{K}$ ) портфеля событий  $\langle \mathfrak{X}, \mathbf{a} \rangle$ , навязываемую структурой *многомерных* зависимостей взвешенных портфельных событий в  $\mathfrak{X}$ .

#### 4.1 Примеры ковариационных пуль трехмерных портфелей событий

Предлагаемые к рассмотрению примеры трехмерных портфелей событий, т.е. портфелей, составленных из трех событий, более, чем просты, но достаточны для демонстрации преимуществ измерения зависимостей высших порядков при решении типичных задач эвентологического портфельного анализа при помощи ковариационных пуль.

Примеры демонстрируют лишь возможности новых инструментов измерения многомерных зависимостей в анализе портфельных событий, сам анализ и классификация структур зависимостей портфельных событий с использованием ковариационных пуль выходит за рамки этой работы.

Вместе с тем даже эти академические примеры позволяют сделать предварительные выводы о практической ценности новых инструментов для качественного анализа многомерных связей портфельных событий.

В первом примере (Рис. 2, Табл. 1) рассматриваются четыре триплета событий:  $\mathfrak{X}_{11}, \mathfrak{X}_{12}, \mathfrak{X}_{13}, \mathfrak{X}_{14}$ , в каждом из которых вероятности событий имеют одни и те же значения:  $p_x = 0.1, p_y = 0.3, p_z = 0.5$ .

В втором примере (Рис. 3, Табл. 2) рассматриваются четыре триплета событий:  $\mathfrak{X}_{21}, \mathfrak{X}_{22}, \mathfrak{X}_{23}, \mathfrak{X}_{24}$ , в каждом из которых вероятности событий имеют одни и те же значения:  $p_x = 0.15, p_y = 0.50, p_z = 0.85$ .

В третьем примере (Рис. 4, Табл. 3) рассматриваются четыре триплета событий:  $\mathfrak{X}_{31}, \mathfrak{X}_{32}, \mathfrak{X}_{33}, \mathfrak{X}_{34}$ , в каждом из которых вероятности событий имеют одни и те же значения:  $p_x = 0.9, p_y = 0.7, p_z = 0.5$ .

На каждом из приведенных в примерах рисунков изображены *три пули Марковица* (справа):

- $\mathfrak{B}_{\mathfrak{X}}(\mathbf{E}, \mathbf{D})$  (черный),

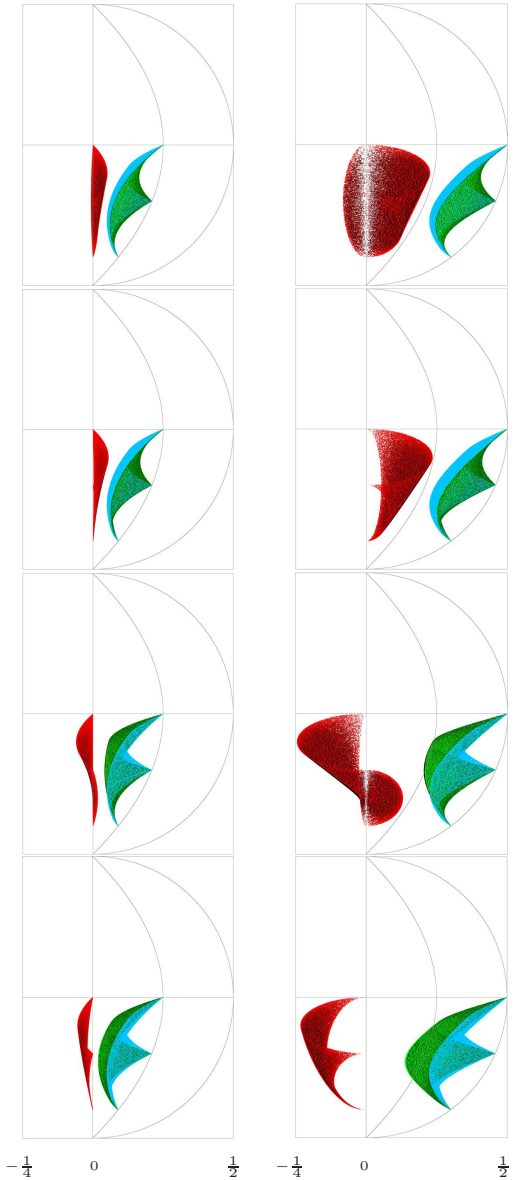


Рис. 1: Ковариационные пули  $\mathfrak{X}_{11}, \mathfrak{X}_{12}, \mathfrak{X}_{13}$  и  $\mathfrak{X}_{14}$  (сверху вниз) в двух масштабах по горизонтали: дисперсионном (слева) и среднеквадратическом (справа). Значения параметров распределений триплетов событий см. в табл. 1.

	$\mathfrak{X}_{11}$ верхний	$\mathfrak{X}_{12}$ второй	$\mathfrak{X}_{13}$ третий	$\mathfrak{X}_{14}$ нижний
$p_x$	0.1	0.1	0.1	0.1
$p_y$	0.3	0.3	0.3	0.3
$p_z$	0.5	0.5	0.5	0.5
$\text{Kov}_{xy}$	0.033	0.012	0.034	-0.026
$\text{Kov}_{xz}$	-0.013	0.006	-0.001	-0.038
$\text{Kov}_{yz}$	0.102	0.110	-0.119	-0.108
$\text{Kov}_{xyz}$	0.018	-0.008	0.012	-0.013

Таблица 1: Вероятности и ковариации триплетов событий  $\mathfrak{X}_{11}, \mathfrak{X}_{12}, \mathfrak{X}_{13}$  и  $\mathfrak{X}_{14}$  (к рис. 2).

- $\mathfrak{B}_{\mathfrak{X}}(\mathbf{E}, \mathbf{D}^{(i)})$  (голубой),
- $\mathfrak{B}_{\mathfrak{X}}(\mathbf{E}, \mathbf{D} + \mathbf{S}^{(3)})$  (зеленый)

и две ковариационные пули (слева):

- $\mathfrak{B}_{\mathfrak{X}}(\mathbf{E}, \mathbf{S}^{(2)})$  (черный),
- $\mathfrak{B}_{\mathfrak{X}}(\mathbf{E}, \mathbf{S}^{(\leq 3)})$  (красный),

которые все расположены в прямоугольнике

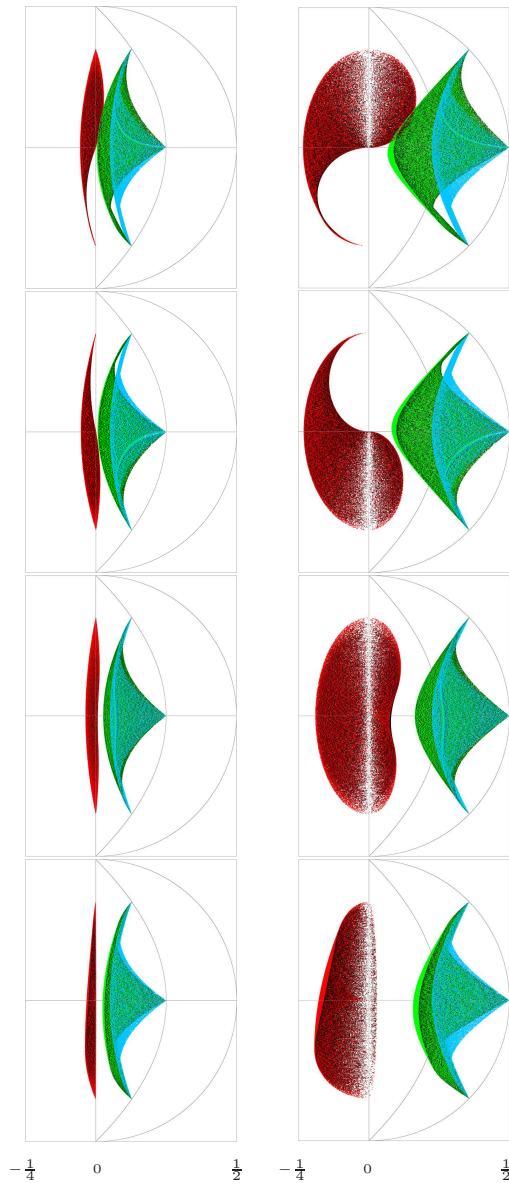


Рис. 2: Ковариационные пули для триплетов событий  $\mathcal{X}_{21}, \mathcal{X}_{22}, \mathcal{X}_{23}$  и  $\mathcal{X}_{24}$  (сверху вниз) в двух масштабах по горизонтали: дисперсионном (слева) и среднеквадратическом (справа). Значения параметров распределений триплетов событий см. в табл. 2.

	$\mathcal{X}_{21}$ верхний	$\mathcal{X}_{22}$ второй	$\mathcal{X}_{23}$ третий	$\mathcal{X}_{24}$ нижний
$p_x$	0.15	0.15	0.15	0.15
$p_y$	0.50	0.50	0.50	0.50
$p_z$	0.85	0.85	0.85	0.85
$\text{Kov}_{xy}$	-0.069	0.030	0.019	-0.075
$\text{Kov}_{xz}$	-0.110	-0.106	-0.073	0.002
$\text{Kov}_{yz}$	0.057	-0.040	0.026	-0.028
$\text{Kov}_{xyz}$	-0.060	-0.056	-0.015	-0.064

Таблица 2: Вероятности и ковариации триплетов событий  $\mathcal{X}_{21}, \mathcal{X}_{22}, \mathcal{X}_{23}$  и  $\mathcal{X}_{24}$  (к рис. 3).

$$[-1/4, 1/2] \times [0, 1].$$

Изображения одних и тех же пуль дается в двух масштабах: дисперсионном (справа) и среднеквадратическом (слева). Для изображения ковариационных пуль в среднеквадратическом масштабе в том случае, если ковариация оказывается меньше нуля, применя-

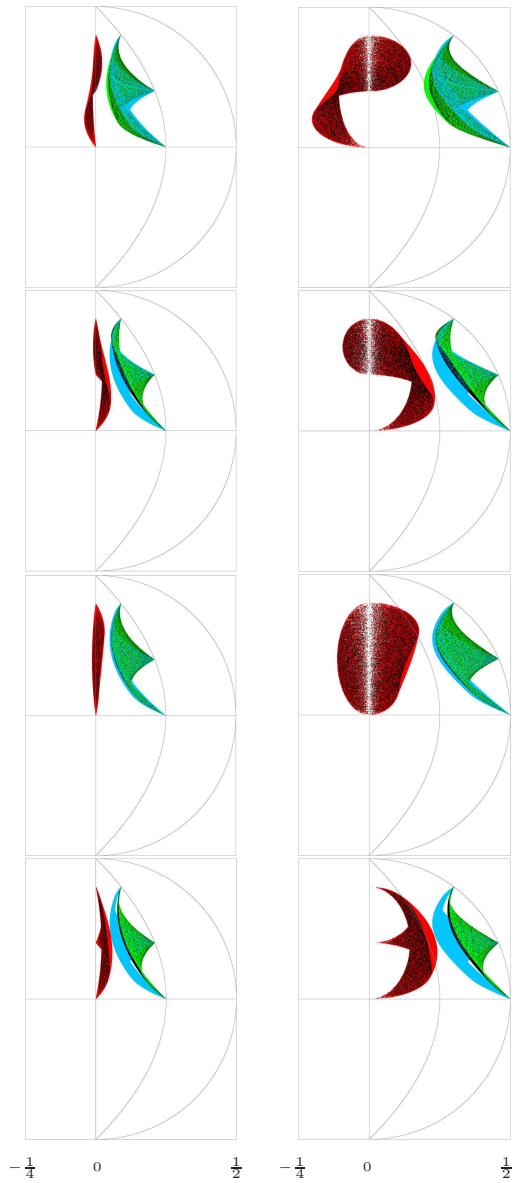


Рис. 3: Ковариационные пули для триплетов событий  $\mathcal{X}_{31}, \mathcal{X}_{32}, \mathcal{X}_{33}$  и  $\mathcal{X}_{34}$  (сверху вниз) в двух масштабах по горизонтали: дисперсионном (слева) и среднеквадратическом (справа). Значения параметров распределений триплетов событий см. в табл. 3.

	$\mathcal{X}_{31}$ верхний	$\mathcal{X}_{32}$ второй	$\mathcal{X}_{33}$ третий	$\mathcal{X}_{34}$ нижний
$p_x$	0.9	0.9	0.9	0.9
$p_y$	0.7	0.7	0.7	0.7
$p_z$	0.5	0.5	0.5	0.5
$\text{Kov}_{xy}$	0.044	-0.018	0.062	0.046
$\text{Kov}_{xz}$	-0.023	0.046	-0.026	0.042
$\text{Kov}_{yz}$	-0.081	0.105	0.022	0.098
$\text{Kov}_{xyz}$	-0.071	0.139	0.052	0.130

Таблица 3: Вероятности и ковариации триплетов событий  $\mathcal{X}_{31}, \mathcal{X}_{32}, \mathcal{X}_{33}$  и  $\mathcal{X}_{34}$  (к рис. 4).

ется масштабирующее преобразование:  $-\sqrt{-\text{Kov}}$ .

По виду ковариационной пули нетрудно заметить как структура зависимостей, отражаемая этим визуальным измерителем, влияет на изменение величины риска. Увеличение риска (при преобладании положительных связей между портфельными событиями) в сравнении с totally независимыми собы-

тиями (голубые пули Марковица) показывают ковариационные пули триплетов  $\mathfrak{X}_{11}, \mathfrak{X}_{12}, \mathfrak{X}_{32}, \mathfrak{X}_{33}, \mathfrak{X}_{34}$ . *Снижение риска* (при преобладании отрицательных связей между портфельными событиями) следует из вида ковариационных пуль триплетов  $\mathfrak{X}_{13}, \mathfrak{X}_{14}, \mathfrak{X}_{21}, \mathfrak{X}_{22}, \mathfrak{X}_{23}, \mathfrak{X}_{24}, \mathfrak{X}_{31}$ .

## 5 Заключение

Отмечу два основных преимущества *эвентологического портфельного анализа* [3, 4].

*Первое*, заключающееся в его *событийной универсальности*, позволяет использовать общую эвентологическую портфельную модель для разного рода частных областей применения, всякий раз подходящим образом интерпретируя портфельные события в рамках выбранной области приложений.

*Второе* заключается в широком ассортименте прозрачных эвентологических измерителей *многомерных зависимостей* портфельных событий [1, 2], которые работают также и в расширенной постановке эвентологических портфельных задач [5, 6], открывая возможность их эффективного использования в практике портфельного анализа *финансово-актуарных, гуманитарных и социо-экономических* рынков.

В предлагаемой работе к этому ассортименту добавлены визуальные измерители — *ковариационные пули множества портфельных событий*, с помощью которых можно анализировать и делать полезные выводы о структуре *многомерных зависимостей* портфельных событий и их влиянии на доходность и риск портфеля событий из любой области, представляющей интерес для теоретика или практика портфельного анализа.

## Список литературы

- [1] О. Ю. Воробьев. Меры рассеяния множества событий по пространству элементарных исходов. *Труды XV Международ. ЭМ конференции по эвентологической математике и смежным вопросам*, Красноярск: СФУ (под ред. Олега Воробьева):54–63, 2011.
- [2] O. Yu. Vorobyev. Measures of scattering of a set of events on a space of elementary outcomes. In. *Proc. of the XV Intern. EM Conf. on Eventological Mathematics*, Krasnoyarsk, SFU (Oleg Vorobyev ed.):54–63, 2011 (in Russian, abstract in English).
- [3] О. Ю. Воробьев. *Эвентология*. Сибирский федеральный университет, Красноярск, 2007, 435с., <http://eventology-theory.ru/0-books/1-VorobyevOleg~2007~Eventology~435p.pdf>.
- [4] O. Yu. Vorobyev. *Eventology*. Siberian Federal University, Krasnoyarsk, Russia, 2007, 435p. (in Russian, abstract in English), <http://eventology-theory.com/0-books/1-VorobyevOleg~2007~Eventology~435p.pdf>.
- [5] О. Ю. Воробьев. Измерение многомерных событийных зависимостей в эвентологическом портфельном анализе финансово-актуарных, гуманитарных и социо-экономических рынков. *Труды XII Международ. ФАМЭБ конференции по финансово-актуарной математике и эвентологии безопасности*, Красноярск: НИИППБ, СФУ (под ред. Олега Воробьева):143–150, 2013.
- [6] O. Yu. Vorobyev. Measuring multivariate event-based dependencies in eventological portfolio analysis of financial and actuarial, humanitarian, and socioeconomic markets. In. *Proc. of the XII Intern. FAMES Conf. on Financial and Actuarial Mathematics and Eventology of Safety*, Krasnoyarsk, SFU (Oleg Vorobyev ed.):143–150, 2013 (in Russian, abstract in English).



# Гиббсовская характеристика множеств событий

Олег Юрьевич Воробьев

Сибирский федеральный университет  
Институт математики и фундаментальной информатики  
Торгово-экономический институт  
Красноярск  
oleg.yu.vorobyev@gmail.com

**Аннотация.** Рассматривается гиббсовская характеристика и другие инструменты гиббсовского анализа вероятностных и ценностных распределений множеств событий.

**Ключевые слова.** Эвентология, теория вероятностей, элементарный исход, событие, вероятность, алгебра событий, множество событий, эвентологическое распределение, гиббсовская характеристика эвентологического распределения, слойный портрет множества событий.

... — Позволю  
Определить неизбежный жребий ...  
М. Ю. Лермонтов, Ночь, 1830.  
“Автопортрет с отрезанным ухом”  
Винсент Ван Гог, 1889.

До сих пор во многих работах по эвентологии, включая мою книгу [1, 2], использовался не совсем удачный термин *гиббсовское распределение* множества событий, или *гиббсовское множество событий*, который обозначал ситуацию, когда в той или иной задаче кроме *вероятностного распределения* множества событий фигурировало и его *ценностное распределение*, которое характеризуется вероятностным распределением. В этой работе я решил, наконец, избавиться от этой стихийно сложившейся и способной ввести в заблуждение терминологии, строго определяя взамен новое понятие: *гиббсовская характеристика* множества событий. Я предлагаю также ввести в практику исследований гиббсовских событийных моделей довольно удобные, на мой взгляд, визуальные инструменты измерения свойств *гиббсовской характеристики* множеств событий, основными из которых служат *слойные портреты* этих множеств событий, создаваемые их вероятностными и ценностными распределениями.

## 1 Гиббсовская характеристика вероятностного распределения множества событий

Пусть  $(\Omega, \mathcal{A}, \mathbf{P})$  — всеобщее вероятностное пространство, а  $\mathfrak{X} \subset \mathcal{A}$  — конечное множество всеобщих событий [1, 2].

**Определение (гиббсовская характеристика множества событий).** Гиббсовской  $\alpha$ -характеристикой вероятностного распределения  $I$ -го рода  $\{p(X//\mathfrak{X}), X \subseteq \mathfrak{X}\}$  множества событий  $\mathfrak{X}$  ценностным распределением  $I$ -го рода  $\{v_\alpha(X//\mathfrak{X}), X \subseteq \mathfrak{X}\}$  называются соотношения, связывающие вероятности и ценности террасных событий  $\text{ter}(X//\mathfrak{X})$  для  $X \subseteq \mathfrak{X}$ , таких что  $p(X//\mathfrak{X}) > 0$ , прямыми:

$$p(X//\mathfrak{X}) = \exp \left\{ \alpha v_\alpha(X//\mathfrak{X}) \right\} \quad (1.1)$$

и обратными формулами:

$$v_\alpha(X//\mathfrak{X}) = \frac{1}{\alpha} \ln p(X//\mathfrak{X}), \quad (1.1')$$

где  $\alpha$  — *гиббсовский параметр*  $I$ -го рода.

Обозначьте,

$$\mathbb{P}(\mathfrak{X}) = \{p \in (0, 1] : p = p(X//\mathfrak{X}), X \subseteq \mathfrak{X}\}, \quad (1.2)$$

$$\mathbb{V}_\alpha(\mathfrak{X}) = \{v \in \mathbb{R} : v = v_\alpha(X//\mathfrak{X}), X \subseteq \mathfrak{X}\} \quad (1.3)$$

— конечные множества значений вероятностей и ценностей террасных событий  $\text{ter}(X//\mathfrak{X})$  для  $X \subseteq \mathfrak{X}$ :

$$p(X//\mathfrak{X}) = \mathbf{P}(\text{ter}(X//\mathfrak{X})), \quad (1.4)$$

$$v_\alpha(X//\mathfrak{X}) = \frac{1}{\alpha} \ln \mathbf{P}(\text{ter}(X//\mathfrak{X})). \quad (1.5)$$

Заметьте, что

$$|\mathbb{P}(\mathfrak{X})| = |\mathbb{V}_\alpha(\mathfrak{X})| \leq 2^{|\mathfrak{X}|}$$

— множества  $\mathbb{P}(\mathfrak{X})$  и  $\mathbb{V}_\alpha(\mathfrak{X})$  равномощны, их мощность не превосходит мощности  $2^{\mathfrak{X}}$  — множества всех подмножеств  $\mathfrak{X}$  — и совпадает с ней, когда все значения вероятностей террасных событий различны и строго больше нуля.

Вероятностно-ценностные соотношения (1.1) определяют *сопровождающее вероятностное распределение* (первого рода)<sup>1</sup>

$$\{p_\alpha^*(v), v \in \mathbb{V}_\alpha(\mathfrak{X})\} \quad (1.6)$$

<sup>1</sup> Аналогичным образом на  $\{v_\alpha(X//\mathfrak{X}), X \subseteq \mathfrak{X}\} \subset \mathbb{R}$  определяются вероятностно-ценностные распределения других (II–VI) родов:  $p_v, p^v, u(v), u_v, u^v$ .

некоторой случайной величины  $\nu_\alpha$  с конечным множеством значений  $\mathbb{V}_\alpha(\mathfrak{X})$ , где

$$p_\alpha^*(v) = \mathbf{P}(\nu_\alpha = v) = \sum_{X: \nu_\alpha(X//\mathfrak{X})=v} p(X//\mathfrak{X}) \quad (1.7)$$

для  $v \in \mathbb{V}_\alpha(\mathfrak{X})$ .

Обозначьте для  $X \subseteq \mathfrak{X}$

$$v_{-1}(X//\mathfrak{X}) = -\ln p(X//\mathfrak{X}) \quad (1.8)$$

—  $\alpha$ -ценности террасных событий  $\text{ter}(X//\mathfrak{X})$  при  $\alpha = -1$  (1.1'), которые, взятые вместе, образуют  $(-1)$ -ценностное распределение

$$\{v_{-1}(X//\mathfrak{X}), X \subseteq \mathfrak{X}\}, \quad (1.9)$$

характеризующее распределение (1.1).

Отметьте, что  $(-1)$ -ценности (1.8) неотрицательны, и нормируйте их к единичной сумме так, чтобы получить то, что называется *нормированными ценностями* гиббсовского распределения (1.1):

$$v(X//\mathfrak{X}) = \frac{v_{-1}(X//\mathfrak{X})}{\sum_{Y \subseteq \mathfrak{X}} v_{-1}(Y//\mathfrak{X})} = \frac{\ln p(X//\mathfrak{X})}{\sum_{Y \subseteq \mathfrak{X}} \ln p(Y//\mathfrak{X})}. \quad (1.10)$$

Вам осталось увидеть в (1.10) формулу

$$v(X//\mathfrak{X}) = \frac{1}{\alpha(p)} \ln p(X//\mathfrak{X}), \quad (1.11)$$

где

$$\alpha(p) = \sum_{Y: p(Y//\mathfrak{X}) > 0} \ln p(Y//\mathfrak{X}) \leq 0, \quad (1.12)$$

и заключить, что образованное величинами (1.10) *нормированное ценностное распределение*<sup>2</sup>

$$\{v(X//\mathfrak{X}), X \subseteq \mathfrak{X}\} = \{v_{\alpha(p)}(X//\mathfrak{X}), X \subseteq \mathfrak{X}\} \quad (1.13)$$

— это  $\alpha(p)$ -ценностное распределение  $I$ -го рода, *гиббсовский параметр*  $\alpha(p)$  которого определяется формулой (1.12) и называется *каноническим гиббсовским параметром  $I$ -рода*.

На рис. 1 показан пример образа совокупности из 50 000 вероятностных распределений  $I$ -го рода дуплетов независимых событий — седлообразная поверхность в вероятностном симплексе, при отображении

<sup>2</sup>Для такого *нормированного ценностного распределения* аналогичным образом определяется *гиббсовское вероятностное распределение* (первого рода)

$$\{p^*(v), v \in \mathbb{V}(\mathfrak{X})\} \quad (1.6')$$

некоторой случайной величины  $\nu$  с конечным множеством значений

$$\mathbb{V}(\mathfrak{X}) = \mathbb{V}_{\alpha(p)}(\mathfrak{X}),$$

где

$$p^*(v) = \mathbf{P}(\nu = v) = \sum_{X: v(X//\mathfrak{X})=v} p(X//\mathfrak{X}) \quad (1.7')$$

для  $v \in \mathbb{V}(\mathfrak{X})$ .

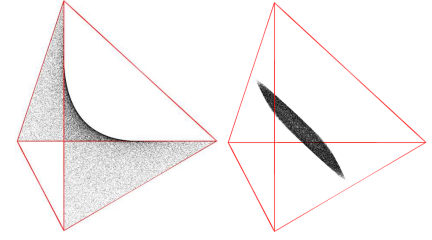


Рис. 1: Совокупность из 50 000 вероятностных распределений  $I$ -го рода дуплетов независимых событий — седлообразная поверхность в вероятностном симплексе (слева); и ее образ при отображении по формулам (1.11) и (1.12) в симплекс нормированных ценностных распределений  $I$ -го рода — квадрат со стороной  $\sqrt{2}/2$  (половина ребра симплекса) (справа).

в симплекс нормированных ценностных распределений  $I$ -рода по формулам (1.11) и (1.12) — квадрат со стороной  $\sqrt{2}/2$  (половина ребра симплекса).

Когда ценностное распределение нормировано, вероятностно-ценностные соотношения (1.1) определяют по аналогии с (1.6) и (1.7) еще и *сопрядающее ценностное распределение* (первого рода)

$$\{v^*(p), p \in \mathbb{P}(\mathfrak{X})\} \quad (1.14)$$

некоторой случайной величины  $\psi$  с конечным множеством значений  $\mathbb{P}(\mathfrak{X})$ , где

$$v^*(p) = \mathbf{P}(\psi = p) = \sum_{X: p(X//\mathfrak{X})=p} v(X//\mathfrak{X}) \quad (1.15)$$

для  $p \in \mathbb{P}(\mathfrak{X})$ .

Итак нормированное ценностное распределение возникает в так называемой *канонической гиббсовской характеристизации*, в которой *канонический гиббсовский параметр* определяется формулой (1.12), т.е. — определяется самим вероятностным распределением<sup>3</sup>.

<sup>3</sup>Любое *вероятностное распределение* на  $\mathbb{R}$ , а не только распределение множества событий  $\mathfrak{X} \subset \mathcal{A}$ , имеет гиббсовскую характеристизацию *нормированным ценностным распределением* вплоть до вероятностных распределений непрерывных случайных величин, *плотности вероятности* которых  $p(r)$  соответствует *нормированная ценностная плотность*  $v(r)$ , определяемая формулой

$$v(r) = \frac{\ln p(r)}{\int_{\{r: p(r) > 0\}} \ln p(r) dr} \quad (1.*)$$

Из (1.\*) следует, что

$$\frac{1}{\alpha(p)} \ln p(r) = \frac{\ln p(r)}{\int_{\{r: p(r) > 0\}} \ln p(r) dr},$$

т.е.

$$\alpha(p) = \int_{\{r: p(r) > 0\}} \ln p(r) dr. \quad (1.**)$$

Формулы (1.\*) и (1.\*\*), соответствующие своим эвентологическим аналогам (1.11) и (1.12).

## 2 Гиббсовская характеристика ценностного распределения множества событий

До сих пор по умолчанию было принято считать, что задание множества событий  $\mathfrak{X}$  эквивалентно заданию его вероятностного распределения  $\{p(X/\mathfrak{X}) : X \subseteq \mathfrak{X}\}$ . Примите теперь по умолчанию эквивалентное двойственное утверждение, что задание множества событий  $\mathfrak{X}$  эквивалентно заданию его *нормированного* ценностного распределения  $\{v(X/\mathfrak{X}) : X \subseteq \mathfrak{X}\}$ , и поставьте задачу нахождения вероятностного распределения, связанного с ценностными формулами (1.10). Иными словами, вам необходимо найти формулы, обратные формулам (1.10), при условии, что ценностное распределение нормировано.

Обозначьте

$$\alpha(X/\mathfrak{X}) = \ln p(X/\mathfrak{X}), \quad (2.1)$$

воспользуйтесь ради удобства сокращенными обозначениями и сначала буквально перепишите систему уравнений (1.10) относительно неизвестных  $\alpha(X)$  для  $X \subseteq \mathfrak{X}$  в виде

$$v(X) = \frac{\alpha(X)}{\sum_{Y \subseteq \mathfrak{X}} \alpha(Y)}, \quad (2.2)$$

а затем — в обычном виде линейной системы однородных уравнений

$$v(X) \sum_{Y \subseteq \mathfrak{X}} \alpha(Y) = \alpha(X). \quad (2.3)$$

для  $X \subseteq \mathfrak{X}$ . После этого введите еще одну неизвестную — константу  $\alpha$ , добавив к системе (2.3) еще одно уравнение

$$\alpha = \sum_{Y \subseteq \mathfrak{X}} \alpha(Y), \quad (2.4)$$

чтобы получить новую систему из  $1 + 2^{|\mathfrak{X}|}$  линейных однородных уравнений:

$$\begin{cases} \alpha(X) - \alpha v(X) = 0, & X \subseteq \mathfrak{X}, \\ \sum_{Y \subseteq \mathfrak{X}} \alpha(Y) - \alpha = 0. \end{cases} \quad (2.5)$$

Прежде, чем решать эту систему в общем виде, попробуйте решить ее для дуплета  $\mathfrak{X} = \{x, y\}$ , т.е. при  $|\mathfrak{X}| = 2$ , когда она состоит из  $1 + 2^2 = 5$  уравнений:

$$\begin{cases} \alpha(\emptyset) - \alpha v(\emptyset) = 0, \\ \alpha(\{x\}) - \alpha v(\{x\}) = 0, \\ \alpha(\{y\}) - \alpha v(\{y\}) = 0, \\ \alpha(\{x, y\}) - \alpha v(\{x, y\}) = 0, \\ \alpha(\emptyset) + \alpha(\{x\}) + \alpha(\{y\}) + \alpha(\{x, y\}) - \alpha = 0, \end{cases} \quad (2.6)$$

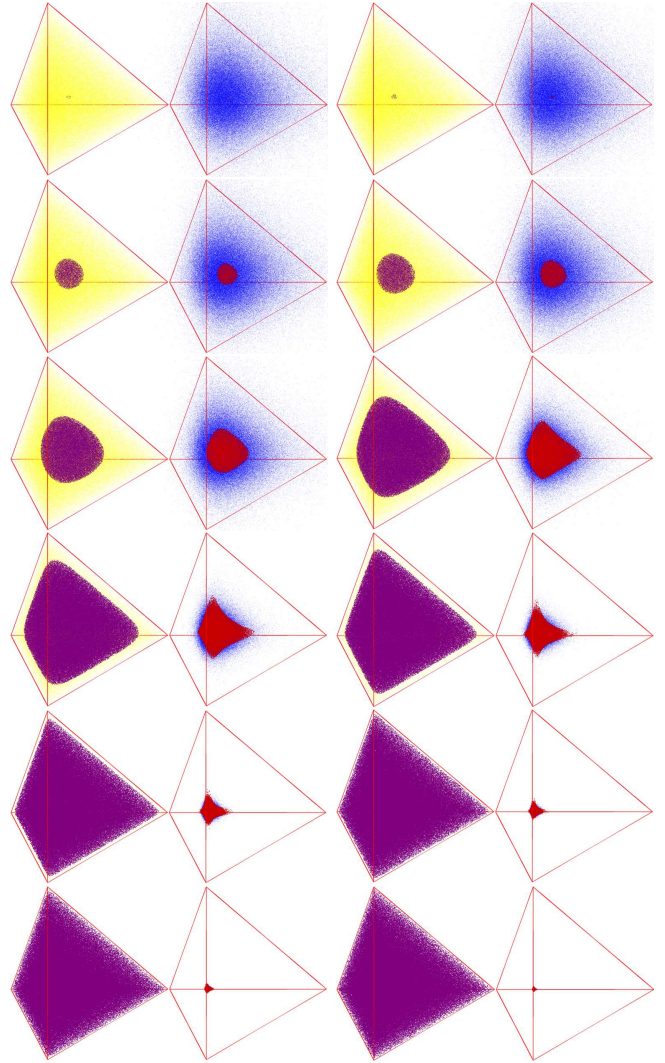


Рис. 2: В 1-й и 3-й колонке подмножества (фиолетовым на желтом) симплекса нормированных *ценностных* распределений дуплета событий, ограниченные замкнутой поверхностью нормированных ценностных распределений, для которых *легитимно* одно и то же значение канонического гиббсовского параметра  $\alpha$ ; во 2-й и 4-й колонке соответствующие подмножества (красным на синем) симплекса *вероятностных* распределений дуплета событий, ограниченные замкнутой поверхностью вероятностных распределений, для того же самого значения канонического гиббсовского параметра  $\alpha$ , который легитимен для ценностного распределения (сосед справа); показаны 12 вариантов (слева направо, сверху вниз) для различных значений гиббсовского параметра:  $-5.550, -5.555; -5.800, -6.000; -7.000, -10.000; -15.000, -25.000; -50.000, -100.000, -200.000, -400.000$ ; в каждом варианте 250 000 распределений в симплексе (фиолетовым и желтым).

или в матричном виде:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & -v(\emptyset) \\ 0 & 1 & 0 & 0 & -v(\{x\}) \\ 0 & 0 & 1 & 0 & -v(\{y\}) \\ 0 & 0 & 0 & 1 & -v(\{x, y\}) \\ 1 & 1 & 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} \alpha(\emptyset) \\ \alpha(\{x\}) \\ \alpha(\{y\}) \\ \alpha(\{x, y\}) \\ \alpha \end{pmatrix} = \mathbf{0}, \quad (2.6')$$

или в сокращенных матричных обозначениях:

$$\mathbf{V}\alpha = \mathbf{0}. \quad (2.6'')$$

Матрица  $\mathbf{V}$  имеет ранг 4, поэтому у однородной системы (2.6) есть нетривиальные решения вида:

$$\alpha = \alpha v. \quad (2.7)$$

где  $\alpha \in \mathbb{R}$  — произвольный параметр, а

$$v = (v(\emptyset), v(\{x\}), v(\{y\}), v(\{x, y\}), 1)^\top \quad (2.8)$$

— вектор, первые четыре компоненты которого определяются известным *нормированным* ценностным распределением дуплета событий  $\mathfrak{X} = \{x, y\}$ .

В силу (2.1) нетривиальные решения (2.7) в рамках поставленной задачи должны определять вероятностное распределение  $\{p(X) = e^{\alpha(X)} : X \subseteq \mathfrak{X}\}$ , т.е. распределение, удовлетворяющее *локальной*:

$$0 \leq e^{\alpha(X)} \leq 1, \quad X \subseteq \mathfrak{X}, \quad (2.9)$$

и *глобальной*:

$$\sum_{X \subseteq \mathfrak{X}} e^{\alpha(X)} = 1 \quad (2.10)$$

вероятностным нормировкам. Поэтому параметр  $\alpha$  обеспечивает локальную вероятностную нормировку лишь когда

$$\alpha < 0, \quad (2.11)$$

а — глобальную, лишь когда является решением трансцендентного уравнения:

$$(e^{v(\emptyset)})^\alpha + (e^{v(\{x\})})^\alpha + (e^{v(\{y\})})^\alpha + (e^{v(\{x, y\})})^\alpha = 1. \quad (2.11')$$

Назовите нетривиальное решение однородной системы линейных уравнений (2.6) *легитимным* для данного ценностного распределения, если параметр  $\alpha$  обеспечивает обе вероятностные нормировки, т.е. удовлетворяет (2.11) и (2.11').

Теперь самое время заметить, что пример дуплета событий рассматривался лишь из-за неудобств обозначений в общей ситуации произвольного множества событий, а не из-за принципиальных трудностей, так как характер решения однородной системы линейных уравнений при росте размерности  $\mathfrak{X}$  не меняется.

В окончательном итоге *легитимное* нетривиальное решение однородной системы линейных уравнений<sup>4</sup>

$$\mathbf{V}\alpha = \mathbf{0}, \quad (2.12)$$

<sup>4</sup>где  $\mathbf{V}$  — квадратная матрица размерности  $(1 + 2^{|\mathfrak{X}|}) \times (1 + 2^{|\mathfrak{X}|})$ , а неизвестный вектор

$$\alpha = (\alpha(\emptyset), \dots, \alpha(X), \dots, \alpha(\mathfrak{X}), \alpha)^\top \quad (2.13)$$

размерности  $1 + 2^{|\mathfrak{X}|}$  составлен из  $2^{|\mathfrak{X}|}$  логарифмов компонент неизвестного вероятностного распределения  $\mathfrak{X}$  (2.1) и неизвестного параметра  $\alpha$ .

имеет тот же вид<sup>5</sup>:

$$\alpha = \alpha v, \quad (2.14)$$

где параметр  $\alpha$  обеспечивает обе вероятностные нормировки, т.е. удовлетворяет (2.11) и

$$\sum_{X \subseteq \mathfrak{X}} (e^{v(X)})^\alpha = 1. \quad (2.16)$$

### 3 Слойные портреты множества событий

Каждое конечное множество событий  $\mathfrak{X} \subset \mathcal{A}$  разбивает  $\Omega$ , вообще говоря, на  $2^{|\mathfrak{X}|}$  событийных фрагментов — *террасных событий*

$$\text{ter}(X//\mathfrak{X}) = \bigcap_{x \in X} x \bigcap_{x \in \mathfrak{X} - X} x^c \subseteq \Omega, \quad (3.1)$$

где  $X \subseteq \mathfrak{X}$ . Поэтому

$$\Omega = \sum_{X \subseteq \mathfrak{X}} \text{ter}(X//\mathfrak{X}).$$

Пусть

$$\check{f} : \Omega \rightarrow \mathbb{R} \quad (3.2)$$

— ограниченная вещественная кусочно-постоянная функция на  $\Omega$ , которая на террасных событиях  $\text{ter}(X//\mathfrak{X}) \subseteq \Omega$  принимает постоянные значения  $f(X//\mathfrak{X})$ , занумерованные подмножествами  $X \subseteq \mathfrak{X}$ :

$$\check{f}(\omega) = \begin{cases} f(\emptyset//\mathfrak{X}), & \omega \in \text{ter}(\emptyset//\mathfrak{X}); \\ \dots, & \dots; \\ f(X//\mathfrak{X}), & \omega \in \text{ter}(X//\mathfrak{X}); \\ \dots, & \dots; \\ f(\mathfrak{X}//\mathfrak{X}), & \omega \in \text{ter}(\mathfrak{X}//\mathfrak{X}). \end{cases} \quad (3.3)$$

Пусть

$$\mathbb{F}(\mathfrak{X}) = \{r \in \mathbb{R} : r = f(X//\mathfrak{X}), X \subseteq \mathfrak{X}\} \quad (3.4)$$

— конечное множество значений функции  $\check{f}$ ,

$$\mathbb{N}(\mathfrak{X}) = \{|X| : X \subseteq \mathfrak{X}\} = \{0, 1, \dots, |\mathfrak{X}|\} \quad (3.4)$$

— множество целых чисел от нуля до  $N = |\mathfrak{X}|$ , составленное из мощностей подмножеств  $\mathfrak{X}$ .

**Определение (слойный портрет множества событий).** Бинарное отношение

$$\mathcal{S}_{\mathfrak{X} \times f} = \left\{ (|X|, f(X//\mathfrak{X})) : X \subseteq \mathfrak{X} \right\} \subseteq \mathbb{N}(\mathfrak{X}) \times \mathbb{F}(\mathfrak{X})$$

называется *слойным  $\check{f}$ -портретом множества событий  $\mathfrak{X}$* .

<sup>5</sup>где

$$v = (v(\emptyset), \dots, v(X), \dots, v(\mathfrak{X}), 1)^\top \quad (2.15)$$

— вектор размерности  $1 + 2^{|\mathfrak{X}|}$ , составленный из  $2^{|\mathfrak{X}|}$  известных компонент ценностного распределения  $\mathfrak{X}$  и единичной константы.



На рис. 3 показан пример *слоистого*  $\check{p}$ -портрета множества событий  $\mathfrak{X}$  (левее вертикальной оси фиолетовым), т.е. портрета полученного при помощи функции

$$\check{p} : \Omega \rightarrow \mathbb{P}(\mathfrak{X}),$$

значениями которой на  $X \subseteq \mathfrak{X}$

$$\check{p}(\omega) = \begin{cases} p(\emptyset // \mathfrak{X}), & \omega \in \text{ter}(\emptyset // \mathfrak{X}); \\ \dots, & \dots; \\ p(X // \mathfrak{X}), & \omega \in \text{ter}(X // \mathfrak{X}); \\ \dots, & \dots; \\ p(\mathfrak{X} // \mathfrak{X}), & \omega \in \text{ter}(\mathfrak{X} // \mathfrak{X}), \end{cases} \quad (3.5)$$

служат вероятности

$$p(X // \mathfrak{X}) = \mathbf{P}(\text{ter}(X // \mathfrak{X})),$$

образующие вероятностное распределение первого рода множества событий  $\mathfrak{X}$ .

На рис. 3 показан пример *слоистого*  $\check{v}$ -портрета множества событий  $\mathfrak{X}$  (ниже вертикальной оси крас-

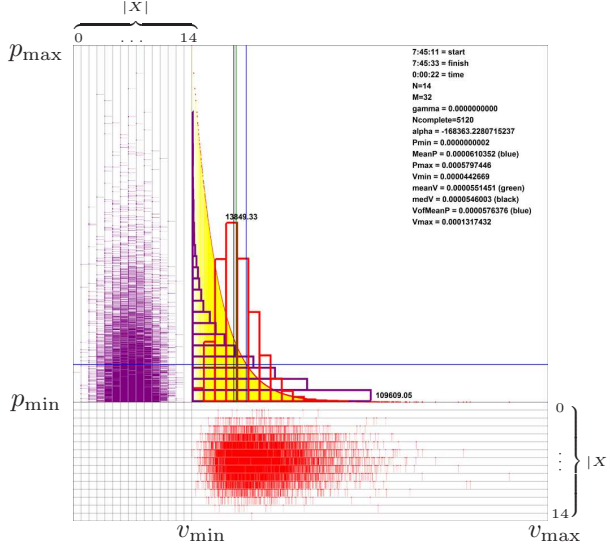


Рис. 3: Диаграмма (красно-желтая) канонической гиббсовской характеристики вероятностей  $\{p(X) = \exp\{\alpha(p)v(X)\}, X \subseteq \mathfrak{X}\}$ ; две гистограммы по  $M = 32$  интервалам: гистограмма  $p_{\check{p}}^*$  (по горизонтали красным цветом) и гистограмма  $v_{\check{p}}^*$  (по вертикали фиолетовым цветом); и два слоистых портрета: *слоистый*  $\check{v}$ -портрет  $S_{\mathfrak{X} \times v}$  14-многообразия событий  $\mathfrak{X}$  (бинарное отношение  $S_{\mathfrak{X} \times v} \subset \mathbb{N}(\mathfrak{X}) \times \mathbb{V}(\mathfrak{X})$  (под горизонтальной осью красным), состоящее из  $N+1=15$  рядов для  $|X| = 0, \dots, N, X \subseteq \mathfrak{X}$ , фактически, — это “расслоение” одного графика из  $2^{14}$  значений вероятностей (1.6) на 15 графиков из  $C_{14}^{|X|}$  значений вероятностей (вид сверху) по  $|X|$ -слоям мощности подмножеств  $X \subseteq \mathfrak{X}$  для  $|X| = 0, 1, \dots, 14$ ); *слоистый*  $\check{p}$ -портрет  $S_{\mathfrak{X} \times p}$  14-многообразия событий  $\mathfrak{X}$  (бинарное отношение  $S_{\mathfrak{X} \times p} \subset \mathbb{N}(\mathfrak{X}) \times \mathbb{P}(\mathfrak{X})$  (левее вертикальной оси фиолетовым), состоящее из  $N+1=15$  рядов для  $|X| = 0, \dots, N, X \subseteq \mathfrak{X}$ , фактически, — это “расслоение” одного графика из  $2^{14}$  значений ценностей (1.6) на 15 графиков из  $C_{14}^{|X|}$  значений ценностей (вид сверху) по  $|X|$ -слоям мощности подмножеств  $X \subseteq \mathfrak{X}$  для  $|X| = 0, 1, \dots, 14$ ); вероятностное распределение  $\mathfrak{X}$  изначально *выбрано равномерно* как точка в 14-вершинном симплексе распределений; параметры интервала  $v_{\alpha}$ -значений:  $v_{\min} \approx 0.00004$ ;  $v_{\max} \approx 0.00013$ ; гиббсовский параметр:  $\alpha(p) \approx -168363.23$ ; параметры интервала значений вероятностей:  $p_{\min} \approx 2 \cdot 10^{-10}$ ;  $p_{\max} \approx 0.00057$ .

ным), т.е. портрета полученного при помощи функции<sup>6</sup>

$$\check{v} : \Omega \rightarrow \mathbb{V}(\mathfrak{X}),$$

значениями которой на  $X \subseteq \mathfrak{X}$

$$\check{v}(\omega) = \begin{cases} v(\emptyset // \mathfrak{X}), & \omega \in \text{ter}(\emptyset // \mathfrak{X}); \\ \dots, & \dots; \\ v(X // \mathfrak{X}), & \omega \in \text{ter}(X // \mathfrak{X}); \\ \dots, & \dots; \\ v(\mathfrak{X} // \mathfrak{X}), & \omega \in \text{ter}(\mathfrak{X} // \mathfrak{X}), \end{cases} \quad (3.6)$$

служат ценности

$$v(X // \mathfrak{X}) = \frac{\ln \mathbf{P}(\text{ter}(X // \mathfrak{X}))}{\sum_{X \subseteq \mathfrak{X}} \ln \mathbf{P}(\text{ter}(X // \mathfrak{X}))}, \quad (3.7)$$

образующие ценностное распределение первого рода множества событий  $\mathfrak{X}$ .

#### 4 Гистограммы сопровождающих распределений в гиббсовском анализе множества событий

Рассмотрите внимательнее для множества событий  $\mathfrak{X}$  сопровождающие его вероятностное  $p^*$  (1.6') и ценностное  $v^*$  (1.14) распределения случайных величин  $\nu$  и  $\psi$ , которые принимают значения из конечных множеств значений  $\mathbb{V}(\mathfrak{X}) \subset \mathbb{R}$  и  $\mathbb{P}(\mathfrak{X}) \subset \mathbb{R}$  соответственно. Для этого сначала обозначьте

$v_{\min} = \min\{v : v \in \mathbb{V}(\mathfrak{X})\}$ ,  $v_{\max} = \max\{v : v \in \mathbb{V}(\mathfrak{X})\}$ ,  
 $p_{\min} = \min\{p : p \in \mathbb{P}(\mathfrak{X})\}$ ,  $p_{\max} = \max\{p : p \in \mathbb{P}(\mathfrak{X})\}$ ,  
— минимальные и максимальные значения из этих множеств, которые определяют минимальные интервалы, их содержащие:

$$\mathbb{V}(\mathfrak{X}) \subset [v_{\min}, v_{\max}], \quad \mathbb{P}(\mathfrak{X}) \subset [p_{\min}, p_{\max}]. \quad (4.1)$$

Если теперь вы разобьете минимальные интервалы (4.1) на  $M$ , к примеру, равных подинтервалов, то вы сможете формально построить для каждого из двух сопровождающих распределений гистограмму, с помощью которой сможете исследовать характер распределения значений сопровождающих случайных величин  $\nu$  и  $\psi$  по соответствующим интервалам (4.1).

Гистограммы распределений, сопровождающих множество событий  $\mathfrak{X}$ , представляют собой наглядные инструменты анализа его гиббсовской характеристики (1.1). Они строятся следующим образом.

Обозначьте  $\mathbb{V}_m$  и  $\mathbb{P}_m$  —  $m$ -е полузамкнутые подинтервалы

$$\left[ v_{\min} + \frac{m-1}{M}(v_{\max} - v_{\min}), v_{\min} + \frac{m}{M}(v_{\max} - v_{\min}) \right)$$

<sup>6</sup>Здесь  $\mathbb{V}(\mathfrak{X}) = \mathbb{V}_{\alpha(p)}(\mathfrak{X})$ , см. (1.3).



$\left[ p_{\min} + \frac{m-1}{M}(p_{\max} - p_{\min}), p_{\min} + \frac{m}{M}(p_{\max} - p_{\min}) \right)$   
интервалов  $[v_{\min}, v_{\max}]$  и  $[p_{\min}, p_{\max}]$  для  $m = 1, \dots, M-1$ , и, наконец,

$$\mathbb{V}_M = \left[ v_{\max} - \frac{1}{M}(v_{\max} - v_{\min}), v_{\max} \right] \subset [v_{\min}, v_{\max}],$$

$$\mathbb{P}_M = \left[ p_{\max} - \frac{1}{M}(p_{\max} - p_{\min}), p_{\max} \right] \subset [p_{\min}, p_{\max}]$$

—  $M$ -е замкнутые подинтервалы. В этих обозначениях

$$[v_{\min}, v_{\max}] = \sum_{m=1}^M \mathbb{V}_m, \quad [p_{\min}, p_{\max}] = \sum_{m=1}^M \mathbb{P}_m.$$

Определите на интервалах  $[v_{\min}, v_{\max}]$  и  $[p_{\min}, p_{\max}]$  гистограммы как кусочно-постоянные функции  $p_{\square}^*$  и  $v_{\square}^*$  соответственно по формулам:

$$p_{\square}^*(v) = \begin{cases} p_1^*, & v \in \mathbb{V}_1; \\ \dots, & \dots; \\ p_m^*, & v \in \mathbb{V}_m; \\ \dots, & \dots; \\ p_M^*, & v \in \mathbb{V}_M; \end{cases} \quad (4.2)$$

$$v_{\square}^*(p) = \begin{cases} v_1^*, & p \in \mathbb{P}_1; \\ \dots, & \dots; \\ v_m^*, & p \in \mathbb{P}_m; \\ \dots, & \dots; \\ v_M^*, & p \in \mathbb{P}_M, \end{cases} \quad (4.3)$$

где значения гистограмм на  $m$ -ных подинтервалах определяются для  $m = 1, \dots, M$  формулами:

$$p_m^* = \sum_{v \in \mathbb{V}_m \cap \mathbb{V}} p^*(v), \quad (4.4)$$

$$v_m^* = \sum_{p \in \mathbb{P}_m \cap \mathbb{P}} v^*(p). \quad (4.5)$$

Примеры двух гистограмм распределений, сопровождающих  $\mathfrak{X}$ , показаны на рис. 3: гистограмма  $p_{\square}^*$  (по горизонтали красным цветом) и гистограмма  $v_{\square}^*$  (по вертикали фиолетовым цветом).

## 5 Справка: гиббсовская характеристизация смесей множеств событий

Представьте себе житейскую ситуацию, когда вас, живущего в гуще событий, разумеется, интересует, какому распределению подчиняются, например, некоторые особо важные для вас  $N$ -событий, образующие  $N$ -множество событий  $\mathfrak{X} \subseteq \mathcal{A}$ . Чтобы это выяснить вам ничего не остается делать, как некоторое

время наблюдать за всеми событиями из  $\mathfrak{X}$ . Сначала убедитесь, что результатом любого такого наблюдения служит ни что иное, как террасное событие  $I$ -го рода

$$\text{ter}(X//\mathfrak{X}) = \bigcap_{x \in X} x \bigcap_{x \in \mathfrak{X} - X} x^c \subseteq \Omega$$

для некоторого  $X \subseteq \mathfrak{X}$ , происходящее тогда, когда происходят только события из  $X$ , а остальные события из  $\mathfrak{X} - X$  не происходят. После этого вы сразу поймете, что результатом  $G$  наблюдений за всеми событиями из  $\mathfrak{X}$  всегда служит  $G$ -выборка террасных событий

$$\begin{aligned} \{\text{ter}(X_g//\mathfrak{X}), g = 1, \dots, G\} = \\ = \{\text{ter}(X_1//\mathfrak{X}), \dots, \text{ter}(X_G//\mathfrak{X})\}, \end{aligned} \quad (5.1)$$

и иного не дано.

Если вы, действительно, хотите разобраться в том, какому распределению подчиняются наблюдаемые события из  $\mathfrak{X}$ , то, имея  $G$ -выборку террасных событий (5.1), вам придется научиться оценивать вероятностное распределение  $I$ -го рода этого множества событий, т.е. набор

$$\{p(X//\mathfrak{X}), X \subseteq \mathfrak{X}\}, \quad (5.2)$$

вероятностей террасных событий

$$p(X//\mathfrak{X}) = \mathbf{P}(\text{ter}(X//\mathfrak{X}))$$

для всех подмножеств  $X \subseteq \mathfrak{X}$ , которых набирается приличное количество  $2^N$ , растущее экспоненциально при росте  $N$ .

Теперь, самое время задуматься о величине  $N$ , которая, конечно, полностью в вашем распоряжении и в зависимости от ваших запросов может достигать нескольких десятков, а то и сотен событий. Если ваши запросы скромны и не превышают, например, полутора десятков событий, то для достаточно надежной оценки<sup>7</sup> вероятностного распределения  $I$ -го рода вам может потребоваться в среднем не менее десятка миллионов наблюдений. Меньшая скромность потребует экспоненциально больше наблюдений. Это обычное дело для каждой  $NP$ -трудной задачи, когда перебор невозможен, а результат нужен.

Однако выход должен быть из любой ситуации, в том числе, и из этой, в которой, кроме вас, может оказаться любой желающий получить надежную оценку вероятностного распределения многомерного множества событий. Да, и ситуация не так уж безнадёжна. Отчетливый проблеск дает следующее соображение.

Наибольшую трудность для статистической оценки представляют такие  $N$ -множества событий, которые

<sup>7</sup>К примеру, при ошибке  $\pm 5\%$  с вероятностью не менее 0.95 для  $N = 14$ .

подчиняются распределениям, близким к равновероятному, т.е. такие, у которых террасные события происходят с вероятностями, близкими к  $1/2^N$ . А наименьшую трудность — такие, у которых лишь вполне приемлемое количество террасных событий происходит с ненулевой вероятностью, а все прочие не наступают никогда<sup>8</sup>. В промежуточной и довольно распространенной в событийной практике ситуации вы столкнетесь с оценкой множеств событий, распределение которых можно представить в виде взвешенной смеси одного из наиболее трудных с одним из наименее трудных для оценки распределений.

Чтобы попытаться воспользоваться этим проблеском, я предлагаю поразмыслить над методами исследования свойств гиббсовской характеристики смесей множеств событий при помощи *слоиных портретов*, которые помогут снять печать *NP*-трудных задач со статистических оценок больших множеств событий. Оценка *слоинового портрета* наблюдаемого множества событий позволит, на мой взгляд, сделать решающие выводы о том, представляет ли оно желанную смесь, а если представляет, то какова структура этой смеси и веса составляющих ее множеств событий.

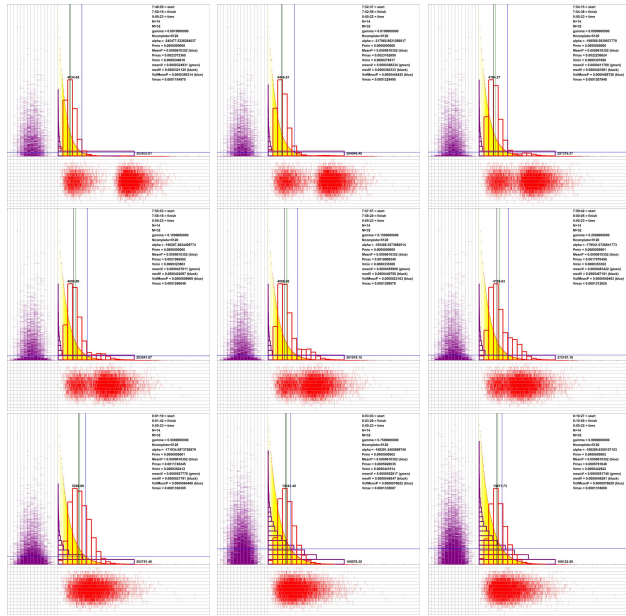


Рис. 4: *Слойные портреты* гиббсовской характеристики смеси 14-множеств событий  $\mathfrak{X}$  (почти равновероятного и неполного) для различных значений параметра смеси  $\gamma = 0.001, 0.010, 0.050; 0.100, 0.150, 0.250; 0.500, 0.750, 0.999$  (слева-направо, сверху-вниз).

В качестве иллюстрации этих идей рассмотрите пример (рис. 4), который демонстрирует возможности *слоиных портретов* в исследовании множеств событий, определяемых смесью почти равновероятного и неполного множеств событий.

*Метод слоиных портретов* может сыграть заметную

<sup>8</sup>Такие множества событий называются *неполными*.

роль во многих применениях эвентологической теории множеств событий. В частности, — в *эвентологическом факторном анализе*, который на событийном уровне решает те же задачи, что и, к примеру, классический метод главных компонент, разделяя террасные события, порождаемые наблюдаемым множеством событий, на *главные террасные события*, порождаемые неполным множеством событий, и *фоновые террасные события*, порождаемые почти равновероятным множеством событий.

## 6 Справка: гиббсовская характеристика вероятностных распределений различного рода

Рассмотрите множество событий  $\mathfrak{X}$  и каноническую гиббсовскую характеристику его вероятностного распределения  $I$ -го рода:

$$p(X//\mathfrak{X}) = \exp\{\alpha(p) v(X//\mathfrak{X})\}, \quad (6.1)$$

$$v(X//\mathfrak{X}) = \frac{1}{\alpha(p)} \ln p(X//\mathfrak{X}), \quad (6.2)$$

где

$$\alpha(p) = \sum_{Y:p(Y//\mathfrak{X})>0} \ln p(Y//\mathfrak{X}) \leq 0, \quad (6.3)$$

нормированным ценностным распределением  $I$ -го рода

$$\{v(X//\mathfrak{X}), X \subseteq \mathfrak{X}\}. \quad (6.4)$$

Как известно [1, 2], каждое множество событий может эквивалентным образом определено одним из вероятностных распределений  $I$ -го,  $II$ -го и т.д.  $VI$ -рода, в то время как остальные могут быть вычислены через него при помощи формул обращения Мебиуса. Для каждого из этих вероятностных распределений каноническая гиббсовская характеристика определяет ценностное распределение соответствующего рода.

Например, для вероятностного распределения  $II$ -рода соотношения канонической гиббсовской характеристики  $II$ -рода имеют вид:

$$p_{X//\mathfrak{X}} = \exp\{\alpha_p v_{X//\mathfrak{X}}\}, \quad (6.5)$$

$$v_{X//\mathfrak{X}} = \frac{1}{\alpha_p} \ln p_{X//\mathfrak{X}}, \quad (6.6)$$

где

$$\alpha_p = \sum_{X:p_{X//\mathfrak{X}}>0} \ln p_{X//\mathfrak{X}} \quad (6.7)$$

— *канонический гиббсовский параметр*  $II$ -го рода. А поскольку вероятностные распределения  $I$ -го и  $II$ -го рода связаны формулами обращения Мебиуса:

$$\begin{cases} p_{X//\mathfrak{X}} &= \sum_{Y \subseteq X} p(Y//\mathfrak{X}), \\ p(X//\mathfrak{X}) &= \sum_{Y \subseteq X} (-1)^{|Y-X|} p_{Y//\mathfrak{X}}, \end{cases} \quad (6.8)$$

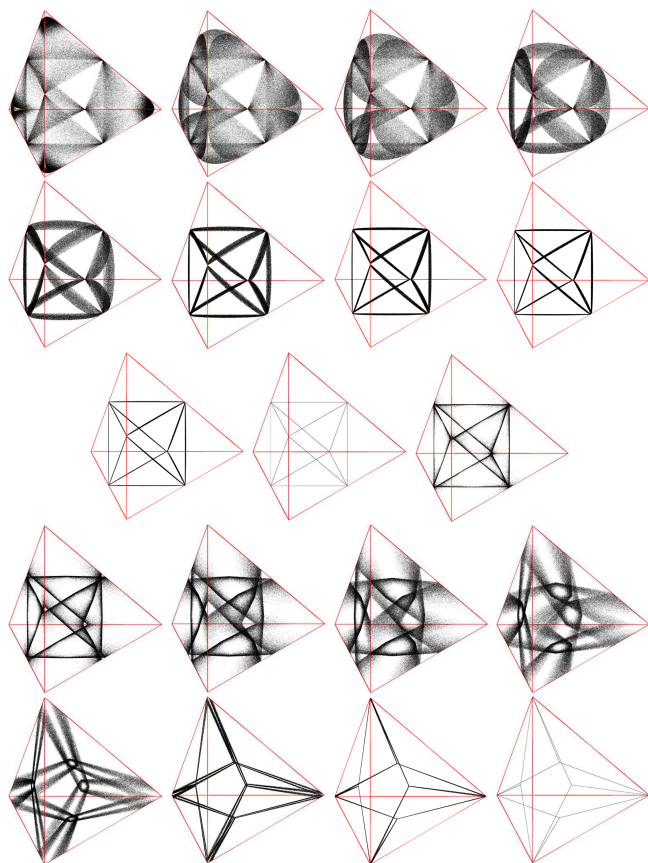


Рис. 5: Образы совокупности из 250 000 вероятностных распределений  $I$ -го рода дуплетов событий при отображении по формулам (6.6) и (6.7) в симплекс нормированных ценностных распределений  $II$ -го рода при различных значениях парной Фреше-корреляции дуплетных событий (слева направо, сверху вниз):  $-0.999999999, -0.99, -0.95, -0.75; -0.50, -0.25, -0.15, -0.10; -0.05, 0.0, 0.01; 0.05, 0.15, 0.25, 0.50; 0.75, 0.95, 0.99, 1.00$ .

то формулы (6.6), (6.7) и (6.8) определяет в итоге *гиббсовское отображение* симплекса вероятностных распределений  $I$ -го рода в симплекс ценностных распределений  $II$ -го рода. Этот факт, очевидно, справедлив и для распределений других родов с  $III$ -го, по  $VI$ -й.

**Задача.** Возможно вас заинтересует исследование структур многообразий ценностных распределений — образов *гиббсовского отображения* симплекса вероятностных распределений  $I$ -го рода в симплекс ценностных распределений того или иного рода.

**Численное решение методом Монте-Карло.** Для решения этой задачи методом Монте-Карло равномерно выберите из единичного  $2^N$ -вершинного симплекса  $2^N$ -мерную точку, которую при помощи случайной перестановки координат интерпретируйте как вероятностное распределение  $I$ -го рода  $N$ -множества событий  $\mathfrak{X}$ . Запаситесь изрядной совокупностью таких распределений. Для каждого распределения первого рода из этой совокупности по формулам обращения Мебиуса вычислите вероятностные

распределения остальных родов, а затем по канонической *гиббсовской* характеристике — соответствующие им нормированные ценностные распределения тех же родов. В итоге получите шесть образов этой совокупности точек из симплекса вероятностных распределений  $I$ -го рода в виде шести совокупностей точек в симплексах нормированных ценностных распределений соответствующих родов.

На рис. 5 показаны полученные этим методом примеры *гиббсовского отображения* совокупности из 500 000 вероятностных распределений  $I$ -го рода дуплетов событий в симплекс их нормированных ценностных распределений  $II$ -рода при различных значениях парной Фреше-корреляции дуплетных событий.

## Список литературы

- [1] О. Ю. Воробьев. *Эвентология*. Сибирский федеральный университет, Красноярск, 2007, 435с., <http://eventology-theory.ru/0-books/1-VorobyevOleg~2007~Eventology~435p.pdf>.
- [2] O. Yu. Vorobyev. *Eventology*. Siberian Federal University, Krasnoyarsk, Russia, 2007, 435p. (in Russian, abstract in English), <http://eventology-theory.com/0-books/1-VorobyevOleg~2007~Eventology~435p.pdf>.

# Аппроксимация множества событий в средневероятном

Олег Юрьевич Воробьев

Сибирский федеральный университет  
Институт математики и фундаментальной информатики  
Торгово-экономический институт  
Красноярск  
oleg.yu.vorobyev@gmail.com

**Аннотация.** Рассматривается аппроксимация множества событий в средневероятном, которая приближает его множеством средневероятных событий.

**Ключевые слова.** Эвентология, теория вероятностей, элементарный исход, состояние, событие, вероятность, множество состояний, множество событий, алгебра состояний, алгебра событий, средневероятное событие, аппроксимация в средневероятном.

*И странной близостью закованный,  
Смотрю за темную вуаль...*  
Александр Блок,  
Незнакомка, 24 апреля 1906, Озерки.

Средневероятное событие, недавно введенное в [1, 2], оказалось удобной событийной характеристикой множества событий для аппроксимации множества событий в средневероятном, которая позволяет оперировать в средневероятном со всем множеством событий как с одним событием.\* Задача приближения множества событий одним событием хоть и проста в формулировке, но требует для своего корректного решения новой и довольно неожиданной событийной интерпретации и затейливых\*\* деталей, добраться до которых мешают обычные для таких обстоятельств нагромождения привычных как в теории вероятностей, так и в эвентологии представлений о свойствах событий. Чтобы избавиться от шор обыденных представлений, полезно сначала основательно разобраться в эвентологическом толковании средневероятного события [1, 2].

Очевидным обобщением аппроксимации в средневероятном является приближение множества событий не одним средневероятным событием, а множеством средневероятных событий, имеющим мощ-

ность, меньшую мощности аппроксимируемого множества. Эту простую на вид\* идею можно развивать в нескольких направлениях.

## 1 Аппроксимация в средневероятном

Пусть  $(\Omega, \mathcal{A}, \mathbf{P})$  — всеобщее вероятностное пространство,  $\mathfrak{A} \subset \mathcal{A}$  и  $\mathfrak{B} \subset \mathcal{A}$  — конечные множества событий.

**Определение (равенство в средневероятном).** Говорят, что два множества событий  $\mathfrak{A}$  и  $\mathfrak{B}$  совпадают в средневероятном, если и только если

$$\hat{\mathfrak{A}} = \hat{\mathfrak{B}} \quad (1.1)$$

— их средневероятные события совпадают.

**Определение (аппроксимация в средневероятном).** Говорят, что множество событий  $\mathfrak{B}$  аппроксимирует в средневероятном множество событий  $\mathfrak{A}$ , если они совпадают в средневероятном и  $|\mathfrak{B}| \leq |\mathfrak{A}|$ ; или в сокращенной записи:<sup>1</sup>

$$\mathfrak{B} \stackrel{mp}{\prec} \mathfrak{A}. \quad (1.2)$$

## 2 Аппроксимация в средневероятном на основе покрытия множества событий

Это направление заключается в том, чтобы, опираясь на некоторое покрытие аппроксимируемого множества событий объединением (или суммой<sup>2</sup>) части его подмножеств, строить аппроксимацию в средневероятном в виде множества средневероятных событий для покрывающих его подмножеств событий.\*\*

\* Эта простота кажущаяся и — всего лишь “эвентологическая простота”. Если надлежащим образом разобраться, то аппроксимация в средневероятном — это не тривиальная замена одного объекта из  $N$  элементов другим объектом из  $M < N$  элементов, это замена вероятностного распределения аппроксимируемого  $N$ -множества событий  $\mathfrak{A}$  вероятностным распределением аппроксимирующего  $M$ -множества средневероятных событий  $\hat{\mathfrak{A}}$ , т.е. замена набора из  $2^N$  вероятностей  $\mathfrak{A}$ -террасных событий на набор из  $2^M$  вероятностей  $\hat{\mathfrak{A}}$ -террасных событий. Проклятие размерности продолжает торжествовать, и задача аппроксимации в средневероятном становится  $NP$ -трудной с ростом  $N$ .

<sup>1</sup> ‘mp’ от англ. the mean probable — средневероятное.

<sup>2</sup> Т.е. объединением непересекающихся подмножеств.

\*\* Видимо, толк от такой аппроксимации в средневероятном будет тогда, когда количество средневероятных множеств меньше количества событий в аппроксимируемом множестве и/или когда вероятностное распределение аппроксимирующего множества средневероятных событий определяется меньшим числом параметров (например, когда средневероятные события totally независимы).

\* Точнее, множеством из одного события.

\*\* Каждая такая “затейливость” помечена сноской\* или\*\*.



**Определение.** Пусть

$$\mathcal{C}_{\mathfrak{A}} \subseteq 2^{\mathfrak{A}} \quad (2.1)$$

— покрытие множества событий  $\mathfrak{A}$ , т.е. совокупность таких подмножеств  $A \subseteq \mathfrak{A}$ , что

$$\mathfrak{A} = \bigcup_{A \in \mathcal{C}_{\mathfrak{A}}} A. \quad (2.2)$$

Множество средневероятных событий

$$\widehat{\mathcal{C}}_{\mathfrak{A}} = \{\widehat{A}, A \in \mathcal{C}_{\mathfrak{A}}\} \subset \mathcal{A} \quad (2.3)$$

называется *аппроксимацией*  $\mathfrak{A}$  в *средневероятном* *относительно покрытия*  $\mathcal{C}_{\mathfrak{A}}$ , если и только если

$$\begin{cases} \widehat{\mathcal{C}}_{\mathfrak{A}} = \widehat{\mathfrak{A}}, \\ |\widehat{\mathcal{C}}_{\mathfrak{A}}| \leq |\mathfrak{A}|; \end{cases} \quad (2.4)$$

или в сокращенной записи:<sup>3</sup>

$$\widehat{\mathcal{C}}_{\mathfrak{A}} \overset{mp}{\prec} \mathfrak{A}, \quad (2.4')$$

которая означает, что множество средневероятных событий  $\widehat{\mathcal{C}}_{\mathfrak{A}}$  *аппроксимирует*  $\mathfrak{A}$  в *средневероятном* *относительно покрытия*  $\mathcal{C}_{\mathfrak{A}}$ .

Рассмотрим два “крайних” вида покрытия (2.1). Первый, когда покрытие (2.1) состоит всего из одного элемента, равного всему множеству  $\mathfrak{A}$ :

$$\mathcal{C}_{\mathfrak{A}} = \{\mathfrak{A}\}. \quad (2.1')$$

Относительно покрытия (2.1') аппроксимация  $\mathfrak{A}$  в средневероятном — это обычная аппроксимация  $\mathfrak{A}$  одним событием\*, т.е. множество из одного элемента — его средневероятного события  $\widehat{\mathfrak{A}}$ :

$$\{\widehat{\mathfrak{A}}\} \overset{mp}{\prec} \mathfrak{A} \quad (2.6)$$

так как

$$\widehat{\mathcal{C}}_{\mathfrak{A}} = \{\widehat{\mathfrak{A}}\}. \quad (2.7)$$

Второе, когда (2.1) состоит из моноплетных слагаемых:

$$\mathcal{C}_{\mathfrak{A}} = \{\{a\} : a \in \mathfrak{A}\}. \quad (2.1'')$$

Относительно покрытия (2.1'') аппроксимация  $\mathfrak{A}$  в средневероятном — это само множество  $\mathfrak{A}$ :

$$\mathfrak{A} \overset{mp}{\prec} \mathfrak{A}, \quad (2.8)$$

так как

$$\widehat{\mathcal{C}}_{\mathfrak{A}} = \{\{\widehat{a}\} : a \in \mathfrak{A}\} = \{a : a \in \mathfrak{A}\} = \mathfrak{A}. \quad (2.9)$$

<sup>3</sup> ‘mp’ от англ. the mean probable — средневероятное.

\* Точнее, *множеством из одного события*

### 3 Аппроксимация в средневероятном на основе вектор-событийного разбиения множества событий

Это направление отражает событийную структуру, типичную для приложений в области безопасности [3, 4, 5, 6, 7, 8], рынка [9, 10], дизайна [11] и многих других, где эвентологической моделью *совокупности однотипных явлений*\* служит совокупность упорядоченных наборов событий одинаковой размерности.

Подобные упорядоченные  $n$ -наборы ( $n$ -плеты) событий я буду здесь называть  *$n$ -вектор-событиями* и обозначать

$$\mathbf{x} = \{x^{(1)}, \dots, x^{(n)}\} \in \mathfrak{X}, \quad (3.1)$$

где

$$\mathfrak{X} = \{\mathbf{x} : \mathbf{x} \in \mathfrak{X}\} = \text{diag} \left( \bigotimes_{i=1}^n \mathfrak{X}^{(i)} \right) \quad (3.2)$$

— диагональ декартова произведения  $N$ -множеств событий

$$\mathfrak{X}^{(i)} = \{x^{(i)} : x \in \mathfrak{X}\} \subset \mathcal{A} \quad (3.3)$$

а  $\mathfrak{X}$  —  $N$ -множество их имен\*\*. Мне еще не раз понадобится обозначение для

$$\mathfrak{A} = \sum_{i=1}^n \mathfrak{X}^{(i)} \subset \mathcal{A} \quad (3.4)$$

— суммы всех  $N$ -множеств событий (3.3), чтобы

- сначала заметить, что

$$\mathfrak{A} = \sum_{\mathbf{x} \in \mathfrak{X}} \mathbf{x} = \sum_{\mathbf{x} \in \mathfrak{X}} \{x^{(1)}, \dots, x^{(n)}\} \quad (3.5)$$

—  $n$ -вектор-события (3.1) из диагонали декартова произведения (3.2) образуют разбиение множества событий  $\mathfrak{A}$  на упорядоченные  $n$ -плеты  $\mathbf{x} \in \mathfrak{X}$ ;

- а затем сформулировать в удобной форме задачу аппроксимации в средневероятном на основе вектор-событийного разбиения (3.5) множества событий (3.4).

Каждый  $n$ -вектор-событий  $\mathbf{x} \in \mathfrak{X}$  — это упорядоченное  $n$ -множество событий ( $n = |\mathbf{x}|$ ), которое порождает разбиение

$$\text{terr}(\mathbf{x}) = \{\text{ter}(\mathbf{x} // \mathbf{x}) : \mathbf{x} \subseteq \mathbf{x}\} \subset \mathcal{A} \quad (3.6)$$

пространства всеобщих элементарных исходов  $\Omega$  на  $\mathbf{x}$ -террасные события:

$$\Omega = \sum_{\mathbf{x} \subseteq \mathbf{x}} \text{ter}(\mathbf{x} // \mathbf{x}). \quad (3.7)$$

\* Представьте себе совокупность субъектов, поведение каждого из которых (явление) характеризуется с событийной точки зрения, присущи ему или не присущи свойства из, например, фиксированной пары свойств, одинаковой для всех субъектов. Тогда каждое такое явление (поведение субъекта) событийно моделируется парой событий, а совокупность всех однотипных явлений — совокупностью таких пар событий.

\*\*  $N$ -множество имен  $\mathfrak{X}$  изоморфно  $N$ -множеству  $n$ -вектор-событий  $\mathfrak{X}$ , которое, с свою очередь, изоморфно некоторой  $N$ -совокупности *однотипных явлений*, каждое из которых событийно моделируется соответствующим  $n$ -вектор-событием  $\mathbf{x} \in \mathfrak{X}$ ; в приведенном выше примере, где явление — это поведение субъекта, множеством имен может служить множество имен, например, субъектов.



Множество всех разбиений (3.6) обозначается

$$t(\mathfrak{X}) = \{\text{terr}(x) : x \in \mathfrak{X}\}. \quad (3.8)$$

В [1, 2], кроме понятия *средневероятного события* для *множества событий*, введено понятие *средневероятного разбиения* для *множества разбиений*, которое обладает более общими, хотя и аналогичными свойствами. По определению *средневероятное разбиение* для множества разбиений  $t(\mathfrak{X})$  (3.7) — это совокупность

$$t(\widehat{\mathfrak{X}}) = \{\widehat{\text{terr}(x)} : x \subseteq \mathfrak{X}\} \quad (3.9)$$

средневероятных террасных событий  $\widehat{\text{terr}(x)}$  для множеств террасных событий\*

$$\text{terr}(x) = \{\text{ter}(x//x) : x \in \mathfrak{X}\}. \quad (3.10)$$

В итоге построена аппроксимация в средневероятном:

$$t(\widehat{\mathfrak{X}}) \stackrel{mp}{\prec} t(\mathfrak{X}) \quad (3.11)$$

$N$ -множества  $(2^n)$ -разбиений<sup>4</sup>  $t(\mathfrak{X})$  одним средневероятным  $(2^n)$ -разбиением  $t(\widehat{\mathfrak{X}})$ .

#### 4 Условная аппроксимация в средневероятном

Пусть  $\mathfrak{U} \subset \mathcal{A}$  — некоторое конечное множество событий<sup>5</sup>.

**Определение (условная аппроксимация в средневероятном).** Условная аппроксимация  $\mathfrak{A}$  в средневероятном относительно  $\mathfrak{U}$  определяется как аппроксимация  $\mathfrak{A}$  множеством *условных средневероятных событий*

$$\mathfrak{A}_{|\mathfrak{U}} = \{\widehat{\mathfrak{A}}|U : U \subseteq \mathfrak{U}\}, \quad (4.1)$$

или в сокращенной записи

$$\mathfrak{A}_{|\mathfrak{U}} \stackrel{mp}{\prec} \mathfrak{A}, \quad (4.2)$$

где

$$\mathfrak{A}|U = \{a|U : a \in \mathfrak{A}\} \quad (4.2)$$

— множество *условных событий* [12, 13] при условии  $\mathfrak{U}$ -террасного события  $\text{ter}(U//\mathfrak{U})$ .

#### 5 Справка: некоторые особенности статистической оценки средневероятного события

Статистическая оценка средневероятного события имеет некоторые особенности, связанные с тем, что

\* Внимательно сравните (3.6) и (3.10), чтобы найти существенные отличия.

<sup>4</sup> Обычно  $n$  достаточно мало и вполне может быть равно 2, а  $N$  чаще всего большое число — сотни, тысячи, а то и больше.

<sup>5</sup>  $\mathfrak{U}$ -террасным событиям предназначено здесь играть роль полной группы событий.

статистика наблюдений за любым множеством событий всегда содержит полиномиальное число наблюдений, в то время как вероятностное распределение множества событий состоит из экспоненциального числа вероятностей, подлежащих статистической оценке. Так что любой алгоритм оценки средневероятного события должен строиться лишь из расчета на полиномиальную статистику.\*

Еще одна особенность возникает из-за того, что за редким исключением средневероятное событие неединственно и поэтому под оценкой приходится понимать либо оценку всего класса событий, которые являются средневероятными, либо оценку какого-то одного из таких средневероятных событий. Статистическая оценка класса всех средневероятных событий сводится к оценке двух событий — минимального и максимального средневероятных событий, которые определяют статистическую оценку всего класса отношениями включения. Однако на практике всегда удобнее иметь единственную оценку, в качестве которой выбирается такое событие из оцениваемого класса, которое определяется по стандартному правилу, приводящему всегда к единственному результату — средневероятному событию, или ближайшему к нему [1, 2].

Исчерпывающий анализ перечисленных особенностей заслуживает отдельного внимания. Для целей же этого исследования достаточно очевидного общего вывода, что в качестве такой статистической оценки достаточно событий из узкого класса\*\*, определяемых как результат некоторой теоретико-множественной операции над данным множеством событий, т.е. — как сумма некоторых террасных событий, порожденных оцениваемым множеством событий; иными словами, для статистической оценки достаточно событий из алгебры, порожденной данным множеством событий. Следовательно,

$$\overline{\mathfrak{A}} \in \mathcal{A}_{\mathfrak{A}} \quad (5.1)$$

— статистическая оценка  $\overline{\mathfrak{A}}$  средневероятного события  $\mathfrak{A}$  для множества  $\mathfrak{A}$  принадлежит алгебре событий  $\mathcal{A}_{\mathfrak{A}}$ , порожденной этим множеством, т.е. служит результатом теоретико-множественной операции над множеством событий  $\mathfrak{A}$ :

$$\overline{\mathfrak{A}} = \text{terop}(\mathcal{O} // 2^{\mathfrak{A}}) = \sum_{A \in \mathcal{O}} \text{ter}(A // \mathfrak{A}), \quad (5.2)$$

которая определяется некоторой совокупностью подмножеств событий  $\mathcal{O} \subseteq 2^{\mathfrak{A}}$ .

\* Подробнее см. [14, 15, стр. 156].

\*\* Это, казалось бы, сильное предположение — вполне естественное следствие неизбежного компромисса между событийными возможностями теории и практики. Если вы заинтересовались множеством событий, именно, как множество событий в строгом эвентологическом смысле, а не как чем-то недостаточно определенным, чтобы об этом имело смысл интересоваться “эвентологически”; то приготовьтесь к тому, что возможности ваших статистических исследований неизбежно будут ограничены довольно простым набором инструментов — террасными событиями, которые порождены этим множеством событий. Если вы сможете “эвентологически” интересоваться большим множеством событий, разрешающая способность ваших событийных инструментов станет возрастать в два раза с каждым новым событием, которое вам удастся удерживать в поле вашего эвентологического зрения.

## Список литературы

- [1] О. Ю. Воробьев. Событийные средние в эвентологии, их асимптотические свойства, интерпретация и визуализация. *Труды XVI Международ. ЭМ конференции по эвентологической математике и смежным вопросам*, Красноярск: СФУ (под ред. Олега Воробьева):50–56, 2012.
- [2] O. Yu. Vorobyev. Event means in eventology, its asymptotic properties, interpretations and visualization. In. *Proc. of the XVI Intern. EM conference on eventological mathematics and related fields*, Krasnoyarsk: SFU (Oleg Vorobyev ed.):50–56, 2012 (in Russian, abstract in English).
- [3] О. Ю. Воробьев. Эвентологический системный анализ безопасности в картинках. *Труды XI Международ. ФАМЭБ конференции по финансово-актуарной математике и эвентологии безопасности*, Красноярск: НИИППБ, СФУ (под ред. Олега Воробьева):113–125, 2012.
- [4] О. Ю. Воробьев. Выразить и измерить: успехи применимой эвентологии. *Труды XVI Международ. ЭМ конференции по эвентологической математике и смежным вопросам*, Красноярск: СФУ (под ред. Олега Воробьева):44–49, 2012.
- [5] О. Ю. Воробьев. Применимая эвентология безопасности: неокончательные итоги. *Труды XII Международ. ФАМЭБ конференции по финансово-актуарной математике и эвентологии безопасности*, Красноярск: НИИППБ, СФУ (под ред. Олега Воробьева):105–115, 2013.
- [6] O. Yu. Vorobyev. Eventological system analysis of safety in pictures. In. *Proc. of the XI Intern. FAMES Conf. on Financial and Actuarial Mathematics and Eventology of Safety*, Krasnoyarsk, SFU (Oleg Vorobyev ed.):113–125, 2012 (in Russian, abstract in English).
- [7] O. Yu. Vorobyev. To express and to measure: successes of applicable eventology. In. *Proc. of the XVI Intern. EM conference on eventological mathematics and related fields*, Krasnoyarsk: SFU (Oleg Vorobyev ed.):44–49, 2012 (in Russian, abstract in English).
- [8] O. Yu. Vorobyev. Applicable eventology of safety: inconclusive totals. In. *Proc. of the XII Intern. FAMES Conf. on Financial and Actuarial Mathematics and Eventology of Safety*, Krasnoyarsk, SFU (Oleg Vorobyev ed.):40–49, 2013.
- [9] Е. Е. Голденок. О безопасности событийного рынка. *Труды XVI Международ. ЭМ конференции по эвентологической математике и смежным вопросам*, Красноярск: СФУ (под ред. Олега Воробьева):78–83, 2012.
- [10] E. E. Goldenok. About safety of event-market. In. *Proc. of the XVI Intern. EM conference on eventological mathematics and related fields*, Krasnoyarsk: SFU (Oleg Vorobyev ed.):78–83, 2012 (in Russian, abstract in English).
- [11] O. Yu. Vorobyev. Eventology of design. *Proc. of the VII Intern. Mathematics & Design M&D-2013 Conference*, Tucumán, Argentina, Universidad Nacional de Tucumán, 2013, September 2-6, 2013.
- [12] О. Ю. Воробьев. Условные события. *Труды XII Международ. ФАМЭБ конференции по финансово-актуарной математике и эвентологии безопасности*, Красноярск: НИИППБ, СФУ (под ред. Олега Воробьева):164–168, 2013.
- [13] O. Yu. Vorobyev. Conditional events. In. *Proc. of the XII Intern. FAMES Conf. on Financial and Actuarial Mathematics and Eventology of Safety*, Krasnoyarsk, SFU (Oleg Vorobyev ed.):164–168, 2013 (in Russian, abstract in English).
- [14] О. Ю. Воробьев. Гиббсовская характеристика множеств событий. *Труды XII Международ. ФАМЭБ конференции по финансово-актуарной математике и эвентологии безопасности*, Красноярск: НИИППБ, СФУ (под ред. Олега Воробьева):151–158, 2013.
- [15] O. Yu. Vorobyev. Gibbsian characterization of sets of events. In. *Proc. of the XII Intern. FAMES Conf. on Financial and Actuarial Mathematics and Eventology of Safety*, Krasnoyarsk, SFU (Oleg Vorobyev ed.):151–158, 2013 (in Russian, abstract in English).

## Портфельный анализ событий: расширенная версия

Олег Юрьевич Воробьев

Сибирский федеральный университет  
Институт математики и фундаментальной информатики  
Торгово-экономический институт  
Красноярск  
oleg.yu.vorobyev@gmail.com

Александр Сергеевич Нифонтов

Сибирский федеральный университет  
Институт математики и фундаментальной информатики  
Красноярск  
nialeksander@gmail.com

**Аннотация.** Рассматриваются две основные задачи портфельного анализа событий — прямая и обратная. Первая из них — событийный аналог задачи Марковица из классического портфельного анализа, а вторая решает задачу Марковица в обратной событийной постановке, когда фиксированы доли капитала, вложенного в портфель событий, а ищется вероятностное распределение доходности этого портфеля при тех же ограничениях, что и в прямой задаче. Постановка обратной эвентологической портфельной задачи — нововведение эвентологии в портфельном анализе: до сих пор ее постановка, по-видимому, вообще не рассматривалась в рамках классического портфельного анализа, хотя так же, как и прямая, обратная портфельная задача имеет очевидное практическое значение.

**Ключевые слова.** Эвентология, теория вероятностей, событие, вероятность, портфельный анализ, портфельный анализ событий.

### 1 Прямая эвентологическая задача Марковица

Рассмотрим прямую эвентологическую задачу Марковица, иными словами, *прямую эвентологическую портфельную задачу*, в которой необходимо найти неизвестные доли событий в портфеле при известном вероятностном распределении множества этих портфельных событий [1, 2, стр. 275–287].

Пусть  $\{\mathfrak{X}(\mathbf{p}), \mathbf{a}\}$  — портфель событий, где  $\mathfrak{X}$  — множество портфельных событий с известным вероятностным распределением

$$\mathbf{p} = \{p(X) : X \subseteq \mathfrak{X}\},$$

$$\mathbf{a} = \left\{ a_x : a_x \geq 0, x \in \mathfrak{X}, \sum_{x \in \mathfrak{X}} a_x = 1 \right\}$$

— множество неизвестных долей этих событий в портфеле.

Предполагается, что доходности портфельных событий  $x \in \mathfrak{X}$  определяются их индикаторами:  $r_x = \mathbf{1}_x$ .

Тогда средняя доходность каждого портфельного события  $x \in \mathfrak{X}$  является вероятностью этого события:

$$\mathbf{E}r_x = \mathbf{E}\mathbf{1}_x = \mathbf{P}(x),$$

дисперсия доходности портфельного события равна

$$\mathbf{D}r_x = \mathbf{D}\mathbf{1}_x = \mathbf{P}(x)(1 - \mathbf{P}(x)) = \sigma_x^2,$$

где  $\sigma_x^2 = \mathbf{D}\mathbf{1}_x$  — дисперсия индикатора портфельного события  $x \in \mathfrak{X}$ . Доходность всего портфеля событий полагается равной линейной комбинации доходностей его событий с долями из  $\mathbf{a}$ :

$$r_{\mathfrak{X}} = \sum_{x \in \mathfrak{X}} a_x \mathbf{1}_x.$$

Тогда средняя доходность портфеля событий определяется вероятностями его событий — это выпуклая комбинация вероятностей событий с коэффициентами, равными долям этих событий в портфеле:

$$\mathbf{E}(r_{\mathfrak{X}}) = \sum_{x \in \mathfrak{X}} a_x \mathbf{P}(x), \quad (1)$$

а дисперсия — это квадратичная форма от  $\mathbf{a}$ :

$$\mathbf{D}(r_{\mathfrak{X}}) = \sum_{x \in \mathfrak{X}} \sum_{y \in \mathfrak{X}} a_x a_y \text{Kov}_{xy}, \quad (2)$$

с коэффициентами, равными  $\text{Kov}_{xy} = \mathbf{P}(x \cap y) - \mathbf{P}(x)\mathbf{P}(y)$  — парной ковариации событий  $x$  и  $y$ .

Постановка прямой эвентологической портфельной задачи выглядит стандартным образом: *найти такое множество долей  $\mathbf{a}$  событий в данном портфеле с известным вероятностным распределением  $\mathbf{p}$ , которое обеспечивает данное фиксированное значение средней доходности портфеля при его минимальном риске (дисперсии портфеля):*

$$\begin{aligned} \mathbf{E}(r_{\mathfrak{X}}) &= \langle r_{\mathfrak{X}} \rangle, \\ \mathbf{D}(r_{\mathfrak{X}}) &\rightarrow \min_{\mathbf{a}}. \end{aligned} \quad (3)$$

### 2 Обратная эвентологическая портфельная задача

Рассмотрим обратную эвентологическую портфельную задачу, в которой необходимо найти неизвестное

вероятностное распределение множества портфельных событий при известных долях этих событий в портфеле [1, 2, стр. 275–287].

Пусть  $\{\mathfrak{X}(\mathbf{p}), \mathbf{a}\}$  — портфель событий, где  $\mathfrak{X}$  — множество портфельных событий с неизвестным вероятностным распределением

$$\mathbf{p} = \{p(X) : X \subseteq \mathfrak{X}\},$$

$$\mathbf{a} = \left\{ a_x : a_x \geq 0, x \in \mathfrak{X}, \sum_{x \in \mathfrak{X}} a_x = 1 \right\}$$

— множество известных долей этих событий в портфеле.

Как и в прямой задаче, предполагается, что доходности портфельных событий  $x \in \mathfrak{X}$  определяются их индикаторами:  $r_x = \mathbf{1}_x$ . Тогда средняя доходность каждого портфельного события  $x \in \mathfrak{X}$  является неизвестной вероятностью этого события:

$$\mathbf{E}r_x = \mathbf{E}\mathbf{1}_x = \mathbf{P}(x),$$

дисперсия доходности портфельного события, равна

$$\mathbf{D}r_x = \mathbf{D}\mathbf{1}_x = \mathbf{P}(x)(1 - \mathbf{P}(x)) = \sigma_x^2,$$

где  $\sigma_x^2 = \mathbf{D}\mathbf{1}_x$  — дисперсия индикатора портфельного события  $x \in \mathfrak{X}$ . Доходность всего портфеля событий полагается равной выпуклой комбинации доходностей его событий с известными долями из  $\mathbf{a}$ :

$$r_{\mathfrak{X}} = \sum_{x \in \mathfrak{X}} a_x \mathbf{1}_x = \sum_{x \in \mathfrak{X}} a_x r_x.$$

Тогда средняя доходность портфеля событий определяется неизвестными вероятностями его событий по формуле:

$$\mathbf{E}(r_{\mathfrak{X}}) = \sum_{x \in \mathfrak{X}} a_x \mathbf{P}(x),$$

а дисперсия — формулой:

$$\mathbf{D}(r_{\mathfrak{X}}) = \sum_{x \in \mathfrak{X}} \sum_{y \in \mathfrak{X}} a_x a_y \text{Kov}_{xy},$$

где  $\text{Kov}_{xy} = \mathbf{P}(x \cap y) - \mathbf{P}(x)\mathbf{P}(y)$  — парная ковариация событий  $x$  и  $y$ .

Стандартная постановка обратной эвентологической задачи выглядит следующим образом: *найти такое вероятностное распределение множества портфельных событий  $\mathbf{p}$ , которое при известном фиксированном множестве долей  $\mathbf{a}$  событий в портфеле обеспечивает данное фиксированное значение средней доходности портфеля при его минимальном риске (дисперсии портфеля):*

$$\begin{aligned} \mathbf{E}(r_{\mathfrak{X}}) &= \langle r_{\mathfrak{X}} \rangle, \\ \mathbf{D}(r_{\mathfrak{X}}) &\rightarrow \min_{\mathbf{p}}. \end{aligned} \quad (4)$$

### 3 Обобщенные постановки: доходность события как линейная функция его индикатора

Рассмотрим линейное обобщение прямой и обратной эвентологических портфельных задач, когда доходность события определяется как линейная функция его индикатора [3].

Пусть  $(\mathfrak{X}, \mathbf{a})$  — портфель событий, где  $\mathfrak{X}$  — множество портфельных событий с вероятностным распределением

$$\mathbf{p} = \{p(X) : X \subseteq \mathfrak{X}\},$$

а

$$\mathbf{a} = \left\{ a_x : a_x \geq 0, x \in \mathfrak{X}, \sum_{x \in \mathfrak{X}} a_x = 1 \right\}$$

— множество долей этих событий в портфеле.

Предполагается, что доходности портфельных событий  $x \in \mathfrak{X}$  определяются линейными функциями их индикатора:

$$r_x = b_x \mathbf{1}_x + d_x.$$

Тогда средняя доходность каждого портфельного события  $x \in \mathfrak{X}$  является линейной функцией вероятности события:

$$\mathbf{E}r_x = b_x \mathbf{E}\mathbf{1}_x + d_x = b_x \mathbf{P}(x) + d_x,$$

дисперсия доходности портфельного события равна

$$\mathbf{D}r_x = b_x^2 \mathbf{D}\mathbf{1}_x = b_x^2 \mathbf{P}(x)(1 - \mathbf{P}(x)) = b_x^2 \sigma_x^2,$$

где  $\sigma_x^2 = \mathbf{D}\mathbf{1}_x$  — дисперсия индикатора портфельного события  $x \in \mathfrak{X}$ .

Доходность всего портфеля событий полагается равной выпуклой комбинации доходностей его событий с долями из  $\mathbf{a}$ :

$$r_{\mathfrak{X}} = \sum_{x \in \mathfrak{X}} a_x b_x \mathbf{1}_x + \sum_{x \in \mathfrak{X}} a_x d_x.$$

Тогда средняя доходность портфеля событий определяется вероятностями его событий по формуле:

$$\mathbf{E}(r_{\mathfrak{X}}) = \sum_{x \in \mathfrak{X}} a_x b_x \mathbf{P}(x) + \sum_{x \in \mathfrak{X}} a_x d_x, \quad (5)$$

а дисперсия — формулой:

$$\mathbf{D}(r_{\mathfrak{X}}) = \sum_{x \in \mathfrak{X}} \sum_{y \in \mathfrak{X}} a_x b_x a_y b_y \text{Kov}_{xy}, \quad (6)$$

где  $\text{Kov}_{xy} = \mathbf{P}(x \cap y) - \mathbf{P}(x)\mathbf{P}(y)$  — парная ковариация событий  $x$  и  $y$ .

Линейно обобщенная постановка прямой эвентологической задачи выглядит стандартным образом: *найти такое множество долей  $\mathbf{a}$  событий в данном портфеле с известным фиксированным вероятностным распределением  $\mathbf{p}$ , которое обеспечивает*



данное фиксированное значение средней доходности портфеля при его минимальном риске (дисперсии портфеля):

$$\begin{aligned} E(r_{\mathfrak{X}}) &= \langle r_{\mathfrak{X}} \rangle, \\ D(r_{\mathfrak{X}}) &\rightarrow \min_{\mathbf{a}}. \end{aligned} \quad (7)$$

Соответственно линейно обобщенная постановка обратной эвентологической задачи также выглядит стандартным образом: найти такое вероятностное распределение множества портфельных событий  $\mathbf{p}$ , которое при известном фиксированном множестве долей  $\mathbf{a}$  событий в портфеле обеспечивает данное фиксированное значение средней доходности портфеля при его минимальном риске (дисперсии портфеля):

$$\begin{aligned} E(r_{\mathfrak{X}}) &= \langle r_{\mathfrak{X}} \rangle, \\ D(r_{\mathfrak{X}}) &\rightarrow \min_{\mathbf{p}}. \end{aligned} \quad (8)$$

#### 4 Визуальные особенности и сравнение портфельных эвентологических пули и зонтика

**Портфельная пуля (прямая задача).** В отличие от стандартной эвентологической прямой портфельной задачи, для которой все “ноги” портфельной пули лежат на одной полуокружности единичного диаметра с центром в точке  $(0, 1/2)$  плоскости  $(\sigma_{\mathfrak{X}}, \mathbf{E}_{\mathfrak{X}})$ , где  $\sigma_{\mathfrak{X}} = \sqrt{D(r_{\mathfrak{X}})}$  — среднее квадратичное отклонение, а  $\mathbf{E}_{\mathfrak{X}} = E(r_{\mathfrak{X}})$  — математическое ожидание доходности портфеля  $r_{\mathfrak{X}}$ ; в линейно обобщенной постановке (7) каждая “нога” портфельной пули, соответствующая событию  $x \in \mathfrak{X}$ , лежит на своей полуокружности радиуса  $b_x/2$  с центром в точке  $(0, a_x b_x/2 + a_x d_x)$  той же плоскости.

Следует заметить, что портфельная пуля на данной плоскости является образом симплекса, состоящего из точек  $|\mathfrak{X}|$ -мерного пространства, которые определяются всеми возможными множествами долей  $\mathbf{a}$  (для какого-то произвольно выбранного их порядка в этом множестве), при соответствующем отображении на плоскость  $(\sigma_{\mathfrak{X}}, \mathbf{E}_{\mathfrak{X}})$ , определяемом формулами (1) и (2)<sup>1</sup>.

**Портфельный зонтик (обратная задача).** В отличие от стандартной эвентологической обратной портфельной задачи, для которой все “спицы” портфельного зонтика лежат на отрезке  $[0, 1]$  вертикальной оси плоскости  $(\sigma_{\mathfrak{X}}, \mathbf{E}_{\mathfrak{X}})$ , а весь зонтик лежит в пределах полукруга единичного диаметра с центром в точке  $(0, 1/2)$  той же плоскости; в линейно обобщенной постановке (8) каждая “спица” портфельного зонтика, соответствующая террасному событию

$\text{ter}(X//\mathfrak{X})$ ,  $X \subseteq \mathfrak{X}$ , лежит на отрезке

$$[\delta_{\mathfrak{X}}, \beta_{\mathfrak{X}}] = \left[ \sum_{x \in \mathfrak{X}} a_x d_x, \sum_{x \in \mathfrak{X}} a_x (b_x + d_x) \right]$$

вертикальной оси этой плоскости и определяется на ней координатами для  $X \subseteq \mathfrak{X}$

$$\alpha_X = \sum_{x \in X} a_x b_x + \sum_{x \in \mathfrak{X}} a_x d_x.$$

Портфельный зонтик формируется полуокружностями соединяющими каждую пару таких “спиц”. При этом полуокружности, соединяющие соседние “спицы”, вместе с полуокружностью, соединяющей пару крайних “спиц” (верхнюю и нижнюю), образуют границу портфельного зонтика.

Портфельный зонтик на данной плоскости является образом симплекса вероятностных распределений, состоящего из точек  $2^{|\mathfrak{X}|}$ -мерного пространства, которые определяются возможными распределениями  $\mathbf{p}$  (для какого-то произвольно выбранного их порядка), при соответствующем отображении на плоскость  $(\sigma_{\mathfrak{X}}, \mathbf{E}_{\mathfrak{X}})$ , определяемом формулами (1) и (2)<sup>1</sup>.

\*\*\*

*Эвентологическая пуля* и обратный ей *эвентологический зонтик* одной и той же портфельной задачи очевидным образом соотносятся друг с другом на плоскости  $(\sigma_{\mathfrak{X}}, \mathbf{E}_{\mathfrak{X}})$ . Так любая точка эвентологической пули, соответствующая некоторому распределению долей событий в портфеле, всегда содержится в пределах эвентологического зонтика обратной задачи, в которой доли событий зафиксированы именно таким образом. Остальные точки этой пули могут, разумеется, оказаться и за пределами данного зонтика. Однако все соответствующие событиям портфеля в прямой задаче полуокруги, на граничных полуокружностях которых лежат “ноги” пули, всегда содержатся в пределах наибольшего полукруга, построенного на “максимальной и минимальной спицах” зонтиков всех обратных задач (т.е. среди зонтиков обратных задач для всех вариантов фиксированных долей).

#### Список литературы

- [1] О. Ю. Воробьев. *Эвентология*. Сибирский федеральный университет, Красноярск, 2007, 435с., <http://eventology-theory.ru/0-books/1-VorobyevOleg~2007~Eventology~435p.pdf>.
- [2] O. Yu. Vorobyev. *Eventology*. Siberian Federal University, Krasnoyarsk, Russia, 2007, 435p. (in Russian, abstract in English), <http://eventology-theory.com/0-books/1-VorobyevOleg~2007~Eventology~435p.pdf>.
- [3] О. Ю. Воробьев. Расширенная версия портфельного эвентологического анализа. *Заметки ЭМ-семинара*, 35:1–10. — Красноярск: СФУ. 28 октября, 2011.

<sup>1</sup>Многочисленные примеры визуализации портфельных эвентологических пули и зонтика для стандартной постановки задач можно найти в [1, 2, стр. 275–287]



## Статистический анализ внутридневных колебаний спреда на российском фондовом рынке

Глик Людмила Андреевна

Кафедра высшей математики и математической физики  
Томский политехнический университет  
Томск  
gtlkla@tpu.ru

Семенов Михаил Евгеньевич

Кафедра высшей математики и математической физики  
Томский политехнический университет  
Томск  
sme@tpu.ru

**Аннотация.** Статья посвящена статистическому анализу внутридневных колебаний спреда на российском фондовом рынке. Приводится обзор существующих моделей, с помощью которых можно спрогнозировать дальнейшую динамику котировок ценной бумаги. Определяются основные свойства, которыми обладает временной ряд прироста спреда и на основании полученных результатов делается вывод о наиболее подходящей модели для прогнозирования динамики цен.

**Ключевые слова.** Рынок ценных бумаг, spread, стационарность, автокорреляция, динамика цен.

### 1 Введение

В экономике, как и в большинстве других научных дисциплин, отсутствуют абсолютно точные утверждения. Любое эмпирическое утверждение имеет вероятностную природу. В частности, экономические измерения содержат различного рода ошибки. В прикладных экономических исследованиях требуется использовать статистические методы.

В связи с экономическими ошибками измерений, анализа или прогноза существует высокий риск неверной интерпретации данных, что обычно приводит к большим финансовым потерям. По данной причине необходимо корректно выбрать математическую модель, что позволит существенно сократить риск убытков.

Методы эконометрики, финансовой математики, теории вероятностей и математической статистики, позволяющие проводить эмпирическую проверку теоретических утверждений и моделей, выступают мощным инструментом для анализа финансовых и фондовых рынков [1].

Появление и развитие электронной торговли через сеть Интернет, как главного средства торговли финансовыми активами, сделало исследование очереди заявок одним из главных механизмов понимания ценообразования на фондовом рынке. Под очередью заявок будем понимать список всех неисполненных либо частично исполненных заявок. В настоящее время

выделяют следующие основные типы заявок: лимитированные, рыночные, условные. Для этого можно анализировать описательные статистики, проверять данные по различным вероятностным показателям (коррелированность данных, нормальность распределения, стационарность ряда и многое другое). По результатам полученных расчетов можно построить математические модели, определить наличие трендов, построить прогнозы движения цены финансового инструмента.

Правильно выполненный технический анализ позволит сделать верные прогнозы по отношению к движению цены на актив, а соответственно получить прибыль от работы на фондовом рынке. Таким образом, риски принятия неверного решения и возможные убытки существенно сокращаются.

На рынке уже существует множество признаков, используемых трейдерами для принятия решения об открытии или закрытии позиций. Однако каждый из них, как правило, работает в строго ограниченных условиях (например, когда идет долгосрочный рост котировки либо ожидается выпуск дополнительной эмиссии компании).

Автоматизация биржевой торговли затруднительна ввиду большого числа факторов, влияющих на динамику рынка и их разнородности. Наряду с количественными факторами, такими как объемы продаж, количество сделок, текущий уровень цен, биржевые индексы, влияние на рынок оказывают качественные показатели, такие как выходящие новости, сезонные тенденции, отчеты ведущих финансовых организаций. В связи с этим, задача вывода является организационно-сложной и использование многих подходов ограничено характером и различной природой доступных ценовых индикаторов, а также тем, что определение значений части признаков возможно лишь с некоторой степенью уверенности.

По этим причинам необходимо детально разобраться с подходами к моделированию очереди заявок, для сокращения числа рассматриваемых факторов и облегчения принятия решения. Данная работа является продолжением исследований [2].

## 2 Обзор подходов к моделированию очереди заявок в таблице котировок

В литературе описаны различные подходы моделирования очереди заявок в таблице котировок: имитационное моделирование [3], методов распознавания образов [4], использование искусственных нейронных сетей [5], моделирование с помощью цепей Маркова и пуассоновских процессов [6], модель с учетом эффекта памяти [7], поведенческая модель [8].

В работе [3] авторы приводят краткое описание имитационного моделирования биржевых торгов. На этапе построения модели авторы используют полную информацию о позициях и операциях трейдеров. Заметим, что этой конфиденциальной информацией обладает только организатор торгов – биржа, что существенно ограничивает применение данного подхода. На выходе вся персональная информация обезличивается в соответствии с требованием сохранения банковской тайны.

В работе [4] производится описание, анализ и прогнозирование котировок ценных бумаг с помощью обобщенного спектрально-аналитического метода. Авторы предлагают математическую процедуру прогнозирования вектора доходности, используя аналитическое описание котировок ценных бумаг.

В статье [5] рассматривается применение искусственных нейронных сетей на американском фондовом рынке. За основу модели взяты карты самоорганизации (сеть Кохонена) с использованием методов главных компонент и различными вариантами обучения. Анализируется возможность прогнозирования фондового рынка на длительном промежутке времени – для этого в модель добавляется синергетическая сеть Хакена и закладываются элементы из теории информации.

Моделирование очереди заявок с помощью цепей Маркова и пуассоновских процессов детально описано в статье [6]. В данной работе приводится математическая модель динамики очереди заявок, модель развития очереди заявок, а также ценовая динамика в предположении, что очередь заявок может быть описана Марковскими случайными процессами. Данное предположение накладывает ряд ограничений на временной ряд, такие как стационарность и эргодичность.

Авторами работы [7] приводится модель динамики спреда доходности и спреда волатильности с учетом эффекта памяти и мультифрактальной природы. На основе автокорреляционной функции и анализа трендов строится прогнозная модель.

В статье [8] описывается поведенческая модель ликвидности и волатильности на основе эмпирических

закономерностей очереди заявок. Данная модель используется для моделирования формирования цен в рамках непрерывного двойного аукциона. Авторы оценивают стратегические факторы, влияющие на величину спреда.

### Постановка цели и задач исследования

Целью данной работы является статистический анализ данных таблицы котировок для выбора корректной математической модели для прогнозирования динамики цены.

Для достижения поставленной цели необходимо решить следующие задачи: 1) найти описательные статистики; 2) проверить данные на автокорреляцию; 3) проверить данные на стационарность.

Для того чтобы определиться с выбором конкретной модели, для дальнейшего прогнозирования динамики цен, необходимо оценить исходный ряд данных и определить его основные свойства.

## 3 Теоретические положения

В данном разделе приведем основные определения и математические формулы, которые необходимы для реализации поставленной цели.

В течение торговой сессии участники торгов в произвольном порядке подают заявки. Каждая заявка характеризуется направленностью (на покупку или на продажу), ценой и объемом. Для каждой поданной заявки торговый автомат просматривает очередь заявок, поданных ранее, пытаясь найти удовлетворяющую ей встречную заявку (цена заявки на покупку должна быть не меньше, чем цена заявки на продажу).

Таблица котировок в торговом терминале позволяет оценить спрос и предложение на рынке в текущий момент времени. Как правило, в торговых терминалах (например, Quik, Transaq) отображается не вся таблица котировок, а 20 самых лучших в данный момент. Важным показателем в таблице котировок является понятие спреда (spread).

**Определение 37.** Под спредом  $S$  будем понимать разницу между лучшей ценой продажи  $P^A$  и лучшей ценой покупки  $P^B$  в текущий момент времени  $t$  [6]:  $S(t) = P^A(t) - P^B(t)$ .

В данной работе анализируется прирост спреда. Под величиной прироста спреда будем понимать величину  $\Delta S$ , такую что:  $\Delta S = S(t) - S(t - 1)$ .

Модель, описанная в статье [6], требует, чтобы временной ряд прироста спреда был стационарным.

**Определение 38.** Статистический процесс  $S(t)$  называется строго стационарным, если взаимное распределение вероятностей  $m$  наблюдений инвариантно по отношению к общему сдвигу временного

аргумента, т.е. совместная плотность распределения случайных величин  $S_{t_1}, S_{t_2}, \dots, S_{t_m}$  такая же, как для величин  $S_{t_1+k}, S_{t_2+k}, \dots, S_{t_m+k}$  при любых целых значениях сдвига  $k$ ,  $t_i$  – набор моментов времени,  $i = 1, 2, \dots, m$ . Совместной плотностью распределения называется функция, которая обладает следующими свойствами [9]:

- 1)  $p(S_{t_1}, S_{t_2}, \dots, S_{t_m}) \geq 0$ ,
- 2)  $\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} p(S_{t_1}, S_{t_2}, \dots, S_{t_m}) dS_{t_1} dS_{t_2} \dots dS_{t_m} = 1$ ,
- 3)  $\int_{-\infty}^{t_1} \dots \int_{-\infty}^{t_m} p(S_{t_1}, S_{t_2}, \dots, S_{t_m}) dS_{t_1} dS_{t_2} \dots dS_{t_m} = F(S_{t_1}, S_{t_2}, \dots, S_{t_m})$ ,

для  $\forall t_i \in (0, \infty)$ ,  $i = 1, 2, \dots, m$ .

Когда  $m = 1$ , из предположения стационарности следует, что безусловное распределение  $p(S(t))$  величины  $S(t)$  одинаково для всех  $t$  и может быть записано как  $p(S)$ .

Требование строгой стационарности являются достаточно жесткими. На практике при изучении случайных процессов ограничиваются моментами первого и второго порядка, тогда говорят о слабой стационарности или стационарности второго порядка. В этом случае процесс имеет постоянные для всех  $t$  моменты первого и второго порядков: среднее значение  $\mu = E(S(t))$ , определяющее уровень, относительно которого он флуктуирует, дисперсию  $\sigma^2 = E(S(t) - \mu)^2$  и автоковариацию  $\gamma_k = E(S(t) - \mu)(S(t+k) - \mu)$  [1].

Автокорреляция  $k$ -го порядка стационарного процесса с ненулевой дисперсией имеет вид [1]:

$$\rho_k = \frac{E(S(t) - \mu)(S(t+k) - \mu)}{\sqrt{E(S(t) - \mu)^2 E(S(t+k) - \mu)^2}} = \frac{\gamma_k}{\gamma_0}.$$

## 4 Иллюстративный пример

В качестве иллюстративного примера приведем результаты расчетов для самой ликвидной ценной бумаги российского фондового рынка – обыкновенная акция ОАО “Сбербанк” (SBER).

Авторами разработан алгоритм и реализована программа ExportData для экспорта данных таблицы котировок из торгового терминала Quik [10]. Программа ExportData позволяет выбрать финансовый инструмент, настроить периодичность, с которой экспортируются данные. При помощи данной программы накоплены данные из таблицы котировок SBER за период с 17.12.2012 по 21.12.2012 (с периодичностью 1 сек.).

Для полученных данных SBER по каждой торговой сессии была проанализирована величина прироста спреда. Описательные статистики для полученных величин прироста спреда приведены в табл. 1:  $E$  – математическое ожидание,  $Me$  – медиана,  $Max$

Стат-ки	17.12	18.12	19.12	20.12	21.12
$E$	0	0	0	0	0
$Me$	0	0	0	0	0
$Max$	0,05	0,11	0,05	0,22	2,22
$Min$	-0,06	-0,07	-0,05	-0,06	-1,77
$\sigma^2$	0,005	0,006	0,006	0,006	0,050
$K_a$	0,45	0,63	0,38	2,52	6,37
$K_e$	19,30	18,58	13,73	96,39	761,41
$n$	30346	30365	30315	30311	30221

Таблица 1: Описательные статистики для ряда прироста спреда

Дата	Наблюдаемое значение, $t_{obs}$	1%	5%	10%
17.12.	-78.53	-2.57	-1.94	-1.62
18.12.	-52.94	-2.57	-1.94	-1.62
19.12.	-55.71	-2.57	-1.94	-1.62
20.12.	-55.62	-2.57	-1.94	-1.62
21.12.	-65.24	-2.57	-1.94	-1.62

Таблица 2: Проверка данных на стационарность с помощью расширенного теста Дики-Фуллера при различных уровнях значимости

– максимальное значение,  $Min$  – минимальное значение,  $\sigma^2$  – стандартное отклонение,  $K_a$  – коэффициент асимметрии,  $K_e$  – коэффициент эксцесса,  $n$  – количество наблюдений.

Осуществим проверку данных на стационарность. Наличие стационарности у ряда дает больший инструментарий для дальнейшего анализа. Воспользуемся для проверки расширенным тестом Дики-Фуллера (табл. 2). Данный тест проводится в пакете эконометрического анализа Eviews и позволяет проанализировать наличие стационарности при различных уровнях значимости.

Анализируя табл. 2, убеждаемся, что временные ряды стационарны, т.к. выполняется условие  $t_{obs} < t_{crit}$  для всех уровней значимости. Кроме того выполняются основные свойства стационарности: математическое ожидание и стандартное отклонение не зависят от времени и очень близки к нулю (табл. 1).

Далее проанализируем, имеется ли зависимость изменения прироста спреда на основе автокорреляционной функции. Если автокорреляция присутствует, то можно достаточно легко предсказать, как будет меняться цена дальше. Для проверки автокорреляции будем использовать статистику Дарбина-Ватсона. Результаты проверки приведены в табл. 3, где  $DW$  – полученное значение статистики Дарбина-Ватсона,  $DW_{up}$  – верхнее теоретическое значение, при котором существует автокорреляция.

Дата	$DW$	$DW_{up}$	Автокорреляция
17.12	2,43	1,69	отсутствует
18.12	2,46	1,69	отсутствует
19.12	2,41	1,69	отсутствует
20.12	2,38	1,69	отсутствует
21.12	2,02	1,69	отсутствует

Таблица 3: Проверка данных на автокорреляцию с использованием статистики Дарбина-Ватсона

Дальнейшие исследования на автокорреляцию были проведены для различных временных лагов: от 1 до 15 сек. В результате вычислений можно сделать вывод, что незначительная автокорреляция присутствует для данных с лагом равным 1 сек., т.е. текущее значение зависит от предыдущего. Это вполне логично, т.к. данные в течение 1-2 сек. не успевают значительно измениться и зависят от предыдущего значения.

## 5 Выводы

В статье приведен краткий обзор подходов к моделированию таблицы котировок. Проанализирована величина прироста спреда на примере обыкновенной акции ОАО «Сбербанк» (SBER). Установлено, что временной ряд величин прироста спреда является стационарным и автокорреляция у данного ряда отсутствует.

На основе полученных результатов можно сделать предположение о видах моделей, которыми могут быть описаны исходные данные. Модель [3] требует полной информации, поэтому сложна в реализации, т.к. данная информация практически не доступна для исследований. Для использования моделей [4], [5], [7], [8] необходимо проводить дополнительные исследования и учитывать дополнительные условия, такие как, эффект памяти и мультифрактальную природу данных, что существенно ограничивает их применимость.

Наиболее адекватной для реализации является модель, приведенная в статье [6], так как с точки зрения моделирования, марковские случайные процессы – это одни из наиболее изученных процессов.

## Список литературы

- [1] В.И. Суслов, Н.М. Ибрагимов, Л.П. Талышева и А.А. Цыплаков. *Эконометрия*. Издательство СО РАН, Новосибирск, 2005.
- [2] М.А. Романюк, Л.А. Глик, М.Е. Семенов. Статистический подход для анализа российского фондового рынка. *Труды IX Международной конференции студентов и молодых ученых «Перспективы развития фундаментальных наук»*, pages 645–647, 2012.
- [3] К.В. Воронцов. Имитационное моделирование реальных биржевых торгов. *ВЦ РАН, Форексис*, 2003.
- [4] И.В. Артюхин, А.К. Бритенков. Описание, анализ и прогнозирование котировок ценных бумаг с помощью обобщенного спектрально-аналитического метода для динамического управления портфелем ценных бумаг. In *Доклады 10-ой Всероссийской конференции «Математические методы распознавания образов»*, pages 165–167, Москва, 2001.
- [5] С.С. Головачев. Использование искусственных нейронных сетей для прогнозирования американского фондового рынка в период кризиса. *Управление экономическими системами*, 2012.
- [6] F. Abergel and A. Jedidi. A mathematical approach to order book modeling. *Social Science Research Network*, pages 93–107, 2011, <http://ssrn.com/abstract=1740889>.
- [7] Tian Qiu, Guang Chen, Li-Xin Zhong, and Xiao-Run Wu. Dynamics of bid-ask spread return and volatility of the chinese stock market. *Physica A: Statistical Mechanics and its Applications*, 8:2656–2666, 2012.
- [8] M. Szabolcs and F. Doyne. An empirical behavioral model of liquidity and volatility. *Journal of Economic Dynamics and Control*, 32:200–234, 2008.
- [9] А.Д. Манита. *Теория вероятностей и математическая статистика*. Издательский отдел УНЦ ДО, Москва, 2001.
- [10] Л.А. Глик, М.Е. Семенов. Экспорт данных таблицы котировок для технического анализ. In *Труды X Международной конференции студентов и молодых ученых «Перспективы развития фундаментальных наук»*, Томск, 2013. Томский политехнический университет.



## Сравнительное исследование влияния семантики политических рекламных материалов на эффективность её восприятия

Гнездилова Маргарита Андреевна

Красноярск  
gretheng.mg@gmail.com

Михаил Георгиевич Садовский

Институт вычислительного моделирования СО РАН  
Красноярск  
msad@icm.krasn.ru

**Аннотация.** Рассматривается сравнительное изучение семантики политических рекламных материалов, для задачи повышения эффективности их восприятия, на примере особенностей восприятия агитационных рекламных материалов предвыборных кампаний: выборы в Законодательное Собрание Красноярского края, 4 декабря 2011 г., выборы в Государственную Думу РФ, 4 декабря 2011 г., выборы Президента РФ 4 марта 2012 г.

**Ключевые слова.** Лозунг, политика, рекламная кампания, выборы, Государственная Дума, Законодательное Собрание, президент.

### 1 Введение

Политическая реклама является одной из важнейших составляющих частей медийного поля, в которое погружён современный человек. Она характеризуется такими особенностями, как характер и тип ее коммуникативного воздействия, и при этом отличается от «обычной» рекламы массовых, брендируемых товаров по таким показателям, как объект рекламирования, охватом целевой аудитории и способами привлечения широкой массы населения.

Несмотря на свою особенность, политическая реклама во многом подчиняется законам, общим для всех видов рекламы. В частности, именно для политической рекламы особую важность приобретает такое свойство, как формирование побуждения к совершению электорального действия (приход к урнам для голосования, выбора кандидата, выбор партии и т. п.).

В настоящее время политическая реклама в России переживает определённый застой. Это связано как с понижением уровня политической жизни в стране, так и с её регуляризацией: многие важные социальные коммуникации в сфере политики переходят в другие каналы и приобретают иной формат.

В связи с тем, что в последнее время принято решение об усилении роли партий в политической жизни и о повышении уровня политической активности избирателей путём возвращения выборов губернаторов

субъектов Федерации, вопросы эффективности политической рекламы приобретают особую остроту.

Анализ материалов политических рекламных кампаний фактически не проводился, ни на примере агитационной рекламы г. Красноярска, в России в целом. Это обуславливает актуальность такого исследования; будем рассматривать её на примере прошедших последних выборов в Законодательное Собрание Красноярского края (4 декабря 2011 г.), выборов в ГД ФС РФ (4 декабря 2011 г.) выборов Президента РФ 4 марта 2012 г.).

### 2 Результаты исследования агитационных политических материалов. Исследование политических карикатур

Сложившаяся в стране практика политической жизни и её стиль привели к широкому распространению карикатур и комиксов с кандидатами на должность президента, а также работы в поддержку митингов оппозиции. Подобные работы редко имеют хорошее качество и как правило не создаются профессиональными художниками или дизайнерами. Чаще всего это медиа вирусы, так называемые «мемы», который может создать любой желающий в простейших приложениях.

Нередко в подписях к подобным картинкам встречаются грубейшие и нелепые орфографические ошибки, что говорит о низком уровне грамотности людей делающих данные работы. Самые популярные герои прошедшего электорального цикла — это В. Жириновский, М. Прохоров и В. Путин. Чаще всего образ В. Путина используют в связке с образом Д. Медведева. Частота использования этих образов определялась по мониторингу ряда интернет-сайтов, наиболее активно используемых политически активными пользователями сети Интернет.

#### 2.1 В. Жириновский

Самый популярный вид комиксов с Жириновским в качестве героя можно найти в социальных сетях. Четыре фотографии указанного политика с текстом разного содержания: будь то политическая или бытовая тематика. По логотипам телеканалов и плохо-



му качеству самих фото ясно, что это скриншоты с видео выпусков новостных передач федеральных каналов.

Первая фотография. Жириновский лицом к читателю, правая рука вытянута вперед, вытянут указательный палец. Рот приоткрыт, становится понятно, что человек что-то говорил в тот момент, когда его сфотографировали. Подпись под фотографией: *Кто ходит в гости по утрам, тот поступает мудро!*. Вторая фотография. Жириновский стоит на трибуне, на месте спикера в палате Государственной Думы. Он смотрит в правую сторону, туда же развёрнут корпус тела, и правая рука. Под фотографией подпись, логично продолжающая предыдущую: *То тут 100 грамм*. На третьей фотографии Жириновский уже смотрит на зрителя, но его левая рука показывает в сторону. Подпись: *То там 100 грамм*. На четвертой, завершающей фотографии Жириновский стоит, широко раскинув руки. Подпись: *На то оно и утро!*.

Как пример, ещё одна работа с идентичными фотографиями, расположенными в том же порядке, но другими подписями.

- 1) Попросил жену разбудить утром на рыбалку...
- 2) Собрал вечером все необходимое...
- 3) А она забыла разбудить...
- 4) Вот такого ляца от меня словила!

На этом комиксе автор добавил интернет адрес группы в социальной сети *Вконтакте*, введя который, можно попасть в группу, где есть целый альбом, посвященный данной тематике.

## 2.2 М. Прохоров

В работах посвящённых Михаилу Прохорову, главным героем юмора становится его рост, а так же все известные увлечения.

Автор следующего примера уподобил кандидата в президенты герою детских стихов «Дяде Степе», автора С. Михалкова. На четырех картинках, использованы иллюстрации произведения, где вместо головы рисованного героя находится фотографии М. Прохорова. В качестве подписей используются стихотворения, как оригинальные, так и стилизованные под авторский текст о герое детской книги.

Примеры:

1. Герой, Прохоров-Дядя Степа, стоит по колено в воде и достает тонущего ребенка. Подпись:  
 — *Это необыкновенно!* —  
*Все кричат ему с моста.*  
 — *вам, товарищ, по колено*  
*все глубокие места!*

2. Улыбающийся Прохоров бежит, и вполоборота машет рукой. На уровне его колен стоят дети и приветственно машут руками. Подпись:

— *Он домой спешит с Арбата.*  
*«Ох, \*\*ать же ты высокий!»* —  
*Вслед кричат ему ребята.*

3. Улыбающийся и довольный, герой идет на стадион, толпа распускается, приветственно распростирает руки приглашая зайти на стадион. Подпись:

*Но зато на биатлон*  
*Проходил бесплатно он:*  
*Пропускали дядю Мишу —*  
*Думали, что чемпион.*

4. Прохоров, с улыбкой, держит на плечах детей с флажками. На заднем плане стены Кремля, на уровне пояса стоят люди, кто-то улыбается, кто-то серьезен. Подпись:

*И того, кто ростом мал,*  
*На Болотной поднимал,*  
*Потому что все должны*  
*Видеть жуликов страны.*

## 2.3 В. Путин

Несомненно, самый популярный герой «низшего искусства» и главный (по числу) адресат насмешек народа — это В. Путин. Его представляют во всевозможных образах. Карикатуры, сделанные на него, чаще всего несут негативный оттенок.

Рассмотрим на примерах.

1. На картинке изображены четыре одинаковых Путина, причем третий держит возле лица маску с лицом Д. Медведева. Фигуры расположены в пол оборота к зрителю, и смотрят в сторону. На заднем плане находится флаг и герб Российской Федерации, а так же Спасская башня Московского Кремля. Здесь есть два равнозначных толкования:  
 — Современная версия знаменитого коммунистического плакат с изображениями К. Маркса, Ф. Энгельса, В. Ленина и И. Сталина, где «отцы» коммунизма смотрят в одну сторону, а под ними толпа с флагами, на которых нанесены коммунистические лозунги. И подпись: «Да здравствует великое, непобедимое знамя Маркса – Энгельса – Ленина – Сталина!». Здесь портреты Путина как продолжение коммунистического прошлого России.  
 — Четыре Путина — это четыре президентских срока. Маска с лицом Медведева — намек, что хоть Д. Медведев формально и находился у власти, все равно В. Путин занимал главенствующую позицию в управлении государством.

2. Автор изобразил крупного аллигатора. У крокодила открыта пасть и видны редкие, но крупные и острые зубы. Хищный оскал напоминает улыбку. На первый взгляд, животное ничем не примечательно, но приглядевшись можно увидеть, что глаза и надбровные дуги идентичны чертам В.Путина. Из левого глаза крокодила стекает крупная слеза. Это автоматически отсылает нас к видео, плачущего Путина в Москве, на Манежной площади, 4.03.2012, на митинге в свою поддержку, когда по предварительным итогам, он уверенно лидировал в выборах на должность Президента.
3. На картинке изображен бурый медведь с лицом В.Путина. У него очень суровое лицо, хмурый взгляд, плотно сжаты губы. На лапах длинные, острые, черные когти. Он сжимает напильник, которым точит когти. По форме, напильник напоминает пистолет.  
  
Бурый медведь — самый популярный образ России за рубежом. К тому же, до ребрендинга в 2005 г., на логотипе партии был бурый медведь. После он стал белым. То, что герой картинки точит когти, символизирует то, что власть никогда не расслабляется и каждый может попасть в «лапы». Пистолет, в свою очередь, подчеркивает, работу Путина в органах ФСБ, а так же установившийся в годы его правления страной так называемый «кровавый режим Путина».
4. На картинке изображен В. Путин. У него усталое осунувшееся лицо, глубокие синяки под глазами. Он устало смотрит прямо на зрителя. Но его фигура отбрасывает тень на стену и в тени отчетливо виден профиль Иосифа Сталина, который смотрит твердо вперед. «Тень Сталина» символизирует то, что режим правления В. Путина отнюдь не демократический, а в какой-то мере тоталитарный и своими методами похожа на автократию И. Сталина.
5. Автор следующего примера — бельгиец Nicolas Vadot. На картинке изображен трехмерный кубик Рубика, в котором, вместо цветных секторов, на каждой стороне сложены портреты В. Путина. Мы кубик видим с трех сторон: они ничем не отличаются, кроме цифр. Каждая картинка подписана датами: 2000, 2004, 2012 — года, когда В. Путин заступал на должность президента России. Сверху картинка подписана: «Russixcub».
6. На картинке изображена змея с лицом В.Путина. Змея закрутилась вокруг себя, поэтому ей становится тяжело дышать, у нее широко раскрыт рот, глаза навыкате и испарина. Подпись: «Удав удушающий самое себя». Это означает, что политика, которую вел Путин,

и партия власти исчерпала себя и теперь они становятся заложниками собственных правил, которые перестали им играть на руку.

7. Нередко можно встретить парные карикатуры Путин — Медведев. Рассмотрим пример такого тандема: на картинке изображены В. Путин и Д. Медведев. Путин выше ростом и его фигура видна из-за стола примерно по грудь, в то время как у Медведева видна голова и плечи. Медведев смотрит на Путина снизу вверх, но не заискивающе, а как равный. В то время, как Путин покровительственно указательным пальцем руки приподнимает кончик носа Медведева. То ли вытирает нос, то ли указывая его место, чтобы не забывал его.

## 2.4 Оппозиция

За последние полгода в стране, во всех городах, прошла масса митингов как оппозиционных, так и поддерживающих партию власти и президента. Люди выходили на площади целыми семьями, с плакатами. Последняя форма — это протестные гуляния. Но в некоторых городах (Санкт Петербург, Самара, Киров) уже прошли митинги новой формы — мимимитинги. Оппозиционеры принесли в парки мягкие игрушки и поставили возле них плакаты, на которых, в мягкой форме написаны все те же лозунги.

Мимими (Интернет сленг, мем) — означает крайнюю степень умиления, все самое дружелюбное и несущую позитивную окраску. Тем самым, мимимитинг — это позитивная форма привлечения общественности к оппозиции.

Рассмотрим пример карикатуры на мимимитинг:

На картинке изображены пять мишек «Metoyou». У первого в руках букет цветов, перевязанных бантом и плакат с надписью: «Президент буэшечка». Второй стоит, прищурив глаза, и держит табличку: «Путин — бяка». Третий стоит на носках и вместе с четвертым мишкой, у которого кисточка, он сам испачкан в краске, рядом тюбики и банка с водой, держат плакат: «Выборы — пичалька». Тут же пририсован грустный смайлик. Последний держит в книжку и плакат: «Нам все не нравится!». Внизу картинки надпись: «Мимимитинг». Все выполнено в розовых тонах, надписи — красные.

Все слова на этом плакате взяты из молодежного Интернет сленга, где и зародилась идея акции.

## 3 Исследование семантики политических лозунгов

Для изучения узнаваемости стилистики политических партий и кандидатов было проведено исследование. Для этого были взяты различные политические

Таблица 1: Ответы респондентов по различным кандидатам (партиям). Цифрами обозначены фамилии кандидатов: I – Зюганов, II – Жириновский, III – Миронов, IV – Прохоров, V – Путин. Прочерк в ряде с фамилиями кандидатов (последний столбец) означает, что респонденты затрудняются соотнести лозунг и кандидата.

Зюганов		I	II	III	IV	V	VI
Вся власть народу России	Власть и собственность — народу!	6	5	4	0	1	
Бороться за власть	За власть надо бороться	6	1	4	5	1	
КПРФ — великие традиции на благо Родины	-	15	1	0	0	0	
КПРФ — реальная оппозиция действующей власти	-	17	2	1	0	0	
КПРФ — сила борьбы	КПРФ — единственная сила, которая борется	17	1	0	0	0	
Не будь пешкой в политической борьбе	-	7	2	5	1	1	1
Победа КПРФ — победа народа	Побеждает Зюганов — побеждает народ	18	0	0	0	0	
Умеем бороться и побеждать	Политика большинства призвана побеждать!	2	1	2	4	9	1
Прохоров		I	II	III	IV	V	VI
А в чем сила, брат?	Сила в правде. Кто прав, тот и сильнее	0	2	3	10	1	2
Ветер перемен	-	0	1	3	12	2	
Новая высота	-	0	0	1	12	4	1
Новые горизонты	-	0	0	1	15	2	
Перемены к лучшему	-	1	3	1	10	2	
Просите достойного	Требуйте большего	0	3	1	12	3	
Решим трудные задачи	-	0	1	2	7	7	
Я вас не подведу	-	0	3	6	8	2	
Миронов		I	II	III	IV	V	VI
Вернём справедливость	-	3	5	11	0	0	
Вместе за справедливость	Добьёмся справедливости вместе	1	1	12	1	3	
Гражданин России — думающий человек	-	0	0	6	6	6	
Долой социальное неравенство	Социальное неравенство — одна из главных бед современной России	9	2	6	0	1	
Искореним коррупцию вместе	Преодолеть монополию власть — значит сломать хребет коррупции	6	3	5	0	3	
Настало время жить по справедливости	Время справедливости	0	0	16	2	0	
Порядочные люди — граждане России	-	3	2	6	2	3	
Скажем «Нет!» социальному неравенству	Преодолеть социальное неравенство	2	3	9	3	0	
Жириновский		I	II	III	IV	V	VI
Жириновский и будет хуже	Жириновский или будет хуже	1	8	3	3	2	1
Жить по-русски	Жить и побеждать по-русски	1	14	3	0	0	0
Поможем русскому народу	-	8	9	4	1	0	
Россия — дом каждого из нас	-	1	7	2	0	9	
Россия — земля наших предков	-	1	7	3	2	3	1
Русские вместе	Вместе с Русским народом	0	11	2	2	2	1
Хватит бездействия	Хватит это терпеть	2	7	5	2	1	
Хватит молчать	Русские, хватит молчать!	1	13	4	0	0	

лозунги с предвыборных кампаний в Государственную Думу Российской Федерации и выборов Президента РФ, которые затем были изменены: в части менялась суть, но оставалась форма или наоборот переписывалась форма лозунга, но оставалась прежняя суть. Всего было взято 4 кандидата (партии): В. Жириновский (ЛДПР), С. Миронов (Справедливая Россия), М. Прохоров, Г. Зюганов (КПРФ). Респондентам были предложены листы с четырьмя различными лозунгами неповторяющихся партий. Задача была соотнести каждый лозунг с одним из пяти кандидатов (партий). Пятым кандидатом был В. Путин (Единая Россия).

В исследовании участвовали 36 респондентов, возраста 18+. Каждый респондент дал ответы на 16 лозунгов. В итоге было получено 576 ответов. Рассмотрим подробнее, какие лозунги использовались. В таблицах представлены лозунги и количество ответов, данные респондентами. Первый столбец — это переделанный лозунг, второй столбец — исходный материал. В случаях, когда не было переделанного лозунга, а он написан самостоятельно, во второй графе стоит прочерк. Цифрами обозначены фамилии кандидатов: 1 — Зюганов, 2 — Жириновский, 3 — Миронов, 4 — Прохоров, 5 — Путин. Прочерк в ряде с фамилиями кандидатов (последний столбец) означает, что респонденты затрудняются соотнести лозунг и кандидата.

#### 4 Результаты и обсуждение

Политическая реклама — важное и сложное направление рекламной индустрии. Само название подразумевает, что используется она в области политики для побуждения людей голосовать, делать свой выбор, отдавать предпочтение. Главное сходство между коммерческой и политической рекламой в принципиально похожем продукте — в политической рекламе таким продуктом является человек (партия). Не существует различий в стиле, текстах, дизайне рекламы и методах, используемых для привлечения внимания. Различия существуют только в объекте рекламы — в политической, это живой человек и рекламировать его сложнее. Политическая реклама, как и реклама коммерческая призвана создать имидж рекламируемому объекту, в данном случае — партии, организации, движению, правительству, или их представителям.

Практически все виды рекламного воздействия могут быть использованы в политических целях. Разница лишь в том, что политические идеи невозможно разложить на прилавке магазина или выставить в витрине. Арсенал политической рекламы, как и рекламы коммерческой, исключительно многообразен, ее виды и формы практически неисчислимы. С древних времен он значительно пополнился, особенно в

XX веке, благодаря научному прогрессу, новым коммуникационным технологиям, появлению телефона, радио, телевидения, компьютерной техники, наконец, Internet.

Важным средством политической рекламы были и остаются плакаты. Не менее эффективная форма рекламы — транспаранты. Определённое рекламное воздействие оказывают и всевозможные тенденциозные иллюстрации и карикатуры на политические темы, которыми изобилуют печатные издания.

Политическая лингвистика является одним из новых исследовательских направлений современного языкознания. Эта область исследования носит ярко выраженный междисциплинарный характер.

Выбор методов маркетинговых исследований обуславливается опытом работы специалистов на рынке, их возможностями, наличием собственного персонала исследователей, видом деятельности, ее технической сложностью и т. д. В общем случае проведения кабинетных или полевых исследований следует иметь в виду, что сбор и обработка данных в процессе исследования осуществляется методами, которые теория маркетинга заимствовала из математики, статистики, психологии, социологии социальной экономики.

Самые популярные герои политических карикатур это В. Путин, В. Жириновский и М. Прохоров. Характерно, что образ Д. Медведева совершенно не популярен, если его и используют, то только вместе с В. Путиным и как второстепенное лицо. Авторы работ показывают самые известные и общепринятые факты о политиках. Так, В. Жириновский всегда выступает как оратор, который не стесняется в выражениях, очень любит говорить и провоцировать оппонентов на скандал. В каждой карикатуре направленной на М. Прохорова обязательно, на первый план ставится его рост и факты из биографии, о политической карьере ничего не говорится. В. Путин, в глазах зрителей и авторов карикатур предстает как отрицательный герой.

Исследование семантики политических лозунгов партий и отдельных кандидатов показало, что в целом люди точно опознают основные идеи партий. Одновременно, процесс «приписывания» лозунга, если они не имеют четкой специфики, показывает внутреннюю адресацию таких политических лозунгов: с кем путают кандидатов если в лозунге попадают слова, которыми чаще апеллируют другие партии. Так, лозунги партии Г. Зюганова узнали 88% респондентов, чаще всего его путали с С. Мироновым (16%). М. Прохорова узнали 60% респондентов, но 16% опрошенных отдали его лозунги В. Путину. 51% респондентов узнал лозунги С. Миронова и его партии, 17%, закономерно, сказали, что это лозунги Г. Зюганова. В. Жириновского в лозунгах узнали 52%

опрашиваемых, 37% отнесли эти лозунги к С. Миронову. Это означает, что зачастую, партии не имеют четкой программы, целей и задач, отличных от избирательных кампаний других кандидатов или они не могут это четко донести до избирателя.

Характерно, что несмотря на то, что партия Единая Россия и В. Путин не участвовали в эксперименте, респонденты стабильно отдавали свои голоса за них. Так, из лозунгов Г. Зюганова к В. Путину отошли 12% голосов, М. Прохорова — 16%, С. Миронов — 11%, В. Жириновский — 11%. При этом опрашиваемые не помнят АПМ от В. Путина и Единой России в президентскую кампанию (по тому, что они практически отсутствовали), но все-таки предполагают, что они видели или слышали эти лозунги и принадлежали они именно ЕР.



## Изучение потребительского спроса с помощью методов ассоциативных правил

Гордеева Ксения Владимировна

Сибирский Федеральный Университет  
Институт математики и фундаментальной информатики  
Красноярск  
kseniagordeeva07@gmail.com

**Аннотация.** В работе рассматриваются понятие, цели и задачи интеллектуального анализа данных, а также области его применения. Описывается алгоритм поиска ассоциативных правил. Рассматривается практический пример применения данного алгоритма для поиска ассоциативных правил в статистике потребительского спроса.

**Ключевые слова.** Интеллектуальный анализ данных, Data Mining, ассоциативные правила, метод Apriori, FP-метод.

### 1 Интеллектуальный анализ данных

Современные предприятия торговли и оказания услуг собирают подробную информацию о каждой отдельной покупке, используя кредитные карточки и компьютеризованные системы контроля. Новейшие технологии хранения и записи данных позволили накопить большой объем данных о совершенных потребителями покупках и заказах. Поэтому одним из актуальных направлений современных информационных технологий является анализ накопленных статистических данных с помощью методов интеллектуального анализа.

Интеллектуальный анализ данных [1] — мультидисциплинарная область, возникшая и развивающаяся на базе прикладной статистики, искусственного интеллекта, теории баз данных и других.

**Определение 39.** *Интеллектуальный анализ данных (от англ. Data Mining) — это процесс обнаружения в сырых данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности.*

К основным задачам интеллектуального анализа данных относятся:

1. Классификация (Classification),
2. Кластеризация (Clustering),
3. Ассоциативный поиск (Associations),

4. Прогнозирование (Forecasting),
5. Оценивание (Estimation),
6. Визуализация (Vizualization, Graph Mining),
7. Подведение итогов (Summarization).

К методам интеллектуального анализа данных (ИАД) относятся искусственные нейронные сети, нечеткая логика, генетические алгоритмы, различные методы классификации, моделирования и прогнозирования.

Одним из самых востребованных методов интеллектуального анализа данных является метод поиска ассоциативных правил.

### 2 Основы ассоциативных правил в интеллектуальном анализе данных

В ходе решения задачи поиска ассоциативных правил отыскиваются закономерности между связанными событиями в наборе данных. Отличие данной задачи от задач классификации и кластеризации заключается в том, что поиск закономерностей осуществляется не на основе свойств анализируемого объекта, а между несколькими событиями, которые происходят одновременно.

Впервые задача поиска ассоциативных правил была предложена для нахождения типичных шаблонов покупок в супермаркетах.

В данной работе рассмотрены два разных алгоритма поиска ассоциативных правил в интеллектуальном анализе данных: метод Apriori и FPG-метод. Приведем основные понятия, связанные с данными методами.

**Определение 40.** *Транзакция — это множество событий, которые произошли одновременно.*

**Определение 41.** *Транзакционная (операционная) база данных представляет собой двумерную таблицу, состоящую из номера транзакции (TID) и перечня событий, происходящих во время этой транзакции.*

**Определение 42.** *Поддержка — количество или процент транзакций, содержащих определенный набор данных.*

### 3 Алгоритм Apriori

Рассмотрим классический ассоциативный метод Apriori [2].

Выявление часто встречающихся наборов элементов — достаточно трудоемкая операция, требующая значительных вычислительных ресурсов и, соответственно, времени. Примитивным подходом к решению данной задачи является выполнение простого перебора всех возможных наборов элементов. Это потребует  $2N$  количества операций, где  $N$  — количество элементов. Для сокращения количества кандидатов метод Apriori использует одно из свойств поддержки, которое носит название анти-монотонности и служит для снижения размерности пространства поиска. Данное правило формулируется следующим образом: «Поддержка любого набора элементов не может превышать минимальной поддержки любого из его подмножеств».

Не имея мы в наличии такого свойства, нахождение многоэлементных наборов было бы практически невыполнимой задачей в связи с экспоненциальным ростом вычислений.

Приведем обозначения, используемые в алгоритме:  
 $D$  — множество транзакций,

$$D = \{T_1, T_2, \dots, T_m\},$$

где  $T_i$  — транзакции из транзакционной (операционной) базы данных.

$L_k$  — множество  $k$ -элементных наборов, чья поддержка не меньше заданной пользователем.

$C_k$  — множество потенциально частых  $k$ -элементных наборов.

Алгоритм поиска ассоциативных правил Apriori имеет следующий вид:

Шаг 1. Присвоить  $k = 1$  и выполнить отбор всех одноэлементных наборов, у которых поддержка больше минимально заданной пользователем  $\text{Suppmin}$ .

Шаг 2.  $k = k + 1$ .

Шаг 3. Если не удастся создавать  $k$ -элементные наборы, то завершить алгоритм, иначе выполнить следующий шаг.

Шаг 4. Создать множество  $k$ -элементных наборов кандидатов из частых наборов. Для этого необходимо объединить в  $k$ -элементные кандидаты  $(k-1)$ -элементные частые наборы. Каждый кандидат будет формироваться путем добавления к  $(k-1)$ -элементному частому набору —  $p$  элемента из другого  $(k-1)$ -элементного частого набора —  $q$ . Причем добавляется последний элемент набора  $q$ , который по порядку выше, чем последний элемент набора  $p$ . При этом все  $k-2$  элемента обоих наборов одинаковы.

№	Транзакция
1	$a, b, c, d, e$
2	$a, b, c$
3	$a, c, d, e$
4	$b, c, d, e$
5	$b, c$
6	$b, d, e$
7	$c, d, e$

Таблица 1: Исходная БД транзакций

Шаг 5. Для каждой транзакции  $T$  из множества  $D$  выбрать кандидатов  $C_t$  из множества  $C_k$ , присутствующих в транзакции  $T$ . Для каждого набора из построенного множества  $C_k$  удалить набор, если хотя бы одно из его  $(k-1)$  подмножеств не является часто встречающимся т.е. отсутствует во множестве  $L_{k-1}$ .

Шаг 6. Для каждого кандидата из  $C_k$  увеличить значение поддержки на единицу.

Шаг 7. Выбрать только кандидатов  $L_k$  из множества  $C_k$ , у которых значение поддержки больше заданной пользователем  $\text{Suppmin}$ . Вернуться к шагу 2.

Результатом работы алгоритма является объединение всех множеств  $L_k$  для всех  $k$ .

### 4 Демонстрация работы алгоритма Apriori на примере простейшей базы данных транзакций

Рассмотрим простейшую базу данных (БД) транзакций — БД покупок. В ней встречается 5 наименований продуктов, обозначенных переменными  $a, b, c, d, e$ . База данных состоит из 7 транзакций, которые перечислены в Табл. 1.

Минимальную поддержку выберем равную 3.

Согласно алгоритма подсчитываем поддержку одноэлементных наборов. Наборы с уровнем поддержки меньше выбранного значения (3) отсекаются. Полученные одноэлементные наборы приведены в Табл. 2.

Далее происходит формирование двухэлементных кандидатов, подсчет их поддержки и отсекающие наборы с уровнем поддержки, меньшим 3. Оставшиеся двухэлементные наборы товаров, считающиеся часто встречающимися двухэлементными наборами, принимают участие в дальнейшей работе алгоритма. Это пары  $ac, bc, bd, be, cd, ce$  и  $de$ . Соответственно, их поддержка равна 3, 4, 3, 3, 4, 4 и 5.

На следующем этапе алгоритм формирует трехэлементные наборы товаров:  $bcd, bce, bde, cde$ , подсчитывает их поддержку и отсекает наборы с уровнем поддержки меньше 3. В результате остается два трех-

Элемент	Поддержка
<i>a</i>	3
<i>b</i>	5
<i>c</i>	6
<i>d</i>	5
<i>e</i>	5

Таблица 2: Одноэлементные наборы

элементных набора *bde* и *cde* с поддержкой 3 и 4, соответственно.

На последнем этапе алгоритм формирует четырехэлементные наборы товаров. Алгоритм завершается, так как создать такие наборы не удастся, поскольку уровень поддержки всех четырехэлементных наборов меньше заданного.

Итогом работы алгоритма становится трехэлементный набор с наибольшим значением поддержки (равным 4) — *cde*.

Узким местом в алгоритме Apriori является процесс генерации кандидатов в частые предметные наборы и необходимость многократного сканирования базы данных транзакций.

## 5 Алгоритм Frequent Pattern-Growth Strategy (FPG)

Одним из наиболее эффективных алгоритмов поиска ассоциативных правил, позволяющих избежать генерации кандидатов и сократить требуемое число сканирований набора данных, является алгоритм FPG, название которого можно перевести как «выращивание популярных (т.е. часто встречающихся) предметных наборов».

В основе метода лежит понятие Frequent-Pattern Tree [3] — дерево популярных предметных наборов (откуда и произошло название алгоритма). В дальнейшем для краткости будем называть эту структуру FP-дерево.

Алгоритм FPG выполняется в два этапа и имеет следующий вид:

1 этап. Строится FP-дерево, которое в компактном виде представляет информацию о частых предметных наборах.

1. Производится первое сканирование транзакций БД, и отбирается множество часто встречающихся предметов, т.е. предметов, у которых поддержка больше или равна минимальной. Обнаруженные частые предметы упорядочиваются в порядке возрастания их поддержки. Будем демонстрировать работу метода на той же БД, и для метода Apriori. В Табл. 3 приведен вид транзакционной базы из предыдущего примера до и после упорядочивания. Уровень под-

№	Исходный набор	Отсортированный набор
1	<i>a, b, c, d, e</i>	<i>c, b, d, e, a</i>
2	<i>a, b, c</i>	<i>c, b, a</i>
3	<i>a, c, d, e</i>	<i>c, d, e, a</i>
4	<i>b, c, d, e</i>	<i>c, b, d, e</i>
5	<i>b, c</i>	<i>c, b</i>
6	<i>b, d, e</i>	<i>b, d, e</i>
7	<i>c, d, e</i>	<i>c, d, e</i>

Таблица 3: Исходная и измененная БД транзакций

держки также установим равным 3.

2. Строится FP-дерево. Сначала упорядочиваются предметы в транзакциях по убыванию значений их поддержки. Для этого создается начальный (корневой) узел FP-дерева (ROOT). И начинается построение дерева с транзакции №1 для упорядоченных предметных наборов.

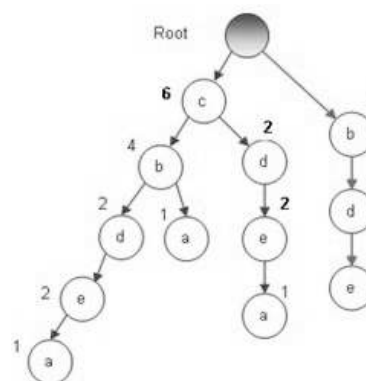


Рис. 1: БД транзакций в виде FP-дерева

Правило построения: если для очередного предмета в дереве встречается узел, имя которого совпадает с именем предмета, то предмет не создает нового узла, а индекс соответствующего узла в дереве увеличивается на 1. В противном случае для этого предмета создается новый узел и ему присваивается индекс 1.

Таким образом, после первого прохода БД и выполнения соответствующих манипуляций с предметными наборами мы построим FP-дерево, которое в компактном виде представляет информацию о частых предметных наборах и позволяет производить их эффективное извлечение, что и выполняется при втором сканировании БД. На Рис. 1 показано построенное FP-дерево для нашей БД транзакций.

Выгода представления базы данных транзакций в виде FP-дерева очевидна. Если в исходной базе данных каждый предмет повторяется многократно, то в FP-дерево каждый предмет представляется в виде узла, а его индекс указывает на то, сколько раз данный предмет появляется. Иными словами, если предмет

в исходной базе данных транзакций появляется 100 раз, то в дереве для него достаточно создать узел и установить индекс 100.

2 этап. Производится процесс извлечения из FР-дерева частых предметных наборов.

Процесс извлечения будет заключаться в следующем:

1. Выбирается предмет (например, а), и в дереве находятся все пути, которые ведут к узлам этого предмета. (Путь — последовательность узлов, которую надо пройти от корневого узла до узла, связанного с данным предметом). Для предмета а это будет набор {cbdea, cba, cdea}.

2. Затем для каждого пути производится подсчет, сколько раз данный предмет встречается в нем, и это записывается в виде (cbdea, 1), (cba, 1) и (cdea, 1).

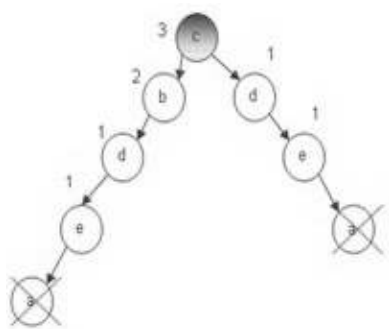


Рис. 2: Условное FР-дерево для элемента а

3. Удаляется сам предмет из ведущих к нему путей. После это останутся пути только с предшествующими ему предметами {cbde, cb, cde}.

4. Выполняется подсчет, сколько раз каждый предмет появляется в префиксах путей, полученных на предыдущем шаге, и производится упорядочение в порядке убывания этих значений. После чего получается новый набор транзакций.

5. На его основе строится новое FР-дерево, которое будем называть условным FР-деревом (conditional FP-tree), поскольку оно связано только с одним объектом (в нашем случае, а). Вид полученного FР-дерева показан на Рис. 2.

6. В этом FР-дереве находятся все предметы (узлы), для которых поддержка (количество появлений в дереве) равна или больше заданной минимальной поддержки. Если предмет встречается два или более раз, то его индексы, т.е. частоты появлений в условном базисе, суммируются.

7. Начиная с верхушки дерева, записываем пути, которые ведут к каждому узлу, для которого поддержка/индекс больше или равны 3. Возвращаем назад предмет, удаленный на шаге 2, и подсчитываем индекс/поддержку, полученную в результате. Таким

образом, и будут появляться **частые предметные наборы**.

В нашем примере для предмета а частым предметным набором будет являться набор (с,а,3).

После извлечения всех частых предметных наборов самым популярным окажется набор (с,д,е,4).

## 6 Сравнение алгоритмов Apriori и FGP

В результате своей работы и алгоритм Apriori, и FPG-алгоритм выдают один или несколько частых предметных наборов, самых длинных в базе данных транзакций.

Преимуществами FPG-алгоритма перед Apriori являются:

1. Сжатие БД транзакций в компактную структуру, что обеспечивает очень эффективное и полное извлечение частых предметных наборов;
2. При построении FР-дерева используется технология разделения и захвата (англ.: divide and conquer), которая позволяет выполнить декомпозицию одной сложной задачи на множество более простых задач.
3. Отсутствие затратной процедуры генерации кандидатов, характерной для алгоритма Apriori

Сравнительные исследования классического алгоритма Apriori и FPG показали, что с увеличением числа транзакций в БД временные затраты на поиск частых предметных наборов растут для FPG намного медленнее, чем для Apriori (см. Рис. 3).

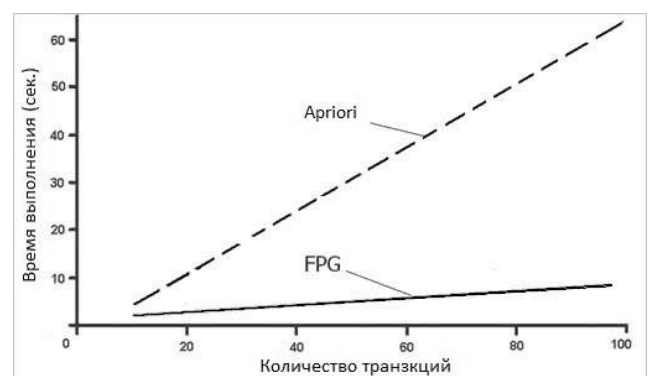


Рис. 3: Сравнение алгоритмов FPG и Apriori

## 7 Практический пример нахождения ассоциативных правил в статистике потребительского спроса

На одном из сайтов, выполняющих online-заказы на покупки товаров и оказание услуг, была собра-

на подробная статистика услуг и товаров, которые заказывали покупатели. Данные были собраны за период с 2010 по 2012 гг. Их целью было определение основных предпочтений потребителей (Интернет-покупателей), актуальных трендов, портрет среднестатистического покупателя.

В статистических данных был представлен следующий перечень товаров (услуг), например, таких, как: услуги турагентств, заказ авиабилетов, заказ железнодорожных билетов, покупка билетов в кино, покупка музыки, книг, оплата услуг связи, заказ одежды (товаров) и т.д. Всего была собрана информация о 1200 пользователях, в возрасте от 20 до 65 лет, проживающих в различных регионах РФ. Каждый пользователь характеризовался тем множеством online-услуг, которые он использовал.

Для решения практической задачи нахождения ассоциативных правил в данной статистике был использован вышеописанный классический метод поиска ассоциативных правил (алгоритм Apriori).

Согласно алгоритму был определен уровень поддержки для данной задачи  $\text{Suppmin}$ . На первом этапе были сформированы одноэлементные кандидаты. Наборы с уровнем поддержки меньше установленного значения (300) были отсечены. Далее были найдены двухэлементные наборы. Затем трехэлементные. После чего работа алгоритма была завершена.

В результате исследования было обнаружено, что преобладающим числом заказов стали все заказы, содержащие в себе авиа и железнодорожные билеты. Прочие товары и услуги в большинстве случаев заказывались комплексно.

Также в работе для этого же практического примера были найдены ассоциативные правила с помощью FPG-алгоритма. После чего был проведен сравнение правил, полученных данными методами.

## Список литературы

- [1] И. А. Чубукова. *Data Mining*. Бином. Лаборатория знаний, М., 2006.
- [2] А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, И.И. Холод. *Технологии анализа данных. Data Mining, Visual Mining, Text Mining, OLAP*. БХВ-Петербург, СПб., 2004.
- [3] В. Г. Матвейкин, Б. С. Дмитриевский, Н.Р. Ляпин. *Информационные системы интеллектуального анализа*. Машиностроение, М., 2008.



## Численная идентификация правой части специального вида системы двух уравнений параболического типа

Горохов Александр Андреевич

Сибирский федеральный университет  
Институт математики и фундаментальной информатики  
Красноярск  
Alexandr\_1991@inbox.ru

Распопов Виталий Евгеньевич

Сибирский федеральный университет  
Институт математики и фундаментальной информатики  
Красноярск

**Аннотация.** В данной работе численно решена обратная коэффициентная задача для системы двух дифференциальных уравнений параболического типа с функциями источника специального вида. Разработан алгоритм численного решения поставленной обратной задачи, создан программный продукт, проведены вычислительные эксперименты.

**Ключевые слова.** Коэффициентная обратная задача, краевые условия, условия переопределения, условия согласования, прямая дифференциальная задача.

В настоящее время известно немало случаев, когда потребности практики приводят к задачам определения коэффициентов дифференциальных уравнений (обыкновенных или в частных производных) по некоторым известным функционалам от его решения. Такие задачи получили название обратных задач математической физики. Прикладная важность обратных задач настолько велика, что ставит их в ряд актуальнейших проблем современной математики. Они возникают в самых различных областях человеческой деятельности, таких как: сейсмология, разведка полезных ископаемых, биология, медицина, контроль качества промышленных изделий и т.д. Обратным коэффициентным задачам посвящено достаточно много работ (в частности смотри [1, 2, 3, 4]). Существуют много методов для нахождения решения таких задач [5, 4].

Рассмотрим следующую задачу. В области  $D = \{(t, x) | 0 \leq t \leq 1, 0 \leq x \leq 1\}$  найти функции  $u(t, x)$ ,  $v(t, x)$  и произведение функций  $f_1(t)g_1(x)$  и  $f_2(t)g_2(x)$  удовлетворяющие задаче:

$$\begin{aligned} u_t(t, x) &= u_{xx}(t, x) + a_1 u_x(t, x) + b_1 v_x(t, x) + f_1(t)g_1(x), \\ v_t(t, x) &= v_{xx}(t, x) + a_2 u_x(t, x) + b_2 v_x(t, x) + f_2(t)g_2(x). \end{aligned} \quad (1)$$

$$\begin{aligned} u(0, x) &= u_0(x), & 0 \leq x \leq 1, \\ v(0, x) &= v_0(x), & 0 \leq x \leq 1, \end{aligned} \quad (2)$$

$$\begin{aligned} u_x(t, 0) &= \psi_1(t), & 0 \leq t \leq 1, \\ v_x(t, 0) &= \psi_3(t), & 0 \leq t \leq 1, \end{aligned} \quad (3)$$

$$\begin{aligned} u_x(t, 1) &= \psi_2(t), & 0 \leq t \leq 1, \\ v_x(t, 1) &= \psi_4(t), & 0 \leq t \leq 1, \end{aligned} \quad (4)$$

$$\begin{aligned} u(T, x) &= \alpha_1(x), & 0 \leq x \leq 1, \\ v(T, x) &= \alpha_2(x), & 0 \leq x \leq 1, \end{aligned} \quad (5)$$

$$\begin{aligned} u(t, \xi) &= \beta_1(t), & 0 \leq t \leq 1, \\ v(t, \xi) &= \beta_2(t), & 0 \leq t \leq 1, \end{aligned} \quad (6)$$

где (1) - исходные уравнения; (2) - начальные условия; (3), (4) - краевые условия; (5), (6) - условия переопределения;  $u_0(x), v_0(x), \psi_1(t), \psi_2(t), \psi_3(t), \psi_4(t), \alpha_1(x), \alpha_2(x), \beta_1(t), \beta_2(t)$  - заданные функции,  $a_1, a_2, b_1, b_2$  - заданные константы,  $\xi$  - фиксированная точка из интервала  $(0, 1)$ .

Считаем, что все известные функции достаточно число раз непрерывно дифференцируемы, и выполняются условия согласования:

$$\begin{aligned} u'_0(0) &= \psi_1(0), & v'_0(0) &= \psi_3(0), \\ u'_0(1) &= \psi_2(0), & v'_0(1) &= \psi_4(0), \\ \alpha'_1(0) &= \psi_1(T), & \alpha'_2(0) &= \psi_3(T), \\ \alpha'_1(1) &= \psi_2(T), & \alpha'_2(1) &= \psi_4(T), \\ \beta_1(0) &= u_0(\xi), & \beta_2(0) &= v_0(\xi), \\ \beta_1(T) &= \alpha_1(\xi), & \beta_2(T) &= \alpha_2(\xi). \end{aligned}$$

Поставленная задача относится к классу коэффициентных обратных задач. Теоремы о существовании и единственности классического решения поставленной задачи на данный момент неизвестны. Поставленную задачу будем решать численно. Аналитическими преобразованиями поставленная задача сведена к прямой:

$$\begin{aligned} W_t(t, x) &= W_{xx}(t, x) + a_1 W_x(t, x) + b_1 V_x(t, x) + \\ &+ \frac{W(0, x) - u'''_0(x) - a_1 u''_0(x) - b_1 v''_0(x)}{\beta'_1(0) - u''_0(\xi) - a_1 u'_0(\xi) - b_1 v'_0(\xi)} \cdot \\ &\cdot \frac{\beta'_1(t) - W_x(t, \xi) - a_1 W(t, \xi) - b_1 V(t, \xi)}{\beta'_1(0) - u''_0(\xi) - a_1 u'_0(\xi) - b_1 v'_0(\xi)}, \\ V_t(t, x) &= V_{xx}(t, x) + a_2 W_x(t, x) + b_2 V_x(t, x) + \\ &+ \frac{V(0, x) - v'''_0(x) - a_2 u''_0(x) - b_2 v''_0(x)}{\beta'_2(0) - v''_0(\xi) - a_2 u'_0(\xi) - b_2 v'_0(\xi)} \cdot \\ &\cdot \frac{(\beta'_2(t) - V_x(t, \xi) - a_2 W(t, \xi) - b_2 V(t, \xi))}{\beta'_2(0) - v''_0(\xi) - a_2 u'_0(\xi) - b_2 v'_0(\xi)}. \end{aligned} \quad (7)$$

$$\begin{aligned}
W(0, x) &= \frac{\beta'_1(0) - u''_0(\xi) - a_1 u'_0(\xi) - b_1 v'_0(\xi)}{\beta'_1(T) - \alpha''_1(\xi) - a_1 \alpha'_1(\xi) - b_1 \alpha'_2(\xi)} \\
&\cdot W(T, x) - \frac{\alpha'''_1(x) + a_1 \alpha''_1(x) + b_1 \alpha'_2(x)}{\beta'_1(T) - \alpha''_1(\xi) - a_1 \alpha'_1(\xi) - b_1 \alpha'_2(\xi)} \\
&\cdot \frac{\beta'_1(0) - u''_0(\xi) - a_1 u'_0(\xi) - b_1 v'_0(\xi)}{\beta'_1(T) - \alpha''_1(\xi) - a_1 \alpha'_1(\xi) - b_1 \alpha'_2(\xi)} + u'''_0(x) + \\
&+ a_1 u''_0(x) + b_1 v''_0(x), \\
V(0, x) &= \frac{\beta'_2(0) - v''_0(\xi) - a_2 u'_0(\xi) - b_2 v'_0(\xi)}{\beta'_2(T) - \alpha''_2(\xi) - a_2 \alpha'_1(\xi) - b_2 \alpha'_2(\xi)} \\
&\cdot V(T, x) - \frac{\alpha'''_2(x) + a_2 \alpha''_1(x) + b_2 \alpha'_2(x)}{\beta'_2(T) - \alpha''_2(\xi) - a_2 \alpha'_1(\xi) - b_2 \alpha'_2(\xi)} \\
&\cdot \frac{\beta'_2(0) - v''_0(\xi) - a_2 u'_0(\xi) - b_2 v'_0(\xi)}{\beta'_2(T) - \alpha''_2(\xi) - a_2 \alpha'_1(\xi) - b_2 \alpha'_2(\xi)} + v'''_0(x) + \\
&+ a_2 u''_0(x) + b_2 v''_0(x).
\end{aligned} \tag{8}$$

$$\begin{aligned}
W(t, 0) &= \psi'_1(t), & V(t, 0) &= \psi'_3(t), \\
W(t, 1) &= \psi'_2(t), & V(t, 1) &= \psi'_4(t).
\end{aligned} \tag{9}$$

Задача (7)-(9) - это неклассическая прямая задача с нелокальными входными данными. После того как задача (7)-(9) решена, функции  $u(t, x)$ ,  $v(t, x)$ ,  $f_1(t)g_1(x)$  и  $f_2(t)g_2(x)$  восстанавливаются по формулам:

$$\begin{aligned}
u(t, x) &= \int_{\xi}^x \int_0^t W(\tau, \eta) d\tau d\eta + u(x) - u_0(\xi) + \beta_1(t), \\
v(t, x) &= \int_{\xi}^x \int_0^t V(\tau, \eta) d\tau d\eta + v(x) - v_0(\xi) + \beta_2(t), \\
f_1(t)g_1(x) &= \\
&= \frac{\int_{\xi}^x W(0, \eta) d\eta + \beta'_1(0) - u''_0(\xi) - a_1 u'_0(\xi) - b_1 v'_0(\xi)}{\beta'_1(0) - u''_0(\xi) - a_1 u'_0(\xi) - b_1 v'_0(\xi)} \\
&\cdot (\beta'_1(t) - \int_0^t W_x(\tau, \xi) d\tau - a_1 \int_0^t W(\tau, \xi) d\tau - b_1 \int_0^t V(\tau, \xi) d\tau - \\
&\quad - u''_0(\xi) - a_1 u'_0(\xi) - b_1 v'_0(\xi)), \\
f_2(t)g_2(x) &= \\
&= \frac{\int_{\xi}^x V(0, \eta) d\eta + \beta'_2(0) - v''_0(\xi) - a_2 u'_0(\xi) - b_2 v'_0(\xi)}{\beta'_2(0) - v''_0(\xi) - a_2 u'_0(\xi) - b_2 v'_0(\xi)} \\
&\cdot (\beta'_2(t) - \int_0^t V_x(\tau, \xi) d\tau - a_2 \int_0^t W(\tau, \xi) d\tau - b_2 \int_0^t V(\tau, \xi) d\tau - \\
&\quad - v''_0(\xi) - a_2 u'_0(\xi) - b_2 v'_0(\xi)).
\end{aligned}$$

Прямая дифференциальная задача аппроксимируется разностной задачей:

$$\begin{aligned}
\frac{P_j^{n+1} - P_j^n}{\tau} &= \frac{P_{j+1}^{n+1} - 2P_j^{n+1} + P_{j-1}^{n+1}}{h^2} + \\
&+ a_1 \frac{P_{j+1}^n + P_{j-1}^n}{2h} + b_1 \frac{R_{j+1}^n + R_{j-1}^n}{2h} +
\end{aligned}$$

$$\begin{aligned}
&+ \frac{P_j^0 - u'''_0(x_j) - a_1 u''_0(x_j) - b_1 v''_0(x_j)}{\beta'_1(0) - u''_0(x_k) - a_1 u'_0(x_k) - b_1 v'_0(x_k)} \\
&\cdot \frac{\beta'_1(t_n) - \frac{P_{k+1}^n - P_{k-1}^n}{2h} - a_1 P_k^n - b_1 R_k^n}{\beta'_1(0) - u''_0(x_k) - a_1 u'_0(x_k) - b_1 v'_0(x_k)}, \\
&j = \overline{1, N-1}, \quad n = \overline{1, M-1}. \\
\frac{R_j^{n+1} - R_j^n}{\tau} &= \frac{R_{j+1}^{n+1} - 2R_j^{n+1} + R_{j-1}^{n+1}}{h^2} + \\
&+ a_2 \frac{P_{j+1}^n + P_{j-1}^n}{2h} + b_2 \frac{R_{j+1}^n + R_{j-1}^n}{2h} + \\
&+ \frac{R_j^0 - v'''_0(x_j) - a_2 u''_0(x_j) - b_2 v''_0(x_j)}{\beta'_2(0) - v''_0(x_k) - a_2 u'_0(x_k) - b_2 v'_0(x_k)} \\
&\cdot \frac{\beta'_2(t_n) - \frac{R_{k+1}^n - R_{k-1}^n}{2h} - a_2 P_k^n - b_2 R_k^n}{\beta'_2(0) - v''_0(x_k) - a_2 u'_0(x_k) - b_2 v'_0(x_k)}, \\
&j = \overline{1, N-1}, \quad n = \overline{1, M-1}.
\end{aligned}$$

$$\begin{aligned}
P_j^0 &= \frac{\beta'_1(0) - u''_0(x_k) - a_1 u'_0(x_k) - b_1 v'_0(x_k)}{\beta'_1(T) - \alpha''_1(x_k) - a_1 \alpha'_1(x_k) - b_1 \alpha'_2(x_k)} \\
&\cdot P_j^N - \frac{\alpha'''_1(x_j) + a_1 \alpha''_1(x_j) + b_1 \alpha'_2(x_j)}{\beta'_1(T) - \alpha''_1(x_k) - a_1 \alpha'_1(x_k) - b_1 \alpha'_2(x_k)} \\
&\cdot \frac{\beta'_1(0) - u''_0(x_k) - a_1 u'_0(x_k) - b_1 v'_0(x_k)}{\beta'_1(T) - \alpha''_1(x_k) - a_1 \alpha'_1(x_k) - b_1 \alpha'_2(x_k)} + \\
&\quad + u'''_0(x_j) + a_1 u''_0(x_j) + b_1 v''_0(x_j), \\
&j = \overline{0, N}, \\
R_j^0 &= \frac{\beta'_2(0) - v''_0(x_k) - a_2 u'_0(x_k) - b_2 v'_0(x_k)}{\beta'_2(T) - \alpha''_2(x_k) - a_2 \alpha'_1(x_k) - b_2 \alpha'_2(x_k)} \\
&\cdot R_j^N - \frac{\alpha'''_2(x_j) + a_2 \alpha''_1(x_j) + b_2 \alpha'_2(x_j)}{\beta'_2(T) - \alpha''_2(x_k) - a_2 \alpha'_1(x_k) - b_2 \alpha'_2(x_k)} \\
&\cdot \frac{\beta'_2(0) - v''_0(x_k) - a_2 u'_0(x_k) - b_2 v'_0(x_k)}{\beta'_2(T) - \alpha''_2(x_k) - a_2 \alpha'_1(x_k) - b_2 \alpha'_2(x_k)} + \\
&\quad + v'''_0(x_j) + a_2 u''_0(x_j) + b_2 v''_0(x_j), \\
&j = \overline{0, N}.
\end{aligned}$$

Разностная задача решается маршевым и итерационным методами.

Для решения разностной задачи создан программный продукт в среде программирования C++. Предложенный алгоритм был апробирован на ряде тестов. Тестовые расчеты показали сходимость итерационного процесса и сходимость численного решения к точному при уменьшении шагов сетки. В частности для теста:

$$u = e^{pt} \sin\left(\frac{\pi x}{6}\right),$$

$$v = e^{pt} \sin\left(\frac{\pi x}{3}\right),$$

$$f_1(t)g_1(x) = e^{pt} (p \sin\left(\frac{\pi x}{6}\right) + \frac{\pi^2}{36} \sin\left(\frac{\pi x}{6}\right) -$$

$$\begin{aligned}
& -a_1 \frac{\pi}{6} \cos\left(\frac{\pi x}{6}\right) - b_1 \frac{\pi}{3} \cos\left(\frac{\pi x}{3}\right), \\
f_2(t)g_2(x) = & e^{pt} \left( p \sin\left(\frac{\pi x}{3}\right) + \frac{\pi^2}{9} \sin\left(\frac{\pi x}{3}\right) - \right. \\
& \left. -a_2 \frac{\pi}{6} \cos\left(\frac{\pi x}{6}\right) - b_2 \frac{\pi}{3} \cos\left(\frac{\pi x}{3}\right) \right).
\end{aligned}$$

при шагах сетки  $\tau = 0.01$   $h = 0.01$  итерационный процесс сошелся за 4 итерации, относительная погрешность для функций  $u$  и  $v$  не превосходит 0.009764%, для функции  $f_1(t)g_1(x)$  не превосходит 0.001978%, для функции  $f_2(t)g_2(x)$  не превосходит 0.003384%.

## Список литературы

- [1] Кучунова Е.В., Расопов В.Е. Численная идентификация коэффициентов параболических уравнений. *Вестник КрасГУ. Серия Физ.-мат. науки*, 5/2:7–14, 2004.
- [2] Мандрик Ю.В., Расопов В.Е. Численная идентификация коэффициентов одномерного параболического уравнения. *Вестник КрасГУ. Серия Физ.-мат. науки*, 1, 2006.
- [3] Жак Т.Ю., Расопов В.Е. . Численная идентификация свободного члена специального вида в параболическом уравнении. *Международная конференция Алгебра и ее приложения: Тезисы докладов*, Красноярск:186, 2007.
- [4] Кабанихин С.И. *Обратные и некорректные задачи*. Новосибирск, Сибирское научное издательство, 2009.
- [5] Самарский А.А., Вабищевич П.Н. . *Численные методы решения обратных задач математической физики*. Москва, 2009.

## Размышления по поводу судьбы нашей вузовской науки

Юрий Дмитриевич Григорьев

Санкт-Петербургский государственный  
электротехнический университет (ЛЭТИ)  
Санкт-Петербург  
yuri\_grigoriev@mail.ru

**Аннотация.** Приводятся соображения по поводу процессов, происходящих в системе вузовской науки. Высказывается мнение о неприемлемости переноса западных наукометрических стандартов на российскую почву.

**Ключевые слова.** самоорганизующаяся система, наукометрия, индекс цитируемости, рейтинг.

**Аннотация.** Приводятся соображения по поводу процессов, происходящих в системе вузовской науки. Высказывается мнение о неприемлемости переноса западных наукометрических стандартов на российскую почву.

**Ключевые слова.** самоорганизующаяся система, наукометрия, индекс цитируемости, рейтинг.

### 1 Введение

В настоящий момент (февраль — апрель 2013 г.) на страницах [1] — электронного издания УБМ (Ин-т проблем управления им. В. А. Трапезникова РАН), развернута оживленная дискуссия по поводу применимости западных стандартов к оценке уровня научного потенциала как отдельных ученых, так и научных коллективов (институтов, университетов).

Предмет дискуссии исключительно актуален, так как затрагивает основополагающие вопросы не только самой науки как культурного феномена, но, по существу, и вопрос значительно более глобальный — вопрос о том, куда мы движемся как цивилизация, противостоящая западной.

Любопытные высказывания на близкую тему можно также найти на форуме [2] (раздел «Образование»). И хотя аудитория участников этого формума и формат дискуссии абсолютно другие, в оригинальности и образности высказываний, которые можно там обнаружить, этому форуму не откажешь.

На страницах данного доклада автор, как участник интернет-конференции [1], воспроизводит свое представленное в этой дискуссии мнение [3], дополняя его новыми соображениями.

### 2 Закономерности сползания отечественной вузовской системы к западным стандартам

В течение последних двадцати лет в системе управления вузовской наукой проявились тенденции, имеющие характер крайне тревожных системных закономерностей. Чтобы сделать формулировки затрагиваемых социологических проблем более выразительными, автор представляет их в форме «социологических теорем» — прямых цитирований академика РАН В. И. Арнольда [4, 5, 6] и выдающегося философа и социолога А. А. Зиновьева [7, 8]. В качестве «доказательств» теорем приводятся конкретные примеры и различные наукометрические положения, рассыпанные в различных источниках типа [9]. Высказываясь по затрагиваемым вопросам несколько иронично, автор надеется, что будет правильно понято, как должны быть правильно поняты и те авторы, на мнение которых он ссылается.

Вузовское образование и вузовская наука — большая самоорганизующаяся система с присущими ей закономерностями. В частности, она обладает инерционностью, направленной на сохранение ее устойчивости [10, с. 282]. Эта и другие закономерности являются также следствием способов управления этой системой.

Автор отлично сознает, что ход эволюции вузовской системы меньше всего зависит от воли и желаний отдельных ее представителей, но сохраняет слабую надежду, что все вместе мы остановим сползание в пропасть. Остановимся на некоторых закономерностях, наиболее ярко характеризующих сегодняшнее состояние нашей вузовской науки.

**1. Наблюдается мировой процесс снижения общей грамотности.** В вузы приходят студенты, уровень школьной подготовки которых предельно низок. Все самое худшее, что есть в школьном образовании Запада, хлынуло в нашу страну. Учить на первых курсах стало невероятно сложно.

**Теорема 1.** [7, с. 14, 139]. *Факт стремительного падения образованности населения западных стран перестали отрицать даже политики и идеологи. Невежество и мракобесие достигли такого уровня, которого не было даже в средние века.*

Доказательство. (1) Французские студенты в наши дни складывают дроби по правилу  $\frac{1}{2} + \frac{1}{3} = \frac{2}{5}$ , с помощью компьютера выясняют, больше единицы дробь  $\frac{4}{7}$  или нет [5, с. 29, 75]. Даже профессора математики стали помножать два на два с помощью вычислительных устройств, совершающих несколько миллиардов операций в секунду [7, с. 25].

(2) Работая в приемной комиссии (конец 90-х—начало 2000-х гг.), я сталкивался с абитуриентами, которые делили дроби по правилу  $\frac{3}{4} : \frac{2}{5} = \frac{3 \cdot 2}{4 \cdot 5} = \frac{6}{20}$ . Абитуриент, заполнявший по моей просьбе анкету (2006 г.), написал: я родился в г. Ленинграде, живу в Калининском районе, диплома с отличием не имею. Не удивляют и такие студенческие ответы: Как называется столица США? — Нью-Йорк. Кто написал «Мертвые души»? — Гривоедов. □

**2.** *Содержание декларируемых целей развития вузовской науки носит пропагандистский характер.* Всякая деятельность является целенаправленной, т. е. направлена на достижение определенной цели. Цели тесно связаны с проблемами, а их в вузовской системе предостаточно. Если цель поставлена, то она порождает проблему ее достижения. Часто из заявлений руководителей вузовской системы истинную цель невозможно вычленишь. Декларации руководства постоянно расходятся с его реальными действиями.

**Теорема 2.** [7, с. 15]. *Реальность нашего общества имеет мало общего с официальной его концепцией и со всем тем, что о нем можно узнать из средств массовой информации, литературы, кино и науки.*

Доказательство. (1) На собрании профессорско-преподавательского состава (ППС) перед началом учебного года ректор сообщает, что во исполнение указа президента заработная плата ППС должна быть доведена до средней по региону. По Ленинградской области она составляет 32,5 тыс. руб. С надеждой на повышение профессор, получающий 20 тыс. руб. в месяц, сообщает жене — с октября зарплата будет повышена. Однако вскоре выясняется, что средняя зарплата по вузу уже составляет 67 тыс. руб., из чего следует, что никакого повышения зарплаты не будет, задачу материального стимулирования руководству удалось решить без всякого реального повышения зарплаты (а зачем? средняя зарплата зав. кафедрой и так уже выше зарплаты профессора в 8-10 раз).

(2) Недавно узнаю, что грядет оптимизация структуры и кадрового состава вузов. Сокращению в известных мне вузах подлежат от 180 до 240 сотрудников, в значительной степени, ППС. Кафедры, занимающие места в «зоне вылета» по внутривузовскому рейтингу, предполагается расформировывать. Непонятны критерии этих рейтингов (о рейтингах — ниже) и судьба сотрудников сокращаемых кафедр. В общем,

«все будут богаты, ... кроме бедных». □

**3.** *Снижается качество управления вузовской наукой за счет ослабления в вузовской системе отрицательных обратных связей.* Многие связи между элементами нормально функционирующих нелинейных систем являются отрицательными. Если их недостаточно или они сознательно разорваны и вместо них преобладают положительные обратные связи, то система теряет устойчивость, начиная эволюционировать к худшему состоянию.

**Теорема 3.** [7, с. 15-16]. *У нас унакомысливших не жгут на кострах, как во времена инквизиции, и не сажают в концентрационные лагеря, как в тоталитарных режимах XX века. У нас их просто игнорируют.*

Доказательство. (1) В процессе происходящей реформы высшего образования профессор настоял на включении в программу обучения специального курса, который, в частности, позволил бы ему заранее отбирать студентов на дипломное проектирование. Результат принятого руководством факультета и кафедры решения следующий: дисциплина стоит в осеннем семестре 6-го курса, вместо экзамена — зачет, контрольных мероприятий (задания, контрольной работы, реферата) нет, соотношение лекционных и семинарских часов не соответствует задачам курса. При принятии решения ни один из этих вопросов с профессором не обсуждается и не согласовывается. Все часы, в том числе на якобы самостоятельную работу студентов, планируются и спускаются сверху. Требования Болонского процесса неукоснительно соблюдаются. В итоге, вообще отпадает надобность в разработке учебно-методических материалов и необходимость самого курса в целом.

(2) В еще одном вузе читаю теорию вероятностей. Аналогично — дисциплина стоит на первом курсе. Осталось математический анализ переставить на четвертый курс, и будет полный порядок. □

**4.** *В вузах действует опасная по своим последствиям многоступенчатая (административная) система управления.* Опасность многоступенчатого управления проявляется в чрезвычайно общей ситуации, т. е. в ситуации, когда подразумеваются не только технические, но и административные системы. Стандартной является следующая система управления: ректор—проректоры (не менее четырех плюс управляющие различными комитетами плюс различные синектуры типа почетный ректор, советник, помощник, референт, консультант и т. п.)—декан—зам. декана (по науке, по магистерской подготовке и т. д.)—зав. кафедрой—зам. зав. кафедрой (по науке, по учебной работе)—исполнитель (доцент, профессор). Как показывают простейшие примеры анализа многоступенчатых систем, поведение руководителя  $i$ -го ранга управляется вышестоящим руководителем ( $i +$



1)-го ранга и т. д. вплоть до ранга  $n$ . При  $n$  достаточно большом это приводит к непредсказуемым результатам.

**Теорема 4.** [6, с. 19], [7, с. 134]. *Справедливы следующие утверждения:*

(1) *Многоступенчатое управление при  $n > 3$  неустойчиво. Управление при  $n = 2$  приводит к периодическим колебаниям, но не вызывает катастрофического нарастания колебаний, происходящего при трех- и более ступенчатом управлении. Настоящую устойчивость обеспечивает только одноступенчатое управление, при котором управляющее лицо более заинтересовано в интересах дела, чем в поощрении со стороны начальства.*

(2) *Наши начальники любят свое положение сравнивать с положением дирижера оркестра. И каждый считает себя дирижером по отношению к нижестоящим. Кто из них дирижер? Все вместе? А что было бы с оркестром, если бы дирижер был не один, а целая иерархия дирижеров в десять и более ступеней?*

Доказательство. (1) (а) Многоступенчатая система заказчик  $\rightarrow$  ректор—проректор—декан—зам. декана—зав. кафедрой—зам. зав. кафедрой  $\rightarrow$  исполнитель ( $n = 6$ ) приводит к такому распределению средств заказчика, при котором они заканчиваются как раз перед носом исполнителя. Финансовые запросы звеньев цепочки не ограничены сверху, каждое звено потребляет свою долю в размере, прямо пропорциональном своему рангу, это — положительная обратная связь (аппетит растет во время еды).

(б) Стандартная схема утверждения рабочей программы учебной дисциплины (подписи) имеет вид: автор (профессор) — рецензент — зав. кафедрой — декан — рук. магистерской программой — председатель методической комиссии — начальник методического отдела ( $n = 6$ ). В советское время использовалось управление исполнитель (доцент, профессор) — зав. кафедрой — декан ( $n = 2$ ). Система работала идеально.

(с) Иерархия ступеней прохождения диссертации в процессе ее подготовки к защите и сама защита: научный руководитель — представление на кафедре — экспертная комиссия диссертационного Совета (ДС) — ведущая организация — оппоненты — отзывы сторонних специалистов — решение демократического большинства членов ДС (защита) — ВАК. Что из этого теперь получается — всем хорошо известно. Слишком много дирижеров.

(2) По вузу вводятся «Правила внутреннего распорядка университета», разработанные администрацией на основе различных федеральных законов, в том числе об образовании (перечисляются — какие). В них особо выделены основные права и обязанности

обучающихся (почему *обучающихся*, а не студентов? может, и мы теперь — *обучающиеся*?). В одном из пунктов сообщается — «обучающиеся имеют право в *свободное от учебы* время работать на предприятиях, в учреждениях и организациях любых организационно-правовых форм». Это нонсенс — студенты давно и по полной программе работают в учебное время в перечисленных выше организациях без всякого на то распоряжения сверху. Речь идет о магистрантах, которые занимаются этим делом вместо написания своих диссертаций. Такое впечатление, что реальной информации о том, что происходит внизу, у руководства нет.  $\square$

5. *Распределение финансируемых бюджетом и другими организациями заказов с помощью грантов и конкурсов неэффективно.* В работе Ракитова [11, с. 16] отмечается, что «... Федеральные органы выделяют средства на проведение международных и отечественных исследовательских проектов на основе создания виртуальных коллективов. Но ни бюджетная, ни конкурсная поддержка не все еще дают ожидаемого эффекта».

**Теорема 5.** [7, с. 92]. *Конкурсы у нас производятся регулярно и по всякому поводу. Это — одно из средств манипулирования массами и оболванивания их, а также источник наживы для бесчисленных паразитов и жуликов.*

Доказательство. (1) На видном месте в здании университета объявление: ректор (фамилия) и декан (фамилия) такого-то факультета выиграли грант на выполнение такой-то научно-исследовательской работы (приводится название работы, сама работа выполняется по поручению министерства такого-то).

(2) Зав. кафедрой гордо сообщает профессору кафедры. — Мы с ректором выиграли конкурс на создание комплекта учебников по такой-то дисциплине. Денежная премия — 300 тыс. руб. Комментарии излишни.

(3) Руководство «просит» автора статьи указать в сноске, что данное исследование поддержано грантом РФФИ, хотя сам автор в нем на самом деле не присутствует. Затем статья, в числе других таких же поддержанных, включается держателями гранта в соответствующий отчет.  $\square$

Давно замечено, что наша бюрократия обладает одним неоспоримым достоинством — отсутствием ума. Вот и в этом случае мы очень рады за наших руководителей. По-видимому, ничего не изменилось с приснопамятных времен (см. раздел «Руководство» в [8, с. 87-91]).

6. *Наблюдается неудержимый рост числа бесполезных «научных» публикаций.* Существующая система подготовки научных кадров приводит к бессмысленному росту числа публикаций и, в результате, — к их

практически полному обесцениванию. Согласно существующим требованиям, аспирант при выходе на защиту должен иметь, как минимум, три публикации в изданиях из перечня ВАК. Каждый доктор наук, чтобы сохранить членство в ДС, должен иметь, в среднем, одну такую публикацию в течение трех последних лет. В конце каждого года научный работник (доцент, профессор) должен представить отчет о научной работе с желательным большим числом публикаций (статей) за отчетный период. Таким образом, изначально программируется количество написания научных статей в единицу времени. Как справедливо отмечается в (Горохов [12]), здесь «необходима корректная интерпретация полученных показателей» (количества статей, особенно, если речь идет о вузах. — Ю. Г.).

**Теорема 6.** [5, с. 29]. *В среднем, число напрасных публикаций, необходимых их авторам для карьеры и трудоустройства, является большим 90%.*

Доказательство. (1) Пример из недавней практики: молодой аспирант (25-26 лет) выходит на защиту, имея более 50 публикаций, из которых 14 из перечня ВАК, плюс 2 монографии.

(2) К похожему выводу о ненужности чрезмерного числа публикаций приходит проф. Орлов [13], считающий, что статья — это необходимый элемент процесса научного творчества. □

Необходимо понимать, что нет никакой необходимости засорять журналы заведомо слабыми работами. Нет ничего зазорного в том, чтобы начинать научную деятельность с публикаций в форме препринтов, депонированных рукописей, отчетов, статей в ведомственных сборниках, тезисов конференций и т. д. Как говорили еще древние греки, кто много доказывает, тот ничего не доказывает. Сколько бы ни представил аспирант ваковских статей (зачастую — в формально рецензируемых изданиях вуза, где он учится), все равно это не доказывает его ученость. Возможно, для всех ваковских журналов следует ввести правило представления статьи одним из членов редколлегии журнала. Это известное правило давно действует в академических журналах.

**7.** *Качество диссертаций, включая регламент их защиты, ниже всякой критики.* Существующие процедуры защиты диссертаций и их утверждения ВАК не защищают от нарушения элементарных этических норм. Это открывает дорогу в науку проходцам, «энергичным людям». Манипуляции с процедурой защиты часто видны невооруженным взглядом (системы наблюдения их не фиксируют, здесь все чинно), но к ним уже все привыкли и не обращают на них внимания.

**Теорема 7.** [7, с. 196]. *Люди настолько натренированы врать, лицемерить, притворяться, что они скрывают свои мысли даже от самих себя, если это*

*нужно.*

Доказательство. (1) Встречаются случаи, когда еще неоперившийся кандидат в кандидаты представляет в ДС до 20 отзывов на автореферат (все отзывы под копирку с замечаниями типа «над символом таким-то не поставлена шляпка») и непостижимое число выступлений на конференциях, на которых диссертант особенно в случае, если она иногородняя, конечно, не выступал. Вместо восхищения талантом соискателя возникает чувство брезгливости к самой процедуре защиты, зарегулированной до такой степени, что она превращается в клоунаду.

(2) Даю отзыв на автореферат докторской диссертации. Значительное место в ней уделено так называемому  $\lambda$ -распределению вероятностей. Распределение интересное, вот только в автореферате выражение для его плотности приводится неверное. Пишу замечание и еще несколько (всего 9), даю положительное заключение. Спустя короткое время узнаю, что странным образом неудобный отзыв в ДС не поступает (приходит в ДС «после» защиты).

(3) Процветает плагиат. В моей практике был случай, когда я просто вызвал руководителя аспиранта и в присутствии их обоих показал страницы, переписанные из книги без ссылок на нее. Теперь мой клиент уже не только кандидат, но и доктор. Совершенно открыто в интернете предлагаются услуги по написанию диссертаций под ключ, которые включают обоснование актуальности, новизны диссертации, полный качественный текст диссертации (введение, главы диссертации, заключение, автореферат, библиография), написание научных статей, монографий и т. д. Диссертации пишутся по экономике, праву (юриспруденции), психологии, педагогике, философии и другим специальностям гуманитарных наук. Существуют и расценки. Выводящие отсюда последствия — см. [14] и сообщения о делах в ВАК и МНПУ по теле- и радиоканалам 3-10 февраля сего года. □

Таких хитростей белого человека придумано сколько угодно, но власть этого не замечает. Главное — честь мундира. Чиновнику от науки очень удобен придуманный им ритуал защиты — подсчитать количество публикаций, вогнать оригинальные свежие мысли в казенные формулировки автореферата и присвоить степень не за полученный результат, а на основе формального подсчета числа публикаций, формальных формулировок (с заранее подготовленными шаблонами) заключения ДС и одобрения диссертации демократическим большинством членов ДС.

В советское время мне приходилось присутствовать на защите диссертаций по физ.-мат. наукам в ИМ СОАН СССР им. С. Л. Соболева — никаких левых отзывов, только оппоненты и ведущая организация.

Более того, бывали случаи, когда оппонент находил ошибки в диссертации непосредственно перед самой защитой, тут же сообщал о них соискателю и вовремя помогал все исправить. Все честно и по-деловому. Ведь можем же, если захотим.

**8. Качество научных кадров стремительно падает.** Непропорциональный рост числа защищаемых докторских диссертаций по отношению к кандидатским приводит к катастрофическому падению качества вузовских научных кадров. Существовавшее в советское время соотношение отклоняемых ВАК диссертаций изменилось с 300 : 1000 на 3 : 1000, а количество защищаемых докторских и кандидатских диссертаций — с 1 : 10 на 1 : 3.

**Теорема 8.** [7, с. 130]. *Главным теперь стало не творчество в старом смысле, а возможность пробиться к делу, считаемому творческим, и быть официально признанным в таком качестве.*

Доказательство. (1) Получив для экспертного заключения докторскую диссертацию, обнаруживаю, что все теоремы в ней пронумерованы по способу нумерации квартир в жилых домах — вместо, скажем, теорема 2.1 читаю — теорема №2.1. Докторант ни разу не открывал ни одной научной монографии и просто не знаком с существующей научной традицией нумерации теорем, лемм и т. д.

(2) Знакомлюсь с учебным пособием (Кундышева [15, с. 97-102]), которое рекомендовано УМО по образованию в области прикладной математики и управления качеством для студентов вузов (один из трех рецензентов — доктор физико-математических наук) и обнаруживаю полную безграмотность автора: (а) как известно, кривая Лоренца  $L(x)$  определена на отрезке  $[0, 1]$ , однако интеграл от нее автор берет почему-то от 0 до  $\infty$ ; (б) приводятся неверные, если не сказать глупые, выражения как для самой кривой  $L(x)$  (доходы распределены по Парето), так и для определяемого с ее помощью индекса Джини; (в) плотность распределения Парето, получаемая дифференцированием функции распределения, записана неверно, т. е. автор просто не умеет дифференцировать.

(3) Любопытный пример являет собой учебник для гидрологов (Сикан [16]), выдержавший два издания. Здесь тоже один из рецензентов — доктор физико-математических наук (остальные два — доктора технических наук). Из этого учебника узнаем [16, с. 20], что математическое ожидание суммы *независимых* случайных величин и дисперсия суммы случайных величин (не уточняется — *каких*) равны сумме их математических ожиданий и дисперсий соответственно. Для вящей убедительности автор ссылается на ни в чем не повинных авторов двух других изданий, где якобы эти утверждения доказываются. □

В качестве обратного примера достаточно напомнить

практику советских лет — никаких липовых рецензентов, а качество учебников и монографий — высочайшее. Попробуйте, например, найти ошибки в классическом курсе дифференциальных уравнений Степанова [17]. И это без всяких рецензентов.

**9. Научные школы и коллективы распадаются и деградируют.** Сохранившие активность научные кадры мигрируют на Запад, а те, что остаются, постепенно теряют творческий потенциал, не получая поддержки со стороны государства. Наука, если и поддерживается, то очень выборочно (академическая — в первую очередь, вузовская и ведомственная — в последнюю). Если ученый не попадает в эту выборку (университет не входит, скажем, в перечень вузов, имеющих статус национального достояния и поэтому не имеет ведомственных или правительственных заказов), то он — ученый второго сорта, т. е., по существу, никто.

**Теорема 9.** [7, с. 147]. *Мои знания избыточны по отношению к тому, что я делаю профессионально. Никакой возможности открыто удовлетворять творческие потребности и тем более получить одобрение в качестве исследователя у меня нет. А заниматься научным творчеством исключительно ради удовлетворения интеллектуального любопытства, причем тайно, — значит, усиливать состояние одиночества и сознание несправедливости судьбы.*

Доказательство. (1) В большинстве университетов наблюдается общая картина — при любом возможности наиболее востребованные доценты и профессора уезжают на Запад. С начала 90-х годов с одной только нашей кафедры уехали четверо, в том числе: в США — 2, в Швецию — 1, в Польшу — 1. А теперь приведем цитату из (Арсенов [18, с. 73]), — «зарплата питерских ученых в несколько десятков раз меньше, чем у их коллег на Западе. Многие ученые уже покинули институт, ведь за рубежом идет самая настоящая борьба за «русские мозги», и большинство молодых специалистов уезжают на Запад. Это основная проблема института (ЛОМИ им. В. А. Стеклова. — Ю. Г.), так как за последнее десятилетие эмигрировали более 40 специалистов. Сегодня здесь осталось 126 сотрудников, из них 51 доктор и 56 кандидатов наук».

(2) Средний возраст ППС на многих кафедрах и в ДС — в диапазоне 55-60 лет (см. табл. 1), в качестве научных семинаров бывают только представления диссертаций, научной среды нет, съездить на конференцию удастся лишь изредка и только за свой счет, общеканцелярских научных хоздоговорных тем нет.

Индекс Джини возрастного состава членов ДС, в котором я состою, равен  $G = 0.3826$ . Случайно он даже

Таблица 1: Распределение по возрастным группам ППС факультета компьютерных технологий и информатики СПбГЭТУ.

Возраст, лет	менее 35	[36, 50]	[51, 65]	более 65
Процент, %	16.7	17.4	27.9	38.0

почти совпал с гармонической константой

$$\Phi^{-2} = \frac{3 - \sqrt{5}}{2} = 0.382,$$

где  $\Phi = (1 + \sqrt{5})/2 = 1.618$  — золотое сечение (Григорьев, Мартыненко [19]). Социологи знают, что если  $G > G_0 \cong 0.4$  (уж не взять ли здесь в качестве  $G_0$  именно 0.382), то распределение возрастов уже можно считать сильно отличающимся от равномерного (слишком велик перевес возрастной группы членов ДС над «молодежной»). Таким образом, положение с профессорами следует считать близким к критическому.

(3) Разрушены многие научные школы, функционировавшие во времена СССР. Скажем, развалились или утратили былую мощь школы по планированию эксперимента, существовавшие в Москве (МГУ, МЭИ), Санкт-Петербурге (СПбГУ), Новосибирске (НГТУ). □

**10.** Пресловутый «Болонский процесс», унижительное обезьянничанье перед Западной Европой в деле организации нашего высшего образования и науки в целом, ничего, кроме чувства раздражения, не вызывают.

**Теорема 10.** [7, с. 431]. *Завоевая мир для себя, Запад истребил все возможные конкурентоспособные «точки роста» цивилизаций иного рода.*

Доказательство. (1) Мы постоянно наступаем на те же грабли, поскольку все, что связано с отношениями с Европой, мы уже давно проходили. Совершенно прав Хруцкий [20], напоминая, что «еще Н. Я. Данилевский [21], нисколько не умаляя достижений европейской (западной) науки, но испытывая к ней «высочайшее уважение» — определяет существо болезни «российского пациента» в его «европейничаньи» и призывает прекратить «смотрение на дела России сквозь европейские очки».

(2) Программа двойных дипломов (уж очень нам хочется стать Европой) абсолютно нереализуема. Как отмечается в (Жмудь [22]), на этом пути встают, как минимум, следующие препятствия: (а) страны, находящиеся в авангарде Болонского процесса, менять свои учебные планы в сторону сближения с учебными планами стран, находящимися в арьергарде, никогда не согласятся; (б) чтобы согласовать стандарты учебных планов, необходимы решения на уровне министерств и ведомств с последующими межгосу-

дарственными соглашениями о взаимном признании дипломов, что крайне сомнительно даже в отдаленной перспективе.

В этой связи вспоминается эпизод с академиком С. Н. Бернштейном. Он в свое время закончил Сорбонну и защитил там диссертацию под руководством выдающегося математика А. Лебега. Так вот, его заслуги у нас в России по чисто формальным признакам не могли быть признаны. Чтобы получить место профессора в Харьковском университете, он должен был заново защитить диссертацию в Совете Харьковского, в то время российского, университета. Н-да, было время, да прошло.

**11.** *Рейтинги университетов и наукометрические индикаторы отечественной науки находятся на неправдоподобно низких уровнях по сравнению с западными.* Наукометрические рейтинги и индексы сопровождают нас всюду. Они дают полезную информацию о тех или иных упорядочиваемых объектах, используемую для выработки соответствующих решений. Но существующие методы построения рейтингов не всегда дают согласованные ответы. В итоге, это приводит к противоречивым, в значительной степени — малодостоверным, выводам.

**Теорема 11.** [7, с. 183]. *В искусственный интеллект человечества загрузили огромную массу глупости, невежества, мракобесия. В понимании своего общества, своей жизни и самих себя мы оказались на уровне наших первобытных предков.*

Доказательство. (1) Как справедливо отмечается в (Горохов [12], см. также [23]), «делать выводы об уровне развития науки на основе библиометрических данных крайне опасно. Не случайно университет Гамбурга отказался выдавать свои данные в международную и национальную рейтинговые организации, хотя и занимает не последнее место в списке ведущих немецких университетов».

(2) Беглый просмотр наиболее известных рейтингов показывает, что в первые пятьсот университетов мира входят от одного до четырех российских вузов [24]. Согласимся, такая издевательская, по-другому не скажешь, оценка нашей вузовской системы задевает за живое, травмируя наше общественное сознание. В общей сложности, в число передовых постоянно попадают четыре университета — МГУ (Москва), СПбГУ (Санкт-Петербург), НГУ (Новосибирск), ТГУ (Томск). Их положение в рейтинге QS—THES показано в табл. 2 [24].

Не будем себя обманывать. Из этого списка можно смело исключить три первых вуза. Дело в том, что эти три университета находятся в городах, где есть отделения РАН. ППС этих вузов напрямую сотрудничает с этими отделениями, тем самым повышая рейтинг своих вузов за счет элитной части нашей науки, т. е. РАН, не относящейся к вузам. Скажем, в



Таблица 2: Положение российских университетов в рейтинге QS–THES [24]. Данные за 2013 год — рейтинг Webometrics (Источник: пресс-служба ТГУ и ТПУ).

Год	Университет			
	МГУ	СПбГУ	НГУ	ТГУ
2007	183	239		
2008	136	224	401 – 450	401 – 450
2009	101	168	312	401 – 450
2010	93	310	315	401 – 450
2012		253		
2013	147	538	604	465

штатном расписании НГУ, вообще нет научных работников (доцентов, профессоров), так как все они на самом деле являются сотрудниками поблизости расположенных академических институтов, т. е. являются преподавателями по совместительству. В самом же штате НГУ имеются разве что преподаватели английского языка и физкультуры. Поэтому, если строго формально судить по этому признаку, то в рейтинге QS–THES останется только ТГУ, за что ему честь и хвала.

В работе (Маркусова и др. [24, с. 124]) справедливо отмечается, что не надо соревноваться с западными и азиатскими университетами, приписывая себе всякими правдами и неправдами несуществующие заслуги. Раз уж нас измеряют там своими мерками, то и мы можем измерить их своими. Просто нужно понять, с какой целью устанавливается рейтинг вуза и как его предполагается использовать. После этого будет ясно, как его строить. Просто смешно, нашими инженерами-физиками, выпускниками МФТИ, МИФИ, и т. д., переполнены все ядерные центры Франции и Швейцарии (западные эмиссары рыщут по Москве в поисках застрявших в столице физиков), а подготовившие их вузы находятся в хвосте западных рейтингов.

(3) Визитной карточкой всякого западного ученого служит его индекс Хирша  $H$ . Очевидно, что чрезмерная абсолютизация этого показателя в наших условиях наносит только вред. Ряд его недостатков отмечен в (Чеботарев [25]). Применяя китайские технологии (взаимное цитирование, размножение количества статей за счет многочисленного соавторства, написание заведомо никому ненужных статей — см. об этом выше, а также (Горохов [12, с. 15])) и, конечно, публикуясь только в западных журналах, (практика перевода наших академических журналов на английский большой погоды не делает) можно сравнительно легко довести индекс Хирша до  $H = 20 - 30$ .

На значение индекса Хирша существенно влияют «модность» научного направления (т. е. численность отряда ученых, работающих в рассматриваемой об-

ласти), малая цитируемость статей экстра-класса (считаю, что тысячи ученых набросятся на результат Г. Перельмана, решившего «проблему А. Пуанкаре» (Арнольд [5, с. 33]; Арсенов [18]), устремляя число ссылок на него к бесконечности). Ряд других факторов также способствуют превращению  $H$  в очень сомнительный показатель.

«Ученых» в возрасте 30–40 лет с  $H = 20 - 30$  на Западе тысячи, хотя известно, что тех же математиков самого высоко класса больше пяти в стране не бывает (Арнольд [5, с. 29]). Стремление к безудержному повышению индекса Хирша превращается в спорт, что приводит к неправдоподобно большим его значениям (больше, чем у нобелевских лауреатов или признанных классиков науки).

Совершенно очевидно, что индекс Хирша, как единственный индикатор силы ученого, не отражает всей палитры его научного потенциала, и в этом смысле он абсолютно бесполезен. Гораздо интереснее и объективнее встраивать его в более сложный наукометрический показатель, основанный на анализе *всей* совокупности публикаций автора, т. е. включая публикации на родном языке автора. Скажем, язык классической философии — это все-таки немецкий, а не английский. И почему это автор должен в этом случае переводить свою статью именно на английский (Горохов [12]).

Мой индекс Хирша равен  $H = 2$  и достигнут всего лишь за счет двух публикаций публикаций, кажется, [26] и [27]. О чем-нибудь это говорит? Абсолютно ни о чем. По одному числу, взятому почти с потолка, невозможно понять, слабый или сильный я ученый. В рассматриваемом случае вообще получается, что мой вьетнамский аспирант имеет, по существу тот же индекс Хирша, что и я. Следовательно, он имеет и тот же научный уровень, что и я, в чем я сильно сомневаюсь.

В табл. 3 приведена примерная структура корпуса научных публикаций выдающегося математика и физика В. П. Козлова [28, с. 484–495] (строка  $K$ ) и, для сравнения, — структура корпуса моих публикаций. (строка  $G$ ). Все публикации разбиты на ядро  $C$  и периферию  $P$  — особо значимые и более второстепенные публикации (разбиение условно и может корректироваться, скажем, книги и главы в коллективных монографиях можно включить в ядро). Объем ядра измеряется индексом  $J = C/(C + P)$ , принятым в стилеметрии (Григорьев, Мартыненко [19, с. 193]). Чем больше значение  $J$ , тем выше потенциал ученого.

Средний импакт-фактор  $F$  публикаций ученого вычисляется по формуле

$$F = N^{-1} \sum_{i=1}^n n_i F_i,$$



Таблица 3: Структура корпуса публикаций В. П. Козлова [28, с. 484-495] и Ю. Д. Григорьева: (1) журналы с ИФ, (2) журналы АН без ИФ, (иностранн. ж-лы), (4) тезисы док-ов, (5) программы, свид-ва, депоненты, (6) книги, разделы в коллективных монографиях, (7) сб-ки трудов, (8) журналы без ИФ,  $N$  — суммарное число публикаций.

	Ядро			Периферия					$N$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	
$K$	29	6	3	30	8	4	11	4	95
$G$	17	6	3	47	5	3	35	8	124

где  $n_i$  — количество публикаций в журнале с ненулевым импакт-фактором  $F_i$ ,  $n$  — всего журналов с  $F_i > 0$ , в которых опубликованы статьи автора. Несложные вычисления дают результат, представленный в табл. 4.

Таблица 4: Сравнение наукометрических показателей корпуса публикаций В. П. Козлова и Ю. Д. Григорьева,  $J$  — индекс объема ядра,  $F$  — средний импакт-фактор.

Ученый	$J$	$F$	$F_0 = \max_{1 \leq i \leq n} F_i$
$K$	0.4000	0.6655	2.362
$G$	0.2097	0.2851	0.829

В данном случае сразу виден подавляющий перевес ученого  $K$ . При более глубоком наукометрическом анализе учет более длинного кортежа показателей потребует более тонких индексов, что в итоге даст более объективную оценку уровня ученого.

По существу, необходима разработка *наукометрического* паспорта ученого (ср. биометрический паспорт), в котором, наряду с индексами  $J$  и  $F$ , учитывались бы многие другие показатели, в том числе полученные с помощью CV: индекс разнообразия творчества (перечень журналов, в которых представлены публикации автора), степень его участия в незримых коллективах (совместные публикации, в том числе с иностранными коллегами), стипендии зарубежных университетов и фондов (например, DAAD, фонд им. Г. Герца, фонд Д. Сороса, Институт актуариев Великобритании и т. п.), членство в оргкомитетах конференций и редакционных коллегиях журналов с высоким импакт-фактором, приглашаемость в различные, отечественные и зарубежные, университеты на стажировки, для чтения лекций и т. д.

Имея для каждого научного работника хорошо обоснованный наукометрический кортеж показателей, можно рассчитать его универсальный рейтинг  $R$ , который и следует принимать во внимание при оценке научного потенциала ученого. Вопрос о статусе уче-

ных-руководителей разного ранга надо рассматривать особо, ибо именно в них воплощаются ум, честь и совесть нашей эпохи.

**Основной вывод** *Принятие западных способов управления наукой — ложный путь развития нашей уникальной цивилизации.* Знакомясь с материалами интернет-дискуссии [1], еще раз убеждаешься, что не по пути нам с Западом. Практически все участники дискуссии находят, что те способы измерения научной продуктивности ученых, которые предлагает Запад, крайне ограничены по своим возможностям, не учитывают творческих особенностей самих ученых (скажем, есть ученые индивидуалисты, т. е. информационные изолянты, а есть коллективисты, т. е. информационные диспетчеры), из наукометрической измерительной процедуры могут выпадать междисциплинарные области (Горохов [12]) и т. д.

Прекрасно то, что в дискуссии (Интернет-конференция [1]) принимают ученые всех профилей и рангов. Знакомясь с ее материалами, не только находишь единомышленников, но и еще раз убеждаешься в том, насколько неординарен наш народ, насколько он боится за свою судьбу. В связи с этим, в качестве логического следствия из возникшего заинтересованного разговора, завершим наше обсуждение наукометрических вопросов общим взглядом на науку как культурный феномен.

### 3 Закономерности, присущие науке как культурному феномену

Многое из сказанного выше становится более понятным при рассмотрении науки как культурного феномена, присущего конкретной цивилизации. Конечно, я не социолог, и вообще не гуманитарий, но есть вещи, которые должны быть понятны каждому. Взаимодействие цивилизаций, в данном случае, в области науки, подчиняется всеобщим законам, и действие этих законов мы ощущаем в нашей повседневной реальности.

**12.** *Навязывание сверху нашему научному сообществу западных стандартов в понимании целей науки и ее роли в обществе.* Никто не спорит, все полезное и нужное для нашего общества *оттуда* надо брать. Но, как говорилось в одной детской игре, должно быть так. — *Бери и помни.*

**Теорема 12.** [8, с. 400]. *Нам, русским, Запад навязывает примитивное и ложное понимание нашей истории и нашего социального строя, высосанную из пальца программу на будущее, совершенно не отвечающую нашим условиям и нашим потребностям.*

Доказательство. (1) В одном своем сочинении, написанном еще в 1832 году (т. е. за четыре года до появления в печати своего наиболее известного перво-

го «философического письма»), но опубликованном впервые лишь в 1908 году, Чаадаев со всей определенностью утверждал (Кожин [29]): «Я полагаю, что на учебное дело в России может быть установлен совершенно особый взгляд, что возможно дать ему национальную основу, в корне расходящуюся с той, на которой оно зиждется в остальной Европе, ибо Россия развивалась во всех отношениях иначе, и ей выпало на долю особое предназначение в этом мире. Мне кажется, что нам необходимо обособиться в нашем взгляде на науку не менее, чем в наших политических воззрениях, и русский народ, великий и мощный, должен, думается мне, вовсе не подчиняться воздействию других народов».

(2) Борьбе русского народа за свободу и достойную жизнь на земле — продолжается. И ныне нам более чем когда-либо подобает верить в Россию, видеть ее духовную силу и своеобразие и выговаривать за нее, от ее лица и для ее будущих поколений ее *творческую идею*. Эту творческую идею нам не у кого и не для чего заимствовать она может быть только русской, национальной. Она должна выражать русское историческое своеобразие и в то же время — русское историческое призвание (Ильин [30, с. 322]).

**13.** *Неприятие западной цивилизацией ценностей других цивилизаций.*

Необходимо понимать, что все, что мы пытаемся перенимать у Запада, давно приспособлено для нужд именно Запада. Мы всегда и везде, с неуклонностью восхода Солнца, будем оказываться в арьергарде всего того, что называется на Западе наукой. Мы для них — не пришей кобыле хвост. Вот если мы уезжаем к ним с почти готовым изобретением (пример: В. Г. Зворыкин, эмиграция из СССР в 1923 г., получение патента на изобретение телевидения в 1928 году, но только после получения гражданства США и т. д.), то тогда, пожалуй ста.

**Теорема 13.** [8, с. 388]. *На Западе русских пускают в число выдающихся деятелей мировой культуры в исключительно редких случаях, да и то при этом стремятся занижить их фактические масштабы.*

Доказательство. (1) Никогда в жизни Запад не признает, что периодическую систему элементов открыл русский ученый Д. И. Менделеев. У них это звучит примерно так. — *The periodic table is a table of the chemical elements in which the elements are arranged by order of atomic number in such a way that the periodic properties (chemical periodicity) of the elements are made clear.* Как и почему с перевесом в один голос Нобелевская премия по химии за 1906 год досталась французскому химику А. Муассану (где он теперь, этот нобелевский лауреат Муассан, со своими исследованиями фтора?), можно прочитать в (Чолаков [31, с. 31, 168]).

(2) Та же всем известная история с изобретением

радио. Ясное дело, что изобретение радио *широкой* западной общественностью приписывается Г. Маркони. А вот что известно *узкой* научной общественности о решении Нобелевского Комитета, касающемся присуждения Нобелевской премии по физике за 1909 год (Чолаков [31, с. 345]). — «Присудить Нобелевскую премию года по физике Г. Маркони и Ф. Брауну за работы по созданию беспроводного телеграфа». Согласитесь, *работы по созданию и изобретение* радио — это далеко не одно и то же. Когда в 1897 году Маркони сунулся в Англию с заявкой на патент об изобретении радио (беспроволочного телеграфа), ему вежливо указали на дверь. Патент ему, конечно, выдали, но только на *применение* электромагнитных волн для беспроводной связи (Чолаков [31, с. 103]).

А где же истинный изобретатель радио А. С. Попов? А нигде, на бороде. Правда, один из своих приборов для проведения физических экспериментов по передаче сообщений без проводов Г. Герц почему-то прислал именно А. С. Попову, а не Г. Маркони.

Когда в 2009 году на VIII Международной конференции ФАМ<sup>2</sup> 2009 я спросил у итальянского математика Э. Орсингера (члена оргкомитета конференции), кто же все-таки *изобрел* радио, он уверенно ответил, — *Маркони*. □

**14.** *Дилемма западничества и славянофильства.* Развернувшаяся дискуссия по поводу использования различных наукометрических показателей, в том числе индекса Хирша, импакт-факторов, рейтингов и т. д., или, более широко, — по поводу принятия западных стандартов образования и науки, имеет на самом деле более глубокий, чем касающийся только науки, смысл. В глобальном аспекте она в чем-то напоминает искушающую дилемму западничества и славянофильства. Давно не секрет (Горохов [12]), что «наши отечественные работы, опубликованные на русском языке, цитируются крайне мало, а подтверждающие их вторичные, скажем, американские, статьи, вышедшие на год-полтора позже, — крайне широко. Иногда американские авторы на нас ссылаются, но делают это крайне невнятно».

**Теорема 14.** (Кожин [32, с. 77]). *Хорошо известны две характернейшие закономерности, присущие освоению русских литературных, научных и вообще культурных ценностей: во-первых, очевидное и нередко разительное «запаздывание» в их понимании и оценке, и, во-вторых, более или менее решительное отделение, отрыв этих ценностей от цельного исторического бытия России.*

Доказательство. (1) Нам в России часто не хватает средств и желания заниматься пропагандой собственных работ (Горохов [12]).

(2) Запад уже давно сделал занятия наукой объектом своих торговых операций. «Джентльмены любят бизнес» (Ильин [33, с. 137]). Соответствующее поло-

жение в науке по отношению к русским, «чья всецело мистическая внутренняя жизнь воспринимает мышление деньгами, как *grex*» (Шпенглер [34]), навязывается нам Западной Европой, куда мы так упорно стремимся, и сегодня.

(3) Чаадаев многократно говорил (Кожин [29]), что русскому взгляду (в том числе взгляду на Запад) присуще уникальное «беспристрастие». И среди его (Чаадаева. — Ю. Г.) формулировок «русской идеи» весомое место занимают утверждения 1835 года о том, «что Провидение создало нас слишком великими, чтобы быть эгоистами; что оно поставило нас вне интересов национальностей и поручило нам интересы человечества; что все наши мысли в жизни, науке, искусстве должны отпираться от этого и к этому приходить; что в этом наше будущее» □.

Я где-то читал, что согласно неофициальным данным до 90% принципиальных научных открытий сделано русскими, включая, естественно, выходцев из России. В свое время в библиотеке СО АН СССР (ГПНТБ, Новосибирск) я держал в руках замечательную книжку пионера космонавтики Ю. В. Кондратюка «Завоевание межпланетных пространств» [35]. Странно, почему-то она есть и в библиотеке Конгресса США. Уж не потому ли, что именно в этой уникальной книге автор разработал теорию промежуточных станций (ракетных баз) в виде спутников планет, используя которую США и осуществили программу «Аполлон» с высадкой человека на Луну.

На Западе идет охота за нашими мозгами. Их общество потребления нуждается в постоянном притоке свежей интеллектуальной силы, и мы в этом им максимально способствуем. Надо постоянно помнить, что с точки зрения своих гипертрофированных запросов Запад идет и будет готов пойти на все, что угодно.

## 4 Основные предложения

Согласно теореме 3, мнение по какому бы то ни было поводу, исходящее от ординарного профессора, ничтожно. Если на минутку допустить противное, то вот что можно было бы предложить от его имени для улучшения состояния вузовской науки:

1. Считать необязательными, кроме отрицательных, сторонние отзывы на автореферат диссертации как потенциально подготовленные заинтересованной стороной; предоставлять диссертанту и его научному руководителю обязательную командировку (при географической необходимости) в ведущую организацию для реального, а не формального рассмотрения диссертации; не считать выступлениями на конференции те доклады, тезисы которых опубликованы при отсутствии реального выступления; упразднить оскорбительную систему видеонаблюдения

заседаний ДС во время защит диссертаций как очень неэффективную по сравнению с существовавшей когда-то системой «шарашек». А если уж и быть в этом вопросе последовательным до конца, то ввести систему видеонаблюдения и для заседаний коллегии ВАК с тем, чтобы при желании диссертант мог увидеть воочию обсуждение своей диссертации. Что это за тайны мадридского двора?

2. Ввести в номенклатуру специальностей высшей школы новые специальности, востребованные жизнью, скажем, такие как математическая социология (Орлов [13]), актуарная и финансовая математика, и т. д. и, наоборот, исключать (возможно, поглощая их другими) те из них, на которые у научного сообщества нет устойчивого спроса. Статистика защит в ДС показывает, что есть специальности, по которым защит проходит очень мало: если число защит по трем специальностям, представленным в одном Совете (не скажу, в каком) за отчетный период, устойчиво имеет, скажем, распределение типа 15.4%, 23.1%, 61.5%, то это должно наводить на определенные размышления.

3. Ввести в структуру управления вузом сектор социологического мониторинга, задача которого — отслеживать с использованием современных наукометрических методов, см., например, (Аллахвердян, Агамова [36]; Мирская [37]; Орлов [38]) параметры функционирования вуза (рейтинги, вычисление индекса Джини, ведение наукометрических паспортов ученых, отслеживание демографических показателей кадрового состава вузов и ДС, анализ миграции научных кадров и выпускников вуза и т. д.); публиковать для всеобщего ознакомления соответствующий вузовский информационный бюллетень, являющийся официальным документом.

4. Руководству кафедр и факультетов вменить в должностную обязанность функции научного менеджмента: поиск заказов на выполнение научно-исследовательских работ вплоть до заключения с работодателями конкретных договоров на подготовку специалистов, включая в число исполнителей не только себя лично (честно скажем, существует порочная практика, когда при подаче заявки на изобретение или на патент в патентном отделе автору предлагают включить в число авторов изобретения ректора или, по крайней мере, декана, иначе, мол, заявка не выйдет из стен университета); доведение до коллективов кафедр информации о имеющихся грантах, оформление всякого рода документации в том числе подачу заявок на гранты и конкурсы, освободив от этой бюрократической рутины непосредственных исполнителей работ.

5. Обязать руководство вузов приглашать для чтения циклов лекций видных профессоров других вузов, в том числе зарубежных и, наоборот, содействовать собственным профессорам выполнять ту же

миссию (недавний пример: приятно было увидеть на фотографии 2012 года, присланной мне из США, стоящих рядом проф. Г. М. Кошкина (ТГУ) и проф. С. Р. Рао); восстановить чтение лекций по специальности для аспирантов там, где эта традиция утрачена.

## 5 Заключение

Есть ли шанс нашей вузовской системе восстановить утраченные позиции и если да, то каким способом? Надежда на спасение, как говорится, умирает последней. «Плохо человеку, который утратил *всякую* надежду и ни на что не надеется. Он как бы отрезает себе перспективу в свое будущее; он перестает верить в возможность лучшего; он ослепляет свой дух и обессиливает свой инстинкт; он отрекается от своего спасения и добровольно *опускается на дно*. Нельзя жить без надежды» (Ильин [33, с. 135]).

Шанс на спасение всегда есть. Вот только способов найти его особо не видно. Среди возможных вариантов существуют или можно предложить следующие:

- ☐ политическая воля руководителей государства;
- ☐ ограничение беспредела в распределении финансовых средств, выделяемых на науку и развитие образования;
- ☐ демографический взрыв, в результате которого страна получит новую плеяду выдающихся ученых уровня Н. А. Дмитриева, П. Л. Капицы, В. П. Козлова, М., Ю. В. Кондратьева, М. В. Ломоносова, Г. Я. Перельмана;
- ☐ вмешательство инопланетян [39].

## Список литературы

- [1] Орлов А. И., Чеботарев П. Ю. Интернет-конференция по проблемам теории и практики управления. In *В кн.: Управление большими системами / Режим доступа: <http://ubs.mtas.ru/about/>*, М., 2013. Ин-т проблем управления им. В. А. Трапезникова РАН.
- [2] Протасов В. И. Сайт коллективного творчества. In *Проект СКОЛТ / Режим доступа: <http://ubs.mtas.ru/about/>*. Ин-т физико-технической информатики, 2013.
- [3] Григорьев Ю. Д. Некоторые закономерности перехода к западной системе управления вузовской наукой. In *Интернет-конференция по проблемам теории и практики управления / Режим доступа: <http://ubs.mtas.ru/about/>*. Ин-т проблем управления им. В. А. Трапезникова РАН, 2013.
- [4] Арнольд В. И. *Нужна ли в школе математика? Стенограмма пленарного доклада (Дубна, 21 сентября 2000 г.)*. МЦНМО, М., 2004.
- [5] Арнольд В. И. *Что такое математика?* МЦНМО, М., 2004.
- [6] Арнольд В. И. *«Жесткие» и «мягкие» математические модели*. МЦНМО, М., 2004.
- [7] Зиновьев А. А. *Глобальный человек. «История XXI века»*. Изд-во Эксмо, М., 2003.
- [8] Зиновьев А. А. *Коммунизм как реальность. Парабеллум*. АСЕ: Астрель, М., 2012.
- [9] Ракитов А. И. *Научно-исследовательские исследования, 2011: Сб. науч. тр. / Ред. А. И. Ракитов*. РАН. ИНИОН. Центр науч. информ. исслед. по науке, образованию и технологиям., М., 2011.
- [10] Налимов В. В. *Облик науки*. Центр гуманитарных инициатив, Изд во МБА., М., 2010.
- [11] Ракитов А. И. Пятая попытка модернизации России: Существует ли стратегия в области образования и науки? In *В кн.: Научно-исследовательские исследования, 2011: Сб. науч. тр. / Отв. ред. А. И. Ракитов*, page 296. РАН. ИНИОН. Центр науч. информ. исслед. по науке, образованию и технологиям., М., 2011.
- [12] Горохов В. Г. Проблема измерения продуктивности отдельных ученых и целых институтов. In *В кн.: Управление большими системами / Режим доступа: <http://ubs.mtas.ru/about/>*. Ин-т проблем управления им. В. А. Трапезникова РАН, М., 2013.
- [13] Орлов А. И. Два типа методологических ошибок при управлении научной деятельностью. In *В кн.: Управление большими системами / Режим доступа: <http://ubs.mtas.ru/about/>*, М., 2013. Ин-т проблем управления им. В. А. Трапезникова РАН.
- [14] Интернет-ресурс. Десятки сфальсифицированных кандидатских и докторских диссертаций выявлены в МПГУ. In *Режим доступа: <http://www.nvpress.ru/incident/2474/>*, М., 2013. [Электронный ресурс].
- [15] Кундышева Е. С. *Математическое моделирование в экономике. Учебное пособие*. Изд. торгов. корпорация «Дашков и К°», М., 2004.
- [16] Сикан А. В. *Методы статистической обработки гидрометеорологической информации. Учебник*. Изд-во РГГМУ, СПб., 2007.
- [17] Степанов В. В. *Курс дифференциальных уравнений*. Гостехиздат, М., 1953.
- [18] Арсенов О. Григорий Перельман и гипотеза Пуанкаре. «Люди науки». ЭКСМО, М., 2010.
- [19] Григорьев Ю. Д., Мартыненко Г. Я. *Типология последовательностей Фибоначчи: теория и приложения. (Введение в математику гармонии)*. LAP LAMBERT Academic Publishing GmbH & Co. KG, Saarbrücken, Deutschland., 2012.
- [20] Хруцкий К. С. Трехмерный биокосмологический процесс к вопросам развития наукометрии в России. In *В кн.: Управление большими системами / Режим доступа: <http://ubs.mtas.ru/about/>*, М., 2013. Ин-т проблем управления им. В. А. Трапезникова РАН.



- [21] Данилевский Н. Я. *Россия и Европа*. «Историко-литературный архив». Книга, М., 1991.
- [22] Жмудь В. А. Перспективы сближения программ подготовки магистров по направлениям подготовки в области автоматизи. *Сб. науч. тр. НГТУ*, Вып. 2(68):133–140, 2012.
- [23] Ловаш Л. *Игра в цифирь, или как теперь оценивают труд ученого. (Сб. статей о библиометрике) / Отв. ред. Л. Ловаш*. МЦНМО, М., 2011.
- [24] Маркусова В. А., Либкинд А. Н., Крылова Т. А. Научная деятельность российских вузов в регионах и их позиции в мировых рейтингах: Библиометрический анализ по статистике информационной системы «web of knowledge». In *В кн.: Научно-исследовательские исследования, 2011: Сб. науч. тр.*, page 296. РАН. ИНИОН. Центр науч.информ. исслед. по науке, образованию и технологиям., М., 2011.
- [25] Чеботарев П. Ю. Наукометрия: как с ее помощью лечить, а не калечить? In *В кн.: Управление большими системами / Режим доступа: <http://ubs.mtas.ru/about/>*, М., 2013. Ин-т проблем управления управления им. В. А. Трапезникова РАН.
- [26] Dette H. and Grigoriev Yu. D. A unified approach to second order optimality criteria in nonlinear design of experiments. *Ann. Inst. Statist. Math.*, 52 (3):574–597, 2000.
- [27] Grigor'ev Yu. D. and Le Dinh Shon. On ruin probability minimization under excess reinsurance. *Automation and Remote Control*, 68 (6):1039–1054, 2007.
- [28] Козлов В. П. *Избранные труды по теории планирования эксперимента и обратным задачам оптического зондирования*. Изд-во С.-Петербур. ун-та, СПб., 2000.
- [29] Кожин В. В. *Пушкин и Чаадаев. К истории русского самосознания*. — В кн.: *Победы и беды России*, page 576. «Политический бестселлер». Алгоритм, Эксмо, М., 2005.
- [30] Ильин И. А. *Наши задачи. Историческая судьба и будущее России. Статьи 1948- 1954 годов. В 2-х томах. Т. 1*. МП «Рарог», М., 1992.
- [31] Чолаков В. *Нобелевские премии. Ученые и открытия*. Мир, М., 1986.
- [32] Кожин В. В. . *Пути русского самосознания*. — С. 69 – 101.— В кн.: *Россия. Век XX*, page 1040. «Русский век». Эксмо: Алгоритм, М., 2011.
- [33] Ильин И. А. *Наши задачи. Историческая судьба и будущее России. Статьи 1948 - 1954 годов. В 2-х томах. Т. 2*. МП «Рарог», М., 1992.
- [34] Шпенглер О. *Закат западного мира. Очерки морфологии мировой истории*. Изд-во АЛЬФА-КНИГА, М., 2010.
- [35] Кондратюк Ю. В. *Завоевание межпланетных пространств*. Типография Сибкрайсоюза, Новосибирск, 1929.
- [36] Аллахвердян А. Г., Агамова Н. С. Эволюция истории науки и науковедения: Тенденция взаимодействия двух социогуманитарных дисциплин. In *В кн.: Научно-исследовательские исследования, 2011: Сб. науч. тр.*, page 296. РАН. ИНИОН. Центр науч. информ. исслед. по науке, образованию и технологиям., М., 2011.
- [37] Мирская Е. З. Новые информационные технологии в российской науке: История, результаты, проблемы и перспективы. In *В кн.: Научно-исследовательские исследования, 2011: Сб. науч. тр.*, page 296. РАН. ИНИОН. Центр науч.информ. исслед. по науке, образованию и технологиям., М., 2011.
- [38] Орлов А. И. *Организационно-экономическое моделирование. Учебник. Часть 1: нечисловая статистика*. Изд-во МГТУ им. Н. Э. Баумана, М., 2009.
- [39] Проханов А. А. *Человек звезды: роман*. Вече, М., 2012.



# Определители Ганкеля и полиномиальная регрессия

Юрий Дмитриевич Григорьев

Санкт-Петербургский государственный  
электротехнический университет (ЛЭТИ)  
Санкт-Петербург  
yuri\_grigoriev@mail.ru

**Аннотация.** На примере полиномиальной регрессии показано применение определителей Ганкеля в задаче рекуррентного вычисления корреляционного отношения и построения оптимальных планов эксперимента.

**Ключевые слова.** определители Ганкеля, корреляционное отношение, полиномиальная регрессия, канонические моменты.

## 1 Введение

К настоящему моменту в теории оптимального планирования эксперимента разработаны методы построения оптимальных планов эксперимента на основе многих новых подходов, включая такие, как функциональный подход Меласа [1] и теория канонических моментов Детте—Стаддена [2]. Первый из них рассматривался в работе [3], а второй — в работах [4] — [5].

Настоящий доклад посвящен краткому изложению второго подхода. На примере полиномиальной регрессии показано, что важную роль при исследовании данной модели играют определители Ганкеля, которые возникают и в другой задаче — рекуррентном вычислении корреляционного отношения.

## 2 Ганкелевы матрицы

Ганкелевой матрицей порядка  $r + 1$  называется матрица вида  $H_r = (h_{i+j})_{i,j=0}^r$ . Из структуры матрицы  $H_r$  видно, что ганкелева матрица всегда симметрична. Более того, ее побочная диагональ, а также все параллельные ей диагонали состоят из равных (своих для каждой диагонали) элементов.

Продолжением ганкелевой матрицы  $H_r$  называется матрица  $H_{r+\nu} = (h_{i+j})_{i,j=0}^{r+\nu}$  ( $\nu = 1, 2, \dots$ ), у которой левый верхний блок совпадает с данной матрицей  $H_r$ . Различные свойства ганкелевых и близких к ним теплицевых матриц рассматриваются в [Иохвидов [6]].

## 3 Полиномиальная регрессия

Пусть  $(X, Y) \in L^2$ , где  $L^2 = L^2(\Omega, \mathcal{B}, \mathbb{P})$  гильбертово пространство случайных величин (с. в.) со скалярным произведением  $\langle X, Y \rangle = \mathbb{M}XY$ . Рассмотрим полиномиальную регрессию  $Y$  на  $X$ , т. е. функцию

$$m_Y(x) := \mathbb{M}(Y|X=x) = \beta_0 + \sum_{i=1}^r \beta_i x^i. \quad (1)$$

Обозначим

$$h_{ij} = \mathbb{M}X^{i+j}, \quad b_j = \mathbb{M}X^j Y, \quad i, j = 0, \dots, r,$$

Отсюда видим, что матрица моментов  $H_r = (h_{ij})$  является ганкелевой. Чтобы оценить коэффициенты регрессии  $\beta_i$ , используем метод наименьших квадратов, согласно которому оценка  $\hat{\beta}$  коэффициентов регрессии  $\beta$  находится из условия

$$\hat{\beta} = \arg \min_{\beta} \mathbb{M} \left( Y - \sum_{i=0}^r \beta_i X^i \right)^2.$$

В результате находим

$$\hat{\beta} = H_r^{-1} b. \quad (2)$$

Отсюда, обозначив  $\sigma_X^2 := \mathbb{D}X$ ,  $\sigma_Y^2 := \mathbb{D}Y$ , при  $r = 1$  получаем оптимальную линейную регрессию

$$m_Y(x) = m_Y + \frac{\sigma_Y}{\sigma_X} \rho(x - m_X).$$

## 4 Корреляционное отношение

1. Понятие корреляционного отношения связано с формулой разложения дисперсии с. в.  $Y$  на две составляющие

$$\mathbb{D}Y = \mathbb{M}[\mathbb{D}(Y|X)] + \mathbb{D}[\mathbb{M}(Y|X)], \quad (3)$$

— дисперсию  $Y$  относительно регрессии  $m_Y(X)$ , измеряемой величиной  $\mathbb{M}[\mathbb{D}(Y|X)]$ , и дисперсию регрессии  $m_Y(X)$  относительно математического ожидания  $m_Y$ , измеряемой величиной  $\mathbb{D}[\mathbb{M}(Y|X)]$ . Первое слагаемое называют *средним групповых дисперсий*, а второе — *межгрупповой дисперсией* (дисперсией групповых средних около общей средней).

В силу разложения (3) возникает понятие *корреляционного отношения*, которым называется с. в.

$$\lambda_{Y|X}^2 = \sigma_Y^{-2} \mathbb{D}\mathbb{M}(Y|X).$$

Коэффициент корреляции  $\rho = \text{cov}(X, Y) / (\sigma_X \sigma_Y)$  является характеристикой стохастической зависимости с. в.  $X$  и  $Y$  в рамках *линейной регрессии*  $m_Y(x)$ . Если функция регрессии  $m_Y(x)$  является нелинейной по  $x$ , то корреляционное отношение  $\lambda_{Y|X}^2$  является более объективной и более качественной характеристикой стохастической зависимости случайных величин. Из определения  $\lambda_{Y|X}^2$  следует, что

$$0 \leq \lambda_{Y|X}^2 \leq 1.$$

**2.** Опишем рекуррентный способ вычисления корреляционного отношения  $\lambda_{Y|X}^2$  в случае полиномиальной регрессии  $Y$  на  $X$ , который будет существенно связан с ганкелевыми матрицами  $H_r$ .

Пусть  $(\xi, \eta) \in L^2$  — стандартизованная с. в.  $(X, Y)$  и

$$\mathbb{M}(\eta|\xi) = \sum_{i=0}^r \beta_i \xi^i. \quad (4)$$

— условная регрессия  $\eta$  на  $\xi$ . В этом случае разложение (3) можно записать в виде равенства

$$\mathbb{M}[\mathbb{D}(\eta|\xi)] = 1 - \lambda_{\eta|\xi}^2,$$

левая часть которого по определению есть

$$\mathbb{M}[\mathbb{D}(\eta|\xi)] = \min_{\beta} \mathbb{M}\left(\eta - \sum_{i=0}^r \beta_i \xi^i\right)^2, \quad (5)$$

что совпадает с (3). Поэтому, подставив (2) в (5), находим

$$\lambda_{\eta|\xi}^2 = b^T H_r^{-1} b, \quad b = (b_i)_{i=0}^r, \quad b_i = \mathbb{M}\xi^i \eta. \quad (6)$$

Наконец, в силу известного равенства (Федоров [7, с. 21])

$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = |A| \cdot |D - CA^{-1}B|,$$

в котором полагаем  $A = H_r$ ,  $B = b$ ,  $C = b^T$ ,  $D = 1$ , из (6) следует

$$\lambda_{\eta|\xi}^2 = 1 - \frac{\begin{vmatrix} H_r & b \\ b^T & 1 \end{vmatrix}}{|H_r|} = 1 - \frac{|B_{r+1}|}{|H_r|}. \quad (7)$$

Выражение (7) позволяет выписать рекуррентное соотношение для  $\lambda_r^2 := \lambda_{\eta|\xi}^2$  в зависимости от степени  $r$  полиномиальной регрессии.

**3.** Положим

$$m_{i+j} = \mathbb{M}\xi^{i+j}, \quad \alpha = m_{2r+2}, \quad \gamma := b_{r+1} = \mathbb{M}\xi^{r+1}\eta,$$

и пусть

$$H_{r+1} = \begin{pmatrix} H_r & a \\ a^T & \alpha \end{pmatrix}, \quad B_{r+1} = \begin{pmatrix} H_r & b \\ b^T & 1 \end{pmatrix},$$

где  $H_{r+1}$  — продолжение матрицы  $H_r$ , в то время как матрица  $B_{r+1}$  таковой не является.

Согласно (7)

$$\lambda_r^2 := 1 - \frac{|B_{r+1}|}{|H_r|}, \quad \lambda_{r+1}^2 := 1 - \frac{|B_{r+2}|}{|H_{r+1}|}.$$

В силу (4) получаем

$$|B_{r+2}| = |H_{r+1}| \cdot (1 - c^T H_{r+1}^{-1} c),$$

где  $c = (b, \gamma) \in R^{r+1}$ . Отсюда, используя формулу Фробениуса для обращения блочных матриц, получаем

$$c^T H_{r+1}^{-1} c = b^T H_r^{-1} b + \chi^{-1}(\gamma - u)^2,$$

где

$$u = a^T H_r^{-1} b, \quad \chi = |H_{r+1}| = \alpha - a^T H_r^{-1} a.$$

Следовательно, согласно (4)

$$\frac{|B_{r+2}|}{|H_{r+1}|} = \frac{|B_{r+1}|}{|H_r|} - \chi^{-1}(\gamma - u)^2.$$

Отсюда вытекает окончательное выражение для корреляционного отношения

$$\lambda_{r+1}^2 = \lambda_r^2 + \chi^{-1}(\gamma - u)^2. \quad (8)$$

Таким образом, доказана следующая

**Теорема 15.** (Григорьев [8]). Пусть для  $r \geq 1$  выполнены следующие условия:

- (1)  $\xi, \eta \in L^2$  — стандартизованные с. в. ;
- (2) существуют моменты  $a_i = \mathbb{M}\xi^i$ ,  $i = 1, \dots, 2r + 2$ ,  $b_j = \mathbb{M}\xi^j \eta$ ,  $j = 0, 1, \dots, r + 1$ ,  $h_{ij} = \mathbb{M}\xi^{i+j}$ ,  $\alpha = \mathbb{M}\xi^{2r+2}$  ;
- (3)  $\chi = \alpha - a^T H_r^{-1} a \neq 0$ ;
- (4) матрица  $H_r = (h_{ij})$  порядка  $r + 1$  невырождена;
- (5)  $m_\eta(\xi) = \sum_{j=0}^r \beta_j \xi^j$  — полиномиальная степени  $r$  регрессия  $\eta$  на  $\xi$ .

Тогда имеют место следующие утверждения:

- (1)  $\hat{\beta} = H_r^{-1} b$ ;
- (2)  $\lambda_r^2 = 1 - b^T H_r^{-1} b$ ;
- (3)  $\lambda_{r+1}^2 = \lambda_r^2 + \chi^{-1}(b_{r+1} - a^T H_r^{-1} b)^2$ .

## 5 Канонические моменты

Для вероятностной меры  $\mu$  на интервале  $[a, b]$  ее начальные моменты определяются как

$$c_k = c_k(\mu) = \int_a^b x^k d\mu(x), \quad k \geq 1.$$

Канонические моменты можно определить следующим геометрическим способом: для заданной вероятностной меры  $\mu$  с моментами  $c_k(\mu)$  положим

$$c_k^- = c_k^-(\mu) = \min_{\eta} c_k^-(\eta), \quad c_k^+ = c_k^+(\mu) = \max_{\eta} c_k^+(\eta),$$

где минимум и максимум берутся по тем вероятностным мерам  $\eta$ , для которых моменты до порядка  $k$  включительно совпадают, т. е.  $c_i(\mu) = c_i(\eta)$ , для  $i \leq k-1$ . Тогда канонические моменты определяются как

$$p_k = \frac{c_k - c_k^-}{c_k^+ - c_k^-}, \quad k \geq 1. \quad (9)$$

Здесь предполагается, что  $c_k^- < c_k^+$ . Таким образом, последовательность  $p_k$  соответствующим образом ограничена.

Канонический момент  $p_k$  зависит от  $c_1, \dots, c_k$ . Если  $M$  — выпуклое множество всех последовательностей начальных моментов, а  $M_k$  — проекция на первые  $k$  координат, то это дает взаимно однозначное соответствие между внутренностью  $M_k$  и кубом  $(0, 1)^k$  для всех  $k \in \mathbb{N}$ . Грубо говоря, канонические моменты  $c_1, \dots, c_{k-1}$  последовательно задают «относительное положение»  $c_k$  в  $M_k$  или  $M$ .

Можно показать, что матрица моментов  $(c_{i+j})_{i,j=0}^m$  неотрицательно определена. Также неотрицательно определена и матрица  $(c_{i+j+1} - c_{i+j+2})_{i,j=0}^{m-1}$ . Таким образом, если  $n = 2m$ , то  $\{c_n\} \in M_n$  тогда и только тогда, когда матрицы

$$(c_{i+j})_{i,j=0}^m, \quad \text{и} \quad (c_{i+j+1} - c_{i+j+2})_{i,j=0}^{m-1}$$

неотрицательно определены. Аналогично, если  $n = 2m+1$ , то  $\{c_n\} \in M_n$  тогда и только тогда, когда матрицы

$$(c_{i+j+1})_{i,j=0}^m, \quad \text{и} \quad (c_{i+j} - c_{i+j+1})_{i,j=0}^m$$

неотрицательно определены.

Определим *определители Ганкеля*

$$\begin{aligned} \underline{H}_{2m} &= \begin{vmatrix} c_0 & \cdots & c_m \\ \vdots & & \vdots \\ c_m & \cdots & c_{2m} \end{vmatrix}, \\ \overline{H}_{2m+1} &= \begin{vmatrix} c_0 - c_1 & \cdots & c_m - c_{m+1} \\ \vdots & & \vdots \\ c_m - c_{m+1} & \cdots & c_{2m} - c_{2m+1} \end{vmatrix}, \\ \underline{H}_{2m+1} &= \begin{vmatrix} c_1 & \cdots & c_{m+1} \\ \vdots & & \vdots \\ c_{m+1} & \cdots & c_{2m+1} \end{vmatrix}, \\ \overline{H}_{2m} &= \begin{vmatrix} c_1 - c_2 & \cdots & c_m - c_{m+1} \\ \vdots & & \vdots \\ c_m - c_{m+1} & \cdots & c_{2m-1} - c_{2m} \end{vmatrix}. \end{aligned}$$

Определение канонических моментов согласно (9) не дает способа их фактического вычисления. Однако это удастся сделать с помощью определителей Ганкеля.

**Теорема 16.** (Детте, Стадден [2, с. 21]). Если  $\{c_{n-1}\} \in \text{Int } M_{n-1}$ , то

$$c_n - c_n^- = \frac{\underline{H}_n}{\underline{H}_{n-2}}, \quad c_n^+ - c_n = \frac{\overline{H}_n}{\overline{H}_{n-2}},$$

и

$$p_n = \frac{\underline{H}_n \overline{H}_{n-2}}{\underline{H}_n \overline{H}_{n-2} + \overline{H}_n \underline{H}_{n-2}},$$

где  $\underline{H}_{-1} = \underline{H}_0 = \overline{H}_{-1} = \overline{H}_0 = 1$ .

Между определителями Ганкеля имеет место следующее *основное соотношение*

$$\underline{H}_n \overline{H}_n = \underline{H}_{n-1} \overline{H}_{n+1} + \overline{H}_{n-1} \underline{H}_{n+1}, \quad n \geq 1, \quad (10)$$

которое приводит к равенствам

$$p_n = \frac{\underline{H}_n \overline{H}_{n-2}}{\underline{H}_{n-1} \overline{H}_{n-1}}, \quad q_n = 1 - p_n = \frac{\underline{H}_{n-2} \overline{H}_n}{\underline{H}_{n-1} \overline{H}_{n-1}}, \quad n \geq 1.$$

С помощью определителей Ганкеля определяется еще одно множество величин, играющих важную роль при построении планов эксперимента, а именно:

$$\zeta_0 = 1, \quad \zeta_1 = p_1, \quad \zeta_k = q_{k-1} p_k, \quad k \geq 2. \quad (11)$$

Из основного соотношения (10) вытекает, что для всех  $n \geq 1$

$$\zeta_n = \frac{\underline{H}_n \underline{H}_{n-3}}{\underline{H}_{n-1} \underline{H}_{n-2}},$$

где  $\underline{H}_{-2} = \underline{H}_{-1} = \underline{H}_0 = \overline{H}_{-2} = \overline{H}_{-1} = \overline{H}_0 = 1$ . Наконец, нам понадобится следующее соотношение, связывающее определители Ганкеля и  $\zeta$ -величины (Детте, Стадден [2, с. 59]):

$$\frac{\underline{H}_{2m}}{\underline{H}_{2m-2}} = 2^{2m+1} \prod_{j=1}^m \zeta_{2j-1} \zeta_{2j}. \quad (12)$$

Для некоторых специальных последовательностей канонических моментов соответствующие им вероятностные меры могут быть выписаны в явном виде (Скибински [9]). Приведем только один пример, необходимый нам в иллюстративных целях. Именно, если вероятностной мере  $\varepsilon$  на интервале  $[-1, 1]$  соответствует последовательность канонических моментов  $\{1/2, p_2, 1/2, 1\}$ , то (Детте, Стадден [2, с. 106])

$$\varepsilon = \begin{pmatrix} -1 & 0 & 1 \\ p_2/2 & 1 - p_2 & p_2/2 \end{pmatrix}. \quad (13)$$

## 6 Пример

Рассмотрим пример построения оптимального плана эксперимента с помощью канонических моментов (Детте, Стадден [2, с. 141]).

Пусть функция регрессии имеет вид  $\eta(x; \theta) = f^T(x)\theta = \theta_0 + \theta_1 x + \theta_2 x^2$ ,  $x \in [-1, 1]$ . Предположим, что нас интересует оценивание старшего коэффициента  $\theta_2$ , или параметрической функции  $e^T \theta$ , где  $e = (0, 0, 1)^T$ . Параметр  $\theta_2$  оцениваем только и только тогда, когда спектр плана  $\varepsilon$  состоит из трех точек. Информационная матрица  $M(\varepsilon)$  является матрицей Ганкеля и имеет вид

$$M(\varepsilon) = \int_{-1}^1 f(x) f^T(x) d\varepsilon(x) = \begin{pmatrix} 1 & c_1 & c_2 \\ c_1 & c_2 & c_3 \\ c_2 & c_3 & c_4 \end{pmatrix}.$$

Поэтому дисперсия оценки Гаусса—Маркова для параметра  $\theta_2 = e^T \theta$  пропорциональна величине

$$e^T M^{-1}(\varepsilon) e = \frac{c_2 - c_1^2}{|M(\varepsilon)|}.$$

Оптимальный план  $\varepsilon^*$  должен минимизировать величину  $e^T M^{-1}(\varepsilon) e$  или максимизировать величину, обратную ей, т. е.

$$\varphi(\varepsilon) = \frac{|M(\varepsilon)|}{c_2 - c_1^2} = \frac{\underline{H}_4}{\underline{H}_2}. \quad (14)$$

Согласно (12) отношение в (14) может быть выражено в терминах канонических моментов плана  $\varepsilon$ , т. е.

$$\varphi(\varepsilon) = 2^4 \prod_{j=1}^2 \zeta_{2j-1} \zeta_{2j} = 2^4 \left( \prod_{j=1}^3 q_j p_j \right) p_4.$$

Максимизация этого выражения в терминах канонических моментов приводит к равенствам  $p_j = \frac{1}{2}$  ( $j = 1, 2, 3$ ),  $p_4 = 1$ . План  $\varepsilon^*$ , соответствующий этой последовательности моментов  $\{p_j\}$ , является оптимальным для оценивания  $\theta_2$  и может быть найден согласно (13):

$$\varepsilon^* = \begin{pmatrix} -1 & 0 & 1 \\ 1/4 & 1/2 & 1/4 \end{pmatrix} \quad (15)$$

Это означает, что экспериментатор должен провести по 1/4 наблюдений в точках  $x_1 = -1$  и  $x_3 = 1$  и 1/2 наблюдений в точке  $x_2 = 0$ . Отметим, что полученный план (15) является одновременно и  $A$ -оптимальным (Федоров [7, с. 153]).

## Список литературы

- [1] Melas V. B. *Functional approach to optimal experimental design*. Springer Science + Business Media, Inc., 2006.

- [2] H. Dette and W. J. Studden. *The Theory of Canonical Moments with Applications in Statistics, Probability, and Analysis*. John Wiley&Sons, Inc., 1997.
- [3] Григорьев Ю. Д., Нгуен Хай Винь. Метод неявной функции в задаче построения двухточечных планов эксперимента. In *Труды IX Международной конференции ФАМЭТ 2010 по финансово-актуарной математике и эвентоконвергенции технологий. Часть 1, 108 – 113*, pages 108 – 113. Красноярск: СФУ, 2010.
- [4] Григорьев Ю. Д., Тамонцев С. Ю. Исследование байесовских D-оптимальных планов для дробно-рациональных моделей. *Сб. научных трудов НГТУ*, Вып. 3 (12):3–10, 1998.
- [5] Григорьев Ю. Д., Нгуен Хай Винь. Преобразование Стилтеса в задачах нелинейного планирования эксперимента. *Сб. научных трудов НГТУ*, Вып. 2 (56):61–70, 2009.
- [6] Иохвидов И. С. *Ганкелевы и теплицевы матрицы и формы*. Наука, М., 1974.
- [7] Федоров В. В. *Теория оптимального эксперимента*. Наука, М., 1971.
- [8] Григорьев Ю. Д. *Теория вероятностей и математическая статистика*. НГТУ, Новосибирск, 1996.
- [9] M. Skibinsky. Extreme  $n$ -th moments for distributions on  $[0, 1]$  and the inverse of a moment space map. *J. Appl. Probab.*, 5:693–701, 1968.
- [10] Григорьев Ю. Д., Щеколдин В. Ю. Канонические моменты вероятностных мер. *Сб. научных трудов НГТУ*, Вып. 3 (27):55–60, 2000.
- [11] Щеколдин В. Ю. *Методы теории канонических моментов в задачах анализа и планирования регрессионных экспериментов*. PhD thesis, Новосибирск, НГТУ, 2002.

## Численная оптимизация со случайными данными

**Борис Станиславович Добронет**

Институт космических и информационных технологий  
Сибирский федеральный университет  
Красноярск  
BDobronets@yandex.ru

**Ольга Аркадьевна Попова**

Институт космических и информационных технологий  
Сибирский федеральный университет  
Красноярск  
OlgaArc@yandex.ru

**Аннотация.** В статье рассматривается новый подход к решению оптимизационных задач со случайными входными параметрами, который определяется как случайное программирование. Данный подход использует численный вероятностный анализ и позволяет строить множество решений оптимизационной задачи на основе совместной функции плотности вероятности.

**Ключевые слова.** Численный вероятностный анализ, случайное программирование, математическое программирование.

### 1 Введение

В статье рассматривается численный вероятностный подход к решению задач оптимизации со случайными входными параметрами. В результате решения подобных задач методами математического программирования получаются оптимальные решения, зависящие от этих параметров. В тех случаях, когда известны плотности вероятности входных параметров, на основе численного вероятностного анализа возможно построение совместной функции плотности вероятности оптимального решения. В отличие от стохастического программирования [1, 2], где оптимальное решение представляет собой некоторое фиксированное решение, данный подход позволяет построить все множество решений оптимизационной задачи, определяемое построенной совместной функцией плотности вероятности.

Методы, позволяющие строить множество решений оптимизационной задачи со случайными входными параметрами на основе применения численного вероятностного анализа, назовем *случайным программированием*.

### 2 Постановка задачи

Сформулируем задачу случайного программирования в следующем виде:

$$f(x, \xi) \rightarrow \min, \quad (1)$$

$$g_i(x, \xi) \leq 0, \quad i = 1, \dots, m. \quad (2)$$

где  $x$  — вектор решения,  $f(x, \xi)$  — целевая функция,  $g_i(x, \xi)$  — функции ограничений.  $\xi$  — случайный вектор параметров. Относительно  $\xi$  известно, что

$$\xi \in \Xi, \quad (3)$$

где  $\Xi$  — случайный вектор.

Точка  $x^*$  — решение задачи (1) – (3), если

$$f(x^*, \xi) = \inf_U f(x, \xi),$$

где

$$U = \{x | g_i(x, \xi) \leq 0, \quad i = 1, \dots, m.\}$$

Множество решений (1)–(3) определим следующим образом

$$\mathcal{X} = \{x | f(x, \xi) \rightarrow \min, g_i(x, \xi) \leq 0, \quad i = 1, \dots, m, \xi \in \Xi\}$$

Заметим, что  $x^*$  — случайный вектор, поэтому в отличие от детерминированной задачи, для  $x^*$  необходимо определять функции плотности вероятности для каждой компоненты  $x_i^*$  как совместную плотность вероятности.

Мы расширим отношение порядка  $*$   $\in \{<, \leq, \geq, >\}$  на случайные переменные:

$$x * y \Leftrightarrow x * y \text{ для всех } x \in \mathbf{x}, y \in \mathbf{y}.$$

Если носители  $\mathbf{x}$ ,  $\mathbf{y}$  пересекаются, можно говорить о вероятности  $\mathbf{x} * \mathbf{y}$

$$P(\mathbf{x} * \mathbf{y}) = \int_{\Omega} p(x, y) dx dy,$$

где  $\Omega = \{(x, y) | x * y\}$ ,  $p(x, y)$  — совместная плотность вероятности  $\mathbf{x}$ ,  $\mathbf{y}$ .

Задача линейного программирования с случайными данными формулируется следующим образом:

$$(c, x) \rightarrow \min, \quad (4)$$

$$Ax = b, x \geq 0. \quad (5)$$

$$A \in \mathbf{A}, b \in \mathbf{b}, c \in \mathbf{c}, \quad (6)$$

где  $\mathbf{A}$  — стохастическая матрица,  $\mathbf{b}$ ,  $\mathbf{c}$  — стохастические векторы размерности  $n$ .

Точка  $x_*$  — решение задачи (5) – (7), если

$$(c, x^*) = \inf_U (c, x),$$



где

$$U = \{x | Ax = b, x \geq 0\}$$

Множество решений (5) – (7)

$$\mathcal{X} = \{x | (c, x) \rightarrow \min, Ax = b, x \geq 0, A \in \mathbf{A}, b \in \mathbf{b}, c \in \mathbf{c}\}$$

### 3 Случайное линейное программирование

Известно, для задачи (5)–(6) оптимальное решение  $x_*$  достигается в угловой точке множества  $U$ .

**Теорема 1.** [3] Пусть множество  $U$  определено условиями (6). Для того чтобы точка  $x = (x_1, \dots, x_n) \in U$  была угловой необходимо и достаточно, чтобы существовали номера  $j_1, \dots, j_r$ :

$$A_{j_1}x_{j_1} + \dots + A_{j_r}x_{j_r} = b; x_j = 0, j \neq j_l, l = 1, \dots, r,$$

причем столбцы  $A_{j_1}, \dots, A_{j_r}$  линейно независимы.

**Пример 1.** Пусть  $U$  определяется матрицей  $A$  и вектором  $b$

$$A = \begin{pmatrix} 1 & 1 & 3 & 1 \\ 1 & -1 & 1 & 2 \end{pmatrix}, b = \begin{pmatrix} 3 \\ 1 \end{pmatrix},$$

тогда столбцам матрицы  $A_1, A_2$  соответствует угловая точка с координатами  $(2, 1, 0, 0)$ ,  $A_1, A_3$  –  $(0, 0, 1, 0)$ ,  $A_2, A_4$  –  $(0, 5/7, 0, 4/3)$ .

Заметим, что из  $n$  столбцов можно выбрать  $r$  линейно независимых не более чем  $C_n^r$  способами. Следовательно, число угловых точек множества  $U$  конечно.

Это значит, что каноническую задачу (5)–(7) можно попытаться решить следующим образом:

- 1) найти все угловые точки  $x$  множества  $U$ ,
- 2) вычислить значение функции  $(c, x)$ , в каждой из угловых точек, и определить наименьшее из них.

Однако такой подход практически не применяется, так как даже в задачах не очень большой размерности число угловых точек может быть очень большим. Тем не менее идея перебора угловых точек множества оказалась весьма плодотворной и послужила основой ряда методов решения задач линейного программирования. Одним из таких методов является так называемый симплекс метод.

Для задачи (5)–(7) построим совместную плотность вероятности вектора  $x^*$ . Для этой цели воспользуемся одним из способов решения детерминированных задач линейного программирования, например, симплекс методом.

Рассмотрим вспомогательную задачу.

$$(c_t, x) \rightarrow \min, \quad (7)$$

$$A_t x = b_t, x \geq 0. \quad (8)$$

$$A_t \in \mathbf{A}, b_t \in \mathbf{b}, c_t \in \mathbf{c}, \quad (9)$$

найдем решение  $x_t^*$  и соответствующую ему угловую точку с номерами  $j_1, \dots, j_r$ .

Решим стохастическую систему линейных алгебраических уравнений численным вероятностным анализом [4]

$$(A_{j_1} \dots A_{j_r})x = b,$$

Совместная плотность вероятности найденного решения будет соответствовать  $x_t^*$ . Если носители входных параметров достаточно малы, то в силу непрерывности  $x_t^*$  будет совпадать с  $x^*$ . В случае произвольных носителей входных параметров процедуру выбора  $A_t \in \mathbf{A}, b_t \in \mathbf{b}, c_t \in \mathbf{c}$ , следует повторить, используя подходы метода Монте Карло или генетических алгоритмов. Если при этом будут получены разные решения  $x_t^*$ , то их можно сравнить вычисляя вероятностные расширения  $f_t = (c, x_t^*)$  [5].

#### 3.1 Численный пример

В качестве численного примера рассмотрим следующую задачу

$$(c, x) \rightarrow \min, \quad (10)$$

$$Ax = b, x \geq 0. \quad (11)$$

$$A \in \mathbf{A}, b \in \mathbf{b}, c \in \mathbf{c}, \quad (12)$$

$A = (a_{ij})$  – равномерная случайная матрица, каждый элемент – равномерная случайная величина с носителем  $[a_{ij}, \bar{a}_{ij}]$ , аналогично  $b, c$  – случайные вектора с элементами в виде равномерных случайных величин.

Носители заданы следующим образом

$$A = \begin{pmatrix} [1-r, 1+r] & [1-r, 1+r] \\ [1-r, 1+r] & [-1-r, -1+r] \end{pmatrix}$$

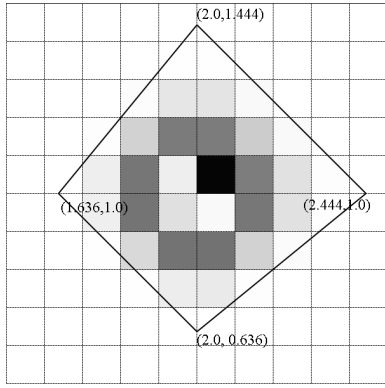
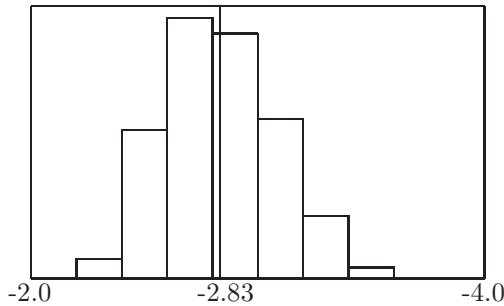
$$\begin{pmatrix} [3-r, 3+r] & [1-r, 1+r] \\ [1-r, 1+r] & [2-r, 2+r] \end{pmatrix},$$

$$b = \begin{pmatrix} [3-r, 3+r] \\ [1-r, 1+r] \end{pmatrix},$$

$$c = (-1, -1, 0, 0).$$

При  $r = 0$ , что соответствует детерминированному случаю, решение  $x^* = (2, 1, 0, 0)$ , столбцы матрицы  $A_1, A_2$  соответствуют угловой точке.

На рис.1 приведена совместная плотность вектора  $x_1, x_2$  при  $r = 0.1$ , компоненты  $x_3 = 0, x_4 = 0$ . Сплошная линия – граница множества решений на плоскости  $(x_1, x_2)$ . Множество решений  $\mathcal{X}$  – четырех угольник с вершинами  $(2.0, 0.636)$ ,  $(2.444, 1.0)$ ,

Рис. 1: Совместная плотность вектора  $x_1, x_2$ Рис. 2: Гистограмма  $c_1x_1 + c_2x_2$ 

(2.0, 1.444), (1.636, 1.0). Как видно из рисунка плотность вероятности распределена крайне неравномерно, самая большая в центре, в окрестности точки (2.0, 1.0).

На рис.2 приведена гистограмма целевой функции  $c_1x_1 + c_2x_2$ , матожидание в точке  $-2.834$ .

Площадь  $\mathcal{X}$  сильно зависит от  $r$ , с увеличением  $r$  она растет и уже при  $r = 1$  становится бесконечной. Это определено тем, среди матриц

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \in \begin{pmatrix} [0, 2] & [0, 2] \\ [0, 2] & [-2, 0] \end{pmatrix},$$

есть линейно-зависимые столбцы.

## 4 Случайное нелинейное программирование

Рассмотрим задачу случайного нелинейного программирования без ограничений в следующем виде

$$\frac{1}{2} (Ax, x) - (b, x) \rightarrow \min. \quad (13)$$

$$A \in \mathbf{A}, b \in \mathbf{b}, \quad (14)$$

где  $\mathbf{A}$  — случайная матрица,  $\mathbf{b}$  — случайный вектор. Задача (13), (14) в случае симметричных положительно определенных матриц  $A$  сводится к решению случайной системы линейных алгебраических уравнений

$$Ax = b. \quad (15)$$

Для решения случайных систем линейных алгебраических уравнений вида (15) в общем случае можно использовать метод Монте-Карло. В ряде случаев удастся использовать численный вероятностный анализ, который эффективнее метода Монте-Карло  $10^2 - 10^3$  раз [4].

В общем случае задачу случайного нелинейного программирования (1), (2) можно свести к решению случайной системы нелинейных уравнений

$$F(x, k) = 0, \quad k \in \mathbf{k}, \quad (16)$$

где  $k \in \mathbf{k}$  — случайный вектор параметров.

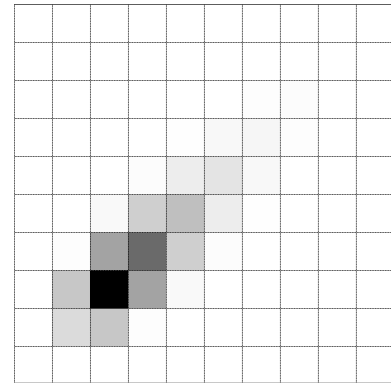
В результате для задачи (15) или (16) получаем совместную плотность вероятности решения  $x$ .

### 4.1 Численные примеры

Пусть в задаче (13)  $A$  — равномерная случайная матрица

$$A = \begin{pmatrix} a_1 & a_2 \\ a_2 & a_1 \end{pmatrix},$$

$b$  — равномерный случайный вектор. Носители  $a_1 = [2, 4]$ ,  $a_2 = [-1, 0]$ ,  $b_1 = b_2 = [0.5, 1]$ .

Рис. 3: Совместная плотность вектора  $x$ 

На рис.3 для задачи (13), (14) приведена кусочно-постоянная аппроксимация совместной плотности вероятности вектора  $x$ . Для сравнения на рис.4 приведены частные решения системы (15) аналогичные квазиМонте-Карло методу [6]. В силу определенной симметрии матрицы  $A$  частные решения системы (15) образуют некоторый порядок.

Добавим к задаче (13) ограничение в виде

$$x_1 + x_2 = a, \quad (17)$$

где  $a \in \mathbf{a}$ ,  $\mathbf{a}$  — равномерная случайная величина с носителем  $[0.9, 1.0]$ . В этом случае решение оптимизационной задачи выписывается в явном виде

$$x_1 = -(b_2 - b_1 - 2 * a * a_2 - 2 * a * a_1) / (8 * a_1)$$

$$x_2 := a - x_1;$$

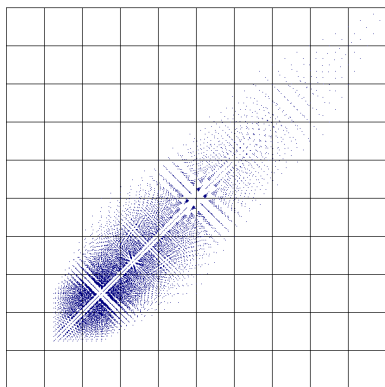


Рис. 4: Частные решения системы (15)

Используя численный вероятностный анализ [7, 5] построим решение  $x$ . На рис.5 на квадрате  $[0.7, 1.0] \times [0, 0.3]$  приведена совместная плотность вероятности вектора решения задачи (13) с ограничениями (17).

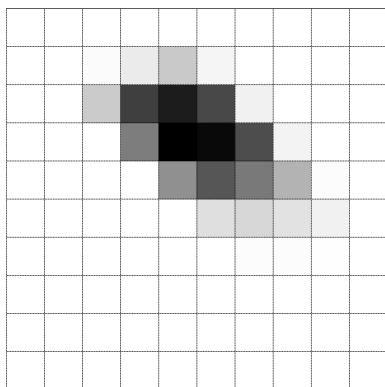


Рис. 5: Решения задачи с ограничениями

## 5 Заключение

Рассмотренные методы решения приведенных задач линейной и нелинейной оптимизации позволяют представить *случайное программирование* как эффективный метод решения оптимизационных задач в условиях неопределенности входных параметров. В дальнейшем планируется разработать алгоритмы выбора наилучших оптимальных решений из построенного множества решений.

## Список литературы

- [1] B. Liu. *Theory and Practice of Uncertain Programming (2nd Edition)*. Springer-Verlag, Berlin, 2009.
- [2] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM, Philadelphia, 2009.
- [3] Ф. П. Васильев. *Численные методы решения экстремальных задач*. Наука, М., 1988.
- [4] Б. С. Добронев и О. А. Попова. Численные операции над случайными величинами и их приложения.

*Journal of Siberian Federal University. Mathematics & Physics*, 4(2):229–239, 2011.

- [5] Б. С. Добронев и О. А. Попова. Численный вероятностный анализ для исследования систем в условиях неопределенности. *Вестник Томского государственного университета. Управление, вычислительная техника и информатика*, 21(4):39–46, 2012.
- [6] С.М. Ермаков. *Метод Монте-Карло в вычислительной математике*. Бином. Лаб. Знаний: Нев. диалект, М.; СПб., 2009.
- [7] Б. С. Добронев и О. А. Попова. Элементы численного вероятностного анализа. *Вестник Сибирского государственного аэрокосмического университета*, 2(42):19–23, 2012.

## Идентификация и прогноз функционирования модели бизнес-процессов дистанционного обучения

Михаил Георгиевич  
Доррер

Сибирский Государственный  
Технологический Университет  
Кафедра системотехники  
Красноярск  
mdorrer@mail.ru

Антон Александрович  
Зырянов

Сибирский Государственный  
Технологический Университет  
Кафедра системотехники  
Красноярск  
ZyryanovAntonA@gmail.com

Евгений Алексеевич  
Ланцев

Сибирский Государственный  
Технологический Университет  
Кафедра системотехники  
Красноярск  
lancevea@gmail.com

**Аннотация.** *Описана технология моделирования, анализа и имитации бизнес-процессов в социальных объектах, позволяющая исключить фактор субъективности эксперта при построении бизнес-модели и снизить трудоемкость построения имитационных моделей и проведения имитационного эксперимента. Технология включает исследование процесса методами Process Mining, построение агентной имитационной модели AnyLogic и прогноз статистических характеристик процесса с использованием GERT-сетей. Работоспособность технологии проверена на процессе дистанционного обучения студентов с использованием информационной системы Moodle.*

**Ключевые слова.** *Бизнес-процесс, Process Mining, GERT-сети, агентные модели, дистанционное обучение.*

### 1 Введение

Процессный подход в современном менеджменте является важнейшей структурообразующей технологией. Он положен в основу таких важнейших стандартов, как ISO 9001:2011 [1], ISO 12 207 [2], PM BOK [3]. Цикл управления бизнес-процессами согласно [4] начинается со стадий:

- Моделирование процесса – описание высокоуровневой модели выполняемых и требуемых ресурсов. Предварительная оценка времени и стоимости выполнения каждого процесса модели.
- Имитация и анализ – полученная высокоуровневая модель используется для прогноза поведения системы при выполнении различных сценариев с целью обнаружения “критических участков” (paths) и “узких мест” (bottlenecks). Результаты проведенного анализа используются для настройки процесса перед его внедрением и развертыванием.

Серьезными проблемами данных стадий являются:

- Для стадии моделирования – субъективность, зависимость адекватности модели от точки зрения и уровня квалификации аналитика.
- Для стадии имитации – высокая трудоемкость исследования при помощи существующих средств имитационного моделирования.

В настоящей работе авторы поставили перед собой цель предложить и обосновать технологию, обеспечивающую решение задач моделирования, анализа и имитации социальных объектов, но лишенную указанных недостатков. Для достижения указанной цели предполагалось включить в технологию следующие методы:

- Сбор данных об объекте непосредственно из автоматизированной системы, поддерживающей его бизнес-процессы;
- Структурная и параметрическая идентификация модели бизнес-процессов на основе данных информационной системы;
- Прогноз значимых для исследователя параметров функционирования системы (например, времени выполнения процесса и его трудоемкости).

При этом проблему субъективности построенных моделей предлагается решать за счет использования технологии и программного обеспечения аналитического исследования бизнес-процессов (Process Mining). Проблема трудоемкости имитационного моделирования решается за счет использования аппарата GERT-сетей и автоматизированной трансляции моделей из формата описания бизнес-моделей (на примере ARIS eEPC) в формат имитационной модели AnyLogic.

### 2 Объект исследования

В качестве объекта исследования в данной работе авторы выбрали процесс обучения студентов по дисциплине “Дискретная математика” с использованием информационной системы дистанционного обучения (СДО) Moodle. Moodle [5], [6] в настоящий момент является одной из самых популярных систем

дистанционного обучения. По уровню предоставляемых возможностей Moodle выдерживает сравнение с известными коммерческими СДО, в то же время выгодно отличается от них тем, что распространяется в открытом исходном коде – это дает возможность перестроить систему под особенности конкретного образовательного проекта, а при необходимости и встроить в нее новые модули. С точки зрения исследования бизнес-процессов СДО Moodle хороша тем, что все данные о прохождении обучающимися курса фиксируются и сохраняются, поэтому появляется возможность восстановления модели процессов дистанционного обучения и ее последующий анализ – Process mining. Дисциплина «Дискретная математика» преподается студентам-заочникам направления 230100.62 «Информатика и вычислительная техника» на 1 курсе. Количество часов, отведенных на изучение дисциплины по учебному плану – 180. В рамках изучения дисциплины предусмотрено выполнение индивидуальных расчетных заданий и курсового проекта. Изучение курса заканчивается экзаменом. Группа, проходившая обучение по курсу в феврале 2012 года, состояла из 70 человек. Дистанционный курс содержит 3 темы:

- Множества. Отношения на множествах.
- Графы.
- Алгоритмы на графах.

Каждая тема курса содержит лекционный и презентационный материал, проверочные работы, тесты. Вступительный раздел курса содержит методические указания по работе с курсом, выполнению расчетных заданий и курсового проекта, задания для работы с интерактивным обучающим практикумом. Методически изучение дисциплины «Дискретная математика» с использованием дистанционной технологии обучения происходит следующим образом. Во время установочной сессии со студентами проводятся лекционные и практические занятия. После этого студентам выдаются логины для доступа к системе дистанционного обучения и далее, до экзаменационной сессии, они работают с дистанционным курсом. Рекомендованная преподавателем последовательность изучения студентом дистанционного курса следующая.

- Проработать лекционный и презентационный материал по темам курса. Выполнить проверочные и тестовые задания по каждой теме.
- Выполнить задания интерактивного практикума и предоставить протоколы их выполнения преподавателю.
- Выполнить расчетные индивидуальные задания по темам «Множества и отношения», «Теория графов», отчеты предоставить на проверку преподавателю.
- Выполнить курсовой проект, отчет предоставить преподавателю.

userid	event id	activity id	activity	Starttime
3	152	26	Лекция 6	1319032327

Таблица 1: Фрагмент журнала событий бизнес-процесса дистанционного обучения студентов курсу дискретная математика.

Однако эта последовательность не является обязательной, и студент может изучать курс в том порядке, в котором ему удобно, вновь возвращаться для повторного изучения отдельных разделов, выполнения тестов, проверочных работ и других заданий. Информация обо всех действиях студента при работе с курсом сохраняются в базе данных системы дистанционного обучения. Оценки, выставяемые системой дистанционного обучения или преподавателем, автоматически заносятся в журнал. Во время экзаменационной сессии студент должен очно защитить выполненные задания и работы, после чего он получает допуск к экзамену; экзамен проводится в форме тестирования, но только аудиторно. Изучение структуры данных системы дистанционного обучения Moodle позволило выявить таблицы, необходимые для составления журнала событий. Для получения данных о процессах обучения в системе Moodle использовалось прямое подключение к рабочей базе данных системы посредством утилиты MySQL Workbench. В результате был сформирован event log (журнал событий), содержащий все необходимые данные для анализа бизнес-процессов. Фрагмент журнала событий приведен в Таблице Табл. 1.

Из значимых полей таблицы можно отметить:

- user id – идентифицирует обучающегося, а соответственно – экземпляр процесса «Дистанционное обучение студента по курсу дискретной математики».
- activity id (и соответствующий activity) – идентифицируют операцию, с которой связано событие.
- time и waitingtime – задают время выполнения операции и время ожидания после нее (в секундах), а поскольку операции выполняются одним студентом, то значение поля time также соответствует трудоемкости выполнения курса для студента.

### 3 Аналитическое исследование процессов

Анализ процессов (Process Mining) представляет собой сравнительно молодую научную дисциплину, занимающую промежуточную позицию между вычислительным интеллектом (computational intelligence) и интеллектуальным анализом данных (data mining) с одной стороны, и моделированием и анализом бизнес-процессов (process modeling and analysis) – с другой [7]. Основная идея анализа процессов заключается в том, чтобы выявить, отследить и произве-



сти улучшение реальных (а не предполагаемых) процессов путем извлечения знаний из журналов событий, легко доступных в современных информационных системах [8]. Анализ процессов включает три основных направления:

- идентификация (выявление, discovery) процессов в виде создания модели процессов из журнала событий (event log);;
- проверка соответствия (conformance) – мониторинг отклонений путем сравнения моделей и журналов событий;;
- улучшение, совершенствование (enhancement) процессов.

В данной работе авторы решали задачу идентификации процессов дистанционного обучения с помощью алгоритмов интеллектуального анализа, реализованных в программном продукте ProM 6 [9]. Встроенные алгоритмы анализа в ProM 6 используют формальные способы описания модели процесса, такие как сети Петри, С-сети и диаграммы переходов состояний. В работе были последовательно использованы:

- нечеткие модели (fuzzy model) для предварительного выявления структуры модели;;
- дискретно-событийные модели ARIS EPC для формализации полученной модели.

Результатом данного этапа исследования явилось построение модели бизнес-процесса “Дистанционное обучение студента по курсу дискретной математики” на основании протокола событий прохождения группой студентов дистанционного обучения в системе Moodle. Кроме структурной идентификации модели была выполнена и параметрическая идентификация. В каждой развилке бизнес-процесса были оценены вероятности исходящих ребер. Для каждой операции был идентифицирован закон распределения времени выполнения и времени ожидания, а также рассчитаны параметры закона распределения. Идентификация закона распределения и его параметров на основе экспериментальных данных осуществлялась с помощью пакета EasyFit [10]. Данный пакет включает 55 распределений, например: нормальное, логнормальное, логистическое, экспоненциальное, Парето, Коши, Лапласа, Вейбулла и т.д. Для идентификации закона распределения времен выполнения и ожидания по операциям бизнес-процесса был использован критерий согласия Колмогорова-Смирнова.

#### 4 Прогноз статистических параметров бизнес-процесса при помощи GERT-сети

Авторы ставили перед собой задачу получения прогноза числовых параметров бизнес-процесса, а именно времени выполнения и общей трудоемкости про-

цесса (дополнительно мог быть получен прогноз движения материальных ресурсов, включая деньги), при условии более высокой точности и более низкой трудоемкости, чем это позволяет сделать имитационный эксперимент. Авторы предположили, что одним из решений данной задачи может быть применение для числовой оценки бизнес-процессов аппарата GERT-сетей (GERT – graphical evaluation and review technique). Классическое описание GERT-сетей приведено в декларации Прицкера [11], на русском языке описание GERT-сетей можно найти в книге Филлипса и Гарсиа-Диаса [12]. GERT-сети являются вариантом полумарковских моделей, но случайные величины в них характеризуются не только дисперсией, но и законом распределения. При использовании GERT-сетей, как правило, определяют значения переменных, связанных с первыми моментами распределения выходной величины относительно начала координат (математическое ожидание, дисперсию, коэффициент асимметрии и эксцесса) [13]. Основные методы расчета GERT-сетей описаны в работах [11], [14], [12], [13]. Данные методы позволяют: преобразовать сеть к эквивалентной дуге, рассчитать вероятностные характеристики стока сети, определить значения переменных, связанных с моментами распределения выходной величины сети (математическое ожидание, дисперсия и др.), построить функцию распределения выходной величины сети. Расчет GERT-сети на основе уравнения Мейсона рассмотрен в работах [11], [12] и представляет нахождение производящей функции моментов эквивалентной дуги сети и вычисления по ней первых моментов сети. Однако серьезным сдерживающим фактором применения GERT-сетей является экспоненциальный рост трудоемкости алгоритмов расчета [15], основанных на использовании топологического уравнения Мейсона, с увеличением числа анализируемых ребер GERT-сети. Данная проблема в работе решена путем нахождения числовых характеристик GERT-сети без использования топологического уравнения Мейсона и вычисления частных производных. Решение основывается на преобразовании GERT-сети к эквивалентной дуге с пересчетом числовых характеристик случайных величин дуг. Под числовыми характеристиками случайных величин будут пониматься математическое ожидание и дисперсия. Алгоритм преобразования однородной GERT-сети к эквивалентной дуге основан на поиске участков сети с тремя типами расположения дуг: последовательные дуги, параллельные дуги, дуга и петля, и их замене на эквивалентную дугу с пересчетом математического ожидания и дисперсии, а также вероятности выполнения дуги. Преобразовывая GERT-сеть на основе данного алгоритма, символично используя в формулах числовые характеристики дуг, можно получить аналитические формулы числовых характеристик GERT-сети. Более подробное описание алгоритма приведено в работе [16].

## 5 Имитационный эксперимент

На основе модели процесса прохождения учебного курса, разработанной с помощью алгоритмов нечеткой логики (плагин Fuzzy Miner для ProM), была создана модель бизнес-процессов в нотации EPC, а на ее основе  $\Pi$  имитационная агентная модель AnyLogic. Агентная модель в AnyLogic [17] реализуется с помощью базового элемента  $\Pi$  активного объекта. Активный объект имеет параметры, переменные, которые можно считать памятью агента, диаграммы состояний (State Chart, стейтчарты) выражают поведение: состояния объекта и изменение состояний под воздействием событий и условий. Агентная модель в AnyLogic, основанная на моделях eEPC, строится на следующих принципах:

- В качестве агентов используются организационные единицы eEPC  $\Pi$  должность, организационная единица и другие, которые задействованы в моделях eEPC. В данной работе агентом является студент.
- Заявки в агентной модели поступают при помощи таймеров (таким образом моделируются начальные события моделей eEPC), которые генерируют их по заданному расписанию, а также сами агенты могут генерировать заявки друг для друга.
- У каждого из типов агентов имеется своя очередь из заявок, которая обрабатывается, когда освобождается необходимое количество агентов. Например, для типа агентов “Грузчик” имеется очередь машин для разгрузки, для разгрузки машин необходимо несколько грузчиков, если освобождается необходимое количество грузчиков, то разгрузка машины начинается.
- Поведение агента задается диаграммой состояний. Состояния соответствуют функциям eEPC. Логика перекрестков реализуется в коде Java при смене состояний.

Диаграмма состояний агентной модели состоит из одного гиперсостояния и нескольких простых состояний. В гиперсостоянии есть одно состояние  $\Pi$  начальное, в котором агент находится в самом начале моделирования, а также переходит после выполнения каждой функции/группы функций. Развернутое описание подхода к трансляции дискретно-событийной модели EPC в агентную модель AnyLogic приведено в работе [18].

## 6 Сравнение результатов эксперимента, оценка предсказательной способности методов

Итак, для объекта исследования – бизнес-процесса дистанционного обучения студентов курсу “Дискрет-

Данные	Мат. ожид.	Мат. ожид. (ошибка)	Дисперсия	Дисперсия (ошибка)
Выборка	2819		14589725	
GERT	3150	11,75%	9773561	33,01%
AnyLogic	3184	11,59%	9814746	32,73%

Таблица 2: Сравнительная оценка предсказательной способности предложенных методов (по показателю общего времени выполнения).

ная математика” была произведена структурная и параметрическая идентификация. На рисунке 1 приведены ее результаты – а) Fuzzy Model, б) GERT-сеть, в) модель ARIS EPC. Легко видеть, что топологически модели абсолютно идентичны.

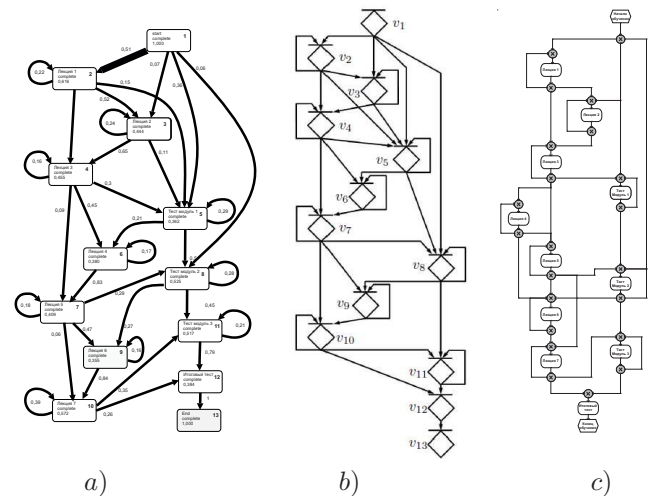


Рис. 1: Результаты описания модели бизнес-процесса “Дистанционное обучение студентов по курсу дискретной математики”.

Далее для бизнес-процесса были получены числовые характеристики модели на основе 2 методов:

- Расчета GERT-сети с помощью методики оценки числовых характеристик на основе эквивалентных преобразований.
- Имитационных экспериментов – одного миллиона прогонов разработанной агентной имитационной модели в AnyLogic.

В Табл. 2, 3. представлены результаты расчетов вероятностно-временных характеристик процесса рассчитанные указанными методами, GERT-сети и в сравнении с аналогичными характеристиками, характеризующими выборку экспериментальных данных по процессу (напомним, размер этой выборки – 70 прецедентов выполнения процесса по числу студентов, прошедших обучение).

Из таблицы 2 видно, что для трудоемкости процесса лучше (погрешность не более 4%) предсказывается математическое ожидание величины, при этом оценка дисперсии завышена на 110 – 150%. Для общей

Данные	Мат. ожид.	Мат. ожид. (ошибка)	Дисперсия	Дисперсия (ошибка)
Выборка	359		196069	
GERT	345	3,71%	495211	152,57%
AnyLogic	352	1,82%	418973	113,69%

Таблица 3: Сравнительная оценка предсказательной способности предложенных методов (по показателю трудоемкости).

длительности процесса предсказание математического ожидания осуществляется с большей погрешностью (около 10%), а дисперсия – с меньшей (около 33%). Причем эти оценки весьма близки для имитационной модели и для GERT-сети, что позволяет предположить наличие некоего общего фактора, вносящего искажение в прогноз. Возможно, речь идет об ограничениях, наложенных на модель при структурной идентификации, но для выяснения этого необходимо проведение дополнительных исследований.

## Список литературы

- [1] Международный стандарт ИСО 9001:2011. *ИСО 9001:2011*. <http://docs.cntd.ru/document/1200093426>, 2011.
- [2] ГОСТ Р ИСО/МЭК 12207-99. <http://standartgost.ru/ГОСТ Р ИСО/МЭК 12207-99>, 2011.
- [3] PMI. *Project Management Body of Knowledge*. <http://www.pmi.org/PMBOK-Guide-and-Standards.aspx>, 2011.
- [4] B. Nainani. Closed loop bpm using standards based tools. *An Oracle White Paper*, november, 2004.
- [5] Описание ПО Moodle. <http://docs.moodle.org>, 2012.
- [6] Описание СДО Moodle. <http://www.web-learn.ru/index.php>, 2011.
- [7] IEEEFP. *Process Mining Manifesto*. <http://www.win.tue.nl/ieetfpm/downloads/Process Mining Manifesto.pdf>, 2011.
- [8] W.M.P. van der Aalst. *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Springer-Verlag, Berlin, 2011.
- [9] Описание ПО ProM. <http://www.processmining.org/prom / start>, 2012.
- [10] MathWave. *Описание ПО EasyFit*. <http://www.mathwave.com/easyfit-distribution-fitting.html>, 2012.
- [11] A. A. B. Pritsker. *GERT: graphical evaluation and review technique*. RAND Corporation, 1966.
- [12] Д. Филлипс, А. Гарсиа-Диаз. *Методы анализа сетей*. Мир, М., 1984.
- [13] А.П. Шибанов. *Обобщенные GERT-сети для моделирования протоколов, алгоритмов и программ телекоммуникационных систем: диссертация д.т.н.: 05.13.13*. PhD thesis, Рязанская государственная радиотехническая академия, 2003.
- [14] K. Neuman. *Stochastic project networks: temporal analysis, scheduling and cost minimization*. Springer-Verlag, Berlin, 1990.
- [15] А.П. Шибанов. Нахождение плотности распределения времени исполнения gert-сети на основе эквивалентных упрощающих преобразований. *An Oracle White Paper*, Автоматика и телемеханика:117–126, 2003.
- [16] А.А.Зырянов. Методика расчета числовых характеристик однородной gert-сети. In *Перспективы развития информационных технологий: сб. мат. XI Межд. научно-практической конференции*, pages 21–27, Новосибирск, 2013. СИБПРИНТ.
- [17] Ю.Г. Карпов. *Имитационное моделирование систем. Введение в моделирование с AnyLogic 5*. БХВ-Петербург, СПб., 2006.
- [18] Е.А. Ланцев. Агентный и дискретно-событийный подходы к имитационному моделированию бизнес-процессов в нотации еерс. *В мире научных открытий*, 1:278–290, 2013.

## Об определении отношения избирателей к кандидату (на примере досрочных выборов главы г. Красноярск)

Мария Вадимовна Елина

Красноярск  
elinamv@gmail.com

Екатерина Борисовна

Логинова

ТРК «Прима»  
Красноярск  
borisovna111@yandex.ru

Михаил Георгиевич Садовский

Институт вычислительного моделирования  
СО РАН  
Красноярск  
msad@icm.krasn.ru

**Аннотация.** *Изложен новый метод определения отношения избирателей к политическому актору. Метод основан на сравнении актора с теми или иными пословицами и поговорками, циркулирующими в современном актору обществе.*

**Ключевые слова.** *Психология, лингвистика, социология, проекция, интеллект*

### 1 Введение

Выявление и описание отношения избирателей к кандидату (либо действующему политику) является важнейшей прикладной задачей; изучение отношения к публичным фигурам имеет и фундаментальный интерес. Для этого требуется специальный подход и специальный инструментарий; мощным средством в этой ситуации могут стать проективные методики.

Специфика самого такого исследования заключается в том, что представления о кандидате либо о действующем политике, либо об ином публичном деятеле (далее — актор) у респондентов носят в основном вербальный характер. Это в первую очередь связано с самим жанром публичной деятельности: она в основе своей является коммуникативной и основывается на речевых образах и представлениях об акторе.

### 2 Новый метод

Нами был предложен способ определения отношения к актору и его восприятия респондентами, основывающийся на новой проективной методике, использующей лишь речевые опорные образы. Этот способ был апробирован на досрочных выборах главы г. Красноярск в июне 2012 года и показал свою высокую эффективность.

Способ исследования базируется на использовании русских пословиц и поговорок в качестве стимульного материала. Респондентам предлагалось оценить, насколько тот или иной кандидат соответствует данной пословице (поговорке) и указать свой выбор.

Идея этого метода основывается на том факте, что пословицы и поговорки представляют собой специальный класс лингвистических явлений, занимая промежуточное положение между (обычным) высказыванием и отдельным словом (лексемой); дело в том, что смысл пословицы (поговорки) не выводим (в отличие от обычного высказывания) из семантики составляющих его слов. Тем самым, пословица (поговорка) становится фактически неделимой языковой единицей, существующей в сознании респондента как целое.

Выбор конкретных пословиц (поговорок) диктуется задачами конкретного исследования: они должны описывать те характеристики акторов, оценки которых респондентами добывается исследователь. Разнообразие и богатство пословиц (поговорок) достаточно велико, чтобы можно было подобрать релевантные единицы лексического материала для самых разных задач исследования.

Мы изучали указанным методом восприятие жителями Красноярска кандидатов на должность главы города, прошедших этап регистрации, на досрочных выборах главы в июне 2012 года. Данные выборы характеризовались, с одной стороны, достаточно сильным общественным напряжением, порождённым прошедшими в декабре 2011 года выборами в ГД ФС РФ и президентскими в марте 2012 года, а с другой — сильным дефицитом сколько-нибудь привлекательных для избирателей акторов от оппозиции. Кандидат от партии власти (Амбулатов Э.Ш.) был к тому моменту уже пять месяцев как исполняющим обязанности главы города, после добровольного ухода со своего поста предыдущего главы. Посредством т.н. первичных выборов был определён «единый» кандидат от оппозиции (Коропачинский А.Ю.), однако в целом в выборах принимало участие 7 кандидатов, из которых оппозиционным был ещё один (Подкорытов А.В.), двое — техническими (Иваных М.С. и Коврова Н.Б.), ещё один (Осколков М.А., формально поддержанный КПРФ, однако со скандалом потерявший эту поддержку) — спойлером и ещё один (Мещеряков А.Н., самовыдвиженец, при поддержке *Справедливой России*) — независимым кандидатом.

Респондентам предлагался набор из 12 пословиц, ко-



Таблица 1: Оценка соответствия актора пословице (в процентах; обозначения см. в тексте)

Кандидат	Наводит тень на ясный день	Видом сокол, а голосом ворона	Люди пахать, а он руками махать	Сказано — сделано	От него, как от козла, — ни молока, ни шерсти	Доверили козлу капусту	Не человек местом красится, а место человеком	Сам тетерев, а хочет выглядеть павлином	Хоть лыком шит, да начальник	Насильно мил не будешь	Был двойкой хорош, а тузом не годится	Лясы точит да людей морочит
Коропачинский А.Ю.	32	16	24	19	10	35	10	27	15	22	15	33
Осколков М.А.	23	34	19	3	37	10	6	30	4	9	23	18
Акбулатов Э.Ш.	13	23	24	26	13	30	29	15	34	31	12	24
Подкорытов А.В.	10	5	7	13	23	6	19	9	12	8	29	6
«затрудняюсь ответить»	22	22	26	40	16	19	35	19	35	30	30	19

торые нужно было соотнести с каким-то кандидатом на пост главы города. В исследовании использовались следующие пословицы: *Наводит тень на ясный день*, *Видом сокол, а голосом ворона*, *Люди пахать, а он руками махать*, *Сказано — сделано*, *От него, как от козла, — ни молока, ни шерсти*, *Доверили козлу капусту*, *Не человек местом красится, а место человеком*, *Сам тетерев, а хочет выглядеть павлином*, *Хоть лыком шит, да начальник*, *Насильно мил не будешь*, *Был двойкой хорош, а тузом не годится*, *Лясы точит да людей морочит*. Результаты представлены в Таблице 1.

Результаты, представленные в Таблице 1, показывают, что наибольшие затруднения при атрибуции кандидатов как субъектов пословиц (поговорок) вызвали поговорки *Сказано — сделано*; *Не человек местом красится, а место человеком*; *Хоть лыком шит, да начальник*.

Следует особо подчеркнуть, что данное исследование носило во многом пионерский характер. Кроме того, существенной трудностью при его проведении были очень сильные ограничения по времени: кампания уже шла к своему завершению и не было достаточного времени, чтобы провести это исследование в полном масштабе. Тем не менее, результаты, представленные в настоящей работе, показывают высокую эффективность данного метода.

Здесь следует сделать несколько важных замечаний.

Для того, чтобы работы типа той, что изложена выше, стали по-настоящему методом исследования отношения избирателей к тому или иному актору, необходимо проведение следующих методологических исследований.

**Во-первых**, нужно подобрать пословицы и поговорки, адекватные как ситуации, так и избирателям. Для этого необходимо знать, какие именно пословицы и поговорки находятся в речевом и ментальном обороте в настоящее время. Кроме того, требуется известная унификация способов из толкования.

**Во-вторых**, сам по себе приём использования пословиц и поговорок в диагностике интеллектуальных способностей и некоторых душевных заболеваний не нов. Им активно пользуются психиатры для диагностики некоторых заболеваний. Здесь однако следует понимать, что условно все пословицы и поговорки можно разделить на высоко частотные и уникальные. Пословицы и поговорки каждого класса используются для диагностики разных способностей и это обстоятельство также должно быть учтено в разработке метода.

**Благодарности.** Авторы выражают благодарность М. Н. Дымшицу за содержательное обсуждение представленного здесь подхода и полученных в его рамках результатов.



# Решение обратной задачи теории риска в рамках модели CVaR

Есин Роман Витальевич

Сибирский Федеральный Университет  
Институт космических и информационных технологий  
Красноярск  
surgeon14@mail.ru

**Аннотация.** В данной работе излагается метод решения обратной задачи относительно CVaR, также рассматривается характеристика отношения инвестора к риску, и как ключевое понятие - неприятие риска. Предлагается процедура для определения параметра неприятия риска на основе развернутого анкетирования.

**Ключевые слова.** Теория риска, когерентные меры риска, неприятие риска, CVaR

## 1 Введение

Прямая задача в теории риска - это задача принятия решения в условиях вероятностной неопределенности. Но что если индивидуум хочет принять решение на основании собственного отношения к риску? Для этого на множестве всевозможных вероятностных распределений необходимо построить некоторый вещественный функционал, на основе которого этот индивидуум будет принимать решение, и этот используемый им функционал должен быть согласован с его предпочтением. Тогда, обратной задачей будем считать построение меры риска, которая строится по каким-либо известным характеристикам предпочтения этого человека. Решим данную задачу в рамках модели CVaR.

## 2 Основные понятия и определения

Рассмотрим вероятностное пространство  $(\Omega, \mathcal{A}, \mathbf{P})$ , где  $\Omega$  - множество элементарных исходов,  $\mathcal{A}$  -  $\sigma$ -алгебра, заданная на  $\Omega$ , а  $\mathbf{P}$  - вероятностная мера, определенная на множествах из  $\mathcal{A}$ .

**Риском**  $X$  на  $(\Omega, \mathcal{A})$  называется произвольное измеримое отображение из  $\Omega$  в  $\mathbf{R}$ , иначе говоря случайная величина. А значения риска будем принимать как доход или убыток некоторого лица, принимающего решение. Тогда  $\mathcal{X}$  - совокупность всех рисков на  $(\Omega, \mathcal{A})$ .

Говорят, что на множестве  $\mathcal{R}$  задано отношение предпочтения  $\preceq$ , если это отношение обладает свойствами:

- полноты: для произвольной пары  $r_1, r_2 \in \mathcal{R}$  либо

$r_1 \preceq r_2$ , либо  $r_2 \preceq r_1$ , либо справедливы оба этих отношения;

- транзитивности: если  $r_1 \preceq r_2$  и  $r_2 \preceq r_3$ , то  $r_1 \preceq r_3$

Элементы  $r_1, r_2 \in \mathcal{R}$ , для которых верно  $r_1 \preceq r_2$ ,  $r_1 \preceq r_2$ , называются эквивалентными:  $r_1 \sim r_2$ . Пусть на заданы отношения (частичного) порядка  $\leq$  и предпочтения  $\preceq$ . Предпочтение  $\preceq$  называют согласованным с порядком  $\leq$ , если  $r_1 \leq r_2 \implies r_1 \preceq r_2$ . Данное отношение предпочтения отражает индивидуальное отношение к риску некоторого индивидуума.

**Мерой риска** называется произвольный функционал  $\rho : \mathcal{X} \rightarrow \mathbf{R}$ . Говорят, что  $\rho$  представляет отношение предпочтения на  $\mathcal{X}$ , кгда

$$\rho(X) \leq \rho(Y), \text{ если } X \preceq Y, \text{ где } X, Y \in \mathcal{X}$$

Отношение предпочтения  $\preceq$  обладает свойством неприятия риска, если для произвольного риска  $\Delta : E\Delta = 0$  и для произвольного  $a \in \mathbf{R}$  имеет место

$$a + \Delta \prec a$$

Иначе говоря, это нежелание человеком заменять детерминированную сумму денег  $a$  на случайное количество со средним значением  $a$ . В терминах меры риска  $\rho$ , задающей предпочтение  $\preceq$  на  $\mathcal{X}$ , это можно записать так:  $\rho(a + \Delta) < \rho(a)$ . Если отношение предпочтения на  $\mathcal{X}$  является регулярным и задано с помощью меры риска  $\rho$ , то  $\rho(a + \Delta) = \rho(a - c)$ ,  $c > 0$ . Величину  $c$  можно использовать в количественной оценке неприятия риска[4].

## 3 Когерентные меры риска

В работе [2], было введено понятие когерентной меры риска. Следуя работе [3] изложим их в терминах супермодулярных функций, который являются эквивалентным описанием когерентных мер риска, точнее, если  $\rho : \mathcal{X} \rightarrow \mathbf{R}$  - когерентная мера риска, то  $f$  - супермодулярная функция, связанная с  $\rho$  посредством соотношения  $f(X) = -\rho(X)$ , для произвольного  $X \in \mathcal{X}$ . Такие супермодулярные функционалы будем называть когерентными мерами риска.

**Когерентной мерой риска** называется функционал  $f : \mathcal{X} \rightarrow \mathbf{R}$ , обладающий свойствами монотон-

ности, супераддитивности, положительной однородности и инвариантности относительно сдвигов, соответственно:

- $X \leq Y \implies f(X) \leq f(Y)$
- $f(X + Y) \geq f(X) + f(Y)$
- $f(\lambda X) = \lambda f(X)$ ,  $\lambda > 0$
- $f(X + aI) = f(X) + a$ ,  $a \in \mathbf{R}$

Здесь  $X, Y$  - произвольные случайные величины из  $\mathcal{X}$ , а  $I = (1, 1, \dots, 1)$  - вектор с единичными компонентами длины  $n$ . Свойство монотонности гарантирует, что когерентные меры риска монотонны относительно естественных порядков на множестве рисков. Супераддитивность отражает стремление к диверсификации портфеля и наличие у задаваемого предпочтения неприятия риска. Положительная однородность требует, чтобы мера риска от  $n$  инструментов с одинаковыми характеристиками в  $n$  раз отличалась от значения меры риска, подсчитанной для одного такого инструмента. Свойство инвариантности относительно сдвига указывает на линейность изменения функционала вдоль вектора, задающего безрисковый актив.

#### 4 Меры риска VaR и CVaR

В последнее время предложен ряд новых мер оценки рыночных рисков. Наиболее распространенной на сегодняшний момент методологией оценивания рыночных рисков является Value-at-Risk (VaR). За последние несколько лет VaR стал одним из самых популярных средств управления и контроля риска. В настоящее время оценка VaR используется организациями, осуществляющими надзор за банками, для определения резервов капитала. Пусть случайная величина  $X$  - убыток.

$$VaR_\alpha(X) = \inf\{x : F_X(x) \geq \alpha\}$$

$F_X(x) = P(X \leq x)$  - функция распределения  $X$ ,  $\alpha \in [0, 1]$  - уровень доверия ( $\alpha \gg 0$ ). Значение меры  $VaR$  - наибольший убыток, который не будет превышен с уровнем доверия  $\alpha$ .

Если значение  $X$  интерпретировать как доход, то  $VaR$  вычисляется следующим образом

$$VaR_\beta(X) = -\inf\{x : F_X(x) \geq \beta\}$$

Здесь  $\beta \ll 1$ . В таком случае,  $VaR$  - наибольший убыток, который не будет превышен с уровнем доверия  $1 - \beta$ .

Мера  $VaR$  проста в вычислении, но имеет ряд серьезных недостатков:  $VaR$  не учитывает возможных

больших потерь, которые могут произойти с маленькими вероятностями, не может различить разные типы хвостов распределения потерь и поэтому недооценивает риск в случае, когда распределение потерь имеет "тяжелые хвосты" (т.е. его плотность медленно убывает), не является когерентной мерой, в частности, он не обладает свойством супераддитивности. Можно привести примеры, когда  $VaR$  портфеля больше, чем сумма  $VaR$ ов двух подпортфелей, из которых он состоит. Это противоречит здравому смыслу. Если рассматривать меру риска как размер капитала, резервируемого для покрытия рыночного риска, то для покрытия риска всего портфеля нет необходимости резервировать больше, чем сумму резервов составляющих подпортфелей из чего следует, что эта мера риска представляет предпочтения, не обладающие свойством неприятия риска.

Мера  $CVaR$  является когерентной альтернативой мере  $VaR$ . Она была предложена в работе [2] под названием Tail conditional expectation. К. Ачерби в работе [1] предложил свой вариант такой меры под названием Expected Shortfall. Итак, пусть  $X$  - убыток,  $\alpha \gg 0$  - уровень доверия

$$CVaR_\alpha(X) = E(X : X(x) \geq VaR_\alpha(X))$$

В [1] показано, что

$$CVaR_\alpha(X) = \frac{1}{1-\alpha} \int_\alpha^1 G_X(p) dp$$

где  $G_X(p) = \inf\{x : P(X \leq x) \geq p\}$  - функция, названная Ачерби обобщенным квантилем. Для абсолютно непрерывных распределений  $G_X(p) = F_X^{-1}(p)$ .

Если случайная величина  $X$  - доход,  $\beta \ll 1$ , то

$$CVaR_\alpha(X) = -E(X : X(x) \leq -VaR_\alpha(X))$$

$$CVaR_\alpha(X) = -\frac{1}{\beta} \int_0^\beta G_X(p) dp$$

Мера  $CVaR$  когерентна и позволяет учитывать большие потери, которые могут произойти с небольшой вероятностью. Она также более адекватно оценивает риск в распространенном на практике случае, когда распределение потерь имеет "тяжелые хвосты".

#### 5 Обратная задача теории риска для модели CVaR

Пусть к мы являемся консультантами, и к нам обратился инвестор, который желает вложить деньги в портфель акций. Но хочет чтобы это вложение согласовывалось с его личным предпочтением и уровнем неприятия риска. Для этого нам нужно определить, согласовывается ли отношение предпочтения инвестора с моделью  $CVaR$ , и определить количественную характеристику неприятия риска у заказчика.

Чтобы определить можно ли описать предпочтения инвестора моделью  $CVaR$  составим и проведем раз-

вернутое анкетирование по методу выбора ответа из нескольких альтернатив. Чтобы это определить будем строить задачи исходя из свойств, которым должны отвечать когерентные меры риска, и, как частный случай, мера  $CVaR$ .

Пусть  $f(x)$  - произвольная когерентная мера риска.

### 5.1 Монотонность

*Задача:* В своем бюджете Вы выделяете 2000 рублей на занятия в бассейне. Вы посещаете бассейн 2 раза в неделю. За прошлый год вы имеете следующую статистику пропусков: 2 месяца Вы пропустили 4 раза, 4 месяца - 2 раза, 5 месяцев - 1 раз и 1 месяц ни разу не пропустили занятия в бассейне. Исходя из этих данных, что для Вас более предпочтительно:

1. купить абонемент на 8 посещений в месяц за 1600 рублей
2. каждый раз платить при посещении 200 рублей

*Обоснование:*

Если инвестор выбирает первый ответ, это означает, что на конец месяца его капитал будет описываться следующим распределением:

$X$	400	400	...	400
$P$	$p_1$	$p_2$	...	$p_n$

$$EX = 400, p_1 + p_2 + \dots + p_n = 1$$

Если инвестор выбирает второй вариант ответа, то на конец месяца его капитал будет описываться следующим распределением:

$Y$	400	600	800	1000	1200	...	2000
$P$	0,08	0,42	0,17	0,33	0	...	0

$$EY = 750,5$$

Таким образом, если предпочтения инвестора согласовываются со свойством монотонности, то он выберет второй вариант, так как  $X < Y$  и  $f(X) < f(Y)$

### 5.2 Положительная однородность

*Задача:* У Вас есть 1000 рублей, Вам предложили сыграть в лотерею, в результате которой Вы с 50%ой вероятностью либо удвоите свой капитал, либо потеряете его. Но поучаствовать в данной лотерее Вы можете разными способами. Выберите более подходящий Вам:

1. вложить всю 1000 рублей сразу
2. разделите сумму по 100 рублей и сыграете в данную лотерею 10 раз

3. разделите сумму по 200 рублей и сыграете в данную лотерею 5 раз
4. разделите сумму по 500 рублей и сыграете в данную лотерею 2 раз

Сколько Вам необходимо взять лотерейных билетов 2-го, 3-го и 4-го типа для того, чтобы ситуации были одинаково предпочтительны?

*Обоснование:*

Значения рисков в данном случае - не капитал, а доход.

При выборе первого ответа, данная игра соответствует риску  $\Delta_1$ , имеющему следующее распределение:

$\Delta_1$	-1000	1000
$P$	0,5	0,5

$$E\Delta_1 = 0, f(\Delta_1) = f(1000)$$

При выборе второго ответа, данная игра соответствует риску  $\Delta_2$ , имеющему следующее распределение:

$\Delta_2$	-100	100
$P$	0,5	0,5

$$E\Delta_2 = 0, \lambda = 10, f(\lambda \cdot \Delta_2) = f(1000) = 10 \cdot f(100)$$

При выборе третьего варианта ответа, данная игра соответствует риску  $\Delta_3$ , имеющему следующее распределение:

$\Delta_3$	-200	200
$P$	0,5	0,5

$$E\Delta_3 = 0, \lambda = 5, f(\lambda \cdot \Delta_3) = f(1000) = 5 \cdot f(200)$$

При выборе четвертого варианта ответа, данная игра соответствует риску  $\Delta_4$ , имеющему следующее распределение:

$\Delta_4$	-500	500
$P$	0,5	0,5

$$E\Delta_4 = 0, \lambda = 2, f(\lambda \cdot \Delta_4) = f(1000) = 2 \cdot f(500)$$

Таким образом, если предпочтения инвестора согласуются со свойством положительной однородности, в данной задаче человек скажет, что ситуации для него будут одинаково предпочтительны при соответствующем выборе количества лотерейных билетов:  $f(1000) = 10 \cdot f(100) = 5 \cdot f(200) = 2 \cdot f(500)$

### 5.3 Супераддитивность

*Задача:* У Вас есть капитал в размере 100 тысяч рублей, Вы можете вложить его в акции следующим образом:

1. вложить весь капитал в акции МТС, тогда через год Ваш капитал увеличится на 75% с вероятностью 0.35, либо уменьшится на 75% с вероятностью 0.65
2. вложить весь капитал в акции Краскон, тогда через год Ваш капитал уменьшится на 50% с вероятностью 0.5, либо уменьшится на 125% с вероятностью 0.5
3. вложить весь капитал в портфель акций МТС и Краскон, тогда через год Вы с вероятностью 0.175 получите 150000 рублей, с тоже вероятностью вы можете ничего не получить, с вероятностью 0.325 получите 75000 рублей, либо с вероятностью 0.325 получите 225000 рублей.

Вам предлагается оценить каждый из этих вариантов по пятибалльной шкале в соответствии с Вашими предпочтениями.

*Обоснование:*

Распределение риска от вложения в акции Краскон имеет следующий вид:

$Y$	50000	-25000
$P$	0,5	0,5

$$EY = 12500$$

Распределение риска от вложения в акции МТС имеет следующий вид:

$X$	25000	175000
$P$	0,65	0,35

$$EX = 77500$$

Распределение риска от вложения в портфель акций Краскон и МТС будет описываться следующим распределением (значения случайных величин считаем независимыми):

$$z_2 = X + Y$$

$z_2$	75000	0	225000	150000
$P$	0,325	0,175	0,325	0,175

$$Ez_2 = 123750$$

Проверку выполнения свойства супераддитивности будем основывать на получении экспертной оценки опрашиваемого. Если его оценки будут удовлетворять условию  $f(z_2) \geq f(X) + f(Y)$ , значит предпочтения инвестора согласовывается с свойством супераддитивности, а это показывает стремление инвестора к диверсификации его портфеля.

#### 5.4 Инвариантность относительно сдвига

*Задача:* Вам предложили на выбор сыграть в одну из двух игр:

1. подбрасывается монетка и, в случае выпадения орла, вы выигрываете 3000 рублей, а в случае выпадения решки, вы проигрываете 2000 рублей
2. подбрасывается монетка и, в случае выпадения орла, вы выигрываете 2000 рублей, а в случае выпадения решки, вы проигрываете 3000 рублей

Какую сумму Вам необходимо заплатить, чтобы оба варианта для Вас были одинаково предпочтительны?

*Обоснование:*

В первой игре возможный доход будет описываться следующим распределением:

$\Delta$	-2000	3000
$P$	0,5	0,5

$$E\Delta = 500$$

Во второй игре доход будет описываться следующим распределением:

$Z$	-3000	2000
$P$	0,5	0,5

$$EZ = -500$$

Таким образом, если предпочтения инвестора согласуются со свойством инвариантности относительно сдвига, в данной задаче человек ответит, что за 1000 рублей риски  $\Delta$  и  $Z$  для него будут одинаково предпочтительны, так как  $\Delta = Z + 1000I$  и  $f(\Delta) = f(Z) + 1000$

Исходя из ответов заказчика на эти вопросы мы сможем отсеять тех, кто слишком рискован и определим согласованность предпочтений инвестора с моделью CVaR. Но выбор "правильных вариантов" во всех предложенных задачах является НЕОБХОДИМЫМ условием того, что предпочтения индивидуума согласуются с моделью когерентных мер риска. Естественно, что для достоверности в опросник мы включаем не по одной задаче на каждое свойство, а по несколько, чтобы результаты проведенного анкетирования были достоверны.

#### 5.5 Неприятие риска

*Задача:* Предлагается такая ситуация: у Вас имеется 1000 рублей, Вам предлагается сыграть в игру, в которой Вы ставите эту сумму, а в результате с 50%ой вероятностью получите:

1.	900 рублей	или	1100 рублей
2.	800 рублей	или	1200 рублей
3.	700 рублей	или	1300 рублей
4.	600 рублей	или	1400 рублей
5.	500 рублей	или	1500 рублей
6.	400 рублей	или	1600 рублей
7.	300 рублей	или	1700 рублей
8.	200 рублей	или	1800 рублей
9.	100 рублей	или	1900 рублей
10.	0 рублей	или	2000 рублей

Предполагается, что перед началом каждого этапа у Вас есть 1000 рублей т.е. приступая к каждой следующей, Вы абстрагируетесь от результатов предыдущей. Согласитесь ли Вы играть в эту игру и на каком этапе готовы остановиться?

*Обоснование:*

В данной задаче мы проверяем желание человека заменять детерминированную сумму денег  $a$  на случайное количество со средним значением  $a$ . Зная на каком этапе инвестор готов остановиться, мы можем определить количественную характеристику неприятия риска. В терминах меры риска  $\rho$  это можно записать так:  $\rho(a + \Delta) < \rho(a)$ . Если отношение предпочтения на  $\mathcal{X}$  является регулярным и задано с помощью меры риска  $\rho$ , то  $\rho(a + \Delta) = \rho(a - c)$ ,  $c > 0$ . В модели CVaR мы получаем:  $CVaR_\alpha(a - c) = CVaR_\alpha(a + \Delta)$ , отсюда и зная что  $CVaR_\alpha(X) = \frac{1}{1-\alpha} \int_\alpha^1 G_X(p)dp$ , мы можем выразить  $\alpha$ .

## Список литературы

- [1] Tasche D. Acerbi C. Expected shortfall: a natural alternative to value-at-risk. *Economic Notes*, 31:379–388, 2001.
- [2] Ebler J.-M. Heath D. Artzner Ph., Delbaen F. Coherent measures of risk. *Mathematical Finance*, 9:203–228, 1999.
- [3] Delbaen F. Coherent risk measures on general probability spaces. *Advance of Finance and Stochastics*, ~:10–38, 2002.
- [4] Новоселов А. А. Неприятие риска: качественный подход и количественные оценки. *Автоматика и телемеханика*, 7:165–177, 2003.
- [5] Новоселов А. А. *Математическое моделирование финансовых рисков. Теория измерения*. Наука, Новосибирск, 2001.
- [6] Кустицкая Т. А. Мера CVaR, ее модификации и место среди когерентных мер риска (Measure CVaR, its modifications and its place among coherent risk measures). *Моделирование и анализ безопасности и риска в сложных системах: Труды Международной Научной школы МА БР ГОУ ВПО "СПбГУАП"*, ~:252–257, 2010.
- [7] Кустицкая Т. А. Решение обратной задачи теории риска в рамках модели обобщенных когерентных мер риска (для евклидовой нормы). *Труды XVI Международ. ЭМ конференции по эвентологической математике и смежным вопросам*, Красноярск: СФУ (под ред. Олега Воробьева):130–135, 2012.



# Расчет числовых характеристик однородной GERT-сети на основе эквивалентных преобразований

Антон Александрович Зырянов

Сибирский государственный технологический университет  
Кафедра системотехники  
Красноярск  
ZyryanovAntonA@gmail.com

**Аннотация.** Рассматривается методика расчета числовых характеристик однородной GERT-сети, в частности математического ожидания и дисперсии. Методика включает алгоритм преобразования GERT-сети к эквивалентной дуге с пересчетом числовых характеристик случайных величин дуг по предложенным формулам.

**Ключевые слова.** GERT, стохастические сети, бизнес-процесс, эквивалентное преобразование.

## 1 Введение

За последние годы для моделирования и оптимизации технических систем все большее распространение получают альтернативные стохастические сети [1], в частности математический аппарат GERT-сетей (GERT – graphical evaluation and review technique).

Подробное описание GERT-сетей представлено в работах Филлипса [2], Neumann [3], Pritsker [4]. В отечественной литературе наибольших научных результатов в развитии аппарата GERT-сетей достиг А.П. Шибанов [5].

Математический аппарат GERT-сетей является одним из инструментов исследования различных классов систем. В настоящее время известны применения аппарата GERT-сетей для исследования вероятностно-временных характеристик локальных сетей, сетей передачи данных, телекоммуникационных систем, моделирования систем обработки изображений при дистанционном зондировании, моделирования распределенных систем обработки информации, оптимизации производственных процессов, моделирования мультиверсионных программных систем. Автор применяет аппарат GERT-сетей для моделирования системы бизнес-процессов [6], [7].

Основные методы расчета GERT-сетей описаны в работах [2], [5], [3], [4]. Данные методы позволяют: преобразовать сеть к эквивалентной дуге; рассчитать вероятность стока сети; определить значения переменных, связанных с моментами распределения выходной величины сети (математическое ожидание, дис-

персия и др.); построить функцию распределения выходной величины сети.

Однако серьезным сдерживающим фактором применения GERT-сетей является экспоненциальная трудоемкость алгоритмов расчета, основанных на использовании топологического уравнения Мейсона (что характерно для классической теории GERT-сетей). При решении данной задачи известными методами возникает проблема вычислительной сложности – необходимо находить петли сети вплоть до  $r$ -ого порядка (число слагаемых в уравнении Мейсона пропорционально числу всевозможных комбинаций петель первого порядка, не имеющих общих вершин), необходимо вычислять частные производные по производящей функции моментов GERT-сети и т.д.

Таким образом, актуальной является задача нахождения числовых характеристик центральных моментов (математического ожидания и дисперсии) случайной величины стока GERT-сети на основе числовых характеристик случайных величин дуг сети минуя решение топологического уравнения Мейсона и вычисления частных производных по производящей функции моментов сети.

## 2 Расчет числовых характеристик GERT-сети

Решение поставленной задачи нахождения числовых характеристик GERT-сети без использования топологического уравнения Мейсона [2], [4] и вычисления частных производных основывается на преобразовании GERT-сети к эквивалентной дуге с пересчетом числовых характеристик случайных величин дуг. Под числовыми характеристиками случайных величин будут пониматься математическое ожидание и дисперсия.

Для GERT-сети существует три основных эквивалентных преобразования дуг: последовательные дуги, параллельные дуги, дуга и петля. Приведем формулы нахождения числовых характеристик для всех трех преобразований в GERT-сети.

Последовательные дуги:

$$M[Y] = M[X_1] + M[X_2] \quad (1)$$

$$D[Y] = D[X_1] + D[X_2] \quad (2)$$

Параллельные дуги:

$$M[Y] = \frac{p_1 M[X_1] + p_2 M[X_2]}{p_1 + p_2} \quad (3)$$

$$D[Y] = \frac{p_1 D[X_1] + p_2 D[X_2]}{p_1 + p_2} + \frac{p_1 p_2 (M[X_1] - M[X_2])^2}{(p_1 + p_2)^2} \quad (4)$$

Дуга и петля:

$$M[Y] = M[X_1] + \frac{p_2}{1 - p_2} M[X_2] \quad (5)$$

$$D[Y] = D[X_1] + \frac{p_2}{1 - p_2} D[X_2] + \frac{p_2}{(p_2 - 1)^2} (M[X_2])^2 \quad (6)$$

$X_1$  и  $X_2$  – случайные величины дуг GERT-сети,  $Y$  – случайная величина стока GERT-сети,  $p_1$  и  $p_2$  – вероятности выполнения соответственно дуг  $X_1$  и  $X_2$ .  $M[Y]$  – математическое ожидание,  $D[Y]$  – дисперсия эквивалентной дуги  $Y$ .

Формулы для параллельных дуг (3, 4) и для дуги и петли (5, 6) приведены для обобщенного случая, когда из первого узла GERT-сети может выходить более двух дуг, то есть  $p_1 + p_2 \neq 1$ .

Данные формулы получены автором путем аналитического вывода на основе уравнения Мейсона – аналитически вычислены частные производные по производящей функции моментов GERT-сети. Формулы (1-6) получены и аналогичны как для непрерывных распределений, так и для дискретных.

Таким образом, получены формулы нахождения числовых характеристик GERT-сети без использования топологического уравнения Мейсона. Аналогично можно вывести формулы для расчета других моментов GERT-сети – коэффициент асимметрии и коэффициент эксцесса.

### 3 Алгоритм преобразования GERT-сети к эквивалентной дуге

Алгоритм преобразования однородной GERT-сети к эквивалентной дуге основан на поиске участков сети с тремя типами расположения дуг: последовательные дуги, параллельные дуги, дуга и петля, и их замене на эквивалентную дугу с пересчетом математического ожидания и дисперсии, а также вероятности выполнения дуги.

Пусть GERT-сеть имеет один источник или приведена к сети с одним источником, и имеет  $m$  стоков,

тогда алгоритм преобразования однородной GERT-сети должен позволять получать  $m$  эквивалентных дуг с числовыми характеристиками по каждому стоку сети.

Структуру GERT-сети  $G$  отобразим в матрице смежности  $A(G)$  с  $n$  вершинами. Каждый элемент матрицы  $a_{ij}$ , для которого  $a_{ij}$ , содержит вероятность  $p_{ij}$  выполнения дуги  $\langle i, j \rangle$  ( $i$  – номер строки,  $j$  – номер столбца,  $i, j = \overline{1, n}$ ) и значения математического ожидания  $M[X_{ij}]$  и дисперсии  $D[X_{ij}]$ .

Поскольку при преобразовании (топологическом упрощении) GERT-сети за один шаг могут вновь образовываться некоторые из типов расположения дуг (например, после преобразования петли могут образоваться последовательные дуги), то алгоритм является итеративным и включает следующие шаги.

Шаг 1. Ищутся простые петли GERT-сети – дуга  $\langle v_k, v_k \rangle$ . Поиск осуществляется по диагонали матрицы  $A_l(G_l)$  путем нахождения элементов  $a_{kk}^l = 1$ , для которых  $i = j$ , где  $l$  – шаг алгоритма,  $l \in \overline{1, N}$  (для первого выполнения Шага 1  $l = 1$ ). Далее находятся все дуги  $\langle v_k, v_j \rangle$ ,  $j = \overline{1, n}$  (элемент  $a_{kj}^1 = 1$ ), и по формулам (5, 6) пересчитывается математическое ожидание  $M[X_{kj}]$  и дисперсия  $D[X_{kj}]$  данных дуг. Вероятность  $p_{kj}^1$  вычисляется по следующей формуле

$$p_{kj}^1 = \frac{p_{kj}}{1 - p_{kk}}$$

После этого дуга  $\langle v_k, v_k \rangle$  исключается из GERT-сети, а соответствующий элемент матрицы  $A_1(G_1)$  становится равным нулю ( $a_{kk}^1 = 0$ ).

После нахождения всех простых петель и их исключения, получим эквивалентную GERT-сеть  $G_2$  и соответствующую матрицу  $A_2(G_2)$ .

На данном шаге ищутся все простые петли GERT-сети и исключаются с пересчетом числовых характеристик соответствующих дуг, тем самым уменьшается топологическая сложность GERT-сети, поскольку уменьшается количество петель  $r$ -ых порядков.

Шаг 2. Ищутся последовательные дуги GERT-сети. Последовательными дугами являются дуги  $\langle v_i, v_j \rangle$  и  $\langle v_j, v_k \rangle$ , для которых  $i, j, k = \overline{1, n}$ ,  $i \neq j \neq k$ , при этом узел  $v_j$  должен иметь только одну входящую дугу из узла  $v_i$ , и только одну исходящую дугу в узел  $v_k$ . То есть элементы матрицы  $A_2(G_2)$   $a_{ij}^2 = 1$ ,  $a_{jk}^2 = 1$ , при этом  $a_{qj}^2 = 0$  и  $a_{jt}^2 = 0$ , где  $q, t = \overline{1, n}$  и  $q \neq i$ ,  $t \neq k$ .

По формулам (1, 2) пересчитывается математическое ожидание  $M[X_{ik}]$  и дисперсия  $D[X_{ik}]$  эквивалентной дуги  $\langle v_i, v_k \rangle$ . Вероятность  $p_{ik}$  вычисляется по следующей формуле

$$p_{ik} = p_{ij} p_{jk}$$

Элемент матрицы  $A_2(G_2)$  становится равным  $a_{ik}^2 = 1$ . Узел  $v_j$  исключается из GERT-сети, в матрице

$A_2(G_2)$  исключается строка  $j$  и столбец  $j$ , тем самым, размерность матрицы  $A_2(G_2)$  уменьшается на единицу.

При преобразовании последовательных дуг GERT-сети могут образовываться участки сети с параллельными дугами. Если при преобразовании дуг  $\langle v_i, v_j \rangle$  и  $\langle v_j, v_k \rangle$  элемент матрицы уже равен единице ( $a_{ik}^2 = 1$ ), то значения полученных числовых характеристик ( $a_{ik}^{2*}$ ) необходимо пересчитать по формулам (3, 4) с характеристиками имеющейся параллельной дуги ( $a_{ik}^2$ ) и заменить их одной эквивалентной. Вероятность  $p_{ik}$  вычисляется по формуле

$$p_{ik} = p_{ik1} + p_{ik2}$$

После нахождения всех последовательных и параллельных дуг и их замене на эквивалентные дуги, получим эквивалентную GERT-сеть  $G_3$  и соответствующую матрицу  $A_3(G_3)$ .

Условие 1. Если количество узлов GERT-сети равно количеству стоков плюс единица, то GERT-сеть преобразована к эквивалентной сети. Если размерность матрицы  $A_l(G_l)$  равна  $n = m + 1$ , где  $m$  – количество стоков, то алгоритм заканчивается.

Условие 2. Если размерность матрицы  $A_l(G_l)$  уменьшилась после выполнения Шага 2 ( $n_l > n_{l+1}$ ), то необходимо проверить GERT-сеть на появление простых петель – дуги типа  $\langle v_k, v_k \rangle$  (переход к Шагу 1), иначе в GERT-сети существует более сложные комбинации из последовательных и параллельных дуг (переход к Шагу 3).

Шаг 3. Ищется узел, в который входит одна дуга, а выходит две или более дуги. Данный узел дублируется по количеству выходящих дуг с копированием входной дуги на каждый продублированный узел так, чтобы из каждой копии узла выходила бы одна дуга.

Узел  $v_j$  ищется подобно тому, как ищутся последовательные дуги  $\langle v_i, v_j \rangle$  и  $\langle v_j, v_k \rangle$  на Шаге 2, при этом  $a_{qj}^3 = 0$ , где  $q = \overline{1, n}$  и  $q \neq i$ , а количество элементов  $a_{jt}^3 = 1$  больше одного, где  $t = \overline{1, n}$ ,  $t \neq j$ . Для каждой дуги  $\langle v_j, v_t \rangle$  создается узел  $v_j^t$ , в который входит дуга  $\langle v_i, v_j^t \rangle$ , и выходит дуга  $\langle v_j^t, v_t \rangle$ . Соответствующие элементы матрицы  $A_3(G_3)$  устанавливаются равными единице  $a_{ijt}^3 = 1$ ,  $a_{jt}^3 = 1$ . Вероятность дуги  $\langle v_i, v_j^t \rangle$  равняется произведению вероятностей дуг  $\langle v_i, v_j \rangle$  и  $\langle v_j, v_t \rangle$ , при этом ее числовые характеристики равны характеристикам дуги  $\langle v_i, v_j \rangle$  ( $a_{ijt}^3 = 1$ ). Вероятность дуги  $\langle v_j^t, v_t \rangle$  равна 1, при этом ее числовые характеристики равны характеристикам дуги  $\langle v_j, v_t \rangle$  ( $a_{jt}^3 = 1$ ).

Узел  $v_j$  исключается из GERT-сети, в матрице  $A_3(G_3)$  исключается строка  $j$  и столбец  $j$ .

Размерность матрицы  $A_3(G_3)$  увеличивается на дан-

ном шаге. Поскольку появляются последовательные дуги, переходим к Шагу 2.

Условие 3. Если на шаге 3 не найден ни один узел, в который входит одна дуга, а выходит две или более дуги, то переходим к Шагу 4, иначе, поскольку появляются последовательные дуги, переходим к Шагу 2.

Шаг 4. Ищется узел, в который входит и выходит более чем по одной дуге. Данный узел дублируется по количеству входящих дуг с копированием выходных дуг на каждый продублированный узел так, чтобы в каждую копию узла входила бы только одна дуга.

Ищется узел  $v_j$ , для которого количество элементов  $a_{qj}^4 = 1$  и  $a_{jt}^4 = 1$  больше одного, где  $q = \overline{1, n}$  и  $q \neq j$ ,  $t = \overline{1, n}$  и  $t \neq j$ . Для каждой дуги  $\langle v_q, v_j \rangle$  создается узел  $v_j^q$ , в который входит дуга  $\langle v_q, v_j^q \rangle$  и выходят дуги  $\langle v_j^q, v_{j+1} \rangle, \dots, \langle v_j^q, v_t \rangle$ . Соответствующие элементы матрицы  $A_4(G_4)$  устанавливаются равными единице  $a_{qj}^4 = 1$ ,  $a_{j+1}^4 = 1, \dots, a_{jt}^4 = 1$ . Вероятность и числовые характеристики дуги  $\langle v_q, v_j^q \rangle$  берутся из дуги  $\langle v_q, v_j \rangle$ . Вероятность и числовые характеристики дуг  $\langle v_j^q, v_{j+1} \rangle, \dots, \langle v_j^q, v_t \rangle$  берутся из соответствующих дуг  $\langle v_j, v_{j+1} \rangle, \dots, \langle v_j, v_t \rangle$ .

Узел  $v_j$  исключается из GERT-сети, в матрице  $A_4(G_4)$  исключается строка  $j$  и столбец  $j$ . Поскольку на шаге 4 появляются узлы, в которые входит одна дуга, а выходит две или более дуги, то переходим к Шагу 3.

Таким образом, общий ход алгоритма преобразования однородной GERT-сети к эквивалентной дуге имеет следующий вид:

Шаг 1. Ищутся простые петли GERT-сети – дуга  $\langle v_k, v_k \rangle$ .

Шаг 2. Ищутся последовательные и параллельные дуги GERT-сети.

Условие 1. Если количество узлов GERT-сети равно количеству стоков плюс единица, то GERT-сеть преобразована к эквивалентной сети.

Условие 2. Если количество узлов GERT-сети уменьшилось после выполнения Шага 2, то необходимо проверить GERT-сеть на появление простых петель (переход к Шагу 1), иначе в GERT-сети существует более сложные комбинации из последовательных и параллельных дуг (переход к Шагу 3).

Шаг 3. Ищется узел, в который входит одна дуга, а выходит две или более дуги. Узел дублируется по количеству выходящих дуг с копированием входной дуги на каждый продублированный узел так, чтобы из каждой копии узла выходила бы одна дуга.

Условие 3. Если на шаге 3 не найден ни один узел, в который входит одна дуга, а выходит две или более дуги, то переходим к Шагу 4, иначе, поскольку появляются последовательные дуги, переходим к Шагу 2.

Шаг 4. Ищется узел, в который входит и выходит более чем по одной дуге. Узел дублируется по количеству входящих дуг с копированием выходных дуг на каждый продублированный узел так, чтобы в каждую копию узла входила бы только одна дуга. Переходим к Шагу 3.

Таким образом, предложен алгоритм преобразования однородной GERT-сети к эквивалентной дуге с помощью разработанных формул (1-6) пересчета математического ожидания и дисперсии. Кроме того, преобразовывая GERT-сеть на основе данного алгоритма, используя в формулах числовые характеристики дуг, можно получить аналитические формулы числовых характеристик GERT-сети.

Следует отметить, что описанный в данной статье итеративный алгоритм отличается от решения, предложенного А.П. Шибановым [8] тем, что адекватнее описывает последовательность эквивалентных преобразований, операции эквивалентных преобразований сгруппированы по шагам (шаги 1, 2, 3, 4), описаны условия (условие 2, 3) перехода к шагам 1, 3, 4 и условие окончания хода алгоритма (условие 1). Кроме того, в предложенном алгоритме описаны критерии поиска последовательных, параллельных дуг, дуги и петли по матрице смежности, также алгоритм может быть применен к GERT-сети с  $m$  стоками.

#### 4 Пример расчета числовых характеристик GERT-сети

В качестве примера рассмотрен бизнес-процесс дистанционного обучающего курса "Дискретная математика". Данный курс разработан в системе дистанционного обучения Moodle.

Курс состоит из 7 лекций с проверочными работами, трех тестов по модулям, в конце курса необходимо пройти итоговый тест.

На основе алгоритма трансляции модели бизнес-процессов в нотации ARIS eEPC в GERT-сеть [7] была получена GERT-сеть бизнес-процесса дистанционного обучающего курса, представленная на Рис. 1. Исследуемыми характеристиками бизнес-процесса являются время выполнения заданий и общее время выполнения курса обучаемым в минутах.

В таблице 1 представлено сопоставление операций бизнес-процесса и узлов GERT-сети.

В таблице 2 представлены параметры данной GERT-сети.

GERT-сеть, представленная на Рис. 1, была рассчитана на основе предложенной методики. Для проверки адекватности разработанной методики оценки числовых характеристик данная GERT-сеть была также рассчитана на основе уравнения Мейсона и с помощью имитационного эксперимента. В табли-

Узел GERT-сети	Операция
$v_1$	—
$v_2$	Лекция 1
$v_3$	Лекция 2
$v_4$	Лекция 3
$v_5$	Тест модуль 1
$v_6$	Лекция 4
$v_7$	Лекция 5
$v_8$	Тест модуль 2
$v_9$	Лекция 6
$v_{10}$	Лекция 7
$v_{11}$	Тест модуль 3
$v_{12}$	Итоговый тест
$v_{13}$	—

Таблица 1: Параметры GERT-сети.

Дуга $\langle i, j \rangle$	$p_{ij}$	$M[X_{ij}]$	$D[X_{ij}]$	Закон распред.
$\langle v_1, v_2 \rangle$	0,51	0	0	—
$\langle v_1, v_3 \rangle$	0,07	0	0	—
$\langle v_1, v_5 \rangle$	0,36	0	0	—
$\langle v_1, v_8 \rangle$	0,06	0	0	—
$\langle v_2, v_2 \rangle$	0,22	12,2	148,6	экспон.
$\langle v_2, v_3 \rangle$	0,52	12,2	148,6	экспон.
$\langle v_2, v_4 \rangle$	0,11	12,2	148,6	экспон.
$\langle v_2, v_5 \rangle$	0,15	12,2	148,6	экспон.
$\langle v_3, v_3 \rangle$	0,24	25	4635,9	логнорм.
$\langle v_3, v_4 \rangle$	0,65	25	4635,9	логнорм.
$\langle v_3, v_5 \rangle$	0,11	25	4635,9	логнорм.
$\langle v_4, v_4 \rangle$	0,16	27	730,9	экспон.
$\langle v_4, v_5 \rangle$	0,3	27	730,9	экспон.
$\langle v_4, v_6 \rangle$	0,45	27	730,9	экспон.
$\langle v_4, v_7 \rangle$	0,09	27	730,9	экспон.
$\langle v_5, v_5 \rangle$	0,29	99,77	146285	логнорм.
$\langle v_5, v_6 \rangle$	0,21	99,77	146285	логнорм.
$\langle v_5, v_8 \rangle$	0,5	99,77	146285	логнорм.
$\langle v_6, v_6 \rangle$	0,17	12,88	166,1	экспон.
$\langle v_6, v_7 \rangle$	0,83	12,88	166,1	экспон.
$\langle v_7, v_7 \rangle$	0,18	16,3	1738,6	логнорм.
$\langle v_7, v_8 \rangle$	0,29	16,3	1738,6	логнорм.
$\langle v_7, v_9 \rangle$	0,47	16,3	1738,6	логнорм.
$\langle v_7, v_{10} \rangle$	0,06	16,3	1738,6	логнорм.
$\langle v_8, v_8 \rangle$	0,28	18,9	1329,1	логнорм.
$\langle v_8, v_9 \rangle$	0,27	18,9	1329,1	логнорм.
$\langle v_8, v_{11} \rangle$	0,45	18,9	1329,1	логнорм.
$\langle v_9, v_9 \rangle$	0,16	15,4	262,4	норм.
$\langle v_9, v_{10} \rangle$	0,84	15,4	262,4	норм.
$\langle v_{10}, v_{10} \rangle$	0,39	12,1	146,4	экспон.
$\langle v_{10}, v_{11} \rangle$	0,35	12,1	146,4	экспон.
$\langle v_{10}, v_{12} \rangle$	0,26	12,1	146,4	экспон.
$\langle v_{11}, v_{11} \rangle$	0,21	92,6	320629	логнорм.
$\langle v_{11}, v_{12} \rangle$	0,79	92,6	320629	логнорм.
$\langle v_{12}, v_{13} \rangle$	1	72,5	31871	логнорм.

Таблица 2: Параметры GERT-сети.



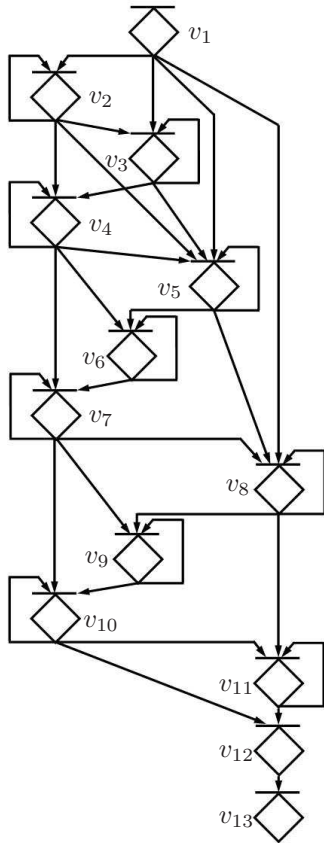


Рис. 1: GERT-сеть бизнес-процесса дистанционного обучающего курса.

Метод расчета	Мат. ож.	Диспер.
Расчет числовых характ. на основе эквив. преобраз.	345,69	495211
Уравнение Мейсона	346,35	496243
Имитац. эксперимент	346,86	504979

Таблица 3: Результаты расчетов GERT-сети по трудоемкости выполнения заданий.

це 3 представлены результаты расчетов GERT-сети для времени выполнения заданий. В таблице 4 представлены результаты расчетов GERT-сети для суммарного времени выполнения курса, включая время ожидания между выполнением заданий.

Таким образом, на основе результатов расчетов (Таблица 3, 4), можно сделать вывод об адекватности полученных формул нахождения числовых характеристик GERT-сети и алгоритма эквивалентных преобразований, по сравнению с известными методами.

Предложенная методика имеет ряд преимуществ по сравнению с имитационным экспериментом – меньшая вычислительная трудоемкость и получение более точного результата, не зависящего от выполнения эксперимента.

Метод расчета	Мат. ож.	Диспер.
Расчет числовых характ. на основе эквив. преобраз.	3150,5	9773561
Уравнение Мейсона	3151,3	9779321
Имитац. эксперимент	3150,7	9765187

Таблица 4: Результаты расчетов GERT-сети по общему времени выполнения курса.

## 5 Заключение

Таким образом, предложена методика оценки числовых характеристик однородной GERT-сети, в частности математического ожидания и дисперсии, основанная на преобразовании GERT-сети к эквивалентной дуге с пересчетом числовых характеристик случайных величин дуг по предложенным формулам. Предложен итеративный алгоритм эквивалентных преобразований GERT-сети. Поскольку топологическое уравнение Мейсона не используется, отсутствует необходимость искать петли GERT-сети и вычислять частные производные для нахождения числовых характеристик.

## Список литературы

- [1] Golenko-Ginzburg D. *Stochastic network models in innovative projecting*. Science Book Publishing House, Voronezh, 2011.
- [2] Филиппс Д., Гарсиа-Диас А. *Методы анализа сетей*. Мир, М., 1984.
- [3] Neumann K. *Stochastic project networks: temporal analysis, scheduling and cost minimization*. Springer-Verlag, Berlin, 1990.
- [4] Pritsker A. A. B. *GERT: graphical evaluation and review technique*. RAND Corporation, 1966.
- [5] А. П. Шибанов. *Обобщенные GERT-сети для моделирования протоколов, алгоритмов и программ телекоммуникационных систем: диссертация д.т.н.: 05.13.13*. Рязань, 2003.
- [6] А. А. Зырянов, М. Г. Доррер. Прогноз динамики событийных моделей бизнес-процессов на основе gert-сетей. *Информатизация и связь*, 7:124–127, 2012.
- [7] А. А. Зырянов, М. Г. Доррер. Трансляция модели бизнес-процессов в нотации aris eerp в gert-сеть. *Труды XI Международной ФАМ конференции по финансово-актуарной математике и смежным вопросам*, Красноярск: СФУ:186–192, 2012.
- [8] А. П. Шибанов. Нахождение плотности распределения времени исполнения gert-сети на основе эквивалентных упрощающих преобразований. *Автоматика и телемеханика*, 2:117–126, 2003.
- [9] Е. С. Вентцель, Л. А. Овчаров. *Теория вероятностей и ее инженерные приложения*. Академия, М., 2003.



# Решение задачи скоринга на основе статистик разнотипных данных

**Жаткина Кристина Владимировна**

Сибирский Федеральный Университет  
Институт математики и фундаментальной информатики  
Красноярск  
Kristina-gut@yandex.ru

**Баранова Ирина Владимировна**

Сибирский Федеральный Университет  
Институт математики и фундаментальной информатики  
Красноярск  
irinabar@yandex.ru

**Аннотация.** В работе рассматривается применение метода двудольных множеств случайных событий для решения задачи эвентологического скоринга. С помощью данного метода решается практическая задача определения характеристик заемщиков, благоприятствующих возврату кредита.

**Ключевые слова.** Скоринговое кредитование, рейтинговый счет, кредитоспособность, многособытийные задачи

## 1 Введение

Выдача крупных кредитов в банковском бизнесе сопровождается предварительной процедурой, которая называется скоринг — определение рейтингового счета клиента. Скоринг представляет собой математическую или статистическую модель, с помощью которой на основе кредитной истории предыдущих клиентов банк пытается определить, насколько велика вероятность, что конкретный потенциальный заемщик вернет кредит в срок. Также в результате решения задачи скоринга выделяются характеристики, которые наиболее тесно связаны с ненадежностью или, наоборот, с надежностью клиента. Скоринг применяется также в маркетинге, менеджменте и других отраслях. В работе изучается метод эвентологического скоринга, предложенный О.Ю. Воробьевым и Е.Е. Голденком [1, 2]. Он позволяет для любого результата произвольного анкетирования получить оценку вероятности наступления события, при условии, что влияющие на него события наступили в определенной известной комбинации. За каждый ответ на вопрос анкеты начисляются баллы, взвешенная сумма которых образует рейтинговый счет, по величине которого судят о вероятности наступления целевого события. В методе эвентологического скоринга вероятности целевого события находятся на основе статистики, состоящей из вопросов с бинарной шкалой ответов.

На практике довольно часто возникает ситуация, в которой ответы на часть вопросов анкеты состоят из одного значения и представляют собой число,

часть других вопросов допускает в ответе перечисление нескольких вариантов. Для работы с подобными статистиками, состоящими из разнотипных данных, И.В. Барановой в работе [3] был предложен метод двудольных множеств случайных событий, который рассматривает статистику как двудольное множество событий, первая доля которого определяется случайными величинами, а вторая — случайными множествами. В дипломной работе Е.С. Родкевич «Применение методов двудольных множеств событий и эвентологического скоринга к статистике актирования» (2009 г.) было предложено объединить методы эвентологического скоринга и двудольных множеств событий, для того, чтобы решать задачи скоринга на основе разнотипной статистики анкетирования.

Целью данной работы является решение практической задачи определения характеристик заемщиков, благоприятствующих возврату кредита. Задача основана на статистике показателей заемщиков кредитов, в которой часть показателей является числовыми, а часть — множественными. Вначале согласно методу двудольного эвентологического скоринга находится вероятность наступления целевого события — возврата кредита клиентами банка. Затем производится анализ влияния различных значений показателей на вероятность наступления целевого события и сравнение полученных результатов для определения событий благоприятствующих увеличению вероятности возврата кредита.

## 2 Двудольное множество случайных элементов и двудольное множество случайных событий

В ситуации, когда поведение каждого элемента сложной системы характеризуется данными, одна часть которых — числовая, а другая — множественная, объект, порождающий данную статистику, может быть представлен как объединение двух долей: случайных величин и случайных множеств событий.

Пусть  $(\Omega, \mathcal{F}, \mathbf{P})$  — вероятностное пространство, где  $\mathcal{F}$  — алгебра событий, случайные события  $x, y, \dots \in \mathcal{F}$  — элементы алгебры  $\mathcal{F}$ , на которой определена вероятность  $\mathbf{P}$ . Множество избранных случайных собы-

тий  $\mathfrak{X} \in \mathcal{F}$ .

**Определение 43.** *Случайное множество событий под  $\mathfrak{X}$  определяется как измеримое отображение*

$$K : (\Omega, \mathcal{F}, \mathbf{P}) \rightarrow (2^{\mathfrak{X}}, 2^{2^{\mathfrak{X}}}),$$

где  $2^{\mathfrak{X}}$  — множество всех подмножеств множества  $\mathfrak{X}$ .

В методе было предложено случайную величину и случайное конечное множество событий рассматривать как реализацию случайного элемента, и объект, порождающий статистику, — рассматривать как двудольное множество случайных элементов.

**Определение 44.** *Двудольным множеством случайных элементов называется множество  $\{\xi, \mathbf{K}\}$ , представимое в следующем виде:*

$$\{\xi, \mathbf{K}\} = \xi \cup \mathbf{K} = \{\xi_a, a \in A, K_\beta, \beta \in B\}, \quad (1)$$

здесь первая доля — это случайные величины  $\xi = \{\xi_a, a \in A\}$ , вторая — случайные множества событий  $\mathbf{K} = \{K_\beta, \beta \in B\}$ ,  $A$  — множество индексов случайных величин,  $B$  — множество индексов случайных множеств событий.

События, которые порождаются двудольным множеством случайных элементов, будем называть множеством случайных событий.

**Определение 45.** *Двудольное множество случайных событий представляет собой объединение двух множеств — множества событий, которое определяется случайными величинами, и множества событий, которое определяется случайными множествами событий:*

$$\{\mathcal{Y}, \mathfrak{X}\} = \{\mathcal{Y}_a, \mathfrak{X}_\beta, a \in A, \beta \in B\}. \quad (2)$$

## 2.1 Двудольная эвентологическая модель сложных систем

Сложная система представляется как совокупность большого числа элементов, обладающая сложной структурой зависимостей между ними. Поэтому поведение всей системы определяется поведением каждого его элемента.

В работе [3] было введено понятие двудольной эвентологической модели сложной системы.

**Определение 46.** *Двудольной эвентологической моделью сложной системы будем называть такую систему, для которой поведение каждого элемента характеризуется двудольным множеством случайных событий  $\{\mathcal{Y}, \mathfrak{X}\}$ , первая доля  $\mathcal{Y}$  определяется случайными величинами  $\xi$ , а вторая доля  $\mathfrak{X}$  — случайными множествами событий.*

Интерпретация множеств событий  $\mathcal{Y}$  и  $\mathfrak{X}$  в различных прикладных областях может быть самой разной.

## 3 Метод эвентологического скоринга

**Определение 47.** *Метод эвентологического скоринга — это метод исчисления по известным вероятностям событий  $P(x)$ ,  $x \in \mathfrak{X}$ , и известными условными вероятностям*

$$P(s | x) = \frac{P(x \cap s)}{P(x)}, \quad P(s^c | x) = \frac{P(x \cap s^c)}{P(x)}, \quad x \in \mathfrak{X}$$

*неизвестных величин*

$$P(s | t_s(X)) = \frac{P(s \cap t_s(X))}{P(t_s(X))}, \quad X \subseteq \mathfrak{X}$$

— условных вероятностей случайного события  $s \in \mathcal{F}$  при условии, что наступило случайное событие  $t_s(X) = s \cap \text{Ter}_X + s^c \cap \text{Ter}_{X^c}$ ,  $X \subseteq \mathfrak{X}$ .

Причем события  $x \in \mathfrak{X}$  называются базовыми случайными событиями, множество  $\mathfrak{X}$  — множеством базовых случайных событий, событие  $s$  — целевым случайным событием, а события  $t_s(X)$ ,  $X \subseteq \mathfrak{X}$  — интерпретирующими случайными событиями.

В скоринге за каждый ответ на вопрос анкеты начисляются баллы, взвешенная сумма которых образует рейтинговый счет (score). По величине этого рейтинга судят (сравнивая с рейтингами заранее произошедших событий) о вероятности наступления целевого события. Все далее рассматриваемые события считаются измеримыми относительно алгебры  $\mathcal{F}$ .

В [3] доказано, что формула нахождения условных вероятностей случайного события  $s$  в методе эвентологического скоринга имеет вид

$$\mathbf{P}(s | t_s(X)) = \frac{\mathbf{P}\left(\bigcup_{x \in X} (x \cap s)\right)}{\mathbf{P}\left(\bigcup_{x \in X} (x \cap s)\right) + \mathbf{P}\left(\bigcup_{x \in X^c} (x \cap s^c)\right)},$$

$$X \subseteq \mathfrak{X}.$$

Условные вероятности события  $s$  для множества базовых событий  $\mathfrak{X}$  с  $s$ -независимой структурой вычисляется по формулам

$$\mathbf{P}(s | t_s(X)) = \frac{1 - \prod_{x \in X} (\mathbf{P}^c(x \cap s))}{1 - \prod_{x \in X} (\mathbf{P}^c(x \cap s)) + 1 - \prod_{x \in X^c} (\mathbf{P}^c(x \cap s^c))},$$

где  $X \subseteq \mathfrak{X}$ ,  $\mathbf{P}^c(x \cap s) = 1 - \mathbf{P}(x \cap s)$ ,  $\mathbf{P}^c(x \cap s^c) = 1 - \mathbf{P}(x \cap s^c)$ .

## 4 Применение метода двудольных множеств событий к эвентологическому скорингу

Пусть респонденты отвечают на вопросы анкеты. Ответы на ряд вопросов состоят из одного значения и представляют собой число (либо могут быть к нему сведены). Например, такие вопросы: «Укажите ваш возраст», «Размер вашей заработной платы» и т.д. Некоторые вопросы допускают в ответе перечисление сразу нескольких вариантов: «Какой собственно-

стью вы владеете» и т.д. Очевидно, что данная статистика является примером статистики двудольного множества событий, порожденной двудольным множеством случайных величин.

В работе Родкевич Е.С. было введено обобщение метода эвентологического скоринга, представляющее собой объединение метода двудольных множеств событий и эвентологического скоринга.

Пусть  $D$  — двудольное множество событий вариантов, содержащее все возможные варианты ответов  $d$  на вопросы анкеты.  $D_x \subseteq D$  — множество событий-вариантов вопроса  $x$ , содержащее все возможные варианты ответа на вопрос  $x$ .  $D_x^+ \subseteq D$  — множество событий-вариантов вопроса  $x$ , благоприятствующих наступлению события, содержащее варианты ответа на вопрос  $x$ , которые благоприятствуют наступлению целевого события.  $D_x^- \subseteq D$  — соответственно благоприятствующих наступлению события  $s^c$ . Результат анкетирования обозначим как  $t_s$ .

Введем еще два обозначения:

$$U_x^+(d \cap s) = \begin{cases} \max_{d \in D_x^+} (d \cap s), \\ \Omega - \bigcap_{d \in D_x^+} (\Omega - (d \cap s)). \end{cases}$$

Соответственно, для случая если события-варианты вложенные, и если события-варианты  $s$ -независимые. И, аналогично, для ситуации с неблагоприятными ответами.

$$U_x^-(d \cap s^c) = \begin{cases} \max_{d \in D_x^-} (d \cap s^c), \\ \Omega - \bigcap_{d \in D_x^-} (\Omega - (d \cap s^c)), \end{cases} \quad (3)$$

Т.е.  $U_x^+$  — анкетные события, ответы на которые в той или иной степени благоприятствуют наступлению целевого события, а  $U_x^-$  — не благоприятствуют. Каждое из этих подмножеств также имеет подмножества, различающиеся структурой вопроса и множества базовых событий.

Так как  $D$  — двудольное множество,

$$D = \mathbf{X} \cup \mathbf{K} = \{\xi_a, a \in A, K_\beta, \beta \in B\},$$

где  $A$  — множество индексов случайных величин,  $B$  — множество индексов случайных множеств. Множество  $D$  имеет 4 подмножества:

$$D_\alpha^+ = \{\xi_a^+, a \in A, \},$$

$$D_\beta^+ = \{K_\beta^+, \beta \in B\},$$

$$D_\alpha^- = \{\xi_a^-, a \in A, \},$$

$$D_\beta^- = \{K_\beta^-, \beta \in B\}.$$

Формула нахождения условных вероятностей в методе двудольного эвентологического скоринга (т.е. для множеств событий с рассмотренной двудольной структурой) имеет следующий вид:

$$\mathbf{P}(s | t_s) = \mathbf{P}_\alpha(s | t_s) \cdot \mathbf{P}_\beta(s | t_s), \text{ где} \quad (4)$$

$$\mathbf{P}_\alpha(s | t_s) = \frac{\sum_{x \in \mathbf{X}} \max_{d \in D_\alpha^+} \mathbf{P}(d \cap s)}{\sum_{x \in \mathbf{X}} \max_{d \in D_\alpha^+} \mathbf{P}(d \cap s) + \sum_{x \in \mathbf{X}} \max_{d \in D_\alpha^-} \mathbf{P}(d \cap s^c)},$$

$$\mathbf{P}_\beta(s | t_s) = \frac{\sum_{x \in \mathbf{X}} \mathbf{P} \left( \Omega - \bigcap_{d \in D_\beta^+} (\Omega - (d \cap s)) \right)}{\sum_{x \in \mathbf{X}} \mathbf{P} \left( \Omega - \bigcap_{d \in D_\beta^+} (\Omega - (d \cap s)) \right) + \sum_{x \in \mathbf{X}} \mathbf{P} \left( \Omega - \bigcap_{d \in D_\beta^-} (\Omega - (d \cap s^c)) \right)}.$$

## 5 Практическая задача определения характеристик заемщиков, благоприятствующих возврату кредита

Решение задачи банковского скоринга осуществлялось на основе статистики показателей заемщиков кредитов, в которой часть показателей является числовыми, а часть — множественными. Всего в статистике были приведены показатели 35 заемщиков. Из них было 18 мужчин и 17 женщин. Всего в статистике рассматривались 7 показателей заемщиков: пол, возраст, заработная плата, социальный статус (рабочий или служащий), имеется ли уже какой оформленный кредит, собственности (квартира, машина или доля квартиры), наличие и количество детей.

Вначале согласно методу двудольного эвентологического скоринга (по формуле 4) была найдена вероятность наступления целевого события — возврата кредита клиентами банка:  $\mathbf{P}(s | t_s) = 0,26$ .

Результат решения прямой задачи показал, что из всех заемщиков банка, только четвертая часть сможет вернуть кредит.

### 5.1 Расчеты вероятности для различных событий-вариантов

Следующая серия расчетов проведена с целью исследования того, как на целевое событие влияют различные показатели события. Результаты исследования представлены в Таблице 1.

Показатель	Значение	Вероятность
Пол	Мужской	0.34
Пол	Женский	0.28
Возраст	(20-30)	0.17
Возраст	(31-40)	0.41
Возраст	(41-60)	0.32
Соц.статус	Рабочий	0.1
Соц.статус	Служащий	0.34
Кредит	Есть	0.35
Кредит	Нет	0.27

Таблица 1: Таблица вероятности различных показателей события

### Анализ полученных результатов

Обработка данных расчетов по вероятности возврата кредита показала, что в общем случае вероятность возврата кредита равна 0,26. Если рассматривать показатель пол, то вероятность возврата кредита у лиц мужского пола на 0,06 выше, чем у лиц женского пола. Для показателя — возраст имеем следующий результат: люди в возрасте от 31 до 40 лет вернут кредит с вероятностью равной 0,41. Это на 0,24 выше, чем для лиц в возрасте от 21 до 30 лет и на 0,09 выше, чем для лиц в возрасте от 41-60. Если сравнить эти два показателя, то возраст с вероятностью 0,07 больше влияет на благоприятный исход целевого события. Для социального статуса вероятность возврата кредита служащего на 0,24 выше, чем у рабочего. Для показателя кредит мы получили, что люди имеющие кредит с вероятностью 0,08 быстрее вернут кредит, чем лица у которого его нет. Такие результаты мы получили для нашей статистики ввиду того, что по совокупности остальных параметров (пола, возраста, социального статуса, наличие собственности, уровню зарплаты и детей) общий рейтинговый счет клиента имеющего кредит, оказался выше, чем у клиента не имеющего кредит. Но в общем случае эти выводы не абсолютны. Таким образом, мы выделили значения показателей, оказывающие наиболее благоприятное влияние на кредитоспособность клиента и, как следствие, на возможность возврата кредита. Исходя из рассматриваемой статистики клиентов следует, что наиболее привлекательным клиентом для банка (с точки зрения его кредитоспособности) будет служащий мужского пола в возрасте от 31-40 лет с наличием кредита.

## 6 Итоги

В работе на основе статистических данных решена практическая задача определения характеристик заемщиков, благоприятствующих возврату кредита. Задача решена при помощи математического аппарата эвентологии и метода двудольных множеств событий. Определена вероятность наступления целево-

го события — возврата кредита. Представлены результаты находений вероятности целевого события по различным показателям. Проведен анализ влияния этих показателей на вероятность наступления целевого события, при прочих равных условиях. Также проведено сравнение полученных результатов для определения характеристик заемщиков, благоприятствующих вероятности наступления целевого события. В результате на основании полученных данных было выявлено наилучшее сочетание значений показателей заемщика для поддержки принятия решения о выдаче кредита.

### Список литературы

- [1] О. Ю. Воробьев. *Эвентология*. СФУ, Красноярск, 2007.
- [2] О. Ю. Воробьев и Е. Е. Голденко. Эвентологический скоринг в экономике, социологии, праве и бизнесе. In *Труды III всероссийской конференции по финансово-актуарной математике и смежным вопросам*, volume т.1, pages 113–128. КрасГУ, ИВМ СО РАН, Красноярск, 2005.
- [3] О. Ю. Воробьев и И. В. Баранова. *Метод двудольных множеств событий в эвентологическом анализе сложных систем*. Ин-т естеств. и гуманитар. наук, Красноярск, 2007.

# О сходимости к экспоненциальному закону и дифференциальный аналог основной леммы теории критических процессов Гальтона-Ватсона

**Аъзам Абдурахимович Имомов**

Каршинский государственный университет  
Улица Кучабег, 17, 180100 Карши, Узбекистан  
imomov\_azam@mail.ru

**Шахло Гулямовна Мусурмонова**

студент Каршинского государственного университета  
Улица Кучабег, 17, 180100 Карши, Узбекистан  
musurmonova91@mail.ru

**Аннотация.** В настоящей заметке исследуются один критерий справедливости теоремы о сходимости к экспоненциальному закону и локально-дифференциальный аналог основной леммы теории критических процессов Гальтона-Ватсона. Обсуждается некоторые их применения.

**Ключевые слова.** Экспоненциальный закон; локально-дифференциальный аналог; процесс Гальтона-Ватсона; Q-процессы.

## 1 Сходимость к экспоненциальному закону

В настоящей заметке мы исследуем один критерий сходимости к экспоненциальному закону, основываясь на идеи работы [3] и [4]. Будет доказан также так называемый локально-дифференциальный аналог основной леммы теории критических ветвящихся процессов Гальтона-Ватсона (ГВП). Мы заметим полезность применения полученных результатов в исследованиях теории ГВП и Q-процессов.

Рассмотрим функцию распределения экспоненциального закона

$$\Gamma_{\alpha}(x) = \begin{cases} 1 - e^{-\alpha x}, & x \geq 0, \\ 0, & x < 0, \end{cases}$$

где  $\alpha > 0$ . Введем следующий класс преобразований Лапласа (ПЛ):

$$\Phi = \left\{ \psi(\theta) : |\psi'(\theta)| = \frac{1}{\alpha}, \theta > 0 \right\}.$$

В классе  $\Phi$  рассмотрим оператор

$$\Delta(\psi(\theta)) := \psi'(\theta) + \frac{1}{\alpha} \psi^2(\theta). \quad (1)$$

Поскольку ПЛ экспоненциального закона  $\psi_{\alpha}(\theta) := \int_0^{\infty} e^{-\theta x} d\Gamma_{\alpha}(x) = 1/[1 + \theta/\alpha]$ , то мы легко заметим, что

$$\Delta(\psi_{\alpha}(\theta)) = 0, \quad (2)$$

то есть оператор  $\Delta(*)$  "аннулирует" ПЛ распределения  $\Gamma_{\alpha}(x)$ .

**Теорема 1.** Пусть  $\{G_n(x), n \in \mathbb{N}\}$  ( $\mathbb{N} = 1, 2, \dots$ ),  $x \geq 0$ , есть последовательность функций распределений и соответствующие им ПЛ  $\psi_n(\theta)$  принадлежат классу  $\Phi$ . Для того, чтобы при  $n \rightarrow \infty$

$$\sup_x |G_n(x) - \Gamma_{\alpha}(x)| \rightarrow 0, \quad (3)$$

необходимо и достаточно, чтобы

$$\sup_{\theta \in \Theta} |\Delta(\psi_n(\theta))| \rightarrow 0, \quad n \rightarrow \infty, \quad (4)$$

при любом  $\Theta > 0$ .

**Доказательство.** Рассуждение по части необходимости условия (4) основывается на свойствах ПЛ. Из (1) и (2) следует, что

$$\begin{aligned} |\Delta(\psi_n(\theta))| &= |\Delta(\psi_n(\theta)) - \Delta(\psi_{\alpha}(\theta))| \\ &\leq |\psi'_n(\theta) - \psi'_{\alpha}(\theta)| \\ &\quad + \frac{1}{\alpha} |\psi_n^2(\theta) - \psi_{\alpha}^2(\theta)|. \end{aligned} \quad (5)$$

Поскольку функция распределения и ПЛ являются ограниченными функциями, то дифференцирование и интегрирование по частям нам дает

$$\begin{aligned} |\psi'_n(\theta) - \psi'_{\alpha}(\theta)| &= \left| \int_0^{\infty} x e^{-\theta x} d(G_n(x) - \Gamma_{\alpha}(x)) \right| \\ &\leq L_1 \cdot \sup_x |G_n(x) - \Gamma_{\alpha}(x)|, \end{aligned} \quad (6)$$

где  $L_1 = L_1(\theta) \equiv \text{const} > 0$ , при любом  $\theta \leq \Theta$ . В свою очередь, очевидно, что

$$|\psi_n^2(\theta) - \psi_{\alpha}^2(\theta)| \leq 2 |\psi_n(\theta) - \psi_{\alpha}(\theta)|. \quad (7)$$

Из соотношений (5)–(7), при наличии условия (3), мы приходим к (4).

Чтобы доказать достаточность условия (4), рассмотрим (1) как дифференциальное уравнение с начальным условием  $\psi(0) = 1$ . Тогда легко убедимся в том, что

$$\psi_n(\theta) - \psi_{\alpha}(\theta) = \psi_n(\theta) \psi_{\alpha}(\theta) \int_0^{\theta} \frac{\Delta(\psi_n(\tau))}{\psi_n^2(\tau)} d\tau.$$

Откуда мы убедимся в справедливости неравенства

$$\sup_{\theta \leq \Theta} |\psi_n(\theta) - \psi_{\alpha}(\theta)| \leq L_2 \cdot \Theta \cdot \sup_{\theta \leq \Theta} |\Delta(\psi_n(\theta))|, \quad (8)$$



для любого  $\Theta > 0$ , здесь  $L_2 = L_2(\theta) \equiv \text{const} > 0$ , при любом  $\theta \leq \Theta$ . Неравенство (8) доказывает достаточность условия (4) для сходимости (3).  $\square$

## 2 Лемма и применение Теоремы 1

В этом пункте мы обсуждаем применение доказанной Теоремы 1.

Обозначим  $Z_n$  ( $Z_0 = 1$ ) число частиц к моменту времени  $n \in \mathbf{N}_0$  ( $\mathbf{N}_0 = \{0\} \cup \{\mathbf{N}\}$ ) в ГВП, заданный производящей функцией (п.ф.)

$$F(x) = \sum_{k \in \mathbf{N}_0} p_k x^k, \quad p_k = \mathbf{P}\{Z_1 = k\}, \quad |x| \leq 1.$$

Будем рассматривать случай, когда среднее число потомков одной частицы за одно поколение равно 1, при котором ГВП называется критическим; см. [1]. Последнее равносильно тому, что  $F'(1) = 1$ ; предположим, что  $B := F''(1)$  конечен.

Пусть  $F_n(x) = \mathbf{E}x^{Z_n}$ . Введем функцию  $R_n(x) := 1 - F_n(x)$ , и заметим, что

$$g_n(x) := \sum_{j \in \mathbf{N}} \mathbf{P}\{Z_n = j | Z_n > 0\} x^j = 1 - \frac{R_n(x)}{R_n(0)}, \quad (9)$$

где величина  $R_n(0) = \mathbf{P}\{Z_n > 0\}$  представляет собой вероятность продолжения ГВП. Утверждение об асимптотическом представлении  $R_n(x)$ , по своей значимости, называется основной леммой теории критических ГВП; см. [1, стр. 19]. В нашем случае немаловажное значение представляет и асимптотическое разложение функции  $R'_n(x)$ . Мы обнаружим это разложение в окрестности точки  $x = 1$ . Имеет место следующий локально-дифференциальный аналог основной леммы.

**Лемма.** Если  $B < \infty$ , то при  $x \rightarrow 1$  справедливо соотношение

$$R'_n(x) \sim -g_n^2(x), \quad n \rightarrow \infty, \quad (10)$$

где п.ф.  $g_n(x)$  задана равенством (9).

*Доказательство.* В условиях теоремы имеем разложение

$$F(t) = t + \frac{B}{2}(t-1)^2(1+o(t-1)^2), \quad t \rightarrow 1. \quad (11)$$

С помощью итерации для  $F(x)$ , с учетом того, что  $\text{Sup}_{|x| \leq 1} |R_n(x)| \rightarrow 0$ ,  $n \rightarrow \infty$  (см. [1, стр. 8]), из (11) получим

$$F_n(F(x)) - F_n(x) = \frac{B}{2} R_n^2(x) (1+o(1)), \quad (12)$$

при  $n \rightarrow \infty$ . Используя теорему Лагранжа, из (12) находим

$$F'_n(c(x)) = \frac{B}{2(F(x)-x)} R_n^2(x) (1+o(1)), \quad (13)$$

при  $n \rightarrow \infty$ , где  $c(x) = x + (F(x) - x)\theta$ ,  $0 < \theta < 1$ . Учитывая асимптотическое представление (11) имеем  $c(x) \sim x$ ,  $x \rightarrow 1$ . Опять-таки, равенство (11) утверждает  $F(x) - x \sim \frac{B}{2}(1-x)^2$ ,  $x \rightarrow 1$ . Сказанные вместе с (13) позволяют написать, что при  $x \rightarrow 1$

$$F'_n(x) \sim \left[ \frac{1}{1-x} R_n(x) \right]^2, \quad n \rightarrow \infty.$$

С помощью основной леммы (см. [1, стр. 19]), последнее представление мы легко преобразуем к виду (10).  $\square$

Теперь используя Теоремы 1 в сочетании с Леммой, приведем новое доказательство следующей известной предельной теоремы Яглома.

**Теорема 2 [1, стр. 20].** Пусть  $B < \infty$ . Тогда для любого  $x > 0$

$$\lim_{n \rightarrow \infty} \mathbf{P} \left\{ \frac{Z_n}{\mathbf{E}[Z_n | Z_n > 0]} \leq x | Z_n > 0 \right\} = 1 - e^{-x}. \quad (14)$$

*Доказательство.* Нам известно, что  $\mathbf{E}[Z_n | Z_n > 0] = 1/R_n(0)$  (см. [1, стр. 19]). Тогда ПЛ рассматриваемого условного распределения

$$\varphi_n(\theta) := \mathbf{E} \left[ e^{-\theta Z_n / \mathbf{E}[Z_n | Z_n > 0]} | Z_n > 0 \right]$$

мы можем записать в виде  $\varphi_n(\theta) = g_n(\theta_n)$ , где  $\theta_n = \exp\{-\theta R_n(0)\}$ . Поскольку  $R_n(0) \sim O(1/n)$ ,  $n \rightarrow \infty$ , то дифференцируя его, с учетом (10), получим

$$\varphi'_n(\theta) \sim -\varphi_n^2(\theta), \quad n \rightarrow \infty.$$

Последнее равносильно тому, что

$$\text{Sup}_{\theta \leq \Theta} |\Delta(\varphi_n(\theta))| \rightarrow 0, \quad n \rightarrow \infty,$$

при любом  $\Theta > 0$ . Согласно Теореме 1, этого достаточно для сходимости (14).  $\square$

В конце мы продемонстрируем одно применение Леммы в теории Q-процессов. Пусть величина  $W_n$  обозначает состояние Q-процесса к моменту времени  $n$ , описываемый последовательностью п.ф. (см. напр., [1, стр. 58] и [2])

$$\left\{ W_n(x) := \mathbf{E}x^{W_n} = -x \frac{R'_n(x)}{\beta^n}, \quad n \in \mathbf{N}_0 \right\},$$

где  $\beta = F'(q)$  и число  $q \in (0; 1]$  есть вероятность вырождения ГВП.

**Теорема 3.** Пусть  $F'(1) = 1$  и  $B < \infty$ . Тогда

$$n^2 W_n(x) = \mu(x) (1+o(1)), \quad n \rightarrow \infty, \quad (15)$$

где предельная п.ф.  $\mu(x) = \sum_{k \in \mathbf{N}} \mu_k x^k$  и последовательность неотрицательных чисел  $\{\mu_n, n \in \mathbf{N}\}$  имеет свойство

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} [\mu_1 + \mu_2 + \dots + \mu_n] = \frac{2}{B^2}. \quad (16)$$

*Доказательство.* В рассматриваемом случае  $q = 1$  и  $\beta = 1$ ; см. [1, стр. 7]. Тогда

$$W_n(x) = -xR'_n(x). \quad (17)$$

Поскольку  $R_n(0) \sim 2/Bn$  при  $n \rightarrow \infty$ , то с помощью (9) и основной леммы (см. [1, стр. 19]), соотношение (10) преобразуем к виду

$$R'_n(x) \sim -\frac{4}{B^2 n^2} \frac{1}{(1-x)^2}, \quad n \rightarrow \infty, \quad (18)$$

при  $x \rightarrow 1$ . Рассмотрев вместе (17) и (18), мы получим (15), где

$$\mu(x) \sim -\frac{4}{B^2} \frac{1}{(1-x)^2}, \quad x \rightarrow 1. \quad (19)$$

Теперь нам остается вспомнить Тауберову теорему Харди-Литтлвуда, согласно которой каждое из соотношений (16) и (19) влечет другое.  $\square$

## Список литературы

- [1] Ш.К. Форманов. Метод Стейна-Тихомирова и неклассическая центральная предельная теорема. *Математические заметки*, 71 (4):604–610, 2002.
- [2] А.Н. Тихомиров. О скорости сходимости в центральной предельной теореме для слабо зависимых величин. *Теория вероятностей и ее применения*, 25 (4):800–818, 1980.
- [3] А.А. Имомов. Q-процессы как ветвящиеся процессы Гальтона-Ватсона с иммиграцией. *Труды IX Международной конференции ФАМЭТ'2013*, Красноярск, Россия:148–152, 2010.
- [4] K.B. Athreya and P.E. Ney. *Branching processes*. Springer, New York, 1972.

# О численном решении изотермической модели фильтрации несжимаемой вязкой жидкости в пористой среде

Киреев Виталий Александрович

Институт математики и фундаментальной информатики СФУ  
Базовая кафедра вычислительных и информационных технологий  
Красноярск  
kireevvit@gmail.com

**Аннотация.** Значительные масштабы поступления нефти в окружающую среду при освоении нефтегазовых ресурсов приводят к тому, что данный вид загрязнения является основным для многих районов нефтедобычи. При этом на первый план выходит необходимость в развитии методов исследования проблем экологической безопасности при нефтяном загрязнении. В данной статье рассматривается математическая модель для мониторинга динамики распространения углеводородного загрязнения. Полученные результаты численного эксперимента согласуются с литературными данными.

**Ключевые слова.** Пористая среда, нефтяное загрязнение, закон Дарси, модель двухфазной фильтрации, неявные итерационные методы.

## 1 Введение

Развитие нефтедобычи способствовало исследованию различных математических моделей и методов, позволяющих описать процесс получения нефти из пласта и разработать новые технологии по увеличению нефтеотдачи и поиску новых ресурсов. Но до последнего времени мало внимания уделялось загрязнению окружающей среды при добыче и транспортировке. Процесс геофильтрации флюида, в частности, рассматривается в работах [1], [2], [3]. Изучение фильтрационных процессов в моделях пористых сред, наиболее адекватных к естественным условиям, является актуальным направлением, которое позволит прогнозировать формирование фронта загрязнения и давать оценку величины загрязненной зоны с целью улучшения экологической ситуации в зоне нефтяного разлива. В данной работе рассматривается процесс фильтрации в пористой среде при разливе нефти на поверхности грунта. Модель построена на основе модели Маскета-Левретта [4] при условии, что свободное пространство в порах грунта занято воздухом, а основной вклад в скорость фильтрации вносят гравитационные силы. На основе законов, управляющих течением углеводородов, были выписаны полные балансовые уравнения для трехмерного

потока с учетом действующих факторов. Оператор модели включает систему дифференциальных нелинейных уравнений параболического и эллиптического типа.

## 2 Постановка задачи

Построение модели будем выполнять при следующих допущениях: рассматривается вязкая несжимаемая жидкость в газонасыщенной пористой среде постоянной пористости. Пористая среда характеризуется двумя основными свойствами: пористостью  $\varphi$  и проницаемостью  $k$  [5]. Пористость среды определяется как отношение объема пустот ко всему объему. Бокковые границы не влияют на процесс фильтрации. Рассмотрим основные законы, описывающие двухфазную фильтрацию. Запишем законы сохранения массы для газовой (воздух, индекс  $a$ ) и нефтяной (индекс  $o$ ) фазы:

$$\frac{\partial}{\partial t} (\varphi \rho_a s_a) + \operatorname{div} (\rho_a u_a) = \rho_a q_a$$

$$\frac{\partial}{\partial t} (\varphi \rho_o s_o) + \operatorname{div} (\rho_o u_o) = \rho_o q_o$$

Здесь  $\varphi$  – пористость среды,  $\rho$  – плотность фазы,  $s$  – насыщенность фазы,  $u$  – скорость фильтрации фазы,  $q$  – интенсивность стока (масса единицы объема в единицу времени). В случае источника  $q$  отрицательно, так как предполагается, что оно положительно для стока. Насыщенность фазы определяется как  $s = V_c/V$ , где  $V_c$  – объем среды, занятый данной компонентой;  $V$  – общий объем пустот среды.

При учете того, что жидкость и газ считаем несжимаемыми, пористость – постоянной, законы переписываются следующим образом:

$$\varphi \frac{\partial s_a}{\partial t} + \operatorname{div} (u_a) = q_a$$

$$\varphi \frac{\partial s_o}{\partial t} + \operatorname{div} (u_o) = q_o$$

Закон Дарси устанавливает связь между перепадом давления (напором) и скоростью потока через пористый столб. Закон Дарси для каждой фазы:

$$u_\alpha = -k \frac{k_\alpha}{\mu_\alpha} \nabla (p_\alpha + \rho_\alpha g z), \quad \alpha = o, a$$

Подставим закон Дарси в закон сохранения массы:

$$\varphi \frac{\partial s_\alpha}{\partial t} - \operatorname{div} \left( k \frac{k_\alpha}{\mu_\alpha} \nabla (p_\alpha + \rho_\alpha g z) \right) = q_\alpha$$

Обозначим  $\lambda_\alpha = \frac{k_\alpha}{\mu_\alpha}$  – фазовая подвижность,  $\lambda_T = \lambda_o + \lambda_a$  – общая подвижность,  $f_\alpha = \frac{\lambda_\alpha}{\lambda_T}$  – доля фазы в двухфазном потоке.

Считаем, что  $q_a = 0$  – у воздуха нет источника и стока:

$$\varphi \frac{\partial s_o}{\partial t} - \nabla (k \lambda_o (\nabla p_o + \rho_o g e_3)) = q_o$$

$$\varphi \frac{\partial s_a}{\partial t} - \nabla (k \lambda_a (\nabla p_a + \rho_a g e_3)) = 0$$

Считаем, что фазы полностью заполняют поровое пространство:

$$s_a + s_o = 1$$

Фазовые давления связаны через капиллярное давление:

$$p_c = p_o - p_a$$

Наличие этого параметра в уравнении означает, что в теории фильтрации загрязнителей в грунте должно учитываться различие фазовых давлений, обусловленное действием капиллярных сил.

Выберем в качестве расчетных переменных нефтенасыщенность  $s = s_o$  и глобальное давление

$$p = p_o - \int_{p_c(s)} f_a(p'_c(\xi)) d\xi$$

Производную от глобального давления можно записать в виде:

$$\nabla p = f_a \nabla p_a + f_o \nabla p_o$$

Получим систему:

$$\varphi \frac{\partial s}{\partial t} - \nabla (k \lambda_o (\nabla p + f_a \nabla p_c + \rho_o g e_3)) = q_o \quad (1)$$

$$\varphi \frac{\partial (1-s)}{\partial t} - \nabla (k \lambda_a (\nabla p - f_o \nabla p_c + \rho_a g e_3)) = 0 \quad (2)$$

Сложив уравнения (1) и (2), получим:

$$\begin{aligned} & \nabla (k \lambda_o (\nabla p + f_a \nabla p_c + \rho_o g e_3)) + \\ & + \nabla (k \lambda_a (\nabla p - f_o \nabla p_c + \rho_a g e_3)) = -q_o \end{aligned}$$

После преобразования полученного равенства получим систему уравнений в терминах давления и нефтенасыщенности:

$$\nabla (k \lambda_T \nabla p + k (\lambda_o \rho_o + \lambda_a \rho_a) g e_3) = -q_o \quad (3)$$

$$\varphi \frac{\partial s}{\partial t} + \nabla (k \lambda_a (\nabla p - f_o \nabla p_c + \rho_a g e_3)) = 0 \quad (4)$$

Здесь  $p_c(s) = \frac{\sigma \cos \theta}{\sqrt{k}} J(s)$  – капиллярное давление, где  $J(s)$  – функция Леверетта,  $\sigma$  – поверхностное натяжение на границе “нефть-газ”,  $\theta$  – угол смачивания. Кривую  $J(s)$  будем аппроксимировать гиперболой:

$$J(s) = J_0 \frac{1-s}{s-s_*}$$

Оценка для  $J_0 = 0.034$  [6]. Величина  $s_*$  есть максимальная неподвижная нефтенасыщенность, примем  $s_* = 0.1$

Для фазовых проницаемостей используем кривые Вико-Босета:

$$k_o(s) = \begin{cases} \left( \frac{s-0.1}{0.9} \right)^{3.5}, & 0.1 < s \leq 1 \\ 0, & 0 < s \leq 0.1 \end{cases}$$

$$k_a(s) = \begin{cases} \left( \frac{0.9-s}{0.9} \right)^{3.5} \cdot (1+3s), & 0 \leq s < 0.9 \\ 0, & 0.9 \leq s \leq 1 \end{cases}$$

Запишем начально-краевые условия:

$$p(0, y, z) = p(l_x, y, z) = p(x, 0, z) = p(x, l_y, z) = p + \rho g z;$$

$$s(x, y, z, 0) = 0;$$

Условие постоянного перетока через верхнюю границу для двухфазной фильтрации:

$$\left. \frac{\partial p}{\partial z} \right|_{z=0} = -\frac{q(x, y, 0)}{k \lambda_T} - (f_o \rho_o + f_a \rho_a) g;$$

Условие непротекания на нижней границе:

$$\left. \frac{\partial p}{\partial z} \right|_{z=l_z} = -(f_o \rho_o + f_a \rho_a) g;$$

Предположим, что источник имеет форму круга радиуса  $\frac{l_x + l_y}{8}$  с центром в точке  $(\frac{l_x}{2}, \frac{l_y}{2}, 0)$ .

### 3 Разностная схема и алгоритм решения

Обозначим через  $\bar{\omega}$  равномерную сетку в области  $\Omega$ :

$$\bar{\omega} = \{X | X(x, y, z), x_\alpha = i_\alpha h_\alpha, i_\alpha = 0, 1, \dots, N_\alpha, N_\alpha h_\alpha = l_\alpha\}$$

Заменим уравнения (3), (4) на разностные аналоги с учетом переменных коэффициентов. Неявная схема для уравнения давления по методу Гаусса-Зейделя:

$$\begin{aligned} & a_{i+1jk} \frac{p_{i+1jk}^n - p_{ijk}^{n+1}}{h_x^2} - a_{ijk} \frac{p_{ijk}^{n+1} - p_{i-1jk}^{n+1}}{h_x^2} + \\ & + a_{ij+1k} \frac{p_{ij+1k}^n - p_{ijk}^{n+1}}{h_y^2} - a_{ijk} \frac{p_{ijk}^{n+1} - p_{ij-1k}^{n+1}}{h_y^2} + \\ & + a_{ijk+1} \frac{p_{ijk+1}^n - p_{ijk}^{n+1} + b_{ijk+\frac{1}{2}} h_z}{h_z^2} - \\ & - a_{ijk} \frac{p_{ijk}^{n+1} - p_{ijk-1}^{n+1} + b_{ijk-\frac{1}{2}} h_z}{h_z^2} = -q_{ijk} \end{aligned}$$

где

$$\begin{aligned} a_{i+1jk} &= k(\lambda_{i+1jk}^o + \lambda_{i+1jk}^a) = \\ &= k \left( \frac{k_o(\frac{s_{i+1jk}^n + s_{ijk}^n}{2})}{\mu_o} + \frac{k_a(\frac{s_{i+1jk}^n + s_{ijk}^n}{2})}{\mu_a} \right), \\ b_{ijk} &= k(\lambda_{ijk}^o \rho_o + \lambda_{ijk}^a \rho_a) g \end{aligned}$$

Отсюда

$$p_{ijk}^{n+1} = \frac{1}{\frac{a_{i+1jk} + a_{ijk}}{h_x^2} + \frac{a_{ij+1k} + a_{ijk}}{h_y^2} + \frac{a_{ijk+1} + a_{ijk}}{h_z^2}} \times$$

$$\begin{aligned} & \times \left( \frac{a_{i+1jk} p_{i+1jk}^n + a_{ijk} p_{i-1jk}^{n+1}}{h_x^2} + \right. \\ & \quad \left. + \frac{a_{ij+1k} p_{ij+1k}^n + a_{ijk} p_{ij-1k}^{n+1}}{h_y^2} + \right. \\ & \quad \left. + \frac{a_{ijk+1} (p_{ijk+1}^n + b_{ijk+1} h_z) + a_{ijk} (p_{ijk-1}^{n+1} - b_{ijk} h_z)}{h_z^2} + \right. \\ & \quad \left. + q_{ijk} \right) \end{aligned} \quad (6)$$

Для метода последовательной верхней релаксации с коэффициентом  $\tau$ :

$$\begin{aligned} p_{ijk}^{n+1} = & \frac{\tau}{\frac{a_{i+1jk} + a_{ijk}}{h_x^2} + \frac{a_{ij+1k} + a_{ijk}}{h_y^2} + \frac{a_{ijk+1} + a_{ijk}}{h_z^2}} \times \\ & \times \left( \frac{a_{i+1jk} p_{i+1jk}^n + a_{ijk} p_{i-1jk}^{n+1}}{h_x^2} + \right. \\ & \quad \left. + \frac{a_{ij+1k} p_{ij+1k}^n + a_{ijk} p_{ij-1k}^{n+1}}{h_y^2} + \right. \\ & \quad \left. + \frac{a_{ijk+1} (p_{ijk+1}^n + b_{ijk+1} h_z)}{h_z^2} + \right. \\ & \quad \left. + \frac{a_{ijk} (p_{ijk-1}^{n+1} - b_{ijk} h_z)}{h_z^2} + q_{ijk} \right) + \\ & \quad + (1 - \tau) p_{ijk}^n \end{aligned}$$

Уравнение для нефтенасыщенности:

$$\begin{aligned} \varphi \frac{s_{ijk}^{n+1} - s_{ijk}^n}{\tau} = & - \left[ c_{i+1jk} \left( \frac{p_{i+1jk}^{n+1} - p_{ijk}^{n+1}}{h_x^2} - \right. \right. \\ & \left. \left. - \frac{f_{i+1jk}^o (p_c (s_{i+1jk}^{n+1}) - p_c (s_{ijk}^{n+1}))}{h_x^2} \right) - \right. \\ & \left. - c_{ijk} \left( \frac{p_{ijk}^{n+1} - p_{i-1jk}^{n+1}}{h_x^2} - \right. \right. \\ & \left. \left. - \frac{f_{ijk}^o (p_c (s_{ijk}^{n+1}) - p_c (s_{i-1jk}^{n+1}))}{h_x^2} \right) + \right. \\ & \left. + c_{ij+1k} \left( \frac{p_{ij+1k}^{n+1} - p_{ijk}^{n+1}}{h_y^2} - \right. \right. \\ & \left. \left. - \frac{f_{ij+1k}^o (p_c (s_{ij+1k}^{n+1}) - p_c (s_{ijk}^{n+1}))}{h_y^2} \right) - \right. \\ & \left. - c_{ijk} \left( \frac{p_{ijk}^{n+1} - p_{ij-1k}^{n+1}}{h_y^2} - \right. \right. \\ & \left. \left. - \frac{f_{ijk}^o (p_c (s_{ijk}^{n+1}) - p_c (s_{ij-1k}^{n+1}))}{h_y^2} \right) + \right. \\ & \left. + c_{ijk+1} \left( \frac{p_{ijk+1}^{n+1} - p_{ijk}^{n+1}}{h_z^2} - \right. \right. \\ & \left. \left. - \frac{f_{ijk+1}^o (p_c (s_{ijk+1}^{n+1}) - p_c (s_{ijk}^{n+1})) + \rho_a g h_z}{h_z^2} \right) - \right. \\ & \left. - c_{ijk} \left( \frac{p_{ijk}^{n+1} - p_{ijk-1}^{n+1}}{h_z^2} - \right. \right. \end{aligned}$$

Где

$$\begin{aligned} c_{ijk} &= k \lambda_{ijk}^a = k \frac{k_a \left( \frac{s_{ijk}^n + s_{i-1jk}^n}{2} \right)}{\mu_a} \\ f_{ijk}^o &= \frac{\lambda_{ijk}^o}{\lambda_{ijk}^o + \lambda_{ijk}^a} \\ p_c (s_{ijk}^n) &= \frac{\sigma \cos \theta}{\sqrt{\frac{k}{\varphi}}} J_0 \frac{1 - s_{ijk}^n}{s_{ijk}^n - s_*} \end{aligned}$$

Аппроксимируем начально-краевые условия:

$$\begin{aligned} p_{0jk} &= p_{N_xjk} = p_{i0k} = p_{iN_yk} = p + \rho g k h_z, \quad \forall n \\ \frac{p_{ij1} - p_{ij0}}{h_z} &= - \frac{q_{ij0}}{k (\lambda_{ij1}^o + \lambda_{ij1}^a)} - (f_{ij1}^o \rho_o + f_{ij1}^a \rho_a) g; \\ \frac{p_{ijN_z} - p_{ijN_z-1}}{h_z} &= - (f_{ijN_z}^o \rho_o + f_{ijN_z}^a \rho_a) g; \\ s_{ijk}^0 &= 0; \end{aligned}$$

Аппроксимируем правую часть:  $q_{ijk} = q_0$  при

$$(5) \quad \left( i - \frac{N_x}{2} \right)^2 h_x^2 + \left( j - \frac{N_y}{2} \right)^2 h_y^2 \leq \left( \frac{N_x h_x + N_y h_y}{8} \right)^2$$

Сформулированная постановка задачи и фильтрации двухфазной несжимаемой жидкости, где искомыми функциями являются  $p$ ,  $s$ , позволяет привести схему решения.

1. По известным на  $j$ -м временном шаге  $p^j$ ,  $s^j$  и краевым условиям для  $p$  определяем  $p^{j+1}$  из уравнения (5) с помощью итерационного метода верхней релаксации.
2. По найденному  $p^{j+1}$ , известному  $s^j$  и краевым условиям для  $s^j$  определяем  $s^{j+1}$  из уравнения (6) с помощью итерационной неявной разностной схемы.
3. Тогда по найденному  $s^{j+1}$  величина  $p^{j+1}$  определяется однозначно. Счет величин на  $(n+1)$ -м шаге окончен, переходим к счету следующего шага и т.д.

## 4 Результаты моделирования

Численные расчеты распространения загрязнителя показывают, что под воздействием только силы тяжести процесс протекает достаточно медленно, что согласуется с экспериментальными данными. Сначала формируется фронт загрязнения, который далее движется вглубь пористого грунта.

При проведении расчетов принято следующее: пористость грунта  $\varphi = 0.06 - 0.4$ ; проницаемость  $k = 1.2 \cdot 10^{-3} - 12 \cdot 10^{-3}$  мкм<sup>2</sup>; плотность загрязнителя



$\rho_o = 1 \text{ г/см}^3$ ; плотность воздуха  $\rho_a = 10^{-3} \text{ г/см}^3$ ; вязкость загрязнителя  $\mu_o = 10 \text{ Па} \cdot \text{с}$ . На рис. 1-2 приведены результаты расчетов распространения загрязнителя в грунтах различной пористости и проницаемости в безразмерных величинах.

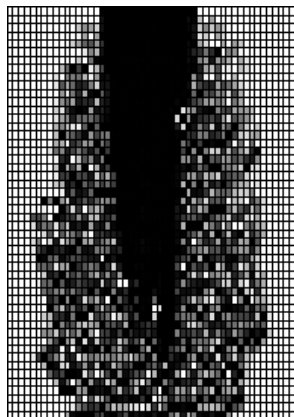


Рис. 1: Распространение нефтезагрязнения в вертикальном разрезе при  $\varphi = 0.06$ ,  $k = 0.12 \text{ мкм}^2$ .

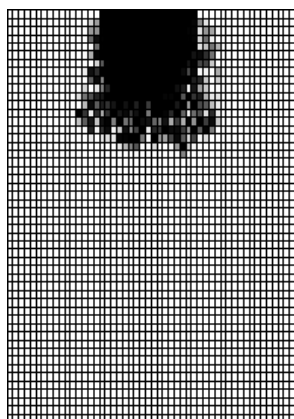


Рис. 2: Распространение нефтезагрязнения в вертикальном разрезе при  $\varphi = 0.12$ ,  $k = 0.06 \text{ мкм}^2$ .

Характер распространения углеводородов в различных почвах согласуется с известными результатами [6], [7].

## Список литературы

- [1] R.W. Lewis, P. Nithiarasu. *Fundamentals of the Finite Element Method for Heat and Fluid Flow*. Wiley, 2004.
- [2] А. В. Костерин, К. А. Поташев, З. В. Харламова, И. П. Бреус. Математическое моделирование фильтрации несмешивающихся с водой органических жидкостей в почвах. *Почвоведение*, Москва: Наука, 7:828–836, 2004.
- [3] А. В. Шишени, А. А. Афонин, А. И. Сухинов. Построение трехмерной модели геофильтрации флюида в многослойных пористых средах. *Известия ЮФУ. Технические науки*, Таганрог: Технологический институт ЮФУ, 97(8):52–62, 2009.
- [4] А.Н. Коновалов. *Задачи фильтрации многофазной несжимаемой жидкости*. Наука. Сиб. отд-ние, Новосибирск, 1988.
- [5] Н.В. Молокова. Решение геофильтрационных задач средствами математического моделирования. *Вестник Сибирского Государственного Аэрокосмического Университета им. Академика М. Ф. Решетнева*, Красноярск: СибГАУ им. акад. М. Ф. Решетнева, 3:31–34, 2008.
- [6] А.И. Гриценко, Г.С. Акопова, В. М. Максимов. *Экология. Нефть и газ*. ВНИИЭГазпром, Москва, 1995.
- [7] Н.В. Молокова. Математическое моделирование процессов нефтезагрязнения пористой среды. *Вестник Сибирского Государственного Аэрокосмического Университета им. Академика М. Ф. Решетнева*, Красноярск: СибГАУ им. акад. М. Ф. Решетнева, 5:142–148, 2010.

## Разработка и применение электронного образовательного ресурса «Алгоритмы быстрых вычислений»

Кириллюк Анна Алексеевна

Сибирский Федеральный Университет  
Институт математики и фундаментальной информатики  
Красноярск  
nutik\_89@list.ru

**Аннотация.** В работе приводится информация о быстрых вычислениях, их назначении, применении и формируемых ими навыках. Также проводится краткий обзор методической литературы по алгоритмам быстрого счета. Перечисляются основные направления изучения математики в среднем звене средней общеобразовательной школы. В работе предлагается электронно-образовательный ресурс в качестве методологического средства, реализующего процесс обучения учащихся среднего звена школы алгоритмам. В работе дается подробное описание состава, назначения данного электронно-образовательного ресурса, а также — его преимуществ по сравнению с традиционными методиками преподавания. Разработанный электронный образовательный ресурс «Алгоритмы быстрых вычислений» позволяет обучать учащихся, предоставляет методическую информацию по проведению занятий преподавателям, а также реализует интерактивную проверку знаний и навыков быстрого счета, приобретенных учащимися в процессе обучения.

**Ключевые слова.** Алгоритм, быстрый счет, электронный образовательный ресурс, тематический элемент, мультимедиа, интерактивность.

### 1 Введение

В современном обществе высоких технологий и всеобщего использования компьютера умение быстро и правильно производить в уме вычисления не утратило своей актуальности. Числа окружают нас повсюду, а выполнение арифметических действий над ними приводит к результату, на основании которого мы принимаем то или иное решение. Поэтому каждый человек должен обладать хорошими навыками устного счета.

Вычислительная культура формируется у учащихся на всех этапах изучения курса математики, но основа ее закладывается в первые 5–6 лет обучения. В этот период школьники обучаются умению осознанно использовать законы математических действий. В

последующие годы полученные умения и навыки совершенствуются и закрепляются в процессе изучения алгебры, физики, химии, черчения и других предметов.

Для повышения вычислительной культуры нужно иметь хорошие вычислительные навыки.

**Определение 48.** *Вычислительный навык — это высокая степень овладения вычислительными приемами.*

Повышению вычислительной культуры способствуют в том числе и алгоритмы быстрых вычислений.

**Определение 49.** *Образовательная среда — совокупность возможностей для развития обучающегося, для проявления и развития его способностей и личностных потенциалов; средство самообучения и саморазвития; предмет проектирования и моделирования.*

**Определение 50.** *Алгоритм — это последовательности действий (шагов), приводящие к результату.*

Сейчас, на этапе стремительного развития информатики и вычислительной техники, понятие алгоритма является одним из центральных. Алгоритмизация считается обязательным этапом в процессе разработки программ и решении задач на ЭВМ. Именно для прикладных алгоритмов и программ принципиально важны детерминированность, результативность и массовость, а также правильность результатов решения поставленных задач. Распространенным критерием оценки алгоритмов является время работы и порядок роста продолжительности работы. Вычислять быстро, подчас на ходу — это требование времени. Понятно, что без вычислений не обойтись в повседневной жизни.

Для быстрого счета необходима рациональность. Для этого ученик должен выбрать те из возможных операций, выполнение которых легче других и быстрее приводит к результату арифметического действия.

Быстрые вычисления имеют большое образовательное, воспитательное и методическое значение [1]:

- способствуют формированию основных математических понятий, более глубокому ознакомле-

нию с составом чисел из слагаемых и сомножителей;

- применяя законы арифметических действий к быстрым вычислениям, можно повторить, закрепить, усвоить их не механически, а сознательно;
- при быстрых вычислениях развиваются внимание, сосредоточенность, наблюдательность, самостоятельность;
- быстрый счет содействует тренировке памяти, развивает быстроту реакции и гибкость мышления.

Сегодняшние школьники устно считают плохо. В наше время изменилась методика обучения детей математике, и многие приемы быстрого счета ушли из современной школы.

Учебники математики ориентированы на общие вычислительные навыки. Так изучение математики в 5–6 классах направлено на формирование действий с натуральными и целыми числами, обыкновенными и десятичными дробями. В 7–9 классах при обучении делается акцент на обобщение и систематизацию знаний о числовых выражениях, решение систем уравнений, умножение и деление степеней с одинаковыми показателями. Также формируются умения выполнять арифметические действия над одночленами и многочленами, вычисление корней линейного и квадратного уравнений и систем уравнений.

К сожалению, в учебниках математики для 5–6 классов способы быстрого счета не приводятся. Также отсутствуют они и в курсе алгебры 7–9 классов. Незначительное число способов быстрого вычисления встречаются лишь в курсе информатики шестого класса: рассматриваются системы счисления и элементы алгоритмизации.

Совершенствование навыков быстрых вычислений зависит не только от методики организации занятий, от форм, контроля, но и во многом от того насколько сами дети проявляют интерес к этой форме работы. Этот интерес можно вызвать, показав учащимся красоту и изящество устных вычислений, используя не совсем обычные вычислительные приемы, помогающие порой значительно облегчить процесс вычисления. Эти приемы вычислений могут быть преподнесены в виде занимательных задач на уроках или на внеклассных занятиях в форме электронных образовательных ресурсов

Целью данной работы является разработка методических материалов по алгоритмам быстрых вычислений, направленных на развивающее обучение учащихся среднего звена общеобразовательной школы с использованием электронных ресурсов и методических руководств по проведению занятий. Для реализации данной цели были решены следующие задачи:

- изучены методики и формы обучения быстрому счету и быстрым вычислениям;
- выявлены основных направлений изучения математики в среднем звене средней общеобразовательной школы;
- произведен выбор и систематизация алгоритмов быстрого счета, подходящих для изучения учащимися среднего звена школы;
- построена схема изучения алгоритмов быстрого счета с помощью мультимедийных средств;
- разработана система практических занятий по быстрому счету;
- разработан электронный образовательный ресурс, реализующий процесс обучения учащихся быстрым алгоритмам с помощью мультимедийных средств, а также выполняющий интерактивную проверку знаний и навыков быстрого счета, приобретенных учащимися в процессе обучения.

## 2 Методические основы быстрых вычислений

В современном обществе личностно-ориентированная педагогика выдвигает на передний план нетрадиционные подходы к организации процесса обучения в современной школе. Содержание образования составляет систему знаний, умений, навыков, черт творческой деятельности, мировоззренческих и поведенческих качеств личности, которые обусловлены требованиями общества и к достижению которых должны быть направлены усилия обучающихся и обучающихся. Осуществление обучения требует знания и умелого использования разнообразных форм организации учебного процесса, их постоянного совершенствования и модернизации.

В современной дидактике организационные формы обучения, включая обязательные и факультативные, классные и домашние занятия, подразделяют на фронтальные, групповые и индивидуальные.

При фронтальном обучении учитель управляет учебно-познавательной деятельностью всего класса. Он организует сотрудничество учащихся и определяет единый для всех темп работы. Педагогическая эффективность фронтальной работы во многом зависит от умения учителя держать в поле зрения весь класс и при этом не упускать из виду работу каждого ученика.

При групповых формах обучения учитель управляет учебно-познавательной деятельностью групп учащихся класса. Их можно подразделить на звеньевые, бригадные, кооперированно-групповые и дифференцированно-групповые.

Индивидуальное обучение учащихся не предполагает их непосредственного контакта с другими учениками. По своей сущности оно есть не что иное, как самостоятельное выполнение одинаковых для всего класса или группы заданий. Факультативы относятся к дополнительным занятиям.

Вычислительная культура является тем запасом знаний и умений, который находит повсеместное применение, является фундаментом изучения математики и других учебных дисциплин. Проблема формирования у учащихся вычислительных умений и навыков всегда привлекала особое внимание психологов, дидактов, методистов, учителей. В методике математики известны исследования Е.С. Дубинчук, А.А. Столяра, С.С. Минаевой, Н.Л. Стефановой, Я.Ф. Чекарева, М.А. Бантовой, М.И. Моро, Н.Б. Истоминой, С.Е. Царевой и других.

Переориентация методической системы на приоритет развивающей функции по отношению к образовательной, характеризующейся изменением характера деятельности учащихся, личностно-ориентированным подходом к обучению, ослабила внимание к развитию и закреплению вычислительных навыков у учащихся.

Отмечается ухудшение качества вычислений учащихся, обучающихся и по обычным, и по развивающим учебникам. Особенно пострадала культура устного счета.

Основная задача технологии формирования вычислительных навыков на уроках математики — задача повышения вычислительной культуры. Повышению вычислительной культуры не в последнюю очередь способствуют алгоритмы быстрых вычислений.

Но в настоящее время вопросы теории и практики быстрых вычислений не нашли еще своего специального отражения в методической литературе. В то время как приемы вычислительной техники могли бы служить учителю прекрасным орудием в его работе с учениками: во-первых, учитель, хорошо, усвоивший эти приемы, имеет возможность быстро проверить вычисления, производимые учащимися, и, во-вторых, вовремя продемонстрировать перед ними тот или иной прием.

Ученики весьма живо реагируют на новое, столь непохожее на обыденные стандартные способы вычислений, и легко воспринимают эти приемы. Вследствие этого сильно возрастает авторитет учителя, с одной стороны, и повышается интерес учащихся к предмету — с другой.

Алгебраические доказательства, обосновывающие приводимые приемы упрощенных вычислений, производятся на базе формул и элементарных алгебраических преобразований, и тем самым учитель имеет возможность закрепить понимание формулы по су-

ществу, иллюстрируя ее соответствующим приемом быстрого счета.

Это полностью отвечает требованиям утвержденной программы по математике для средней школы, где указано, что «для закрепления навыков в устном счете необходимо использовать формулы сокращенного умножения» и вообще «на всем протяжении курса арифметики должно уделяться внимание устному решению задач, устным вычислениям».

Таким образом, можно выделить три основных составляющих навыка быстрого счета:

1. способность концентрировать внимание;
2. знание специальных алгоритмов и умение оперативно подобрать нужный);
3. тренировка и опыт.

Совершенствование навыков устных вычислений зависит не только от методики организации занятий, от форм, контроля, но и во многом от того насколько сами дети проявляют интерес к этой форме работы. Этот интерес можно вызвать, показав учащимся красоту и изящество устных вычислений, используя не совсем обычные вычислительные приемы, помогающие порой значительно облегчить процесс вычисления. Эти приемы вычислений могут быть преподнесены в виде занимательных задач на уроках или на внеклассных занятиях в форме электронных образовательных ресурсов [2].

### **3 Методические основы применения информационно - коммуникативных технологий для изучения алгоритмов быстрого счета**

Нами были разработаны методические материалы по обучению учащихся среднего звена общеобразовательной школы алгоритмам быстрых вычислений с использованием мультимедийных средств и электронных ресурсов, а также — методическое руководство по проведению занятий.

Применение компьютерных технологий в учебном процессе дает возможность использовать в педагогической практике психолого-педагогические разработки, позволяющие интенсифицировать учебный процесс, реализовывать идеи развивающего обучения [3].

Современные сетевые информационные-коммуникативные технологии позволяют осуществлять контроль за собственным образованием, выстраивать индивидуальную траекторию обучения, то есть самостоятельно осуществлять выбор темпа, времени и места обучения.



Главным для учителя и ученика в интерфейсе компьютерного средства обучения является эффективность и удобство восприятия учебной информации.

**Определение 51.** *Интерфейс — это сторона программного продукта, которая организует обучение "глазами пришедшее на смену обучения" "руками то есть является одним из наиболее существенных компонентов компьютерного средства обучения.*

Данный подход к интерфейсу компьютерного средства обучения выдвигает на первый план восприятие глазами информации, представляемой на экране монитора персонального компьютера.

**Определение 52.** *Информационная среда — система средств общения с человеческим знанием (служит как для хранения, структурирования и представления информации, составляющей содержание накопленного знания), так и для ее передачи, переработки и обогащения).*

Под визуальной средой обучения мы понимаем совокупность условий обучения, в которых акцент ставится на использование и развитие визуального мышления ученика. Эти условия предполагают наличие как традиционно наглядных, так и специальных средств и приемов, позволяющих активизировать работу зрения.

**Определение 53.** *Мультимедиа — современная компьютерная информационная технология, позволяющая объединить в компьютерной системе текст, звук, видеоизображение, графическое изображение и анимацию (мультипликацию).*

Мультимедиа как средство обучения обладает определенными технико-педагогическими и дидактическими возможностями и рассматривается в трех аспектах: информационном, коммуникативном и как средство индивидуализации обучения. Ведущие функции мультимедиа, применяемые в процессе образования: разъясняющая, информационная, коммуникативная, эвристическая, моделирующая, систематизирующая, мотивирующая и развивающая.

Использование мультимедийных средств позволяет активизировать процесс обучения за счет усиления наглядности и сочетания логического и образного способов усвоения информации. Интерактивность мультимедийных технологий предоставляет широкие возможности для реализации личностно-ориентированных моделей обучения.

**Определение 54.** *Электронными образовательными ресурсами (ЭОР) называют учебные материалы, для воспроизведения которых используются электронные устройства.*

Здесь и далее мы рассматриваем электронные образовательные ресурсы, для воспроизведения которых требуется компьютер.

Классификация ЭОР:

1. текстографические. Они отличаются от книг в основном базой предъявления текстов и иллюстраций — материал представляется на экране компьютера, а не на бумаге;
2. текстографические с нелинейной навигацией по тексту — фрагменты текста просматриваются в произвольном порядке, определяемом логической связностью и собственным желанием, при нажатии гипертекстовой ссылки происходит переход в нужное место;
3. мультимедийные ресурсы, целиком (или частично) состоящие из визуального или звукового фрагмента.

Наиболее мощными и интересными для образования продуктами являются мультимедийные ЭОР.

Когда мы говорим о мультимедийных ЭОР, имеется в виду возможность одновременного воспроизведения на экране компьютера некоторой совокупности объектов, представленных различными способами. Разумеется, все представляемые объекты должны быть связаны логически, подчинены определенной дидактической идее, и изменение одного из них вызывает соответствующие изменения других.

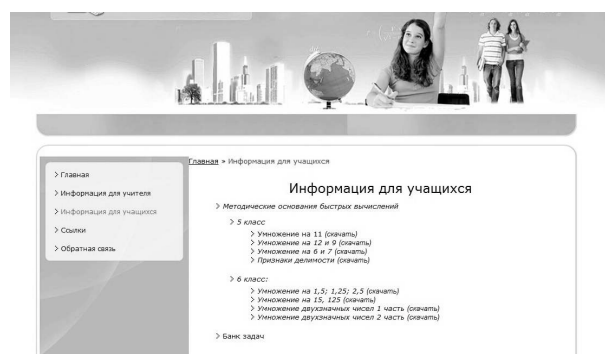


Рис. 1: Вид первой страницы сайта в блоке для учащихся

К основным инновационным качествам электронных образовательных ресурсов относятся:

- обеспечение всех компонентов образовательного процесса: получение информации;
- проведение практических занятий;
- аттестация (контроль учебных достижений);
- интерактивность, которая обеспечивает резкое расширение возможностей самостоятельной учебной работы за счет использования активно-деятельностных форм обучения
- возможность более полноценного обучения вне аудитории.



Электронный образовательный ресурс «Алгоритмы быстрых вычислений» представляет собой мультимедийный электронный образовательный ресурс с интерактивными возможностями.

Информационная схема изложения материала ресурса состоит из блоков (для пятых и sixth классов). В каждом блоке выделены учебные фрагменты (алгоритмы). Используются два языка описания информации (знаки и текст), что позволяет быстро ориентироваться в содержании ресурса. Материал представлен в виде электронных информационных тетрадей.

**Определение 55.** *Электронная информационная тетрадь — это серия визуальных блоков, выстроенных в определенной последовательности и по смыслу тесно связанных между собой.*



Рис. 2: Представление одной из презентаций ресурса

Тетрадь разбивает процесс овладения темой на отдельные шаги, выделяющие каждый поворот мысли. Причем для каждого из них представлен свой способ визуально-схематического отражения, как его содержания, так и его связи с предыдущим материалом (это реализовано с помощью гиперссылок).

#### 4 Практическое применение алгоритмов в быстром счете

Электронный образовательный ресурс «Алгоритмы быстрых вычислений» представляет собой полноценный Интернет-ресурс. Он предназначен для дополнительной помощи при проведении занятий в пятых и sixth классах, а также для проведения элективного курса. Рассчитан на четыре академических часа в каждом классе. С помощью сайта излагается информация для пятого класса по теме «Натуральные числа», в sixth классе при повторении материала в начале года и при изучении темы «Арифметические операции с дробями». По реализационному принципу ресурс «Алгоритмы быстрых вычислений» относится к мультимедийно-презентационному типу ресурсов.

Электронно-образовательный ресурс состоит из двух блоков: информации для учителя и информации для

учащихся. Каждый блок содержит модули быстрых вычислений для пятых и sixth классов. В блоке для учителя представлены конспекты уроков и презентации для проведения уроков. В блоке для учащихся минимальной структурной единицей является тематический элемент. Для каждого тематического элемента имеется три типа электронных учебных модулей: модуль получения информации, модуль практических занятий, модуль контроля.

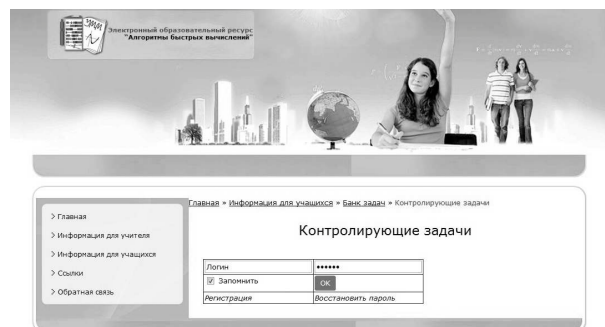


Рис. 3: Авторизация пользователя ресурса

Модуль практических знаний построен на интерактивности. Признак интерактивности информационного взаимодействия: активность в процессе осуществления информационного взаимодействия образовательного назначения проявляют не только обучающий и обучающийся, но и средство обучения, функционирующее на базе информационно-коммуникационных технологий, благодаря таким возможностям, как обеспечение интерактивного диалога, компьютерной визуализации, обработки значительно объема информации.

Практические задания представлены в виде банка задач, которые подразделяются на обучающие и контролируемые.

Обучающие задачи направлены на формирование элементов теоретических знаний и связанных с ними умений. Данные задачи разбиты на темы. Задачи представлены в виде теста открытого типа (не содержит набор ответов для выбора). На прохождение теста отводится пятнадцать минут. Компьютер показывает правильность введенного ответа. Результат тестирования выводится в виде процентного соотношения.

Контролирующие задачи представлены в виде теста закрытого типа (содержит набор ответов для выбора). Функциональное назначение контролируемых задач — осуществление контроля со стороны учителя или учащихся уровня сформированности знаний, умений и навыков (позволяет оценить знания учащихся). Задачи разделены на задания для пятого и sixth класса. Задания выбираются произвольным образом из общего банка задач. Для просмотра контролируемых задач необходимо пройти авто-

ризацию. Авторизация позволяет хранить результаты каждого ученика, и позволяет отследить время прохождения теста, количество набранных баллов и названия тем, в которых допущены ошибки.

Электронный ресурс реализован на языке гипертекстовой разметки HTML. Web-приложение, выполняющее работу по тестированию учащихся (интерактивную проверку знаний), написано на языке PHP [4]. На Рис. 1 представлен вид первой страницы электронного образовательного ресурса «Алгоритмы быстрых вычислений» в блоке для учащихся.

На Рис. 2 показана одна из презентаций ресурса. На Рис. 3 и Рис. 4 демонстрируются выполнение авторизации пользователя и интерактивная проверка знаний в виде тестирования, соответственно.

Умножение на 12

Линия времени: 00:14:46

Задание 1 из 10

1. Условие: Количество баллов: 1

22 \* 12 =

[Проверить](#)

Таблица лучших: Умножение на 12

Место	Имя	Записано	Баллы	Результат
1	admi	2013/10/09 9:13 дп	1	100 %

Умножение на 9

описание

[Начать тест](#)

Таблица лучших: Умножение на 9

Рис. 4: Интерактивное тестирование учащегося

В качестве макета сайта был выбран двухколоночный макет с плавающей шириной и с шапкой. В шапке находится логотип и название ресурса. В левой колонке располагается навигационное меню, а правой – основная информация. Навигация по сайту осуществляется с помощью меню, в котором расположены ссылки на соответствующие страницы.

Электронный ресурс можно использовать для коллективной или индивидуальной деятельности в условиях смоделированных ситуаций, и направленных на формирование профессиональных умений и навыков.

## Список литературы

- [1] Г.В. Репкина, Е.В. Заика. *Оценка уровня сформированности учебной деятельности*. Пеленг, Томск, 1993.
- [2] Э. Катлер, Р. Мак-Шейн. *Система быстрого счета по Трахтенбергу*. Просвещение, М., 1967.
- [3] Н. А. Резник. *Методические основы обучения математике в средней школе с использованием средств развития визуального мышления*. МПУ, М., 1997.
- [4] Л. Веллинг, Л. Томсон. *Разработка Web-приложений с помощью PHP и MySQL*. Вильямс, М., 2007.

# О распределении расстояний до ближайшего соседа в марковских цепях

Ксения Андреевна Клокова

ИКИТ СФУ, Красноярск  
ksklokova@yandex.ru

Михаил Георгиевич Садовский

Институт вычислительного моделирования СО РАН  
Красноярск  
msad@icm.krasn.ru

**Аннотация.** Представлены строгие результаты по определению распределения расстояний до ближайшего соседа в Марковских цепях. Сформулирована и доказана лемма, проведены вычислительные эксперименты. Для экспериментов использовались Марковские цепи порядка 2, 3, 4 и 6, для символьных последовательностей из четырёхбуквенного алфавита.

**Ключевые слова.** Частота, эргодичность, релаксация, генетика.

## 1 Введение

Рассмотрим бесконечную символьную последовательность из алфавита  $\aleph$  мощности  $m = 4$ :  $\aleph = \{A, C, G, T\}$ . Зафиксируем два слова этого алфавита  $\omega_1$  и  $\omega_2$  каждое длины  $k = 3$ . Найдём первое попавшееся вхождение слова  $\omega_1$  в рассматриваемую последовательность и будем искать ближайшее к нему справа вхождение слова  $\omega_2$ . Количество символов от начала  $\omega_1$  до начала  $\omega_2$  будем называть расстоянием  $n$  между словами  $\omega_1$  и  $\omega_2$ ,  $\omega_2$  назовем ближайшим соседом к  $\omega_1$ . Например,

$$\omega_1 = ACC, \quad \omega_2 = AAA$$

$$\dots TAACCACTGTAATG \dots$$

$$n = 8$$

Рассмотрим ситуацию, когда последовательность задается Марковским процессом порядка  $t$ , то есть вероятность появления того или иного символа алфавита  $\aleph$  в определенном месте последовательности зависит от того, какие символы были на предыдущих  $t$  местах, рассматриваются оба варианта  $t \leq k$  и  $t > k$ .

Необходимо найти функцию  $f_{\omega_1, \omega_2}(n)$  распределения расстояний  $n$  до ближайшего соседа для слов  $\omega_1$  и  $\omega_2$ , т.е. функцию, определяющую вероятность того, что слово  $\omega_2$  встретится на расстоянии  $n$  справа от слова  $\omega_1$  и не раньше.

## 2 Основные утверждения

### 2.1 Случай первый: $t \leq k$

В случае  $t \leq k$  вероятности перехода из одного  $k$ -символьного слова в другое можно объединить в матрицу  $P$ :

$$P = \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1r} \\ p_{21} & p_{22} & \dots & p_{2r} \\ \dots & \dots & \dots & \dots \\ p_{r1} & p_{r2} & \dots & p_{rr} \end{pmatrix}$$

Здесь  $i$ -той строке соответствуют вероятности перехода из слова под номером  $i$  в любое другое, а  $j$ -тому столбцу — вероятности перехода в слово под номером  $j$  из любого другого слова. Индексы  $i$  и  $j$  меняются от 1 до  $r$  — количество всевозможных слов длины  $k$  равно мощности алфавита в степени длины слов:  $r = m^k = 4^3 = 64$ . То есть матрица  $P$  — квадратная, размера  $r \times r = 64 \times 64$ .

Но, так как на самом деле из каждого слова за один шаг можно перейти не во все возможные слова, а только в  $m = 4$ , то в каждой строке (и, аналогично, в каждом столбце) только  $m = 4$  из  $m^k = 64$  ( $m^{-(k-1)} = 1/16$ ) элементов будут отличны от нуля. Например, из  $AAA$  за один шаг можно перейти в  $AAA, AAC, AAG, AAT$ , но не в  $ACA$ . Аналогично, в  $AAA$  можно перейти только из  $AAA, CAA, GAA, TAA$ .

Для Марковского процесса сумма элементов каждой строки матрицы перехода будет равна единице.

$$P = \begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & p_{25} & p_{26} & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & p_{rr} \end{pmatrix}$$

$$\forall i = 1, 2, \dots, r \quad \sum_{j=1}^r p_{ij} = 1$$

Возводя такую матрицу в степень  $n$ , в каждой ее ячейке  $(i, j)$  получаем вероятность встретить слово  $j$  на расстоянии  $n$  от слова  $i$ . Действительно, при возведении  $P$  во вторую степень в  $(i, j)$ -ячейке получим  $\sum_{s=1}^r p_{is}p_{sj}$ , то есть сумму вероятностей попасть сначала из  $i$  в  $s$ , а затем из  $s$  в  $j$  через всевозможные  $s$ .

Далее, если рассуждать по индукции, получим, что в каждой  $(i, j)$ -ячейке  $n$ -ной степени матрицы  $P$  находится сумма вероятностей всевозможных маршрутов из  $i$  в  $j$  длиной  $n$ .

Но всевозможные маршруты могут включать в себя и те, в которых встретится слово  $\omega_2$  раньше, чем на  $n$ -ном шаге. Для того, чтобы исключить данные случаи, столбец матрицы  $P$ , соответствующий переходу в слово  $\omega_2$ , занулим, то есть заменим все  $m = 4$  его ненулевых элемента нулями.

$$P_{\omega_2} = \begin{matrix} & \omega_2 \\ \begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} & \dots & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & 0 & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & 0 & \dots & p_{rr} \end{pmatrix} \end{matrix}$$

Докажем по индукции, что в каждой  $(i, j)$ -ячейке  $n$ -ной степени такой матрицы  $P_{\omega_2}$  находится сумма вероятностей **всевозможных** маршрутов из  $i$  в  $j$  длиной  $n$  таких, что слово  $\omega_2$  не встречается ни разу.

База индукции. Возведем матрицу  $P_{\omega_2}$  в квадрат. В большинстве своем такая матрица совпадет с квадратом матрицы  $P$ . Элементы же, в получении которых участвовали  $m = 4$  искусственно зануленных ячейки  $P_{\omega_2}$ , будут как раз не учитывать вероятности перехода **из** и **в** слово  $\omega_2$ . То есть при умножении строки  $i$  с зануленным элементом на обычный столбец  $j$  получится сумма вероятностей пройти по маршруту  $i \rightarrow s \rightarrow j$  для всех  $s = 1, \dots, n$  кроме  $s = \omega_2$ , т.к.  $p_{i\omega_2} := 0$ , тем самым «лишний» ноль поможет не учесть вероятность перехода **в** слово  $\omega_2$ . А при умножении на столбец, соответствующий переходу в слово  $\omega_2$ , получится просто ноль, так как переход в это слово запрещен.

Шаг индукции. Допустим, матрица  $P_{\omega_2}^l$  в каждой своей ячейке имеет нужную, правильную вероятность. Домножим ее справа на  $P_{\omega_2}$ . Тем самым при умножении строки  $i$  на обычный столбец  $j \neq \omega_2$  получится вероятность пройти любым маршрутом длины  $l + 1$  из  $i$  в  $j$ , не побывав ни разу в  $\omega_2$ . А при умножении на зануленный столбец получится ноль.

Если домножить  $P_{\omega_2}^{n-1}$  в конце не на  $P_{\omega_2}$  еще раз, а на исходную матрицу  $P$ , то в последней точке маршрута будет разрешен переход в слово  $\omega_2$ , так как соответствующий столбец не занулен. Таким образом матрица

$$P_{\omega_2}^{n-1} \cdot P$$

содержит вероятности попадания из одного слова в другое маршрутом длины  $n$  с условием возможности посещения пункта  $\omega_2$  только на последнем шаге. То есть в ее столбце под номером  $\omega_2$  содержатся значения искомой функции  $f_{\omega_1, \omega_2}(n)$  для всевозможных слов  $\omega_1$ .

Чтобы получить одну единственную вероятность того, что слово  $\omega_2$  впервые встретится на расстоянии  $n$  от слова  $\omega_1$  можно просто взять нужный элемент полученной матрицы  $P_{\omega_2}^n$  с координатами  $(\omega_1, \omega_2)$ . А можно воспользоваться формулой

$$f_{\omega_1, \omega_2}(n) = a_{\omega_1} \cdot P_{\omega_2}^{n-2} \cdot b_{\omega_2}, \quad (1)$$

где  $a_{\omega_1}$  — это строка перехода из слова  $\omega_1$ , выписанная из матрицы  $P_{\omega_2}$ , а  $b_{\omega_2}$  — изначальный столбец перехода в слово  $\omega_2$ , взятый из матрицы  $P$ , который был потом занулен.

Например, для слов  $\omega_1 = AAA$  и  $\omega_2 = AAC$  выглядеть это будет так:

$$f_{\omega_1, \omega_2}(n) = (p_{11} \ 0 \ p_{13} \ p_{14} \ 0 \ \dots \ 0) \times \\ \times \begin{pmatrix} p_{11} & 0 & p_{13} & p_{14} & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & p_{25} & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ \dots & 0 & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & 0 & \dots & p_{rr} \end{pmatrix} \times \\ \times \begin{pmatrix} p_{12} \\ 0 \\ 0 \\ \dots \\ 0 \end{pmatrix} \quad (2)$$

Умножение слева на конкретную строку  $a_{\omega_1}$  вместо матрицы  $P_{\omega_2}$  дает в итоге строку перехода из конкретного слова  $\omega_1$  в любое другое, а умножение этой строки справа на столбец  $b_{\omega_2}$  перехода из любого слова в  $\omega_2$  (вместо умножения на матрицу  $P$ ) дает в результате один единственный элемент — искомую вероятность, равную значению конкретной ячейки  $(\omega_1, \omega_2)$  матрицы  $P_{\omega_2}^{n-1} \cdot P$ .

## 2.2 Случай второй: $t > k$

Ситуация, когда порядок процесса  $t$  больше длины  $k$  слов  $\omega_1$  и  $\omega_2$ , аналогична предыдущей. Но теперь матрица  $P$  будет иметь размерность  $r * r$ , где  $r = m^t = 4^t$ . В каждой ее строчке и в каждом столбце будет все так же по  $m = 4$  ненулевых элемента, но каждому слову будет соответствовать не одна, а  $m^{t-k} = 4^{t-3}$  строчек и столько же столбцов. Например, если  $t = 5$ , то переходу из слова  $\omega_1 = AAA$  в любое другое будут соответствовать  $m^{t-k} = 4^2 = 16$  строчек, связанных со словами  $AAAAA, ACAAA, AGAAA, ATAAA, CAAAA, CCAAA, CGAAA, CTA AA, GAAAA, GCAAA, GGAAA, GTAAA, TAAAA, TCAAA, TGAAA, TTA AA$ .

Аналогично, чтобы запретить переход в какое-нибудь слово  $\omega_2$ , нужно в матрице  $P$  обнулить соответствующий квазистолбец, то есть все  $m^{t-k} = 4^{t-3}$  столбцов, соответствующих переходу **в** слово  $\omega_2$ . Получив таким образом матрицу  $P_{\omega_2}$  можем опять-таки возвести ее в степень  $n - 1$  и домножить на  $P$ , либо

воспользоваться формулой из предыдущего пункта:

$$f_{\omega_1, \omega_2}(n) = a_{\omega_1} \cdot P_{\omega_2}^{n-2} \cdot b_{\omega_2},$$

где  $a_{\omega_1}$  теперь уже не строка, а своеобразная горизонтальная матрица (квазистрока) размера  $m^{t-k} * m^t$ , составленная из строк перехода из нужного слова  $\omega_1$  в любое другое, а  $b_{\omega_2}$  - вертикальная матрица (квазистолбец) размера  $m^t * m^{t-k}$ , состоящая из тех самых обнуленных столбцов.

В результате применения такой формулы теперь уже получится не один элемент, а квадратная матрица размера  $m^{t-k} * m^{t-k}$ , содержащая в себе вероятности всевозможных переходов из слова  $* \dots * \omega_1$  в слово  $* \dots * \omega_2$  (оба длины  $k$ , символом  $*$  обозначен произвольный символ алфавита  $\aleph$ ) длиной в  $n$  символов, при условии, что слово  $\omega_2$  встретилось впервые на расстоянии  $n$  от слова  $\omega_1$ .

### 2.3 Понятие индикаторной матрицы

Назовем **индикаторной** матрицу  $P^{\text{ind}}$ , в которой элемент  $(i, j)$  равен единице, если переход из слова  $i$  в слово  $j$  за один шаг фактически возможен, и нулю, если это невозможно.

$$P^{\text{ind}} = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix}$$

Такая матрица будет состоять только лишь из нулей и единиц и будет очень похожа на обычную матрицу  $P$ , в которой все ненулевые элементы заменены на единицы (за исключением тех случаев, когда само значение вероятности перехода  $p_{ij} = 0$ , хотя переход возможен). Очевидно, разным  $P$  может соответствовать одна и та же  $P^{\text{ind}}$ , так как в ней по сути не важны сами значения вероятностей, роль играет только сам факт того, возможен переход или нет. Более того, для фиксированных значений  $m, k, t$  можно получить только одну матрицу  $P^{\text{ind}}$  — размера  $m^k * m^k$  в случае  $t \leq k$  или  $m^t * m^t$  при  $t > k$ .

Возведение такой матрицы  $P^{\text{ind}}$  в некоторую натуральную степень  $n$  даст матрицу, в каждой  $(i, j)$ -ячейке которой будет содержаться количество всевозможных путей из  $i$  в  $j$  длиной  $n$ . Докажем это.

Так как, рассуждая по индукции, получим, что при умножении  $P^{\text{ind}^l}$  на  $P^{\text{ind}}$  значение  $(i, j)$ -ячейки матрицы  $P^{\text{ind}^{l+1}}$  становится равным  $\sum_{s=1}^x p_{is}^l p_{sj}$  (здесь за  $p_{is}^l$  обозначен соответствующий элемент матрицы  $P^{\text{ind}^l}$ ). То есть, если предположить, что  $p_{is}^l$  — это действительно количество всевозможных путей из  $i$  в  $s$ , тогда значение произведения  $p_{is}^l p_{sj}$  обратится в ноль, если переход из  $s$  в  $j$  за один шаг не возможен, и останется равным  $p_{is}^l$ , если такой переход может быть

осуществлён. Таким образом  $\sum_{s=1}^x p_{is}^l p_{sj}$  соберет в себе только некоторые значения  $p_{is}^l$ , причем значение этой суммы в точности окажется равным количеству всевозможных путей из  $i$  в  $j$  за  $l+1$  шагов.

Если в индикаторной матрице обнулить столбец перехода в слово  $\omega_2$  (матрица  $P_{\omega_2}^{\text{ind}}$ ), обозначить за  $a_{\omega_1}^{\text{ind}}$  индикаторную строку перехода из  $\omega_1$ , за  $b_{\omega_2}^{\text{ind}}$  — индикаторный столбец перехода в  $\omega_2$ ; тогда можно получить формулу числа всевозможных цепочек длины  $n$  таких, что в начале стоит слово  $\omega_1$ , в конце —  $\omega_2$ , и нигде внутри слово  $\omega_2$  не встречается:

$$g_{\omega_1, \omega_2}(n) = a_{\omega_1}^{\text{ind}} \cdot P_{\omega_2}^{\text{ind}^{n-2}} \cdot b_{\omega_2}^{\text{ind}}$$

Справедливость этой формулы, а также справедливость соответствующих утверждений для случая  $t > k$  доказывается совершенно аналогично — методом математической индукции.

### 2.4 Собственные значения

Сумма диагональных элементов любой квадратной матрицы называется следом этой матрицы и равна сумме ее собственных значений. Это свойство будет выполняться и для исследуемых матриц  $P$ ,  $P_{\omega_2}$ ,  $P^{\text{ind}}$  и  $P_{\omega_2}^{\text{ind}}$ .

Рассмотрим для начала индикаторные матрицы. На диагонали матрицы  $P^{\text{ind}}$  будет  $m = 4$  единицы и все остальные нули. И она будет иметь только одно ненулевое собственное значение, равное, очевидно,  $m$ .

В зависимости от того, какую строку индикаторной матрицы будем обнулять, получая  $P_{\omega_2}^{\text{ind}}$ , диагональные элементы либо не изменятся, либо один из них обнулится. Соответственно, след такой матрицы будет равен либо  $m$ , либо  $m-1$ . Но такая матрица, что интересно, будет иметь уже несколько (предположительно  $k$ ) ненулевых собственных значений.

Теперь рассмотрим матрицы, задающие исследуемые Марковские процессы. Основываясь на результатах вычислений для некоторых примеров матриц  $P$  и  $P_{\omega_2}$ , можно утверждать следующее. Все собственные значения матрицы  $P$  будут ненулевыми. А после обнуления одного из столбцов среди собственных значений уже появятся нули. Первое утверждение в общем случае (когда порядок процесса больше длины слова  $t \geq k$ ) верно, так как исходная матрица  $P$  не будет содержать одинаковых строк, а значит в нашем случае можно утверждать, что ее определитель  $\det(P) \neq 0$ , следовательно, матрица  $P$  не имеет нулевых собственных значений.



### 3 Эксперимент

#### 3.1 Подготовка

Для вычисления значений функции (1) выберем порядок Марковского процесса  $t = 3$  и длину слов  $k = 3$ , а также определим максимальное расстояние между словами  $n = 2000$ .

Результаты эксперимента хотелось бы сравнить с данными, полученными на реальных генетических последовательностях. Действительно, имея реальный геном в качестве конечной последовательности символов длины  $N$  алфавита  $\aleph = \{A, C, G, T\}$  для любых выбранных слов  $\omega_1$  и  $\omega_2$  длины  $k$  можем посчитать, сколько раз слово  $\omega_2$  встретилось справа от  $\omega_1$  на расстоянии  $n$  ( $n \ll N$ ) и вычислить вероятность такого события.

Поэтому генерируемая матрица перехода  $P$  процесса, должна быть как-то связана с процессом, который мы собираемся смоделировать. То есть необходимо допустить предположение, что реальный процесс построения генома вызван Марковским процессом  $k$ -го порядка, в рамках такого процесса найти значения теоретической функции распределения  $f_{\omega_1, \omega_2}(n)$  и сравнить их со значениями экспериментальной функции  $f_{\omega_1, \omega_2}^{\text{real}}(n)$ , полученными при исследовании реальной конечной последовательности нуклеотидов.

Для исследования был выбран геном бактерии *Streptococcus pyogenes*, содержащий 1899877 нуклеотидов (579618 A; 368403 C; 366567 G; 585289 T; 0 other).

Рассмотрим геном как конечную последовательность символов алфавита  $\aleph = \{A, C, G, T\}$ . По этой последовательности символов построим частотный словарь для слов длины  $k + 1 = 4$ , где каждому слову  $\omega$  сопоставим частоту  $F_\omega$ , с которой оно встречается в данной последовательности. По такому частотному словарю можно составить матрицу перехода для Марковского процесса порядка  $k = 3$ . Вероятность перехода из слова  $\omega_1 = \nu_1 \nu_2 \nu_3$  в слово  $\omega_2 = \nu_2 \nu_3 \nu_4$  (слова пересекаются по символам  $\nu_2$  и  $\nu_3$ , иначе из  $\omega_1$  невозможно перейти в  $\omega_2$ , то есть  $p_{\omega_1 \omega_2} = 0$ ) можно вычислить по формуле

$$p_{\omega_1 \omega_2} = \frac{F_{\nu_1 \nu_2 \nu_3 \nu_4}}{\sum_{x \in \{A, C, G, T\}} F_{x \nu_2 \nu_3}} \quad (3)$$

Действуя таким образом построим матрицу перехода  $P$ .

Для вычислительного эксперимента выберем несколько пар слов  $\omega_1, \omega_2$  по определенному принципу. Например, рассмотрим частотный словарь слов длины  $k = 3$ , отсортированный по возрастанию частот, и выберем слова, которые встречаются наиболее и наименее часто, а также слова из середины

этого списка. Кроме того в выборку должны войти не только обычные пары, но и комплиментарные палиндромы — такие пары слов, которые читаются одинаково в противоположных направлениях с учетом замены нуклеотидов по правилу Чаргаффа (например, AGT-ACT). Комплиментарные палиндромы отличаются тем, что вероятности появления слов из такой пары в исходной последовательности примерно равны.

Для эксперимента нами были отобраны 6 пар: AAA-TTT, CAA-TTA, ACT-AGT, AGT-CAG, CCC-GGG, ACG-CGT — по 2 из часто, средне и редко встречающихся, причем каждая вторая — комплиментарный палиндром.

Далее для каждой пары выберем из матрицы  $P$  столбец  $b_{\omega_2}$ , в самой матрице этот столбец обнулим, получив таким образом матрицу  $P_{\omega_2}$ , а из нее выпишем строку  $a_{\omega_1}$ .

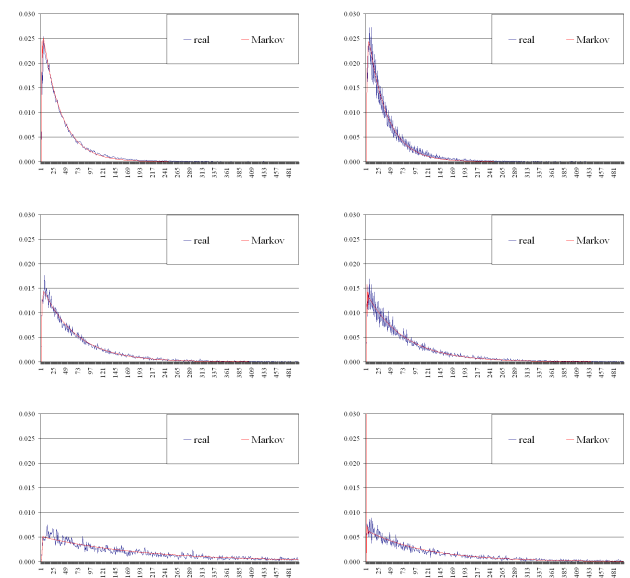


Рис. 1: Графики теоретической  $f_{\omega_1, \omega_2}(n)$  и экспериментальной  $f_{\omega_1, \omega_2}^{\text{real}}(n)$  функций распределения расстояний между словами, выбранными в качестве соседей. По оси абсцисс отложено расстояние между словами, по оси ординат — значения функций. Графики соответствуют парам слов (слева-направо, сверху-вниз): часто встречающиеся AAA-TTT, CAA-TTA, средне ACT-AGT, AGT-CAG, редко CCC-GGG, ACG-CGT.

#### 3.2 Особенности программной реализации

Для проведения вычислительного эксперимента были реализовано конструирование матрицы перехода  $P$  по соответствующему частотному словарю и вычисление значений функции (1) для некоторых пар слов  $\omega_1, \omega_2$ . В целом, программная реализация этих алгоритмов не является сложной. Но стоит заострить

внимание на некоторых моментах.

Во-первых, в следствии того, что вычисление значений (1) требует обработки больших матриц и возведения их в степень, были введены некоторые ограничения. Принято решение рассматривать не все возможные пары  $\omega_1, \omega_2$ , а лишь несколько (критерии, по которым отбирались слова, описаны выше). Также пришлось ограничиться довольно небольшим значением максимального расстояния между словами, выбранными в качестве соседей, так как это и есть максимальная степень, в которую придется возвести матрицу  $P$ .

Во-вторых, следует заметить, что в каждой строке матрицы перехода  $P$  всего  $m = 4$  ненулевых элемента. То есть в случае, когда длина слова равна порядку процесса  $k = t = 3$  доля ненулевых элементов в матрице  $P$  равна  $\frac{m \cdot m^t}{m^{2t}} = \frac{256}{4096} = \frac{1}{16}$ , а с увеличением порядка процесса процент нулевых элементов только возрастет. Матрицу, содержащую такую большую долю нулевых элементов можно назвать **разреженной**.

Хранение лишних нулей не целесообразно, а возведение такой матрицы в степень по обычным правилам содержит слишком много ненужных умножений на ноль. Для оптимизации программы каждая разреженная матрица  $P$  была представлена в виде элемента структуры, состоящего из трех массивов:  $P.\text{el}$ ,  $P.\text{row}$  и  $P.\text{col}$ .  $P.\text{el}$  хранит сами значения ненулевых элементов,  $P.\text{row}$  — их номера строк, а  $P.\text{col}$  — номера столбцов. Тогда  $k$ -тый ненулевой элемент матрицы  $P$   $p[i][j]$  представляется в виде  $i = P.\text{row}[k]$ ,  $j = P.\text{col}[k]$ ,  $p[i][j] = P.\text{el}[k]$ .

Таким образом объем памяти, выделяемый под каждую исходную матрицу может быть уменьшен примерно в 10 раз. Хотя, конечно, при возведении рассматриваемой разреженной матрицы в степень количество ненулевых элементов заметно сокращается.

К тому же использование специальной функции умножения на разреженную матрицу позволяет сократить количество затратных операций умножения в несколько раз (за счет дополнительных операций сравнения).

### 3.3 Результаты

На рис.1 представлены графики теоретической функции  $f_{\omega_1, \omega_2}(n)$  распределения расстояний между словами  $\omega_1$  и  $\omega_2$  и значения экспериментальной функции  $f_{\omega_1, \omega_2}^{\text{real}}(n)$  для тех же слов.

Как мы видим, характер теоретической и реальной функций совпадает для каждой пары слов. Причем, пары графиков для слов, которые встречаются с примерно одинаковой частотой, похожи друг на друга. Также стоит заметить, что график распределения, построенного на реальной последовательности, име-

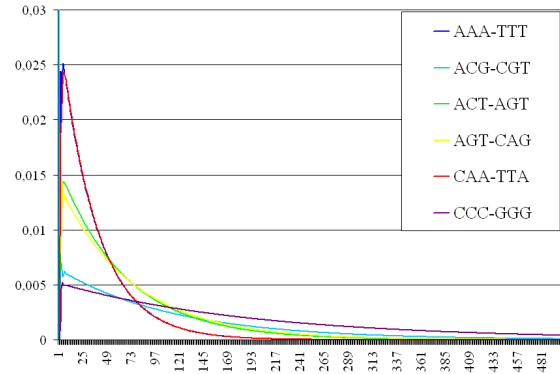


Рис. 2: Значения функции распределения расстояния между словами  $f_{\omega_1, \omega_2}(n)$ , полученные в ходе вычислительного эксперимента. Графики соответствуют выбранным парам слов  $\omega_1, \omega_2$ .

$\omega_1$	$\omega_2$	$a$	$b$
AAA	TTT	-0.0271	-3.77
CAA	TTA	-0.0271	-3.76
ACT	AGT	-0.0153	-4.25
AGT	CAG	-0.0139	-4.35
CCC	GGG	-0.0051	-5.30
ACG	CGT	-0.0087	-5.12

Таблица 1: Коэффициенты  $a$  и  $b$  в формуле аппроксимации теоретической функции распределения  $f_{\omega_1, \omega_2}(n) \approx e^{an+b}$  для исследуемых пар слов  $\omega_1, \omega_2$ .

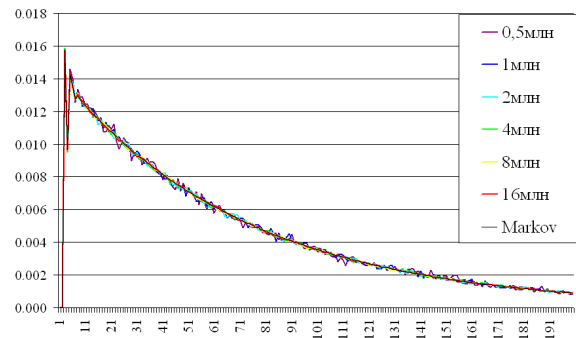


Рис. 3: График экспериментальных значений функции  $f_{\omega_1, \omega_2}^{\text{real}}(n)$  для разных длин исходных последовательностей и теоретической функции  $f_{\omega_1, \omega_2}(n)$ .

ет характерные отклонения, которые нельзя назвать просто флуктуациями.

Собственно, эти пики и представляют основной интерес. Заметим, что незначительное увеличение порядка процесса (с Зеге до бго, например) не поможет смоделировать их все.

Можно предположить, что реальная функция  $f_{\omega_1, \omega_2}^{\text{real}}(n)$  складывается из суммы теоретической

$N$	$A$	$\lambda$
0.5	0.0013	0.007
1	0.0009	0.007
2	0.0006	0.007
4	0.0005	0.007
8	0.0003	0.007
16	0.0002	0.007

Таблица 2: Соответствие длины исходной последовательности  $N$  и параметров  $A$  и  $\lambda$  в формуле  $Ae^{-\lambda n}$  аппроксимации стандартного отклонения

функции  $f_{\omega_1, \omega_2}(n)$ , определяющий общий характер, и еще одного, пока неизвестного закона. Причем основной закон начиная с некоторого расстояния (порядка длины слова  $k$ ) начинает вести себя практически как показательная функция, то есть хорошо аппроксимируется функцией вида  $f_{\omega_1, \omega_2}(n) \approx e^{an+b}$ . Параметры  $a$  и  $b$  для каждой пары  $\omega_1, \omega_2$  вычисляются отдельно на основании матрицы  $P_{\omega_2}$ .

Рис. 2 представляет собой сводный график теоретических значений функции  $f_{\omega_1, \omega_2}(n)$  для всех исследуемых пар слов  $\omega_1, \omega_2$ . В таблице 1 представлены значения коэффициентов  $a$  и  $b$  для выбранных пар  $\omega_1, \omega_2$ . Можно заметить, что  $f_{AAA, TTT}(n)$  и  $f_{CAA, TTA}(n)$  практически совпадают, и в целом значения коэффициентов аппроксимации  $a$  и  $b$  очень близки для пар слов, имеющих схожую частоту встречаемости в исходной последовательности.

### 3.4 Погрешности

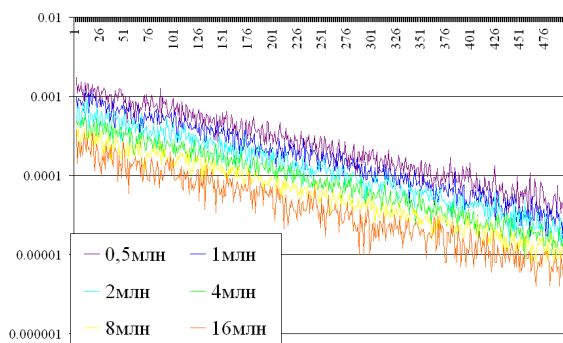


Рис. 4: Стандартное отклонение для разных длин исходной последовательности в логарифмических координатах. По оси абсцисс отложено расстояние между словами.

Первостепенным является вопрос, не укладываются ли наблюдаемые флуктуации в коридор погрешностей, вызванных тем, что реальная последовательность не бесконечна.

Попробуем выяснить, как именно зависит погрешность экспериментальных значений функции

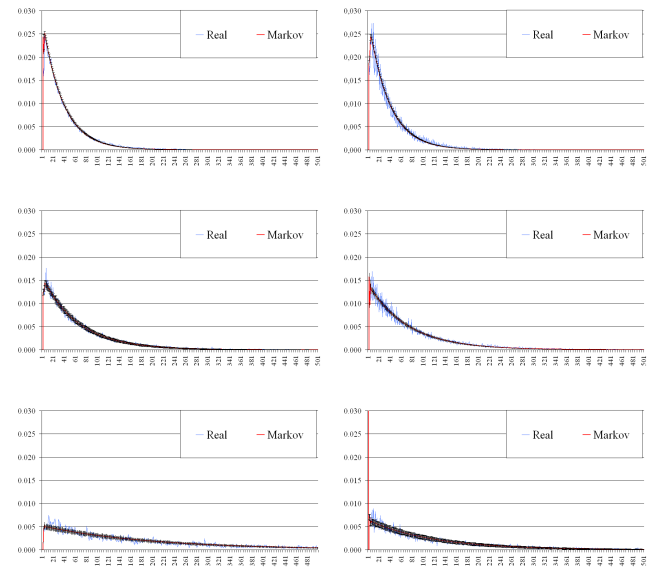


Рис. 5: Графики теоретической  $f_{\omega_1, \omega_2}(n)$  и экспериментальной  $f_{\omega_1, \omega_2}^{real}(n)$  функций распределения расстояний между словами, выбранными в качестве соседей. По оси абсцисс отложено расстояние между словами, по оси ординат — значения функций. На графике теоретической функции отмечена погрешность, вызванная конечностью исходного гена. Графики соответствуют парам слов (слева-направо, сверху-вниз): часто встречающиеся AAA-TTT, CAA-TTA, средне ACT-AGT, AGT-CAG, редко CCC-GGG, ACG-CGT.

$f_{\omega_1, \omega_2}^{real}(n)$  от длины  $N$  исходной последовательности.

Для этого выберем несколько значений  $N$  — длины исходной последовательности. Например,  $N_0 = 500000$ ,  $N_1 = 1000000$ ,  $N_2 = 2000000$ ,  $N_4 = 4000000$ ,  $N_8 = 8000000$ ,  $N_{16} = 16000000$ . Сгенерируем по 16 последовательностей, заданных Марковским процессом 3 порядка, для всех длин  $N_i$ .

По этим последовательностям найдем значения функций  $f_{\omega_1, \omega_2}^{real}(n)$ . Для каждой последовательности  $N_i$  в каждой точке  $n$  найдем среднее значение по 16 вариантам значений  $f_{\omega_1, \omega_2}^{real}(n)$  и стандартное отклонение от этого среднего.

На рис.3 представлены графики средних  $f_{\omega_1, \omega_2}^{real}(n)$  для одной и той же пары слов  $\omega_1, \omega_2$  для разных длин  $N_i$  исходных последовательностей. Они, очевидно, сходятся к гладкой кривой, а именно — к  $f_{\omega_1, \omega_2}(n)$ . На рис.4 — величины стандартного отклонения для разных длин  $N_i$  в логарифмических координатах. Как мы можем заметить, с ростом  $n$  СКО убывает экспоненциально, и, если аппроксимировать точки отклонения законом  $Ae^{-\lambda n}$ , то параметр  $A$  будет экспоненциально зависеть от длины исходной последовательности  $N$ , а  $\lambda$  — не будет зависеть никак (см. таб.2).

Аналогичным образом сгенерируем последовательности, совпадающие по длине с исследуемым геном, которые будут задаваться тем же частотным словарем, что и матрица  $P$ . Вычислим по ним среднее значение и стандартное отклонение. На рис.5 видно, что величина стандартного отклонения не всегда будет превышать величину флуктуаций реального графика, то есть некоторые пики “выпрыгивают” за коридор погрешности.

## 4 Выводы

Вычислительный эксперимент по определению поведения функции  $f_{\omega_1, \omega_2}(n)$  показал, что характеры теоретической и экспериментальной зависимостей для одинаковых пар слов совпадают. Слова, имеющие близкие вероятности появления в исходной последовательности, имеют близкий характер распределения расстояний между соседями. Погрешность вычислительного эксперимента, вызванная конечностью исходной последовательности, экспоненциально убывает с увеличением длины исходной последовательности. Экспериментальные значения  $f_{\omega_1, \omega_2}(n)$ , вычисленные для реального гена, не укладываются в этот коридор погрешностей.

# Когнитивное моделирование в управлении рисками от лесных пожаров

**Коморовский Витольд Станиславович**

Сибирская пожарно-спасательная академия - филиал  
Санкт-Петербургского университета ГПС МЧС России  
Центр НИОКР  
Красноярск  
komorovski.w@mail.ru

**Доррер Георгий Алексеевич**

Сибирский государственный технологический университет  
Факультет автоматизации и информационных технологий  
Красноярск  
g\_a\_dorrer@mail.ru

**Аннотация.** Данная работа посвящена применению когнитивных моделей в управлении лесопожарными рисками. Рассматриваются вопросы построения когнитивной карты, анализа системы с ее помощью, а также разработка сценариев развития лесопожарной ситуации.

**Ключевые слова.** Когнитивные модели, Когнитивные карты, Лесные пожары, Риски, Устойчивость систем, Управляемость систем.

## 1 Введение

Одна из наиболее значительных проблем лесопользования - это лесные пожары. В России ежегодно происходит от 10 до 40 миллионов лесных пожаров [1]. Они причиняют огромный экономический и экологический ущерб, а также представляют опасность для населения. Любая хозяйственная деятельность в лесах требует оценки лесопожарного риска и, более того, управления лесопожарными рисками.

Оценка риска - это не тривиальная проблема. Обычно, риск определяется как произведение вероятности неблагоприятного события на возможный ущерб от него. Для лесопожарных рисков будем рассматривать следующие вероятности: вероятность возникновения пожара, вероятность позднего обнаружения пожара, вероятность распространения пожара и вероятность неэффективного тушения пожара. На перечисленные вероятности влияет множество факторов, учесть которые довольно трудно.

Целью данной статьи является рассмотрение возможного подхода к решению проблемы оценки вероятностей, связанных с лесопожарным риском. Подход базируется на рассмотрении системы «лесной пожар - объекты экономики», представленной в виде когнитивной карты, создаваемой на основе обработки мнений экспертов.

## 2 Построение когнитивной модели

Когнитивные карты и основанный на их использовании метод когнитивного моделирования - относительно новый, но уже широко используемый метод экспертного моделирования слабоструктурированных систем. Это связано с некоторыми его преимуществами. Метод позволяет представить рассматриваемую систему как взвешенный ориентированный граф. Соответственно, для исследования системы мы можем использовать методы теории графов. Кроме того, представление системы становится наглядным и формализованным.

Метод когнитивного моделирования, основан на использовании мнений экспертов, сформулированных в виде когнитивной карты нашли широкое применение при исследовании многих сложных слабоструктурированных систем [2], [3]. Методология когнитивного моделирования включает в себя два этапа:

- структуризацию изучаемой ситуации;
- модель предоставления знаний эксперта в виде знакового орграфа (когнитивной карты)  $M = \{X, W\}$ , где  $X = [x_1, x_2, \dots, x_n]^T$  — вектор множества базисных факторов (концептов) ситуации, оцениваемых, как правило, в безразмерных относительных единицах,  $W = \|W_{i,j}\|$ ,  $i, j = 1 \dots n$  — множество экспертно определяемых причинно-следственных отношений между факторами ситуации. Значок  $T$  означает транспонирование матрицы.

Поскольку рассматриваемая в работе предметная область пожарных рисков является высокодинамичной, мы, в отличие от [2], будем рассматривать базисные факторы как функции времени, т.е. полагать  $X(t) = [x_1(t), x_2(t), \dots, x_n(t)]$ . Элементы матрицы  $W_{i,j}$  в общем случае также зависят от времени. Они определяют влияние фактора  $x_i(t)$  на фактор  $x_j(t)$  и представляют собой частные производные:

$$W_{i,j}(t) = \frac{\partial x_i(t)}{\partial x_j(t)} \quad (1)$$

Рассмотрим изменение одного из факторов  $x_i(t)$  за временной шаг  $h$ . С учетом вышесказанного, имеем:



$$x_i(t + \Delta t) = x_i(t) + \sum_{j=1}^n x_j(t) \cdot W_{i,j}(t) \cdot h \quad (2)$$

Модель 2 может быть использована при расчете системы как в дискретном времени, в пошаговом режиме, так и в режиме непрерывного времени. В первом случае выбирается шаг расчета  $h$ , удобный для анализа системы, и время измеряется тактами:  $t = 0, 1, 2, \dots, n$ , а во втором случае величина  $h$  устремляется к нулю, и мы получаем дифференциальное уравнение.

В дискретном времени система 2 может быть представлена в более компактном матричном виде:

$$\begin{aligned} X(t + \Delta t) &= X(t) + W(t) \cdot X(t) \cdot h = \\ &= (I + W(t) \cdot h) \cdot X(t) \end{aligned} \quad (3)$$

где  $I$  — единичная  $n$ -матрица.

При расчете системы в непрерывном времени перенесем  $x_i(t)$  в левую часть уравнения и поделим его на  $h$ , при  $h \rightarrow 0$ , получим уравнение динамики фактора  $x_i(t)$ :

$$\frac{dx_i}{dt} = \sum_{j=1}^n x_j(t) \cdot W_{i,j}(t), i = 1, \dots, n \quad (4)$$

которое в матричной записи имеет вид:

$$\frac{dx}{dt} = X(t) \cdot W(t) \quad (5)$$

Системы 2, 3, 4 рассматриваются при заданных начальных условиях:

$$X(0) = [x_{10}, x_{20}, \dots, x_{n0}]^T \quad (6)$$

Построим когнитивную карту, оценивающую динамику рисков от лесных пожаров. На основе литературных данных и анализа мнений экспертов был выбран определенный набор факторов. Положим, что выходными параметрами будут вероятность возникновения пожара, вероятность позднего обнаружения пожара, вероятность распространения пожара и вероятность неэффективного тушения пожара. Представим эти вероятности в виде графических элементов и укажем прямые связи между ними. Далее, добавим в модель факторы, непосредственно влияющие на выходные переменные. И наконец, добавим на схему факторы, оказывающие косвенное влияние на выходные параметры. Взаимосвязи на схеме представлены

стрелками, а каждой стрелке присвоен вес, определяющий степень взаимосвязи между параметрами. Естественно, что веса стрелок нормированы. Таким образом, мы получили взвешенный ориентированный граф, который в дальнейшем будем называть когнитивной картой. Полученная когнитивная карта представлена на Рис. 1.

Полученная когнитивная карта может быть интерпретирована как система дифференциальных уравнений с дискретным или непрерывным временем. Соответствующие линейные дифференциальные уравнения приведены в таблице 1.

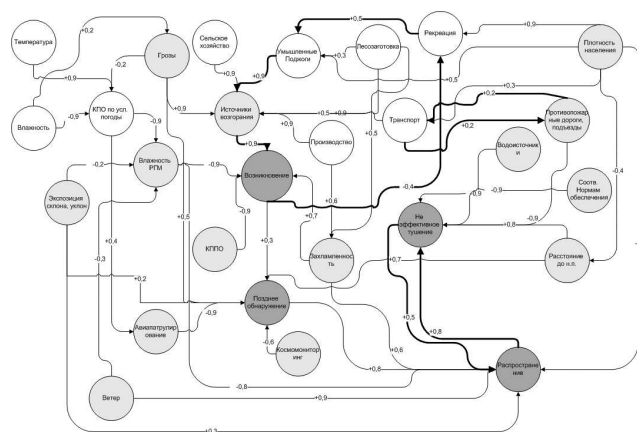


Рис. 1: Когнитивная карта для управления лесопожарными рисками.

### 3 Исследование системы с помощью когнитивной карты

Рассмотрим основные характеристики полученной системы. На Рис. 1 жирными стрелками выделены петли обратной связи. Есть два усиливающих цикла на графе (с положительной обратной связью) и один стабилизирующий (с отрицательной). Исследуем управляемость системы, для этого построим матрицу управляемости. В системе можно выделить 9 управляющих воздействий, это  $X_3, X_5, X_6, X_9, X_{10}, X_{11}, X_{12}, X_{20}, X_{21}$ , перечисленные в таблице 1. Перенумеровав, если нужно, факторы таким образом, чтобы входные факторы располагались в начале списка, представим вектор-столбец  $X(t)$  в следующем виде:

$$X(t) = [U(t) \ S(t)]^T \quad (7)$$

где  $U(t)$  —  $m$ -вектор входных факторов,  $S(x)$  —  $n$ - $m$ -вектор управляемых факторов. Матрица весов  $W$  может быть представлена в виде

Обозначение	Имя переменной	Уравнение
$X_1$	Температура воздуха	—
$X_2$	Грозы	$\frac{dX_2}{dt} = 0.2 \cdot X_{23}$
$X_3$	Производство	—
$X_4$	Поджоги	$\frac{dX_4}{dt} = 0.5 \cdot X_6 + 0.5 \cdot X_7 + 0.3 \cdot X_{21}$
$X_5$	Транспорт	$\frac{dX_5}{dt} = 0.3 \cdot X_7 + 0.2 \cdot X_9$
$X_6$	Рекреация	$\frac{dX_6}{dt} = 0.9 \cdot X_7 - 0.4 \cdot Y_1$
$X_7$	Население	—
$X_8$	Водоисточники	—
$X_9$	Противопожарные дороги и проезды	$\frac{dX_9}{dt} = 0.2 \cdot X_5$
$X_{10}$	Соответствие нормам обеспечения	—
$X_{11}$	Авиапатрулирование	$\frac{dX_{11}}{dt} = 0.4 \cdot X_{17}$
$X_{12}$	Космомониторинг	—
$X_{13}$	Расстояние до населенного пункта	$\frac{dX_{13}}{dt} = -0.4 \cdot X_7$
$X_{14}$	Влажность РГМ	$\frac{dX_{14}}{dt} = -0.3 \cdot X_{15} - 0.9 \cdot X_{17} - 0.2 \cdot X_{18}$
$X_{15}$	Ветер	—
$X_{16}$	Захламленность	$\frac{dX_{16}}{dt} = 0.6 \cdot X_3 + 0.5 \cdot X_{21}$
$X_{17}$	КПО по условиям погоды	$\frac{dX_{17}}{dt} = 0.9 \cdot X_1 - 0.2 \cdot X_2 - 0.9 \cdot X_{23}$
$X_{18}$	Экспозиция склона, уклон	—
$X_{19}$	Источники возгорания	$\frac{dX_{20}}{dt} = 0.2 \cdot X_1 + 0.9 \cdot X_2 + 0.9 \cdot X_3 +$ $+0.9 \cdot X_4 + 0.5 \cdot X_5 + 0.9 \cdot X_{20} + 0.9 \cdot X_{21}$
$X_{20}$	Сельское хозяйство	—
$X_{21}$	Лесозаготовки	—
$X_{22}$	КППО	—
$X_{23}$	Влажность	—
$X_{24}$	Вероятность возникновения	$\frac{dY_1}{dt} = -0.9 \cdot X_{14} + 0.7 \cdot X_{16} - 0.9 \cdot X_{22}$
$X_{25}$	Вероятность позднего обнаружения	$\frac{dY_2}{dt} = 0.5 \cdot X_2 - 0.9 \cdot X_{11} - 0.6 \cdot X_{12} +$ $+0.7 \cdot X_{13} + 0.2 \cdot X_{18} + 0.3 \cdot Y_1$
$X_{26}$	Вероятность распространения	$\frac{dY_3}{dt} = -0.2 \cdot X_7 - 0.8 \cdot X_{14} + 0.9 \cdot X_{15} +$ $+0.6 \cdot X_{16} + 0.3 \cdot X_{18} + 0.8 \cdot Y_2 + 0.5 \cdot Y_4$
$X_{27}$	Вероятность низкоэффективного тушения	$\frac{dY_4}{dt} = -0.9 \cdot X_8 - 0.9 \cdot X_9 - 0.9 \cdot X_{10} +$ $+0.8 \cdot X_{13} + 0.8 \cdot Y_3$

Таблица 1: Переменные когнитивной карты.

$$W = \begin{bmatrix} D & C \\ B & A \end{bmatrix}. \quad (8)$$

Здесь  $B$  — матрица размерности  $m \cdot (n - m)$ , определяющая воздействие входных факторов на управляемые;  $A$  — матрица размерности  $(n - m) \cdot (n - m)$ , определяющая взаимное влияние управляемых факторов;

$C$  — матрица размерности  $(n - m) \cdot m$ , определяющая влияние управляемых факторов на входные;

$D$  — матрица размерности  $m \cdot m$ , определяющая взаимное влияние входных факторов.

Примем, что матрица  $W$  неизменна за весь период исследования. Тогда динамику системы можно представить уравнением:

$$\begin{bmatrix} \frac{dU(t)}{dt} \\ \frac{dS(t)}{dt} \end{bmatrix} = \begin{bmatrix} D & C \\ B & A \end{bmatrix} \cdot \begin{bmatrix} U(t) \\ S(t) \end{bmatrix} = [B \cdot U(t) + A \cdot S(t) \quad D \cdot U(t) + C \cdot S(t)]^T \quad (9)$$

Матрица управляемости имеет вид:

$$Q_y = [B \quad A \cdot B \quad A^2 \cdot B \quad \dots \quad A^{n-1} \cdot B] \quad (10)$$

Если ранг матрицы  $Q_y = n$ , то система полностью управляема, т.е. мы можем с помощью некоторого ограниченного управляющего воздействия  $U(t)$  перевести систему за конечное время  $t_f$  из любого начального состояния  $X(0)$  в заданное конечное состояние  $X(t_f)$ . В нашем случае  $n = 23$ , а  $\text{rank}(Q_y) = 5$ . Система не является полностью управляемой.

Далее, исследуем устойчивость системы. Линейная система устойчива тогда и только тогда, когда действительная часть всех корней характеристического уравнения является отрицательной. Запишем характеристическое уравнение нашей системы:

$$p^8 - 0.44 \cdot p^6 + 0.178 \cdot p^4 - 0.0713 \cdot p^2 + 0.0026 = 0$$

Корни данного уравнения равны:

$$p_1 = 0.6325$$

$$p_2 = -0.6325$$

$$p_3 = 0.4486 + 0.4486i$$

$$p_4 = 0.4486 - 0.4486i$$

$$p_5 = -0.4486 + 0.4486i$$

$$p_6 = -0.4486 - 0.4486i$$

$$p_7 = 0.2000$$

$$p_8 = -0.2000$$

У характеристического уравнения два положительных вещественных корня и два комплексных корня с положительной действительной частью. Таким образом, наша система не устойчива. Неустойчивость обусловлена присутствием в системе циклов с положительной обратной связью.

Исследуем поведение системы в случае дискретного и непрерывного времени. Зададим вектор начальных условий  $X(0)$  размерности  $1 \times n$ . Пусть у нас изменяется только один параметр - лесозаготовки, а именно увеличиваются на 20%. Тогда вектор  $X(0)$  будет содержать нули, за исключением  $X(0)_{21} = 0.2$ . Результаты вычисления выходных параметров в дискретном времени с шагом 1 представлены на Рис. 2. По оси абсцисс представлено время (шаги), по оси ординат — изменения параметров в логарифмическом масштабе:  $\ln |X(t)|$ . Как мы можем видеть из Рис. 2,

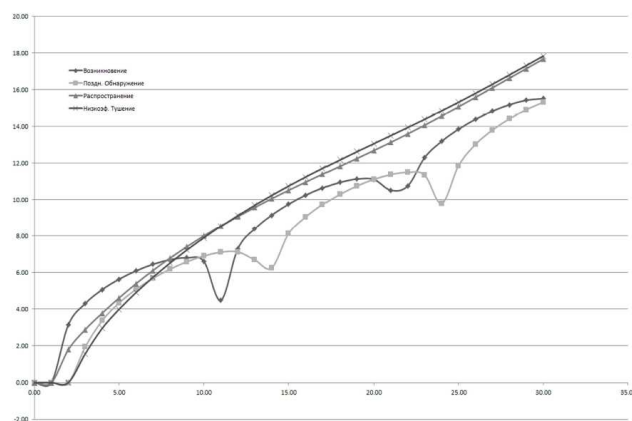


Рис. 2: Изменения выходных параметров. Модель с дискретным временем

система неустойчива и происходит нарастание колебаний выходных параметров.

Такая же картина наблюдается если рассматривать данную систему, как систему с непрерывным временем. Решим уравнения из таблицы 1 при начальных условиях  $X_{21} = 0.2$ , остальные  $X_i = 0$ . График численного решения методом Рунге-Кутты четвертого порядка до  $t = 30$  приведен на Рис. 3. В системе присутствует выраженный тренд, обусловленный наличием положительных корней характеристического уравнения и колебания, объясняемые наличием комплексных корней. Как можно видеть из Рис. 3, система неустойчива, колебания выходных параметров нарастают. Качественно картины на Рис. 2 и 3 не отличаются.

В данной работе рассмотрен простейший случай, когда на систему влияет только один параметр и изучается отклик системы. Когнитивные карты, естественно, пригодны для моделирования и более сложных сценариев, учитывающих много параметров и управляющих воздействий.

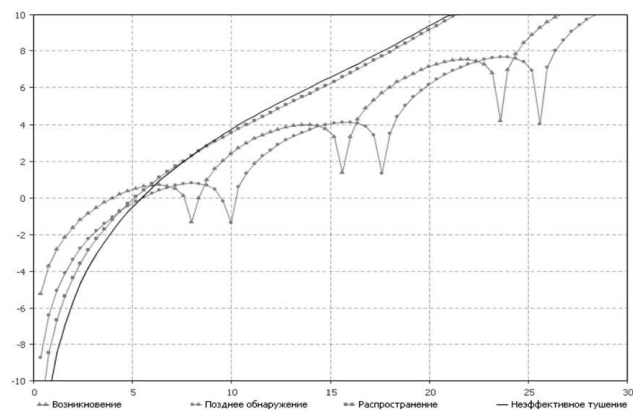


Рис. 3: Изменения выходных параметров. Модель с непрерывным временем

## 4 Заключение

Преимуществом когнитивных моделей является их дуализм: с одной стороны, это вербальные модели, хорошо объясняющие рассматриваемую систему. С другой стороны, это готовое, хотя и упрощенное, математическое описание объекта. Недостатком когнитивного подхода является то, что результат и выводы очень сильно зависят от правильности построения исходной карты. Существуют разные подходы к составлению и верификации когнитивных карт, таким образом, одна и та же система может быть описана по-разному. Тем не менее, на наш взгляд, когнитивное моделирование является весьма перспективным подходом в решении многих прикладных и теоретических задач, в том числе и для оценки лесопожарных рисков.

## Благодарности

Авторы выражают признательность Ю. А. Андрееву и А. В. Брюханову, как экспертам в области лесных пожаров.

## Список литературы

- [1] ФГУ Авиалесоохрана. *Информационная система дистанционного мониторинга федерального агентства лесного хозяйства [Электронный ресурс]*. <http://www.pushkino.aviales.ru>, Пушкино, 2012.
- [2] Ю. М. Плотинский. *Теоретические и эмпирические модели социальных процессов: Учебное пособие для высших учебных заведений*. Издательский дом «Логос», М., 1998.
- [3] Roberts F. *Applied Combinatorics: Edition 2*. CRC Press, Inc., Boca Ration, USA, 2009.

# Применение методов факторного анализа для исследования социально-экономического положения населения

Кучеров Матвей Николаевич

Сибирский Федеральный Университет  
Институт математики и фундаментальной информатики  
Красноярск  
matrix-elf@yandex.ru

**Аннотация.** В работе рассматривается понятие факторного анализа, его назначение и основные проблемы. Излагаются наиболее известные методы факторного анализа: метод главных компонент, центроидный метод и метод минимальных остатков. В работе решается практическая задача исследования многомерных статистических данных с помощью перечисленных методов — исследования социально-экономического положения населения Красноярского края.

**Ключевые слова.** Факторный анализ, центроидный метод, метод главных компонент, метод минимальных остатков, корреляция, фактор

## 1 Введение

Факторный анализ [1] — многомерный статистический метод, применяемый для изучения взаимосвязей между значениями переменных. Делается предположение о том, что несколько измеряемых переменных коррелируют между собой. Эта ситуация возникает в ситуации, когда они взаимно определяют друг друга или связь между этими переменными обуславливается какой-то третьей величиной, которую непосредственно измерить нельзя. Возникает задача, можно ли по данным переменным выделить величину, так называемый фактор, который объяснил бы наблюдаемые связи. Под термином «фактор» подразумевается математическая величина, получаемая на основе наблюдений.

Факторный анализ позволяет решить две важные проблемы исследователя: описать объект измерения *всесторонне* и в то же время *компактно*. С помощью факторного анализа возможно выявление скрытых латентных переменных факторов, отвечающих за наличие линейных статистических связей корреляций между наблюдаемыми переменными.

## 2 Области применения факторного анализа

Впервые факторный анализ был применен в психологии, но в последние годы методы факторного анали-

за стали все более широко применяться и в других областях знаний: социологии, метеорологии, медицине, географии, экономике, системном анализе, теории катастроф, технике связи, биологии, медицине, геологии и метеорологии.

Из этого краткого перечня областей знаний, использующих факторный анализ, наглядно видно, как много исследователей уже оценили мощь и возможности методов этого раздела многомерной статистики.

## 3 Основные понятия факторного анализа

В основе методики факторного анализа лежит анализ коэффициентов корреляции наблюдаемых переменных. Приведем некоторые определения, необходимые для постановки задачи факторного анализа и описания его методов.

При решении задачи измерения у  $n$  изучаемых объектов (или субъектов), измеряются два признака (или переменные  $x$  и  $y$ ). В результате имеем отдельные значения  $x_1; x_2; \dots; x_j; \dots; x_n$  и  $y_1; y_2; \dots; y_j; \dots; y_n$ .

**Определение 56.** Сумма квадратов отклонений отдельных значений наблюдаемой переменной от среднего значения:

$$S_{xx} = \sum_{j=1}^n (x_j - \bar{x})^2 = \sum_{j=1}^n x_j^2 - \frac{(\sum_{j=1}^n x_j)^2}{n},$$

где  $\bar{x} = \frac{1}{n} \sum_{j=1}^n x_j$  — среднее значение переменной.

**Определение 57.** Дисперсия случайной величины — мера разброса данной случайной величины, т.е. ее отклонения от математического ожидания:

$$s_x^2 = \frac{S_{xx}}{(n-1)}.$$

**Определение 58.** Стандартное отклонение наблюдаемой переменной

$$s_x = \sqrt{\frac{S_{xx}}{(n-1)}}$$



— является мерой среднего отклонения отдельных значений от их средней величины.

Сумма произведений отклонения находится по формулам:

$$S_{xy} = \sum_{j=1}^n (x_j - \bar{x})(y_j - \bar{y}) = \sum_{j=1}^n x_j y_j - \frac{\sum_{j=1}^n x_j \sum_{j=1}^n y_j}{n}.$$

**Определение 59.** Ковариация — мера линейной зависимости двух случайных величин:

$$s_{xy} = \frac{S_{xy}}{n-1}.$$

**Определение 60.** Коэффициент корреляции — мера линейной взаимосвязи между двумя измеряемыми величинами:

$$r_{xy} = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{s_{xy}}{s_{xx}s_{yy}},$$

где  $S_{xy}$  — сумма произведений отклонений, а  $s_{xy}$  — ковариация.

$r_{xy}$  может принимать значения между +1 и -1.

**Определение 61.** Факторным отображением (factor pattern) называется матрица  $A = (a_{ij})$ , элементы которой являются факторными нагрузками (factor loadings). Каждый фактор характеризуется столбцом, каждая переменная — строкой матрицы  $A$ .

**Определение 62.** Фактор называется генеральным (general factor), если все его нагрузки значительно отличаются от нуля. Фактор называется общим (common factor), если хотя бы две его нагрузки значительно отличаются от нуля. Фактор называется характерным (unique factor), если только одна его нагрузка значительно отличается от нуля.

**Определение 63.** Сумма квадратов нагрузок общих факторов называется общностью (communality). Обозначается:  $h^2 = a_{11}^2 + \dots + a_{ir}^2$ . Общность представляет собой часть единичной дисперсии переменной, которую можно приписать общим факторам. Она равна квадрату коэффициента множественной корреляции между переменной и общими факторами.

**Определение 64.** Доля дисперсии, которая соответствует квадрату нагрузки определенного характерного фактора называется характерностью. Обозначается  $u_i^2 = 1 - h_i^2$

### 3.1 Схема решения и основные проблемы факторного анализа

На Рис. 1 представлена схема последовательных действий, выполняемых при решении задач факторного анализа [2].

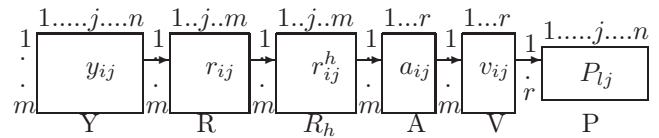


Рис. 1: Схема факторного анализа

Здесь  $Y = (y_{ij})$  — матрица исходных данных,  $i = 1, \dots, m$  — переменные,  $j = 1, \dots, n$  — индивидуумы (объекты),  $R = (r_{ik})$  — корреляционная матрица,  $i, k = 1, \dots, m$  — коэффициенты корреляции,  $R_h = (r_{ik}^h)$  — редуцированная корреляционная матрица,  $i, k = 1, \dots, m$  — остаточные коэффициенты корреляции,  $A = (A_{ij})$  — матрица отображения, элементами которой являются факторные нагрузки,  $l = 1, \dots, r$  — факторы,  $i = 1, \dots, m$  — переменные,  $V = (v_{ij})$  — факторная матрица после поворота,  $l = 1, \dots, r$  — факторы,  $i = 1, \dots, m$  — переменные,  $P = (p_{ij})$  — матрица значений факторов,  $l = 1, \dots, r$  — факторы,  $j = 1, \dots, n$  — индивидуумы (объекты).

Любой метод факторного анализа начинается с  $Y$  — матрицы исходных данных. По ней вычисляется корреляционная матрица  $R$ . По главной диагонали корреляционной матрицы проставляются оценки общностей и получают  $R_h = (r_{ik}^h)$ . Это составляет проблему общности, которая заключается в установлении оценок общностей  $h_i^2$ . Это самая первая проблема, которая возникает в ходе факторного анализа.

Из  $R_h$  с помощью определенных способов извлекают факторы, получая в результате матрицу  $A$ . Столбцы матрицы  $A$  ортогональны и занимают произвольную позицию в отношении переменных, определяемую методом выделения факторов. Возможно большое число матриц  $A$ , которые будут одинаково хорошо воспроизводить  $R_h$  по равенству  $R_h = A \cdot A'$ . Из них должна быть выбрана одна, что составляет проблему вращения. Решение проблемы вращения одним из нескольких способов приводит к матрице  $V$ . И, наконец, последняя проблема касается оценки значений факторов для каждого индивидуума.

## 4 Метод главных компонент

### 4.1 Геометрическая интерпретация метода

При решении задачи измерения трех нормально распределенных параметров  $y_n$  индивидуумов, получается ситуация, изображенная на Рис. 2, где  $n$  точек сосредоточены в трехмерном пространстве с тремя осями переменными  $X; Y; Z$  в облаке вокруг общего центра тяжести. Это облако точек наблюдений в общем случае имеет овальную форму и называется эллипсоидом. На Рис. 2 изображено овальное тело с тремя секущими плоскостями (различно заштрихованными), проходящими через центр тяжести.



виде — характеристическим уравнением. Все  $m$  корней этого уравнения действительны, т. е.  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m \geq 0$  являются возможными, иногда совпадающими решениями. После подстановки в (3) найденного значения  $\lambda_1$  получим вектор частного, нетривиального решения  $(\alpha_{11}, \dots, \alpha_{m1})$ , который (для выполнения условия максимизации дисперсии) необходимо разделить на корень из суммы их квадратов и умножить на  $\sqrt{\lambda}$ . В итоге имеем:

$$a'_{j1} = \frac{a_{j1}\sqrt{\lambda}}{\sqrt{a_{11}^2 + a_{21}^2 + \dots + a_{m1}^2}} \quad (j = 1, \dots, n),$$

которые являются нормированными коэффициентами первого фактора. После мы повторяем данные вычисления для  $\lambda_2$  и т.д. В результате получаем в качестве решения матрицу, строками которой будут коэффициенты при соответствующем факторе.

### 4.3 Процедура вычисления

Компонентный анализ можно провести с любой корреляционной матрицей с общностями на главной диагонали. Принимается решение о количестве факторов, подлежащих выделению. Для вычисления собственных значений и собственных векторов можно использовать метод вращений (метод Якоби) решения полной проблемы собственных значений для симметричной матрицы. Подставив полученные данные в систему уравнений и пронормировав их, получаем матрицу  $A$ , где столбцами являются найденные нами факторы.

## 5 Центроидный метод

До недавнего времени из-за большого объема вычислительных работ и отсутствия быстродействующих ЭВМ, метод главных компонент применялся лишь в исключительных случаях, исследователи ограничивались центроидным методом, хотя он дает неточное и неоднозначное решение.

### 5.1 Геометрическая интерпретация метода

Синонимом названия «центроидный метод» является «метод центра тяжести». Это название объясняет принцип метода. Положение первой координатной оси должно быть определено так, чтобы она проходила через центр тяжести скопления точек. Факторное отображение можно рассматривать как размещение  $m$  точек-переменных в  $r$ -мерном пространстве, причем отдельные точки или векторы представляют переменные. На Рис. 3.А схематично изображены несколько точек-переменных в двумерной системе координат. Указана нулевая точка, в которой начинаются все векторы.

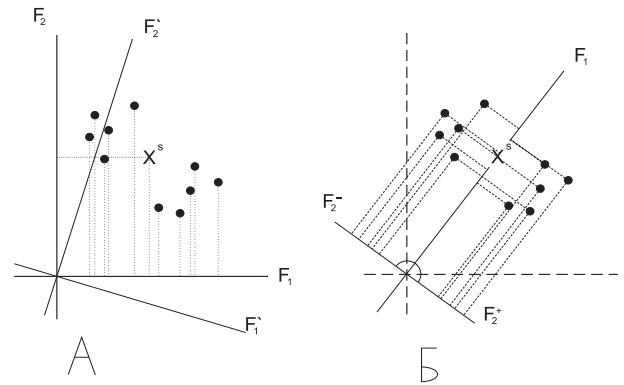


Рис. 3: Определение положения координатных осей с помощью центроидного метода двумерной системе координат.

### 5.2 Геометрическая интерпретация метода

На Рис. 3.А схематично изображены несколько точек-переменных в двумерной системе координат. Кроме того, указана нулевая точка, в которой начинаются все векторы. Это соответствует типичной ситуации перед началом выделения факторов. Переменные представлены  $m$  точками в  $r$ -мерном пространстве, положение нулевой точки известно. Разумеется, точное значение необходимой размерности пространства неизвестно.

Чтобы получить однозначное положение системы координат, уславливаются, что первая ось должна проходить через центр тяжести  $S$  скопления точек-переменных. Вторая ось  $F_2$  перпендикулярна к первой. Это положение осей соответствует позиции факторов в центроидном решении.

Проекции точек на оси координат на Рис. 3.Б определяют факторные нагрузки  $a_{il}$ , которые рассчитываются по корреляционной матрице. Координаты центра тяжести можно вычислить по координатам отдельных точек. Если в общем случае рассматривать  $r$ -мерную систему координат, то координатами центра тяжести являются выражения:

$$\frac{1}{m} \sum_i a_{i1}; \quad \frac{1}{m} \sum_i a_{i2}, \quad \dots, \quad \frac{1}{m} \sum_i a_{ir};$$

т.е. средние значения координат отдельных точек дают координаты центра тяжести. Это можно увидеть, рассматривая Рис. 3.Б, на котором координаты центра тяжести отмечены пунктиром. Если теперь система координат выбрана так, что первая ось проходит через центр тяжести, то сумма проекций точек на все остальные ортогональные к ней оси равны нулю (это следует из определения центра тяжести) и тогда координаты центра тяжести  $S$  становятся равными:

$$\frac{1}{m} \sum a_{i1}, 0, 0, \dots, 0, \quad (4)$$

т.е.

$$\sum a_{i2} = \sum a_{i3} = \dots = \sum a_{ir} = 0,$$

в чем можно убедиться по Рис. 3.Б для случая двумерной задачи. Сумма проекций на ось  $F_2$  равна нулю, так как положительные и отрицательные значения проекций взаимно компенсируются.

### 5.3 Вычислительная процедура

Вычисление нагрузок первого фактора осуществляется по формуле:

$$a_{k1} = t \cdot \sum_i r_{ik}, \quad k = 1, \dots, m, \quad (5)$$

где

$$t = \frac{1}{\sqrt{T}}, \quad (6)$$

$$T = \sum_{ik} r_{ik}. \quad (7)$$

После вычисления нагрузок первого фактора по (5) определяют остаточные корреляции:  $R_h - a_1 \cdot a_1' = R_1$ , где  $a_1$  является вектор-столбцом факторных нагрузок. Матрица  $a_1 \cdot a_1' = R^+$  содержит так называемые воспроизведенные корреляции.  $R_1$  дает остаточные корреляции, которые остаются после выделения первого фактора ( $R_1$ - остаточная матрица). Если принимают решение выделить второй фактор, то повторяется та же самая вычислительная процедура по матрице остатков  $R_1$ .

При этом возникает затруднение, заключающееся в том, что поскольку после выделения фактора сумма проекции всех точек на другие ортогональные оси равна нулю, то нельзя определить второй центр тяжести и приступить к выделению другого фактора. Например, по Рис. 3.Б видно, что сумма проекций на ось  $F_2$  равна нулю и совпадает с началом координат. Для устранения этой проблемы необходимо изменить знаки некоторых переменных таким образом, чтобы новый центр тяжести был удален от начала координат, так чтобы все точки-переменные находились по одну сторону от оси. На последующем этапе расчета изменение знака аннулируется.

## 6 Метод минимальных остатков

Метод минимальных остатков — это итерационный метод факторного анализа.

К элементам любой строки  $j$  матрицы  $A$  добавляется приращение  $\epsilon_p$ ,  $p = 1, 2, \dots, m$ :

$$a_{j1} + \epsilon_1, a_{j2} + \epsilon_2, \dots, a_{jp} + \epsilon_p, \dots, a_{jm} + \epsilon_m.$$

Новые факторные нагрузки равны:

$$b_{jp} = a_{jp} + \epsilon_p, \quad p = 1, 2, \dots, m,$$

обозначение  $b_{jp}$  принято для большей ясности: после получения окончательного набора факторных нагрузок им снова присваивается прежнее обозначение  $a_{jp}$ . Запишем выражение для вычисленного коэффициента корреляции между некоторым фиксированным параметром  $j$  и произвольным параметром  $k$ :

$$r_{jk} = \sum_{p=1}^m a_{kp} b_{jp}$$

и суммы квадратов остаточных коэффициентов корреляции этого параметра:

$$f_j = \sum_{k=1; k \neq j}^n \left( r_{jk} - \sum_{p=1}^m a_{kp} b_{jp} \right)^2,$$

где номер  $j$  фиксирован. Выделив в последнем выражении исходные коэффициенты корреляции и приращение, получим

$$f_j = \sum_{k=1; k \neq j}^n \left( r_{jk}^* - \sum_{p=1}^m a_{kp} \epsilon_p \right)^2,$$

где номер  $j$  фиксирован,  $r_{jk}^*$  — исходные остаточные коэффициенты корреляции параметра  $j$  с параметром  $k$  (без учета приращений в факторных нагрузках), т.е.

$$r_{jk}^* = r_{jk} - \sum_{p=1}^m a_{kp} a_{jp}, \quad k = 1, 2, \dots, j-1, j+1, \dots, n.$$

Для определения значений  $\epsilon_q$ , приносящих минимум целевой функции  $f$ , возьмем частную производную от  $f_j$  по этим величинам  $\epsilon_q$ :

$$(f_j)'_{\epsilon_q} = 2 \sum_{k=1; k \neq j}^n \left( r_{jk}^* - \sum_{p=1}^m a_{kp} \epsilon_p \right) (-a_{hq}),$$

$$q = 1, 2, \dots, m.$$

Приравняем эти выражения нулю, получив тем самым набор неявных уравнений относительно  $\epsilon_p$ :

$$\left( \sum_{p=1}^m a_{kp} a_{jp} \right) \epsilon_p = \sum_{k=1; k \neq j}^n r_{jk}^* a_{kq}, \quad q = 1, 2, \dots, m,$$

или в матричной форме

$$\epsilon_j \bar{A}'_j \bar{A}_j = r_j^0 A,$$

где  $\epsilon_j$  — вектор-строка приращений факторных нагрузок параметра  $j$ ,  $A_j$  — факторная матрица, в которой элементы  $j$  заменены нулями,  $r_j^0$  — вектор-строка остаточных коэффициентов корреляции параметра  $j$  со всеми остальными параметрами (остаточных коэффициент корреляции с собой равен нулю). Теперь набор факторных нагрузок данного параметра, минимизирующих целевую функцию, определяется из

$$\epsilon_j = r_j^0 A (\bar{A}' \bar{A})^{-1}.$$



Описанный процесс повторяется для всех параметров поочередно. Последовательно получаются приближения строк факторных нагрузок, приносящих минимальное значение функции  $f$  (с заданной степенью точности).

## 7 Критерии оценки числа факторов, подлежащих выделению

При применении как метода главных факторов, так и центроидного метода возникает вопрос: когда должен быть закончен процесс выделения факторов (каким числом факторов нужно ограничиться). Общепризнанный метод определения числа факторов, подлежащих выделению, не существует. Наиболее известным является *критерий Кайзера и Дикмана*: факторы, вклады которых (сумма квадратов факторных нагрузок) в полную дисперсию меньше единицы, имеют долю дисперсии, меньшую единичной дисперсии переменных. Такие факторы не должны выделяться. Также можно использовать *критерий собственных значений*, предложенный Каттелом [3]:

- рассматриваются собственные значения корреляционной матрицы;
- они сортируются в порядке убывания и находится  $t$  число собственных значений больше  $\lambda > 1$ , следовательно, число выделяемых факторов должно быть не меньше  $t$ .

С другой стороны, по одному из простых правил, должно быть выделено меньше  $m = 2$  факторов. Таким образом, число выделяемых факторов лежит между  $t$  и  $m = 2$ . Оба критерия не противоречат друг другу.

## 8 Описание статистики

В работе решается практическая задача исследования многомерных статистических данных с помощью двух из описанных выше методов — метода главных компонент и метода минимальных остатков.

Территориальным органом федеральной службы государственной статистики по Красноярскому краю была предоставлена статистика показателей социально-экономического положения населения края. Данные показатели подразделяются на 9 категорий: социальная поддержка по оплате жилых помещений и коммунальных услуг, местный бюджет, учреждения здравоохранения, социальное обслуживание населения, население, занятость и заработная плата, здравоохранение, промышленное производство и финансовая деятельность. Всего рассматриваются около 80 показателей. В статистике содержатся данные по 45 районам края — Абанскому, Ачинскому, Балахтинскому,

Березовскому, Бирилюсскому, Боготольскому, Богучанскому, Большемурутинскому, Большеулуйскому, Дзержинскому, Емельяновскому, Енисейскому, Ермаковскому, Идринскому, Иланскому, Ирбейскому, Казачинскому, Канскому, Каратузскому, Кежемскому, Козульскому, Краснотуранскому, Курагинскому, Манскому, Минусинскому, Мотыгинскому, Назаровскому, Нижнеингашскому, Новоселовскому, Партизанскому, Пировскому, Рыбинскому, Саянскому, Северо-енисейскому, Сухобузимскому, Таймырскому Долгано-Ненецкому, Таймырскому, Тасеевскому, Туруханскому, Түхтөтскому, Ужурскому, Уярскому, Шарыповскому, Шушенскому и Эвенкийскому районам. При обработке статистики города края были вынесены в отдельные территориальные единицы, поскольку их население достаточно велико и зачастую больше многих районов (особенно население Красноярска и Норильска). Таким образом, в статистике также представлена информация по показателям 14 крупных городов — Ачинска, Боготола, Бородино, Дивногорска, Енисейска, Канска, Красноярска, Лесосибирска, Минусинска, Назарово, Норильска, Сосновоборска, Шарыпово и поселка Кедровый.

На основе статистических данных была составлена корреляционная матрица. Затем, согласно методу главных компонент, была получена матрица, в которой каждая строчка представляет собой фактор, ячейками которого являются нагрузки.

Также на основе той же первоначальной корреляционной матрицы происходит вычисление приближенной матрицы факторных нагрузок, строки которой минимизируют значение функции  $f$  (описанной в алгоритме метода минимальных остатков).

Производится сравнение факторов, найденных обоими методами.

## Список литературы

- [1] К. Иберла. *Факторный анализ*. Наука, М., 1972.
- [2] Г. Хартман. *Современный факторный анализ*. Статистика, М., 1972.
- [3] L. Thurstone. *Multiple factor analysis*. University of Chicago, Chicago, 1961.



# Нахождение эллиптического конуса приемлемых рисков для некоторых норм Гельдера

Татьяна Алексеевна Кустицкая

Сибирский федеральный университет  
Институт космических и информационных технологий  
Красноярск  
m-tanika@yandex.ru

**Аннотация.** В работе излагается метод нахождения эллиптического конуса приемлемых рисков в модели обобщенных когерентных мер риска по значению неприятия риска в нуле для равномерного распределения. Получены формулы для нахождения осевой функции по значению неприятия риска для наиболее распространенных норм Гельдера.

**Ключевые слова.** Обобщенные когерентные меры риска, отношение предпочтения, неприятие риска, эллиптический конус приемлемых рисков.

## 1 Введение

В работах [1] и [2] был предложен класс обобщенных когерентных мер риска (ОКМР), расширяющий класс представляемых индивидуальных предпочтений по сравнению с классическими когерентными мерами риска [3]. Это связано с введением нового параметра, отвечающего за отношение к риску - нормы в пространстве рисков.

При этом для конструирования меры риска, представляющей конкретное предпочтение, необходимо построить соответствующее множество приемлемых рисков. Для предпочтений, согласованных со стохастическим доминированием и обладающих свойством неприятия риска предлагается строить его в виде эллиптического конуса.

В работе [4] изложен метод нахождения параметров эллиптического конуса для евклидовой нормы. В данной работе предлагается метод нахождения конуса с помощью неприятия риска для других часто встречающихся норм Гельдера - нормы  $\|\cdot\|_1$ ,  $\|\cdot\|_\infty$ .

## 2 Обобщенные когерентные меры риска

Рассмотрим вероятностное пространство  $(\Omega, \mathcal{A}, P)$ .

**Риском**  $X$  на  $(\Omega, \mathcal{A})$  называется произвольное измеримое отображение из  $\Omega$  в  $\mathbf{R}$ . Возможные значения риска интерпретируем как доход некоего лица.

Обозначим  $\mathcal{X}$  совокупность всех рисков на  $(\Omega, \mathcal{A})$ .

Предположим, что  $|\Omega| = n$ . Тогда риск  $X = (X_1, \dots, X_n)$  - вектор из  $\mathbf{R}^n$ .

Пусть на  $\mathcal{X}$  задано отношение предпочтения  $\preceq$ , отражающее индивидуальное отношение к риску некоторого инвестора.

**Приемлемым** назовем такой риск, на который инвестор готов пойти при вложении нулевого капитала.

Множество всех приемлемых рисков обозначим  $A$ .

Пусть  $A$  удовлетворяет следующим аксиомам:

**Аксиома А1:**  $C_+ \subset A$ ,  $C_+ = \{X \in \mathcal{X} : X \geq 0\}$

**Аксиома А2:**  $A \cap C_- = \emptyset$ ,  $C_- = \{X \in \mathcal{X} : X < 0\}$

**Аксиома А3:**  $A$  - выпуклый конус (если  $X \in A$ ,  $Y \in A$ , то  $\alpha_1 X + \alpha_2 Y \in A$  при  $\alpha_1, \alpha_2 \geq 0$ ).

По заданному конусу  $A$  строится **обобщенная когерентная мера риска** (ОКМР)  $f_A[1]$ :

$$f_A(X) = f_{A, \|\cdot\|}(X) = \delta_A(X) \inf_{Y \in \partial A} \|X - Y\|, \quad (1)$$

$$\text{где } \delta_A(X) = \begin{cases} 1, & X \in A, \\ -1, & X \in A^c, \end{cases}$$

здесь  $\partial A$  - это граница множества  $A$ .

Функционал  $f_A(X)$  обладает свойствами:

**М** (монотонность):

$$f_A(X) \leq f_A(Y), \quad \forall X, Y \in \mathcal{X}, X \leq Y;$$

**ПО** (положительная однородность):

$$f_A(\lambda X) = \lambda f_A(X), \quad \forall \lambda \geq 0, X \in \mathcal{X};$$

**С** (супераддитивность):

$$f_A(X + Y) \geq f_A(X) + f_A(Y), \quad \forall X, Y \in \mathcal{X}$$

**Свойство КП** (кратчайшего пути):

а)  $\forall X \in \mathcal{X} \exists X'(X) \in \partial A$

такой, что  $\|X - X'(X)\| = \inf_{Y \in \partial A} \|X - Y\|$ ;

б)  $f_A(X + \lambda u(X)) = f_A(X) + \lambda$ ,  $-\infty < \lambda \leq \lambda_A(X)$ ,

где  $\lambda_A(X) > 0$ , а вектор кратчайшего пути  $u(X)$  определяется следующим образом:

- для случая строго выпуклой нормы  $\|\cdot\|$

$$u(X) = \delta_A(X) \frac{X - X'_A(X)}{\|X - X'_A(X)\|};$$

- для нормы  $\|\cdot\|_\infty$   $u(X) = I = (1, 1, \dots, 1)$ ;

- для нормы  $\|\cdot\|_1$   $u(X) = e_i \forall i = 1, 2, \dots, n$ , где

$e_i$  - вектор из стандартного базиса пространства  $\mathbf{R}^n$ .

Для заданной ОКМР  $f(X)$  можно найти со-

ответствующее множество приемлемых рисков:

$$A_f = \{X \in \mathcal{X} : f(X) \geq 0\}. \quad (2)$$

Граница множества приемлемых рисков, соответствующего мере  $f$  определяется как

$$\partial A_f = \{X \in \mathcal{X} : f(X) = 0\}.$$

Когерентные меры риска являются частным случаем ОКМР, соответствующим норме  $\|\cdot\| = \|\cdot\|_\infty$ .

## 2.1 Предпочтения, согласованные со стохастическим доминированием

Обозначим  $\mathcal{F}$  совокупность всех вещественных функций распределения;  $F_X$  - функцию распределения случайной величины  $X$ :  $F_X(x) = P(X \leq x)$ .

Пусть  $\mathcal{F}_1$  - множество всех функций распределения, обладающих конечным 1-м моментом.

Пусть  $F, Q \in \mathcal{F}_1$ . Будем говорить, что  $F$  предшествует  $Q$  в смысле **стохастического доминирования** первого порядка ( $F \leq_1 Q$ ), если

$$F(x) \geq Q(x), \quad -\infty < x < \infty.$$

Пусть  $F, Q \in \mathcal{F}_1$ . Считаем, что  $F$  предшествует  $Q$  в смысле **строго стохастического доминирования** первого порядка ( $F <_1 Q$ ), если

$$F \leq_1 Q \quad \text{и} \quad \exists x \in \mathbf{R} : F(x) > Q(x).$$

Зададим частичный порядок на  $\mathcal{X}$ : Риск  $Y$  (строго) **стохастически доминирует** над риском  $X$ :  $X \leq_1 Y$  ( $X <_1 Y$ ), если

$$F_X \leq_1 F_Y \quad (F_X <_1 F_Y).$$

## 2.2 Неприятие риска в модели обобщенных когерентных мер риска

Отношение предпочтения  $\preceq$  обладает свойством **неприятия риска**, если для произвольного невырожденного риска  $\Delta : E\Delta = 0$  и произвольного  $a \in \mathbf{R}$  имеет место

$$a + \Delta \prec a. \quad (3)$$

В терминах меры риска  $\rho$ , задающей предпочтение  $\preceq$  на  $\mathcal{X}$ , это можно записать так:  $\rho(a + \Delta) < \rho(a)$ .

Для регулярного предпочтения на  $\mathcal{X} \exists c > 0$

$$\rho(a + \Delta) = \rho(a - c).$$

Величину  $c$  можно использовать в количественной оценке неприятия риска, которая была введена в [5].

Для обобщенных когерентных мер риска

$$c_{a,\Delta} = \begin{cases} a - \frac{f(aI+\Delta)}{f(I)}, & aI + \Delta \in A, \\ a + \frac{f(aI-\Delta)}{f(-I)}, & aI + \Delta \notin A \end{cases}$$

В частности,

$$c_{0,\Delta} = \frac{f_A(\Delta)}{f_A(-I)} \quad (4)$$

Так как  $f(aX) = af(X)$ ,  $a \geq 0$ , то  $c_{0,a\Delta} = a \cdot c_{0,\Delta}$ , значит для исследования функции неприятия риска в нуле можно ограничиться  $\Delta : \|\Delta\| = 1$ .

## 3 Свойства конуса приемлемых рисков для некоторых предпочтений

В работах [6] и [2] было доказано, что для предпочтения, согласованного со *стохастическим доминированием*, конус приемлемых рисков обладает следующими свойствами:

**I. Симметричность для равномерного распределения.** Если  $P = (\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n})$  и вектор  $X = (X^1, X^2, \dots, X^n) \in A$ , то любой вектор  $Y$ , компоненты которого получены перестановкой компонент вектора  $X$  также принадлежит  $A$  (конус симметричен относительно осей координат).

**II. Зависимость от вероятностной меры.** Обозначим за  $\mathcal{X}_P$  и  $\mathcal{X}_Q$  множества всех рисков, соответствующих вероятностным мерам  $P$  и  $Q$ , множества приемлемых рисков -  $A_P$  и  $A_Q$ . За  $X_P$  обозначим вектор  $(X_1, \dots, X_n) \in \mathcal{X}_P$ , а за  $X_Q$  вектор  $(X_1, \dots, X_n) \in \mathcal{X}_Q$ .

Тогда если  $P \neq Q$  и  $\exists X_P$  такой, что  $X_P <_1 X_Q$  (или  $X_Q <_1 X_P$ ), то  $A_P \neq A_Q$ .

**III. Уменьшение размерности конуса.** Для конуса приемлемых рисков  $A_P$ , соответствующего вероятностной мере  $P = (p_1, \dots, p_{k-1}, 0, p_{k+1}, \dots, p_n)$ :

$$X = (X^1, \dots, X^{k-1}, X^k, X^{k+1}, \dots, X^n) \in A_P \Rightarrow \\ Y = (X^1, \dots, X^{k-1}, y, X^{k+1}, \dots, X^n) \in A_P \quad \forall y \in \mathbf{R}.$$

**IV. Вырождение конуса в полупространство.** Конус приемлемых рисков  $A_P$ , соответствующий  $P = (p_1, \dots, p_n)$ , где  $p_k = 1$ ;  $p_i = 0, i \neq k$ , может быть задан неравенством

$$X^k \geq 0.$$

Для предпочтения, обладающего свойством *неприятия риска*, имеет место

**V. Свойство положительности математического ожидания приемлемых рисков.** Для всех  $X \in A, X \neq 0$  выполняется  $EX > 0$ .

## 4 Эллиптический конус приемлемых рисков

Будем рассматривать предпочтения, согласованные со стохастическим доминированием и обладающие свойством неприятия риска. Как видно из свойств II, III, IV, множество приемлемых рисков не является постоянным - оно меняется при изменении размерности пространства рисков и значений вероятности.

Идея заключается в том, чтобы построить модель множества приемлемых рисков, которая бы позволяла, однажды вычислив необходимые параметры для конкретного индивидуального предпочтения, перестраивать конус приемлемых рисков при изменении значений вероятностной меры и размерности задачи.

Вполне очевидно, что чем больше вероятность некоторого исхода ( $p_i$ ), тем менее желательны для инвестора отрицательные значения дохода, соответствующего этому исходу ( $X_i$ ). Можно предположить, что наиболее приемлемые для инвестора риски находятся на луче  $\lambda P$ ,  $\lambda \geq 0$ .<sup>1</sup> Будем считать, что риск тем «лучше», чем ближе он к риску, лежащему в той же гиперплоскости  $(P, X) = \lambda$  на луче  $\lambda P$ . Чтобы было учтено влияние  $p_i$  на приемлемость значения  $X_i$ , нужно чтобы допустимое отклонение значений приемлемых рисков в направлении  $i$ -й координаты уменьшалось при росте  $p_i$ .

Представим  $A$  в виде выпуклого конуса, сечения которого плоскостями  $(P, X) = a$ ,  $a \geq 0$  - эллипсоиды с полуосями  $r(p_i)$ ,  $i = 1, \dots, n$ .

В соответствии с геометрическим толкованием функцию  $r(p)$ ,  $p \in [0, 1]$  назовем **осевой функцией**. Это единственная характеристика конуса, отражающая индивидуальное отношение к риску и описывающая зависимость такого отношения от вероятности.

Если в качестве осевой функции  $r(p)$  брать функцию из некоторого однопараметрического семейства, то для исчерпывающего определения эллиптического конуса приемлемых рисков достаточно знать значение этой функции всего в одной точке.

В качестве такой точки возьмем  $p = \frac{1}{n}$ . Значение

$$r_0 = r\left(\frac{1}{n}\right)$$

будем определять, исследуя эллиптический конус для равномерного распределения  $P = \left(\frac{1}{n}, \dots, \frac{1}{n}\right)$  (такой конус можно назвать **сферическим**).

Для норм  $\|\cdot\|_1, \|\cdot\|_2, \|\cdot\|_\infty$  сферический конус определяется неравенством

$$\left\| X - \frac{(I, X)}{n} I \right\| \leq \frac{(I, X)}{n} r_0.$$

<sup>1</sup>Строго говоря, вектор  $P = (p_1, \dots, p_n)$  значений вероятности является вектором из пространства  $\mathcal{X}^*$ , двойственного к  $\mathcal{X}$ , поэтому здесь и далее под  $P$  мы будем понимать вектор из  $\mathcal{X}$ , имеющих координаты, равные координатам вектора вероятности из  $\mathcal{X}^*$ .

## 4.1 Эллиптический конус приемлемых рисков для нормы $\|\cdot\|_2$

Для случая  $\|\cdot\| = \|\cdot\|_2$  эллиптический конус может быть задан с помощью неравенства

$$\sum_{i=1}^n \frac{(X^i - (P, X)np_i)^2}{r^2(p_i)} \leq (P, X)^2, \quad X \in \mathbf{R}^n \quad (5)$$

В работе [2] была доказана следующая

**Теорема 17.** При выполнении условий

$$\begin{aligned} 1. \quad & (P, X) \geq 0; \\ 2. \quad & r(p) \geq \sqrt{\frac{n}{p^2} + n^3 p^2}. \end{aligned} \quad (6)$$

эллиптический конус  $A$ , задаваемый неравенством (5), является конусом приемлемых рисков для некоторого предпочтения.

Также было показано, что

$$r_0 = \frac{\sqrt{1 - nc_{0,\Delta}^2}}{c_{0,\Delta}} \quad (7)$$

Зная неприятие риска в нуле, можно найти значение  $r\left(\frac{1}{n}\right)$ , определить отсюда параметр функции  $r(p)$  и построить конус приемлемых рисков.

В работе [4] показано, что в качестве осевых функций для конуса (5) можно предложить следующие:

1. *Степенная осевая функция*

$$r_1(p, m_1) = \frac{\sqrt{n + n^3}}{p^{m_1}}, \quad m_1 \geq 1$$

2. *Экспоненциальная осевая функция*

$$r_2(p, m_2) = \sqrt{n + n^3} \cdot e^{\frac{m_2(1-p)}{p}}, \quad m_2 \geq M \approx 0.203$$

3. *Логарифмическая осевая функция*

$$r_3(p, m_3) = \frac{n^2}{p} \ln \left( \frac{m_3(1-p)}{p} + e \right), \quad m_3 \geq 0$$

Все рассматриваемые функции убывают по  $p$  и возрастают по  $m_i$ .

## 4.2 Эллиптический конус приемлемых рисков для нормы $\|\cdot\|_1$

Рассмотрим множество

$$\sum_{i=1}^n \frac{|X_i - (P, X) \cdot n \cdot p_i|}{r(p_i)} \leq (P, X) \quad (8)$$

Предположим, что оно удовлетворяет условиям

$$\begin{aligned} 1. \quad & (P, X) \geq 0 \\ 2. \quad & r(p) \geq \frac{n}{p}(1 + np^2) \end{aligned} \quad (9)$$

**Теорема 18.** Множество рисков  $A$ , определяемое неравенствами (8) и (9) является множеством приемлемых рисков для некоторого предпочтения.

*Доказательство*

1. Докажем, что  $C_- \cap A = \emptyset$ .

Пусть  $X \in C_-$ , тогда  $(P, X) < 0 \Rightarrow X \notin A$ .

2. Докажем, что  $A$  - выпуклый конус.

Пусть  $X : (P, X) = a > 0$  и  $X \in A$ , т.е.

$$\sum_{i=1}^n \frac{|X_i - a \cdot n \cdot p_i|}{r(p_i)} \leq a.$$

Рассмотрим  $A_a = \{Y \in A : (P, Y) = a\}$ .

Покажем, что  $A_a$  - выпуклое множество.

Возьмем  $\forall X, Y \in A : (P, X) = (P, Y) = a$  и рассмотрим риск  $(1 - \alpha)X + \alpha Y$ ,  $\alpha \in [0, 1]$ .

$$\begin{aligned} \sum_{i=1}^n \frac{|(1 - \alpha)X_i + \alpha Y_i - (P, (1 - \alpha)X + \alpha Y) \cdot n \cdot p_i|}{r(p_i)} &= \\ &= \sum_{i=1}^n \frac{|(1 - \alpha)(X_i - anp_i) + \alpha(Y_i - anp_i)|}{r(p_i)} \leq \\ &\leq (1 - \alpha) \sum_{i=1}^n \frac{|X_i - anp_i|}{r(p_i)} + \alpha \sum_{i=1}^n \frac{|Y_i - anp_i|}{r(p_i)} = a \end{aligned}$$

Значит  $(1 - \alpha)X + \alpha Y \in A_a \Rightarrow A_a$  - выпукло.

Пользуясь свойством положительной однородности для норм, можно показать, что если  $X \in A$  то и  $\lambda X \in A$  при условии  $\lambda \geq 0$ .

Значит,  $A$  - конус, сечения которого гиперплоскостями  $(P, X) = a > 0$  - выпуклые множества.

Поэтому,  $A$  - выпуклый конус.

3. Докажем, что  $C_+ \subset A$ .

Так как  $A$  - выпуклый конус, то если  $e_1, \dots, e_n \in A$ , то и  $\lambda_1 e_1 + \dots + \lambda_n e_n \in A \quad \forall \lambda_i \geq 0$ .

Покажем, что  $e_1 \in A$  (Для остальных  $e_i$  доказательство аналогично):

Подставив координаты  $e_1$  в (8) обозначив результат подстановки  $D$  получим:

$$\begin{aligned} D &= \frac{|1 - np_1^2|}{r(p_1)} + \frac{np_1 p_2}{r(p_2)} + \dots + \frac{np_1 p_n}{r(p_n)} = \\ &= \frac{|1 - np_1^2|}{r(p_1)} + np_1 \left( \frac{p_2}{r(p_2)} + \dots + \frac{p_n}{r(p_n)} \right) \end{aligned}$$

$$\text{Обозначим } D_1 = \frac{|1 - np_1^2|}{r(p_1)}.$$

Предположим, что  $1 - np_1^2 \geq 0$ .

$$D_1 \leq \frac{1 - np_1^2}{\frac{n}{p_1}(1 + np_1^2)} < \frac{p_1}{n}.$$

$$\text{Если } 1 - np_1^2 < 0, \text{ то } D_1 \leq \frac{np_1^2 - 1}{\frac{n}{p_1}(np_1^2 + 1)} < \frac{p_1}{n}.$$

Рассмотрим функцию

$$h(p) = \frac{p}{\frac{n}{p}(1 + np^2)} = \frac{p^2}{n(1 + np^2)}.$$

$$h'(p) = \frac{2pn(1 + np^2) - 2n^2 p^3}{n^2(1 + np^2)^2} = \frac{2p}{n(1 + np^2)^2}$$

$$h'(p) = 0 \iff p = 0, \quad h'(p) > 0 \quad p \in (0, 1]$$

Функция  $h(p)$  является возрастающей на  $(0, 1]$  и достигает своего максимума в точке  $p = 1$ .

$$h(1) = \frac{1}{n(1 + n)} \rightarrow h(p) \leq h(p) \leq \frac{1}{n(1 + n)}$$

$$\begin{aligned} D &\leq \frac{p_1}{n} + np_1 \left( \frac{p_2}{r(p_2)} + \dots + \frac{p_n}{r(p_n)} \right) \leq \\ &\leq \frac{p_1}{n} + np_1 \frac{n - 1}{n(1 + n)} = p_1 \frac{n^2 + 1}{n^2 + n} < p_1 = (P, e_1) \end{aligned}$$

Следовательно,  $e_1 \in A$ .  $\square$

Пусть  $P = (\frac{1}{n}, \dots, \frac{1}{n})$ . Рассмотрим  $I = (1, 1, \dots, 1)$ . Его норма  $\|I\| = n$ .

**Лемма 1.** Для риска  $I$  ближайший (в смысле нормы  $\|\cdot\|_1$ ) риск  $I'$  из  $\partial A$ , получается смещением  $I$  вдоль любого из векторов стандартного базиса, а именно:

$$f(I) = \|I - I'\|,$$

$$\text{где } I' = I - \frac{r_0 n}{2(n - 1) + r_0} e_j \quad (j \in \{1, 2, \dots, n\}).$$

*Доказательство*

Обозначим

$$\phi = \frac{r_0 n}{2(n - 1) + r_0} \quad (10)$$

Зафиксируем произвольный индекс  $j$ . Рассмотрим вектор  $I' = I - \phi e_j$ .

$$\|I - I'\| = \phi.$$

$$\begin{aligned} \left\| I' - \frac{(I', I)}{n} I \right\| &= \left\| I - \phi e_j - \frac{(I - \phi e_j, I)}{n} I \right\| = \\ &= \left\| I - \frac{(I, I)}{n} I + \frac{\phi}{n} I - \phi e_j \right\| = \frac{\phi}{n} \|I - n e_j\| = \end{aligned}$$

$$= \frac{\phi}{n}(|1-n|+1+\dots+1) = \frac{\phi \cdot 2(n-1)}{n} = \frac{2r_0(n-1)}{2(n-1)+r_0}.$$

$$\begin{aligned} \frac{(I', I)}{n} r_0 &= \frac{(I - \phi e_j, I)}{n} r_0 = \frac{(I, I)}{n} r_0 - \frac{\phi}{n} r_0 = \\ &= \frac{r_0}{n} (n - \phi) = r_0 \cdot \frac{2(n-1)}{2(n-1)+r_0} \end{aligned}$$

Получаем, что  $\left\| I' - \frac{(I', I)}{n} I \right\| = \frac{(I', I)}{n} r_0 \Rightarrow I' \in \partial A$ .

Докажем, что  $I'$  - ближайший к  $I$  вектор из  $\partial A$ .

Возьмем произвольный вектор

$$Y = I - \alpha_1 e_1 - \dots - \alpha_n e_n,$$

такой, что  $\|I - Y\| = \|\alpha_1\| + \dots + \|\alpha_n\| < \phi$ .

$$\begin{aligned} \left\| Y - \frac{(Y, I)}{n} I \right\| &= \left\| I - \alpha_1 e_1 - \dots - \alpha_n e_n - \frac{(I - \alpha_1 e_1 - \dots - \alpha_n e_n, I)}{n} I \right\| = \\ &= \left\| I - \frac{(I, I)}{n} + \frac{\alpha_1 + \dots + \alpha_n}{n} I - \alpha_1 e_1 - \dots - \alpha_n e_n \right\| = \\ &= \left| \frac{(1-n)\alpha_1 + \alpha_2 + \dots + \alpha_n}{n} \right| + \dots + \\ &\quad + \left| \frac{\alpha_1 + \dots + \alpha_{n-1} + (1-n)\alpha_n}{n} \right| \leq \\ &\leq \frac{2(n-1)}{n} (|\alpha_1| + \dots + |\alpha_n|) < \frac{2(n-1)}{n} \phi. \end{aligned}$$

$$\begin{aligned} \frac{(Y, I)}{n} r_0 &= \frac{(I - \alpha_1 e_1 - \dots - \alpha_n e_n, I)}{n} r_0 = \\ &= \frac{r_0}{n} (n - (\alpha_1 + \dots + \alpha_n)) \geq \frac{r_0}{n} (n - (|\alpha_1| + \dots + |\alpha_n|)) > \\ &> \frac{r_0}{n} (n - \phi) = \frac{2(n-1)\phi}{n-\phi} \cdot \frac{n-\phi}{n} = \frac{2(n-1)}{n} \phi \end{aligned}$$

Таким образом,

$$\left\| Y - \frac{(Y, I)}{n} I \right\| < \frac{(Y, I)}{n} r_0,$$

а значит  $Y \in A \setminus \partial A$ .  $\square$

$$\text{Итак, } f(I) = \frac{r_0 n}{2(n-1) + r_0}.$$

Теперь перейдем к нахождению зависимости  $r_0$  от неприятия риска. Пусть для определенности  $e_j = e_n$ .

$$I' = I - f(I) e_n$$

$$I' = (1, \dots, 1, 1 - f(I)); \quad \|I'\| = n + f(I) - 2.$$

$$(I, I') = n - 1 + 1 - f(I) = n - f(I)$$

По свойству кратчайшего пути  $f(I + \beta e_n) = f(I) + \beta e_n$ ,  $\beta \leq 0$ .

Пусть  $I_\Delta = I + \beta e_n$ :  $(I, I_\Delta) = 0$  (см. рис.1):

$$(I, I + \beta e_n) = 0 \Rightarrow \beta = -n, \quad I_\Delta = (1, \dots, 1, 1 - n)$$

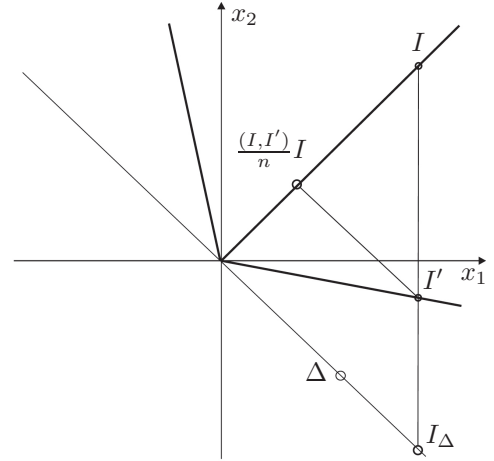


Рис. 1: Нахождение  $r_0$  через  $\tilde{c}$

В то же время  $I_\Delta = I' - \|I' - I_\Delta\| e_n$ , поэтому  $f(I_\Delta) = f(I') - \|I' - I_\Delta\|$ .

По определению неприятия риска для ОКМР

$$\tilde{c} = c_{0, I_\Delta} = \frac{f(I_\Delta)}{f(-I)} = \frac{\|I_\Delta - I'\|}{n} \Rightarrow \|I_\Delta - I'\| = \tilde{c} n$$

$$\|I_\Delta\| = n - 1 + |1 - n| = 2(n - 1).$$

$$\|I - I_\Delta\| = n$$

$$\|I - I'\| = \|I - I_\Delta\| - \|I_\Delta - I'\| = n(1 - \tilde{c})$$

Но ранее мы выяснили, что  $\|I - I'\| = f(I) = \frac{r_0 n}{2(n-1) + r_0}$ . Отсюда

$$r_0 = \frac{2(n-1)(1 - \tilde{c})}{\tilde{c}}.$$

### 4.3 Эллиптический конус для нормы $\|\cdot\|_\infty$

Рассмотрим множество

$$\max_{i=1, \dots, n} \frac{|X_i - (P, X) \cdot n \cdot p_i|}{r(p_i)} \leq (P, X) \quad (11)$$

Предположим, что оно удовлетворяет условиям

$$\begin{aligned} 1. \quad & (P, X) \geq 0 \\ 2. \quad & r(p) \geq \frac{1 + np^2}{p} \end{aligned} \quad (12)$$

**Теорема 19.** Множество рисков  $A$ , определяемое неравенствами (11) и (12) является множеством приемлемых рисков для некоторого предпочтения.

*Доказательство*

1.  $C_- \cap A = \emptyset$ . Доказательство аналогично доказательству п.1. в теореме 18



2.  $A$  - выпуклый конус.

Пусть  $X : (P, X) = a > 0$  и  $X \in A$ , т.е.

$$\max_{i=1, \dots, n} \frac{|X_i - a \cdot n \cdot p_i|}{r(p_i)} \leq a.$$

Рассмотрим  $A_a = \{Y \in A : (P, Y) = a\}$ .

Покажем, что  $A_a$  - выпуклое множество.

Возьмем  $\forall X, Y \in A : (P, X) = (P, Y) = a$  и рассмотрим риск  $(1 - \alpha)X + \alpha Y$ ,  $\alpha \in [0, 1]$ .

$$\begin{aligned} & \max_{i=1, \dots, n} \frac{|(1 - \alpha)X_i + \alpha Y_i - anp_i|}{r(p_i)} = \\ & = \max_{i=1, \dots, n} \left| \frac{(1 - \alpha)(X_i - anp_i)}{r(p_i)} + \frac{\alpha(Y_i - anp_i)}{r(p_i)} \right| \leq \\ & \max_{i=1, \dots, n} (1 - \alpha) \frac{|X_i - anp_i|}{r(p_i)} + \max_{i=1, \dots, n} \alpha \frac{|Y_i - anp_i|}{r(p_i)} = a \end{aligned}$$

Значит  $(1 - \alpha)X + \alpha Y \in A_a \Rightarrow A_a$  - выпукло.

Пользуясь свойством положительной однородности для норм, можно показать, что если  $X \in A$  то и  $\lambda X \in A$  при условии  $\lambda \geq 0$ .

Значит,  $A$  - конус, сечения которого гиперплоскостями  $(P, X) = a > 0$  - выпуклые множества.

Поэтому,  $A$  - выпуклый конус.

3. Докажем, что  $C_+ \subset A$ .

Так как  $A$  - выпуклый конус, то если  $e_1, \dots, e_n \in A$ , то и  $\lambda_1 e_1 + \dots + \lambda_n e_n \in A \quad \forall \lambda_i \geq 0$ .

Покажем, что  $e_1 \in A$  (Для остальных  $e_i$  доказательство аналогично). Подставляем  $e_1$  в (11) и обозначим результат подстановки  $D$ :

$$D = \max \left\{ \frac{|1 - np_1^2|}{r(p_1)}, \frac{np_1 p_2}{r(p_2)}, \dots, \frac{np_1 p_n}{r(p_n)} \right\}.$$

Если  $D = \frac{|1 - np_1^2|}{r(p_1)}$ , то по условию на  $r(p)$  из (12):

$$D \leq \frac{|1 - np_1^2|}{1 + np_1^2} p_1 \leq p_1.$$

Если же  $D = \frac{np_1 p_k}{r(p_k)}$ , где  $k = 2, \dots, n$ , то

$$D \leq p_1 \frac{np_k^2}{1 + np_k^2} \leq p_1.$$

Получаем, что  $e_1$  удовлетворяет (11).

Перейдем к вычислению  $r_0$ . Пусть  $P = (\frac{1}{n}, \dots, \frac{1}{n})$ .

Рассмотрим риск  $\Delta : E\Delta = 0$ ,  $\|\Delta\|_\infty = 1$ . Так как для для нормы  $\|\cdot\|_\infty$  вектором кратчайшего направления является вектор  $I = (1, \dots, 1)$ , то ближайшим вектором к  $\Delta$ , принадлежащим  $\partial A$  является  $\Delta' = \Delta + f(\Delta)I$ .

$$(\Delta', I) = (\Delta, I) + f(\Delta)(I, I) = f(\Delta)n = c_{0, \Delta}n$$

$$\text{Пусть } \tilde{\Delta} = \frac{(I, \Delta')}{n}I$$

Так как  $\Delta' \in \partial A$ , то

$$\|\Delta' - \tilde{\Delta}\| = \frac{(\Delta', I)}{n}r_0 = c_{\Delta}r_0$$

Так как  $(\Delta, \Delta - \Delta') = 0$ ,  $(\tilde{\Delta}, \tilde{\Delta} - \Delta') = 0$ ,  $(\tilde{\Delta}, \Delta) = 0$ , то

$$\begin{aligned} \Delta = \Delta' - \tilde{\Delta} \Rightarrow \|\Delta\| &= \|\Delta' - \tilde{\Delta}\| = c_{\Delta}r_0 \Rightarrow \\ r_0 &= \frac{1}{c_{\Delta}} \end{aligned}$$

## 5 Итоги и выводы

В работе была определена зависимость между значениями осевой функции эллиптического конуса  $r(\frac{1}{n})$  и неприятием риска в нуле для норм  $\|\cdot\|_1$  и  $\|\cdot\|_\infty$ . Эти результаты позволяют по заданной величине неприятия риска строить соответствующий эллиптический конус приемлемых рисков и, следовательно, саму меру риска, которую затем можно применять в задачах принятия решений.

Моделью эллиптического конуса можно воспользоваться, если известно, что предпочтения инвестора согласованы со стохастическим доминированием и обладают свойством неприятия риска.

## Благодарности

Выражаю благодарность О.Ю. Воробьеву, А.А. Новоселову и другим участникам ФАМ-семинара за плодотворные обсуждения данной темы и конструктивную критику.

## Список литературы

- [1] Новоселов А.А. Обобщенные когерентные меры риска. *Тр. IV Всеросс. ФАМ-конференции, Красноярск, 2005*, 1:325–339, 2005.
- [2] Kustitskaya T. A. Representation of preferences by generalized coherent risk measures. *Журн. СВУ. Сер. Матем. и физ.*, 5:451–461, 2012.
- [3] Artzner Ph., Delbaen F., Eber J.-M., Heath D. Coherent measures of risk. *Mathematical Finance*, 9:203–228, 1999.
- [4] Кустицкая Т.А. Решение обратной задачи теории риска в рамках модели обобщенных когерентных мер риска (для евклидовой нормы). *Труды XVI международной ЭМ'2012 конференции. Под ред. Олега Воробьева. — Красноярск: СВУ, НИИППБ*, pages 100–105, 2012.
- [5] Новоселов А.А. Неприятие риска: качественный подход и количественные оценки. *Автоматика и телемеханика*, 7:165–177, 2003.
- [6] Kustitskaya T. Acceptance cone for some preferences. *Proceedings of the X International FAMET'2011 Conference / Oleg Vorobuov ed. - Krasnoyarsk, KSTEL, SFU*, pages 192–196, 2011.

# Применение process mining для разработки агентных имитационных моделей на примере процесса дистанционного обучения

Евгений Алексеевич Ланцев

ФБГОУ ВПО Сибирский государственный технологический университет  
Кафедра системотехники  
Красноярск  
lancevea@gmail.com

**Аннотация.** В работе приведено описание агентной имитационной модели обучения дискретной математике, созданной на основе результатов анализа журнала событий соответствующего процесса СДО Moodle. Приведены результаты имитационных экспериментов реализованной агентной модели.

**Ключевые слова.** Событийный анализ, агентное моделирование, AnyLogic, process mining.

## 1 Введение

В настоящее время большинство организаций используют информационные системы для поддержки выполнения бизнес-процессов. Информационные системы обычно обладают функциональностью регистрации фактов выполнения операций в организации. Они ведут журналы, содержащие данные о случаях, произошедших в организации. Эти журналы являются исходными данными для анализа процессов, как правило, такие журналы называют журналами событий. Анализ процессов (process mining) нацелен на автоматическое извлечение информации из журналов событий [1].

Дисциплина «Дискретная математика» преподается студентам-заочникам направления 230100.62 «Информатика и вычислительная техника» на 1 курсе. Количество часов, отведенных на изучение дисциплины по учебному плану – 180. В рамках изучения дисциплины предусмотрено выполнение индивидуальных расчетных заданий и курсового проекта. Изучение курса заканчивается экзаменом. Группа, проходившая обучение по курсу в феврале 2013 года, состояла из 50 человек.

Дистанционный курс содержит 3 темы:

1. Множества,
2. Отношения на множествах,
3. Графы, Алгоритмы на графах.

Каждая тема курса содержит лекционный и презентационный материал, проверочные работы, тесты. Вступительный раздел курса содержит методические указания по работе с курсом, выполнению расчетных заданий и курсового проекта, задания для работы с интерактивным обучающим практикумом. Методически изучение дисциплины «Дискретная математика» с использованием дистанционной технологии обучения происходит следующим образом.

Во время установочной сессии со студентами проводятся лекционные и практические занятия. После этого студентам выдаются логины для доступа к системе дистанционного обучения (СДО) и далее, до экзаменационной сессии, они работают с дистанционным курсом. Студент может изучать курс в том порядке, в котором ему удобно, вновь возвращаться для повторного изучения отдельных разделов, выполнения тестов, проверочных работ и других заданий.

Информация обо всех действиях студента при работе с курсом сохраняется в базе данных системы дистанционного обучения. Оценки, выставляемые системой дистанционного обучения или преподавателем, автоматически заносятся в журнал.

Предполагается использовать журнал системы дистанционного обучения Moodle для анализа и построения модели прохождения учебного курса. Полученная модель процесса обучения будет использоваться с построения имитационной модели процесса обучения, которая будет вычислять трудоемкость и общее время выполнения процесса обучения, а в перспективе и другие показатели процесса обучения для вычисления степени его эффективности

## 2 Построения модели процесса при помощи process mining

Для построения модели обучения дискретной математике на основании полученного журнала событий системы дистанционного обучения Moodle будем использовать программный продукт ProM 6 [2]. Модель процесса была разработана с помощью алгоритма нечеткой логики (плагин Fuzzy Miner [3] для Prom) и представлена на рисунке 1.

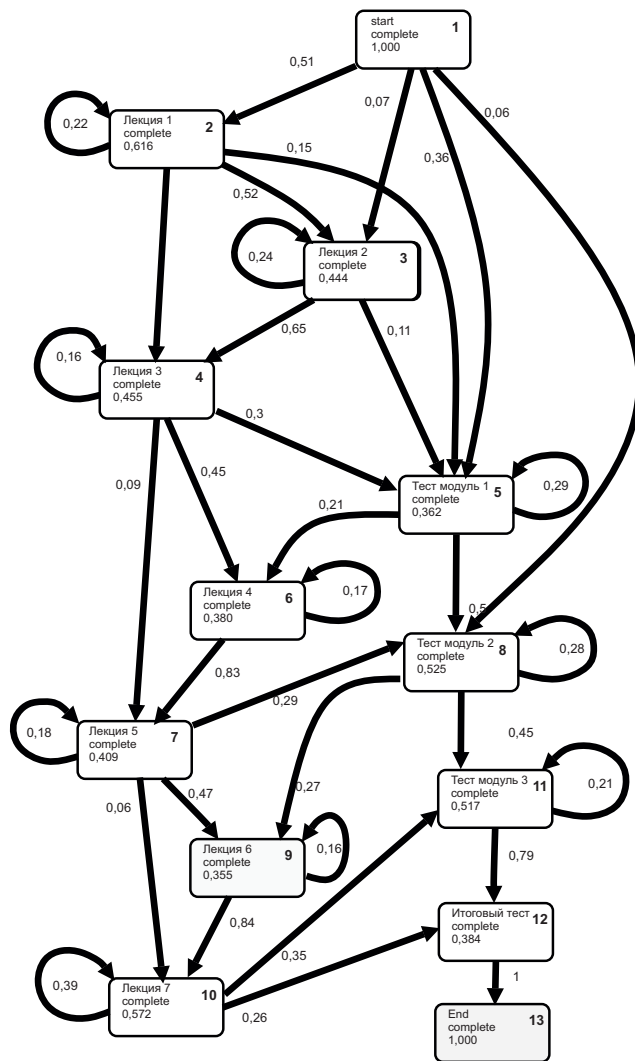


Рис. 1: Модель прохождения учебного курса разработанная с помощью алгоритмов нечеткой логики (плагин Fuzzy Miner для Prom)

### 3 Подбор параметров распределений для времени выполнения процессов

Подбор подходящего распределения для времени выполнения и времени ожидания процессов на основе экспериментальных данных осуществлялся с помощью пакета EasyFit, предназначенного для автоматического определения классов распределения и вычисления параметров, соответствующих классу распределения (Таблица 1, 2). Данный пакет включает 55 распределений, например: нормальное, логнормальное, логистическое, экспоненциальное, Парето, Коши, Лапласа, Вейбулла и т.д.

В данном случае, в качестве критерия для выбора распределения случайных величин, был использован критерий Колмогорова-Смирнова.

Процесс	Распределение и параметры
Лекция 1	Вейбулла; $\alpha = 0,56955; \beta = 11,679$
Лекция 2	Логнормальное; $\sigma = 1,4595; \mu = 2,1539$
Лекция 3	Экспоненциальное; $\lambda = 0,03699$
Лекция 4	Экспоненциальное; $\lambda = 0,07759$
Лекция 5	Логнормальное; $\sigma = 1,4211; \mu = 1,7821$
Лекция 6	Нормальное; $\sigma = 16,199; \mu = 15,384$
Лекция 7	Экспоненциальное; $\lambda = 0,08265$
Тест за модуль 1	Логнормальное; $\sigma = 1,6593; \mu = 3,2263$
Тест за модуль 2	Логнормальное; $\sigma = 1,2445; \mu = 2,1668$
Тест за модуль 3	Логнормальное; $\sigma = 1,9099; \mu = 2,7045$
Итоговый тест	Логнормальное; $\sigma = 1,3978; \mu = 3,3073$

Таблица 1: Параметры распределений для времени выполнения.

Процесс	Распределение и параметры
Лекция 1	Бета; $\alpha = 0,50334; \beta = 3,2459$ ; $a = -454,22; b = 6636,1$
Лекция 2	Бета; $\alpha = 0,03104; \beta = 0,47505$ ; $a = -36,283; b = 7277,9$
Лекция 3	Бета; $\alpha = 0,57029; \beta = 3,8224$ ; $a = -557,38; b = 7050,5$
Лекция 4	Бета; $\alpha = 0,09906; \beta = 1,0952$ ; $a = -63,817; b = 2114,1$
Лекция 5	Бета; $\alpha = 0,07788; \beta = 1,0816$ ; $a = -35,5; b = 2909,7$
Лекция 6	Бета; $\alpha = 0,05087; \beta = 0,58316$ ; $a = -34,717; b = 6714,8$
Лекция 7	Бета; $\alpha = 0,06641; \beta = 0,82094$ ; $a = -15,883; b = 4118,1$
Тест за модуль 1	Бета; $\alpha = 2,8511; \beta = 5,5235$ ; $a = -531,35; b = 1356,3$
Тест за модуль 2	Бета; $\alpha = 0,45604; \beta = 1,9273$ ; $a = -581,27; b = 7107,1$
Тест за модуль 3	Бета; $\alpha = 0,79715; \beta = 3,4763$ ; $a = -589,25; b = 4405,8$

Таблица 2: Параметры распределений для времени ожидания.

### 4 Реализация агентной модели

На основе модели процесса прохождения учебного курса, разработанной с помощью алгоритмов нечеткой логики (плагин Fuzzy Miner для Prom), была со-

здана имитационная агентная модель AnyLogic. Все переходы между процессами осуществляются по логике исключающего ИЛИ (XOR rule) и происходят последовательно. Агентом в имитационной модели является студент, для которого моделируется процесс прохождения учебного курса. Стейтchart для моделирования поведения агента представлен на рисунке 2.

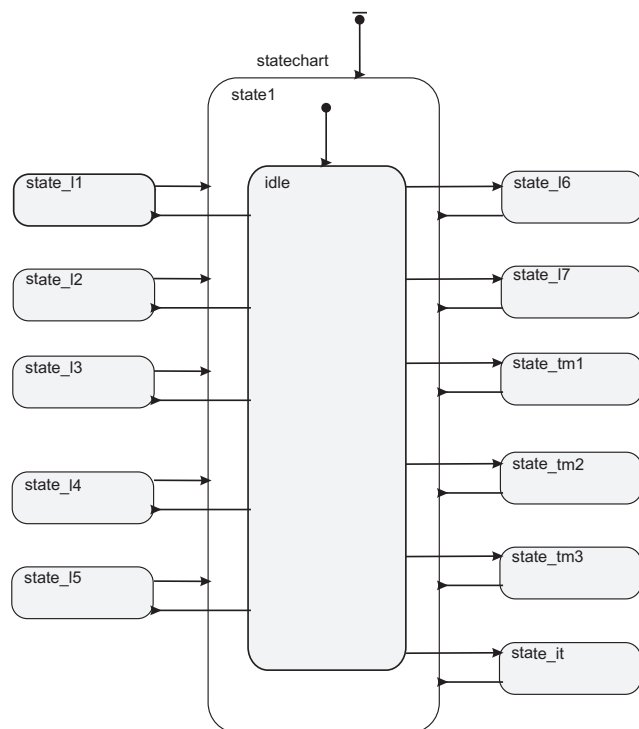


Рис. 2: Стейтchart для агентной имитационной модели процесса прохождения учебного курса.

Для созданной агентной имитационной модели был проведен один миллион экспериментальных прогонов и по полученным в результате данным рассчитаны математическое ожидание и дисперсия для трудоемкости, и общего времени прохождения учебного курса студентом (Таблица 3).

Показатель	Трудоемкость, мин	Общее время выполнения, мин
Математическое ожидание	352,71	3 184,58
Дисперсия	418 973,53	9 814 746,73

Таблица 3: Результаты имитационного эксперимента для трудоемкости и общего времени выполнения процесса, в минутах.

## 5 Заключение

Таким образом, на примере модели прохождения учебного курса показана возможность создания агентных имитационных моделей на основе схемы

процесса, восстановленного по журналу событий СДО. А также показано, что получившаяся агентная имитационная модель позволяет спрогнозировать трудоемкость и общее время прохождения учебного курса на основе статистических данных извлеченных из СДО.

## Список литературы

- [1] et al van der Aalst, Wil M.P. Process mining manifesto. In *BPM 2012 Workshops Proceedings*, Springer-Verlag, Campus des Cezeaux, Clermont-Ferrand:169–194, 2012.
- [2] W.M.P. van der Aalst. *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Springer-Verlag, Berlin, 2011.
- [3] C.W. Gunther and W.M.P. van der Aalst. Fuzzy mining: Adaptive process simplification based on multi-perspective metrics. In G. Alonso, P. Dadam, and M. Rosemann, editor, *International Conference on Business Process Management (BPM 2007)*, volume 4714, pages 328–343. Springer, Berlin, Germany, 2007.

# Энтропия множества событий и его подмножества

Наталья Александровна Лукьянова

Сибирский федеральный университет  
Институт фундаментальной подготовки  
Красноярск  
nata00sfu@gmail.com

**Аннотация.** В работе рассмотрена идея равновероятной детализации вероятностного распределения подмножества событий до размерности всего множества, чтобы расширить понятие относительной энтропии. Введено определение относительной квазиэнтропии. Рассмотрена относительная энтропия полного и неполного вероятностных распределений множества событий и его подмножества и проиллюстрирована на простом примере произвольного триплета событий.

**Ключевые слова.** Энтропия множества событий, энтропия подмножества событий, относительная энтропия, относительная квазиэнтропия.

## 1 Предварительные сведения

### 1.1 Вероятностное распределение множества событий

Рассмотрим всеобщее вероятностное пространство  $(\Omega, \mathcal{A}, \mathbf{P})$ . Пусть  $\mathfrak{X} \subset \mathcal{A}$  – конечное множество событий, выбранное из алгебры  $\mathcal{A}$  этого пространства, и которое определяется вероятностным распределением (эвентологическим распределением<sup>1</sup>) – набором

$$\{p(X//\mathfrak{X}), X \subseteq \mathfrak{X}\}$$

вероятностей

$$p(X//\mathfrak{X}) = \mathbf{P}(\text{ter}(X//\mathfrak{X})), X \subseteq \mathfrak{X}$$

террасных событий<sup>2</sup>

$$\begin{aligned} \text{ter}(X//\mathfrak{X}) &= \\ &= \left( \bigcap_{x \in X} x \right) \cap \left( \bigcap_{x \in X^c} x^c \right) = \bigcap_{x \in X} x \bigcap_{x \in X^c} x^c, \end{aligned} \quad (1)$$

которые образуют по всем  $X \subseteq \mathfrak{X}$  разбиение  $\Omega$ , порожденное  $\mathfrak{X}$ :

$$\Omega = \sum_{X \subseteq \mathfrak{X}} \text{ter}(X//\mathfrak{X})$$

© 2013 Н.А. Лукьянова

Олег Воробьев (ред.), Труды ХII ФАМЭБ'2013, Красноярск: СФУ

<sup>1</sup>Эвентологическим распределением [1] множества событий  $\mathfrak{X}$  называется набор из  $2^{|\mathfrak{X}|}$  вероятностей террасных событий, порожденных этим множеством событий.

<sup>2</sup>Здесь  $x^c = \Omega - x$ ,  $X^c = \mathfrak{X} - X$  – два вида дополнений: события  $x \subseteq \Omega$  до  $\Omega$ , и множества событий  $X \subseteq \mathfrak{X}$  до  $\mathfrak{X}$ .

и обеспечивают вероятностную нормировку:

$$\sum_{X \subseteq \mathfrak{X}} p(X//\mathfrak{X}) = 1.$$

### 1.2 Вероятностное распределение подмножества событий

Чем больше событий содержит избранное множество событий, тем более сложными структурами вероятностных зависимостей событий оно обладает. В работах [2, 3, 4] введено определение эвентологического подраспределения множества событий  $\mathfrak{X}$ , как вероятностного распределения подмножества событий  $W \subseteq \mathfrak{X}$ , мощности  $m = |W|$ ,  $m = 1, \dots, N$ ,  $N = |\mathfrak{X}|$ , которые являются наборами из  $2^m$  вероятностей. Это продиктовано необходимостью анализировать не всё множество  $\mathfrak{X}$  избранных событий целиком, а только её часть – подмножество событий из исходного множества избранных событий  $W \subseteq \mathfrak{X}$ .

Вероятностным распределением подмножества событий (эвентологическое подраспределение множества событий  $\mathfrak{X}$ )  $W \subseteq \mathfrak{X}$  мощности  $m = |W|$ ,  $m = 1, \dots, N$ ,  $N = |\mathfrak{X}|$ , называется набор из  $2^m$  вероятностей  $\{p(Y//W), Y \subseteq W\}$ , в котором

$$p(Y//W) = \mathbf{P}(\text{ter}(Y//W)), Y \subseteq W,$$

– вероятности террасных событий, порожденных этим множеством событий

$$\text{ter}(Y//W) =$$

$$= \left( \bigcap_{y \in Y} y \right) \cap \left( \bigcap_{y \in Y^c} y^c \right) = \bigcap_{y \in Y} y \bigcap_{y \in Y^c} y^c, \quad (2)$$

где  $Y^c = W - Y$  – дополнение множества событий  $Y \subseteq W$  до  $W$ . Подмножество событий  $W \subseteq \mathfrak{X}$  порождает ещё одно террасное разбиение  $\Omega$  по всем  $Y \subseteq W$ :

$$\Omega = \sum_{Y \subseteq W} \text{ter}(Y//W).$$

**Замечание 1.** Каждое террасное событие вероятностного распределения подмножества событий  $W \subseteq \mathfrak{X}$  есть сумма из  $2^{|\mathfrak{X}| - |W|} = 2^{N - m}$  слагаемых террасных событий исходного вероятностного распределения множества событий  $\mathfrak{X}$

$$\text{ter}(Y//W) = \sum_{Z \subseteq \mathfrak{X} - W} \text{ter}(Y + Z//\mathfrak{X}), Y \subseteq W, \quad (2')$$



где при  $Y \cap Z = \emptyset$ ,  $Y + Z \subseteq \mathfrak{X}$

$$\begin{aligned} \text{ter}(Y + Z // \mathfrak{X}) &= \left( \bigcap_{x \in Y+Z} x \right) \cap \left( \bigcap_{x \in \mathfrak{X} - (Y+Z)} x^c \right) = \\ &= \bigcap_{x \in Y+Z} x \bigcap_{x \in (Y+Z)^c} x^c = \bigcap_{x \in Y+Z} x \bigcap_{x \in Y^c \cap Z^c} x^c. \end{aligned}$$

Таким образом, для  $Y \subseteq W$

$$\text{ter}(Y // W) = \sum_{Z \subseteq \mathfrak{X} - W} \left( \bigcap_{x \in Y+Z} x \bigcap_{x \in Y^c \cap Z^c} x^c \right).$$

Тогда вероятностное распределение подмножества событий  $W \subseteq \mathfrak{X}$  определяется следующим образом:

$$\begin{aligned} p(Y // W) &= \mathbf{P}(\text{ter}(Y // W)) = \\ &= \mathbf{P} \left( \sum_{Z \subseteq \mathfrak{X} - W} \text{ter}(Y + Z) \right) = \\ &= \sum_{Z \subseteq \mathfrak{X} - W} p(Y + Z), \end{aligned} \quad (3)$$

с выполнением вероятностной нормировки:

$$\begin{aligned} \sum_{Y \subseteq W} p(Y // W) &= \sum_{Y \subseteq W} \sum_{Z \subseteq \mathfrak{X} - W} p(Y + Z) = \\ &= \mathbf{P}(\Omega) = \sum_{X \subseteq \mathfrak{X}} p(X) = 1. \end{aligned}$$

### 1.3 Террасное представление множества событий $\mathfrak{X}$ через ее подмножества событий

Здесь имеют место два случая:

- $\mathfrak{X} = W + W^c$

Пусть из множества событий  $\mathfrak{X}$  выбрано некоторое его непустое подмножество  $W \subset \mathfrak{X}$ , дополнение  $W^c = \mathfrak{X} - W$  этого подмножества до  $\mathfrak{X}$ . Рассмотрим теоретико-множественные соотношения для соответствующих террасных событий для  $Y \subseteq W$  и  $Z \subseteq W^c$ , учитывая формулу (3)

$$\begin{aligned} \text{ter}(Y // W) &= \bigcap_{x \in Y} x \bigcap_{x \in W - Y} x^c = \bigcap_{x \in Y} x \bigcap_{x \in Y^c} x^c. \\ \text{ter}(Z // W^c) &= \bigcap_{x \in Z} x \bigcap_{x \in W^c - Z} x^c = \bigcap_{x \in Z} x \bigcap_{x \in Z^c} x^c. \end{aligned}$$

Учитывая, что  $W \cap W^c = \emptyset$  имеем

$$Y^c + Z^c = W - Y + W^c - Z = (W + W^c) - (Y + Z) = \mathfrak{X} - (Y + Z) = (Y + Z)^c.$$

Таким образом,

$$\begin{aligned} \text{ter}(Y // W) \cap \text{ter}(Z // W^c) &= \\ &= \left( \bigcap_{x \in Y} x \bigcap_{x \in Y^c} x^c \right) \cap \left( \bigcap_{x \in Z} x \bigcap_{x \in Z^c} x^c \right) = \\ &= \bigcap_{x \in Y+Z} x \bigcap_{x \in Y^c+Z^c} x^c = \bigcap_{x \in Y+Z} x \bigcap_{x \in (Y+Z)^c} x^c = \\ &= \text{ter}(Y + Z // \mathfrak{X}). \end{aligned}$$

Следовательно для случая  $\mathfrak{X} = W + W^c$

$$\text{ter}(Y // W) \cap \text{ter}(Z // W^c) = \text{ter}(Y + Z // \mathfrak{X}), \quad (4)$$

где  $Y + Z \subseteq \mathfrak{X}$ .

- $\mathfrak{X} = W \cup U$

Пусть из множества событий  $\mathfrak{X}$  выбраны два непустых подмножества  $W \subset \mathfrak{X}$  и  $U \subset \mathfrak{X}$ . Рассмотрим теоретико-множественные соотношения для соответствующих террасных событий для  $Y \subseteq W$  и  $Z \subseteq U$

$$\begin{aligned} \text{ter}(Y // W) \cap \text{ter}(Z // U) &= \\ &= \left( \bigcap_{x \in Y} x \bigcap_{x \in Y^c} x^c \right) \cap \left( \bigcap_{x \in Z} x \bigcap_{x \in Z^c} x^c \right) = \\ &= \begin{cases} \emptyset, & (Y^c \cap Z \neq \emptyset) \text{ or} \\ & (Y \cap Z^c \neq \emptyset) \\ \text{ter}(Y \cup Z // \mathfrak{X}), & \text{otherwise,} \end{cases} \end{aligned} \quad (5)$$

где  $W - Y = Y^c$  и  $U - Z = Z^c$ .

Формула (4) является частным случаем (5):

учитывая, что  $W \cap W^c = \emptyset$  и  $Y \subseteq W$ ,  $Z \subseteq W^c$  имеем

$$\begin{aligned} Y^c \cap Z &= (W - Y) \cap Z = W \cap Y^c \cap Z = \\ &= (W \cap Z) \cap Y^c = \emptyset \cap Y^c = \emptyset. \end{aligned}$$

Аналогично получаем  $Z^c \cap Y = \emptyset$ .

## 2 Энтропия событий

*Энтропия* – мера неопределенности множества событий<sup>3</sup>. Если  $\mathfrak{X} \subset \mathcal{A}$  – конечное множество событий, выбранное из алгебры  $\mathcal{A}$  всеобщего вероятностного пространства  $(\Omega, \mathcal{A}, \mathbf{P})$  и происходящее в виде террасных событий

$$\{\text{ter}(X // \mathfrak{X}), X \subseteq \mathfrak{X}\}$$

<sup>3</sup>или в работах [1], [5] интерпретировалась как мера неопределенности эвентологического распределения  $\{p(X // \mathfrak{X}), X \subseteq \mathfrak{X}\}$  множества событий.

с вероятностями

$$\{p(X//\mathfrak{X}), X \subseteq \mathfrak{X}\},$$

где  $\text{ter}(X//\mathfrak{X})$  определяется по формуле (1), а

$$p(X//\mathfrak{X}) = \mathbf{P}(\text{ter}(X//\mathfrak{X})), X \subseteq \mathfrak{X},$$

то *энтропия множества событий*  $\mathfrak{X}$  определяется по классической формуле так же, как определяется в теории вероятностей энтропия вероятностного распределения, или энтропия случайной величины с конечным числом значений

$$H(p(X//\mathfrak{X})) = - \sum_{X \subseteq \mathfrak{X}} p(X//\mathfrak{X}) \ln p(X//\mathfrak{X}). \quad (6)$$

как средняя неопределенность одного террасного значения  $\mathfrak{X}$ .

По аналогии с (6) определим *энтропию подмножества событий*  $W \subseteq \mathfrak{X}$  мощности  $|W| = m$ ,  $m = 1, \dots, N$ ,  $N = |\mathfrak{X}|$ , как энтропию вероятностного распределения  $\{p(Y//W), Y \subseteq W\}$  на подмножестве событий  $W$ , как среднюю неопределенность одного террасного значения  $W$

$$H(p(Y//W)) = - \sum_{Y \subseteq W} p(Y//W) \ln p(Y//W).$$

## 2.1 Энтропийные свойства вероятностного распределения подмножества событий

Энтропия вероятностного распределения  $\{p(Y//W), Y \subseteq W\}$  подмножества событий  $W \subseteq \mathfrak{X}$  мощности  $|W|$ , такой что  $|W| = 1, \dots, N$ ,  $N = |\mathfrak{X}|$ :

### Свойство 1.

$$H(p(Y//W)) \geq 0 \quad (7)$$

— всегда неотрицательна.

### Свойство 2.

$$H(p(Y//W)) = - \sum_{Y \subseteq W} \frac{1}{2^{|W|}} \cdot \ln \frac{1}{2^{|W|}} = \ln 2^{|W|} \quad (8)$$

— достигает максимума равного  $\ln 2^{|W|}$  на равномерном распределении  $p(Y//W) = 2^{-|W|}$ .

### Свойство 3.

$$0 \leq H(p(Y//W)) \leq H(p(X//\mathfrak{X})). \quad (9)$$

— не превышает энтропию вероятностного распределения  $\{p(X//\mathfrak{X}), X \subseteq \mathfrak{X}\}$  исходного множества.

### Замечание 2.

$$H(p(X//\mathfrak{X})) = 0 \Rightarrow H(p(Y//W)) = 0,$$

когда одно из террасных событий множества  $\mathfrak{X}$  достоверно  $p(X//\mathfrak{X}) = 1$ , а другие — невозможны, т.е. неопределенность в вероятностном распределении  $p(X//\mathfrak{X})$  множества событий  $\mathfrak{X}$  отсутствует.

Обратно, собственно говоря, неверно, т.е. из  $H(p(Y//W)) = 0$  не следует, что  $H(p(X//\mathfrak{X})) = 0$ .

**Замечание 3.** Для множеств событий  $U \subseteq W \subseteq \mathfrak{X}$  справедливо:

$$H(p(Z//U)) \leq H(p(Y//W)). \quad (10)$$

## 2.2 Пример для равномерного распределения триплета событий и его подмножеств 1-ой и 2-ой мощности

Рассмотрим равномерное распределение триплета событий  $\mathfrak{X} = \{x, y, z\}$

$$p(X//\mathfrak{X}) = \{p(\emptyset//\mathfrak{X}), p(\{x\}//\mathfrak{X}), p(\{y\}//\mathfrak{X}),$$

$$p(\{z\}//\mathfrak{X}), p(\{x, y\}//\mathfrak{X}), p(\{x, z\}//\mathfrak{X}), p(\{y, z\}//\mathfrak{X}),$$

$$p(\{x, y, z\}//\mathfrak{X})\} = \left\{ \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8} \right\}.$$

Энтропия вероятностного распределения триплета событий  $\mathfrak{X} = \{x, y, z\}$  равна:

$$H(p(X//\mathfrak{X})) = \ln 2^{|\mathfrak{X}|} = \ln 2^3 = \ln 8.$$

Для равномерного распределения триплета событий  $\mathfrak{X} = \{x, y, z\}$  рассмотрим вероятностное распределение произвольного дуплета событий  $W = \{x, y\} \subset \mathfrak{X}$ ,  $|W| = 2$

$$p(Y//W) = p(Y//\{x, y\}) = \{p(\emptyset//\{x, y\}),$$

$$p(\{x\}//\{x, y\}), p(\{y\}//\{x, y\}), p(\{x, y\}//\{x, y\})\},$$

где

$$p(\emptyset//\{x, y\}) = p(\emptyset//\mathfrak{X}) + p(\{z\}//\mathfrak{X}) = \frac{1}{4};$$

$$p(\{x\}//\{x, y\}) = p(\{x\}//\mathfrak{X}) + p(\{x, z\}//\mathfrak{X}) = \frac{1}{4};$$

$$p(\{y\}//\{x, y\}) = p(\{y\}//\mathfrak{X}) + p(\{y, z\}//\mathfrak{X}) = \frac{1}{4};$$

$$p(\{x, y\}//\{x, y\}) = p(\{x, y\}//\mathfrak{X}) + p(\{x, y, z\}//\mathfrak{X}) = \frac{1}{4}.$$

Энтропия равномерного распределения дуплета событий  $W \subset \mathfrak{X}$  равна:

$$H(p(Y//\{x, y\})) = \ln 2^{|W|} = \ln 2^2 = \ln 4.$$

Рассмотрим вероятностное распределение произвольного моноплета событий  $U = \{z\} \subseteq \mathfrak{X}$ ,  $|U| = 1$ ,

$$p(Z//U) = p(Z//\{z\}) = \{p(\emptyset//\{z\}), p(\{z\}//\{z\})\}$$

где

$$p(\emptyset//\{z\}) = p(\emptyset//\mathfrak{X}) + p(\{x\}//\mathfrak{X}) +$$

$$+ p(\{y\}//\mathfrak{X}) + p(\{x, y\}//\mathfrak{X}) = \frac{1}{2};$$

$$p(\{z\}//\{z\}) = p(\{z\}//\mathfrak{X}) + p(\{x, z\}//\mathfrak{X}) +$$

$$+p(\{y, z\} \parallel \mathfrak{X}) + p(\{x, y, z\} \parallel \mathfrak{X}) = \frac{1}{2}.$$

Энтропия равновероятного распределения моноплета событий  $U \subset \mathfrak{X}$  равна:

$$H(p(Z \parallel \{z\})) = \ln 2^{|U|} = \ln 2^1 = \ln 2.$$

Получаем,

$$\ln 2 \leq \ln 4 \leq \ln 8,$$

$$H(p(Z \parallel \{z\})) \leq H(p(Y \parallel \{x, y\})) \leq H(p(X \parallel \mathfrak{X})).$$

### 2.3 Пример для произвольного триплета событий и его подмножеств 2-ой мощности

Рассмотрим разность энтропии триплета событий  $\mathfrak{X} = \{x, y, z\}$ ,  $|\mathfrak{X}| = 3$  и энтропии дуплета событий  $W = \{x, y\}$ ,  $|W| = 2$ :

$$\begin{aligned} H(p(X \parallel \mathfrak{X})) - H(p(Y \parallel W)) &= \\ &= - \sum_{X \subseteq \mathfrak{X}} p(X \parallel \mathfrak{X}) \ln p(X \parallel \mathfrak{X}) + \\ &+ (p(\emptyset \parallel \mathfrak{X}) + p(\{z\} \parallel \mathfrak{X})) \ln p(\emptyset \parallel W) + \\ &+ (p(\{x\} \parallel \mathfrak{X}) + p(\{x, z\} \parallel \mathfrak{X})) \ln p(\{x\} \parallel W) + \\ &+ (p(\{y\} \parallel \mathfrak{X}) + p(\{y, z\} \parallel \mathfrak{X})) \ln p(\{y\} \parallel W) + \\ &+ (p(\{x, y\} \parallel \mathfrak{X}) + p(\{x, y, z\} \parallel \mathfrak{X})) \ln p(\{x, y\} \parallel W) = \\ &= p(\emptyset \parallel \mathfrak{X}) \ln \left( \frac{p(\emptyset \parallel \mathfrak{X}) + p(\{z\} \parallel \mathfrak{X})}{p(\emptyset \parallel \mathfrak{X})} \right) + \\ &+ p(\{z\} \parallel \mathfrak{X}) \ln \left( \frac{p(\emptyset \parallel \mathfrak{X}) + p(\{z\} \parallel \mathfrak{X})}{p(\{z\} \parallel \mathfrak{X})} \right) + \\ &+ p(\{x\} \parallel \mathfrak{X}) \ln \left( \frac{p(\{x\} \parallel \mathfrak{X}) + p(\{x, z\} \parallel \mathfrak{X})}{p(\{x\} \parallel \mathfrak{X})} \right) + \\ &+ p(\{x, z\} \parallel \mathfrak{X}) \ln \left( \frac{p(\{x\} \parallel \mathfrak{X}) + p(\{x, z\} \parallel \mathfrak{X})}{p(\{x, z\} \parallel \mathfrak{X})} \right) + \\ &+ p(\{y\} \parallel \mathfrak{X}) \ln \left( \frac{p(\{y\} \parallel \mathfrak{X}) + p(\{y, z\} \parallel \mathfrak{X})}{p(\{y\} \parallel \mathfrak{X})} \right) + \\ &+ p(\{y, z\} \parallel \mathfrak{X}) \ln \left( \frac{p(\{y\} \parallel \mathfrak{X}) + p(\{y, z\} \parallel \mathfrak{X})}{p(\{y, z\} \parallel \mathfrak{X})} \right) + \\ &+ p(\{x, y\} \parallel \mathfrak{X}) \ln \left( \frac{p(\{x, y\} \parallel \mathfrak{X}) + p(\{x, y, z\} \parallel \mathfrak{X})}{p(\{x, y\} \parallel \mathfrak{X})} \right) + \\ &+ p(\{x, y, z\} \parallel \mathfrak{X}) \ln \left( \frac{p(\{x, y\} \parallel \mathfrak{X}) + p(\{x, y, z\} \parallel \mathfrak{X})}{p(\{x, y, z\} \parallel \mathfrak{X})} \right) = \\ &= p(\emptyset \parallel \mathfrak{X}) \ln \left( 1 + \frac{p(\{z\} \parallel \mathfrak{X})}{p(\emptyset \parallel \mathfrak{X})} \right) + \\ &+ p(\{z\} \parallel \mathfrak{X}) \ln \left( \frac{p(\emptyset \parallel \mathfrak{X})}{p(\{z\} \parallel \mathfrak{X})} + 1 \right) + \\ &+ p(\{x\} \parallel \mathfrak{X}) \ln \left( 1 + \frac{p(\{x, z\} \parallel \mathfrak{X})}{p(\{x\} \parallel \mathfrak{X})} \right) + \end{aligned}$$

$$\begin{aligned} &+ p(\{x, z\} \parallel \mathfrak{X}) \ln \left( \frac{p(\{x\} \parallel \mathfrak{X})}{p(\{x, z\} \parallel \mathfrak{X})} + 1 \right) + \\ &+ p(\{y\} \parallel \mathfrak{X}) \ln \left( 1 + \frac{p(\{y, z\} \parallel \mathfrak{X})}{p(\{y\} \parallel \mathfrak{X})} \right) + \\ &+ p(\{y, z\} \parallel \mathfrak{X}) \ln \left( \frac{p(\{y\} \parallel \mathfrak{X})}{p(\{y, z\} \parallel \mathfrak{X})} + 1 \right) + \\ &+ p(\{x, y\} \parallel \mathfrak{X}) \ln \left( 1 + \frac{p(\{x, y, z\} \parallel \mathfrak{X})}{p(\{x, y\} \parallel \mathfrak{X})} \right) + \\ &+ p(\{x, y, z\} \parallel \mathfrak{X}) \ln \left( \frac{p(\{x, y\} \parallel \mathfrak{X})}{p(\{x, y, z\} \parallel \mathfrak{X})} + 1 \right) > 0, \end{aligned}$$

т.к. в каждом слагаемом  $0 < p(X) < 1$  и дробь под знаком логарифма всегда больше единицы, поскольку знаменатель входит в сумму числителя. Таким образом,

$$H(p(X \parallel \mathfrak{X})) > H(p(Y \parallel W)), W \subseteq \mathfrak{X}.$$

## 3 Об относительной энтропии между множеством событий и его подмножеством

### 3.1 Равновероятная детализация вероятностного распределения подмножества событий

О.Ю.Воробьевым была предложена идея *равновероятно детализировать* вероятностное распределение подмножества событий до размерности всего множества (так, чтобы сохранить вероятностное распределение подмножества), а затем расширить понятие относительной энтропии, считая по определению относительную энтропию между равновероятно детализированным подмножеством и множеством событий равной относительной энтропии между множеством событий и его данным подмножеством.

Пусть из множества событий  $\mathfrak{X}$ , имеющего *полное*<sup>4</sup> вероятностное распределение, выбрано некоторое его непустое подмножество  $W \subset \mathfrak{X}$ , которое имеет вероятностное распределение для  $Y \subseteq W$  вида

$$p(Y \parallel W) = \mathbf{P}(\text{ter}(Y \parallel W)) = \mathbf{P} \left( \bigcap_{y \in Y} y \bigcap_{y \in W - Y} y^c \right).$$

Дополнение  $W^c = \mathfrak{X} - W$  этого подмножества до  $\mathfrak{X}$  имеет *полное* вероятностное распределение для  $Z \subseteq W^c$  вида

$$p(Z \parallel W^c) = \mathbf{P}(\text{ter}(Z \parallel W^c)) = \mathbf{P} \left( \bigcap_{z \in Z} z \bigcap_{z \in W^c - Z} z^c \right).$$

<sup>4</sup>Вероятностное распределение множества событий  $\mathfrak{X}$ , как и само  $\mathfrak{X}$ , называется *полным*, если все террасные события  $\text{ter}(X \parallel \mathfrak{X})$ , порожденные  $\mathfrak{X}$ , имеют ненулевые вероятности, т.е. отличны от невозможного события, так как вероятность  $\mathbf{P}$  считается *регулярной*.

Вероятностное распределение

$$p(X//\mathfrak{X}) = \mathbf{P}(\text{ter}(X//\mathfrak{X})) = \mathbf{P}\left(\bigcap_{x \in X} x \bigcap_{x \in X^c} y^c\right)$$

всего множества событий  $\mathfrak{X}$  определяет эти два распределения по симметричным формулам для  $Y \subseteq W$  и  $Z \subseteq W^c$ :

$$\begin{aligned} p(Y//W) &= \sum_{Z \subseteq W^c} \mathbf{P}(\text{ter}(Y//W) \cap \text{ter}(Z//W^c)) = \\ &= \sum_{Z \subseteq W^c} p(Y + Z//\mathfrak{X}), \end{aligned}$$

$$\begin{aligned} p(Z//W^c) &= \sum_{Y \subseteq W} \mathbf{P}(\text{ter}(Y//W) \cap \text{ter}(Z//W^c)) = \\ &= \sum_{Y \subseteq W} p(Y + Z//\mathfrak{X}), \end{aligned}$$

которые следуют из аналогичных теоретико-множественных соотношений для соответствующих террасных событий:

$$\begin{aligned} \text{ter}(Y//W) &= \sum_{Z \subseteq W^c} \text{ter}(Y//W) \cap \text{ter}(Z//W^c) = \\ &= \sum_{Z \subseteq W^c} \text{ter}(Y + Z//\mathfrak{X}), \end{aligned} \quad (11)$$

$$\begin{aligned} \text{ter}(Z//W^c) &= \sum_{Y \subseteq W} \text{ter}(Y//W) \cap \text{ter}(Z//W^c) = \\ &= \sum_{Y \subseteq W} \text{ter}(Y + Z//\mathfrak{X}), \end{aligned} \quad (12)$$

Эти формулы помогают наглядно формализовать идею равновероятной детализации вероятностного распределения подмножества  $W \subset \mathfrak{X}$ , дополняющей его до размерности вероятностного распределения всего множества событий  $\mathfrak{X}$ .

*Равновероятная детализация*  $p_W(X//\mathfrak{X})$  вероятностного распределения  $p(Y//W)$  подмножества  $W$  определяется для всех  $Y \subseteq W$  и  $Z \subseteq W^c$  формулами:

$$p_W(X//\mathfrak{X}) = p_W(Y + Z//\mathfrak{X}) = \frac{1}{2^{|W^c|}} p(Y//W). \quad (13)$$

Аналогичные формулы справедливы для равновероятной детализации  $p_{W^c}(X//\mathfrak{X})$ :

$$p_{W^c}(X//\mathfrak{X}) = p_{W^c}(Y + Z//\mathfrak{X}) = \frac{1}{2^{|W|}} p(Z//W^c). \quad (14)$$

**Замечание 4.** Формулы (13) и (14), справедливые для множеств, имеющих полные вероятностные распределения, могут быть обобщены для множеств событий  $\mathfrak{X}$  с *неполным* вероятностным распределением и  $W \subseteq \mathfrak{X}$  следующим образом:

$$p_W(Y + Z//\mathfrak{X}) = \frac{1}{N_{W^c}(Y)} p(Y//W). \quad (13')$$

$$p_{W^c}(Y + Z//\mathfrak{X}) = \frac{1}{N_W(Z)} p(Z//W^c), \quad (14')$$

где

$$N_{W^c}(Y) = |\{Z \subseteq W^c : \text{ter}(Y + Z//\mathfrak{X}) \neq 0\}| \leq 2^{|W^c|},$$

$$N_W(Z) = |\{Y \subseteq W : \text{ter}(Y + Z//\mathfrak{X}) \neq 0\}| \leq 2^{|W|}$$

— два числа подмножеств  $Z \subseteq W^c$  и  $Y \subseteq W$  соответственно, для которых террасные события  $\text{ter}(Y + Z//\mathfrak{X}) \subseteq \Omega$  отличны от невозможного события  $\emptyset \subseteq \Omega$ .

### 3.2 $H(p_W(X//\mathfrak{X})) - H(p(Y//W))$

Пусть энтропия вероятностного распределения подмножества  $W$ :

$$H(p(Y//W)) = - \sum_{Y \subseteq W} p(Y//W) \ln p(Y//W),$$

а энтропия для равновероятной детализации  $p_W(X//\mathfrak{X})$  вероятностного распределения  $p(Y//W)$  подмножества  $W$  определяется для всех  $Y \subseteq W$  и  $Z \subseteq W^c$  следующим образом:

$$\begin{aligned} H(p_W(X//\mathfrak{X})) &= - \sum_{X \subseteq \mathfrak{X}} p_W(X//\mathfrak{X}) \ln p_W(X//\mathfrak{X}) = \\ &= -2^{|W^c|} \sum_{Y \subseteq W} \frac{1}{2^{|W^c|}} p(Y//W) \ln \left( \frac{1}{2^{|W^c|}} p(Y//W) \right) = \\ &= - \sum_{Y \subseteq W} p(Y//W) \ln \left( \frac{1}{2^{|W^c|}} p(Y//W) \right). \end{aligned}$$

Тогда разность этих энтропий:

$$\begin{aligned} H(p_W(X//\mathfrak{X})) - H(p(Y//W)) &= \\ &= \ln 2^{|W^c|} = |W^c| \ln 2. \end{aligned} \quad (15)$$

### 3.3 Относительная квазиэнтропия

Далее, рассмотрим разность энтропий вероятностных распределений всего множества  $\mathfrak{X}$  и его подмножества  $W$ :

$$\begin{aligned} H(p(X//\mathfrak{X})) - H(p(Y//W)) &= \\ &= - \sum_{X \subseteq \mathfrak{X}} p(X//\mathfrak{X}) \ln p(X//\mathfrak{X}) + \\ &\quad + \sum_{Y \subseteq W} p(Y//W) \ln p(Y//W) = \\ &= - \sum_{Y \subseteq W} \sum_{Z \subseteq W^c} p(Y + Z//\mathfrak{X}) \ln p(Y + Z//\mathfrak{X}) + \\ &\quad + \sum_{Y \subseteq W} \sum_{Z \subseteq W^c} p(Y + Z//\mathfrak{X}) \ln \left( \sum_{Z \subseteq W^c} p(Y + Z//\mathfrak{X}) \right) = \end{aligned}$$

$$\begin{aligned}
&= \sum_{Y \subseteq W} \sum_{Z \subseteq W^c} p(Y + Z \| \mathfrak{X}) \ln \left( \frac{\sum_{Z \subseteq W^c} p(Y + Z \| \mathfrak{X})}{p(Y + Z \| \mathfrak{X})} \right) = \\
&= \sum_{Y \subseteq W} \sum_{Z \subseteq W^c} p(Y + Z \| \mathfrak{X}) \ln \left( \frac{p(Y \| W)}{p(Y + Z \| \mathfrak{X})} \right) = \\
&= - \sum_{Y \subseteq W} \sum_{Z \subseteq W^c} p(Y + Z \| \mathfrak{X}) \ln \left( \frac{p(Y + Z \| \mathfrak{X})}{p(Y \| W)} \right) = H_{\mathfrak{X}|W}.
\end{aligned}$$

Назовем полученную величину *относительной квазиэнтропией*:

$$H(p(X \| \mathfrak{X})) - H(p(Y \| W)) = H_{\mathfrak{X}|W}, \quad (16)$$

как величину отклонения между вероятностным распределением исходного множества событий и вероятностным распределением его подмножества.

**Замечание 5.** Относительная квазиэнтропия  $H_{\mathfrak{X}|W} > 0$ .

### 3.4 Относительная энтропия

Теперь можно определить относительную энтропию<sup>5</sup> между множеством событий  $\mathfrak{X}$  и его подмножеством  $W \subset \mathfrak{X}$  как относительную энтропию между вероятностными распределениями  $p(X \| \mathfrak{X})$  и  $p_W(X \| \mathfrak{X})$ , которые имеют одинаковую размерность и одну и ту же проекцию на подмножестве  $W \subset \mathfrak{X}$ .

$$H_{\mathfrak{X}|W} = \sum_{X \subseteq \mathfrak{X}} p(X \| \mathfrak{X}) \ln \frac{p(X \| \mathfrak{X})}{p_W(X \| \mathfrak{X})} \quad (17)$$

или

$$H_{W|\mathfrak{X}} = \sum_{X \subseteq \mathfrak{X}} p_W(X \| \mathfrak{X}) \ln \frac{p_W(X \| \mathfrak{X})}{p(X \| \mathfrak{X})}. \quad (18)$$

#### 3.4.1 Относительная энтропия вида $H_{\mathfrak{X}|W}$

Опираясь на (16) и условие глобальной нормировки

$$\sum_{Y \subseteq W} \sum_{Z \subseteq W^c} p(Y + Z \| \mathfrak{X}) = \sum_{X \subseteq \mathfrak{X}} p(X \| \mathfrak{X}) = 1$$

рассмотрим подробнее формулу (17):

$$\begin{aligned}
H_{\mathfrak{X}|W} &= \sum_{X \subseteq \mathfrak{X}} p(X \| \mathfrak{X}) \ln \frac{p(X \| \mathfrak{X})}{p_W(X \| \mathfrak{X})} = \\
&= \sum_{Y \subseteq W} \sum_{Z \subseteq W^c} p(Y + Z \| \mathfrak{X}) \ln \left( \frac{2^{|W^c|} p(Y + Z \| \mathfrak{X})}{p(Y \| W)} \right) = \\
&= \ln 2^{|W^c|} - H_{\mathfrak{X}|W} = \\
&= \ln 2^{|W^c|} - H(p(X \| \mathfrak{X})) + H(p(Y \| W)) = \\
&= H(p_W(X \| \mathfrak{X})) - H(p(Y \| W)) - H(p(X \| \mathfrak{X})) + \\
&\quad + H(p(Y \| W)) = H(p_W(X \| \mathfrak{X})) - H(p(X \| \mathfrak{X})).
\end{aligned}$$

<sup>5</sup> Нессимметричное отклонение.

Таким образом, разность энтропий равновероятно детализированного вероятностного распределения  $p_W(X \| \mathfrak{X})$  подмножества событий  $W$  и вероятностного распределения  $p(X \| \mathfrak{X})$  множества событий  $\mathfrak{X}$  есть относительная энтропия между множеством событий  $\mathfrak{X}$  и его подмножеством  $W$ :

$$H(p_W(X \| \mathfrak{X})) - H(p(X \| \mathfrak{X})) = H_{\mathfrak{X}|W}. \quad (17')$$

#### 3.4.2 Относительная энтропия вида $H_{W|\mathfrak{X}}$

Поскольку для относительной энтропии не выполняется свойство симметричности, т.е.  $H_{\mathfrak{X}|W} \neq H_{W|\mathfrak{X}}$ , рассмотрим теперь подробнее формулу (18):

$$\begin{aligned}
H_{W|\mathfrak{X}} &= \sum_{X \subseteq \mathfrak{X}} p_W(X \| \mathfrak{X}) \ln \frac{p_W(X \| \mathfrak{X})}{p(X \| \mathfrak{X})} = \\
&= \sum_{Y \subseteq W} \sum_{Z \subseteq W^c} \frac{1}{2^{|W^c|}} p(Y \| W) \ln \left( \frac{\frac{1}{2^{|W^c|}} p(Y \| W)}{p(Y + Z \| \mathfrak{X})} \right) = \\
&= \frac{1}{2^{|W^c|}} \sum_{Y \subseteq W} \sum_{Z \subseteq W^c} \left[ p(Y \| W) \ln \left( \frac{p(Y \| W)}{p(Y + Z \| \mathfrak{X})} \right) \right] - \\
&\quad - \ln 2^{|W^c|} \sum_{Y \subseteq W} \sum_{Z \subseteq W^c} \frac{1}{2^{|W^c|}} p(Y \| W),
\end{aligned}$$

где

$$\sum_{Y \subseteq W} \sum_{Z \subseteq W^c} \frac{1}{2^{|W^c|}} p(Y \| W) = \sum_{X \subseteq \mathfrak{X}} p_W(X \| \mathfrak{X}) = 1.$$

Введем обозначение для первого слагаемого и продолжим его дальше расписывать

$$\begin{aligned}
&\frac{1}{2^{|W^c|}} H^* = \\
&= \sum_{Y \subseteq W} \sum_{Z \subseteq W^c} \left[ \sum_{K \subseteq W^c} p(Y + K \| \mathfrak{X}) \right] \cdot \ln \left( \frac{p(Y \| W)}{p(Y + Z \| \mathfrak{X})} \right) = \\
&= \sum_{Y \subseteq W} \sum_{Z \subseteq W^c} p(Y + Z \| \mathfrak{X}) \cdot \sum_{K \subseteq W^c} \left[ \ln \left( \frac{p(Y \| W)}{p(Y + K \| \mathfrak{X})} \right) \right] = \\
&= \sum_{Y \subseteq W} \sum_{Z \subseteq W^c} p(Y + Z \| \mathfrak{X}) \cdot \ln \prod_{K \subseteq W^c} \left( \frac{p(Y \| W)}{p(Y + K \| \mathfrak{X})} \right) = \\
&= \sum_{Y \subseteq W} \sum_{Z \subseteq W^c} p(Y + Z \| \mathfrak{X}) \cdot \ln \left( \frac{(p(Y \| W))^{2^{|W^c|}}}{\prod_{K \subseteq W^c} p(Y + K \| \mathfrak{X})} \right) = \\
&= \sum_{Y \subseteq W} \sum_{Z \subseteq W^c} p(Y + Z \| \mathfrak{X}) \cdot \ln \left( \frac{(p(Y \| W))^{2^{|W^c|}}}{\prod_{K \subseteq W^c} p(Y + K \| \mathfrak{X})} \right).
\end{aligned}$$

Итак,

$$H^* =$$



$$\begin{aligned}
&= \sum_{Y \subseteq W} \sum_{Z \subseteq W^c} p(Y + Z // \mathfrak{X}) \cdot \ln \left( \frac{p(Y // W)}{\prod_{K \subseteq W^c} p(Y + K // \mathfrak{X})} \right) = \\
&= \sum_{Y \subseteq W} \sum_{Z \subseteq W^c} p(Y + Z // \mathfrak{X}) \cdot \ln \left( \sum_{Z \subseteq W^c} p(Y + Z // \mathfrak{X}) \right) - \\
&- \sum_{Y \subseteq W} \sum_{Z \subseteq W^c} p(Y + Z // \mathfrak{X}) \cdot \ln \left( \prod_{K \subseteq W^c} p(Y + K // \mathfrak{X}) \right) = \\
&= -H(p(Y // W)) - \sum_{X \subseteq \mathfrak{X}} p(X // \mathfrak{X}) \ln p(X // \mathfrak{X}) - \\
&- \sum_{Y \subseteq W} \sum_{Z \subseteq W^c} p(Y + Z // \mathfrak{X}) \cdot \ln \left( \prod_{K \subseteq W^c - Z} p(Y + K // \mathfrak{X}) \right).
\end{aligned}$$

Таким образом, получаем следующий вид для относительной энтропии вида (18):

$$\begin{aligned}
\mathcal{H}_{W|\mathfrak{X}} &= -\ln 2^{|W^c|} - H(p(Y // W)) + H(p(X // \mathfrak{X})) - \\
&- \sum_{Y \subseteq W} \sum_{Z \subseteq W^c} p(Y + Z // \mathfrak{X}) \cdot \ln \left( \prod_{K \subseteq W^c - Z} p(Y + K // \mathfrak{X}) \right)
\end{aligned}$$

или

$$\begin{aligned}
\mathcal{H}_{W|\mathfrak{X}} &= H(p(X // \mathfrak{X})) - H(p_W(X // \mathfrak{X})) - \\
&- \sum_{Y \subseteq W} \sum_{Z \subseteq W^c} p(Y + Z // \mathfrak{X}) \cdot \ln \left( \prod_{K \subseteq W^c - Z} p(Y + K // \mathfrak{X}) \right).
\end{aligned}$$

Отсюда, получаем  $J$ -дивергенцию:

$$\begin{aligned}
J(\mathfrak{X}, W) &= \mathcal{H}_{W|\mathfrak{X}} + \mathcal{H}_{\mathfrak{X}|W} = \\
&- \sum_{Y \subseteq W} \sum_{Z \subseteq W^c} p(Y + Z // \mathfrak{X}) \cdot \ln \left( \prod_{K \subseteq W^c - Z} p(Y + K // \mathfrak{X}) \right).
\end{aligned}$$

### 3.4.3 Пример для относительной энтропии вида $\mathcal{H}_{\mathfrak{X}|W}$

Рассмотрим вероятностное распределение триплета событий  $\mathfrak{X} = \{x, y, z\}$

$$\begin{aligned}
p(X // \mathfrak{X}) &= \{p(\emptyset // \mathfrak{X}), p(\{x\} // \mathfrak{X}), p(\{y\} // \mathfrak{X}), \\
&p(\{z\} // \mathfrak{X}), p(\{x, y\} // \mathfrak{X}), p(\{x, z\} // \mathfrak{X}), p(\{y, z\} // \mathfrak{X}), \\
&p(\{x, y, z\} // \mathfrak{X})\}.
\end{aligned}$$

и вероятностное распределение произвольного моноплета событий  $W = \{z\} \subseteq \mathfrak{X}$ ,  $|W| = 1$ ,

$$p(Z // W) = p(Z // \{z\}) = \{p(\emptyset // \{z\}), p(\{z\} // \{z\})\}$$

где

$$\begin{aligned}
p(\emptyset // \{z\}) &= p(\emptyset // \mathfrak{X}) + p(\{x\} // \mathfrak{X}) + \\
&+ p(\{y\} // \mathfrak{X}) + p(\{x, y\} // \mathfrak{X});
\end{aligned}$$

$$\begin{aligned}
p(\{z\} // \{z\}) &= p(\{z\} // \mathfrak{X}) + p(\{x, z\} // \mathfrak{X}) + \\
&+ p(\{y, z\} // \mathfrak{X}) + p(\{x, y, z\} // \mathfrak{X}).
\end{aligned}$$

Тогда равновероятная детализация  $p_W(X // \mathfrak{X})$  вероятностного распределения  $p(Y // W)$  подмножества  $W$ :

$$\begin{aligned}
p_W(\emptyset // \mathfrak{X}) &= \frac{1}{4} p(\emptyset // \{z\}); \\
p_W(\{x\} // \mathfrak{X}) &= \frac{1}{4} p(\emptyset // \{z\}); \\
p_W(\{y\} // \mathfrak{X}) &= \frac{1}{4} p(\emptyset // \{z\}); \\
p_W(\{z\} // \mathfrak{X}) &= \frac{1}{4} p(\{z\} // \{z\}); \\
p_W(\{x, y\} // \mathfrak{X}) &= \frac{1}{4} p(\emptyset // \{z\}); \\
p_W(\{x, z\} // \mathfrak{X}) &= \frac{1}{4} p(\{z\} // \{z\}); \\
p_W(\{y, z\} // \mathfrak{X}) &= \frac{1}{4} p(\{z\} // \{z\}); \\
p_W(\{x, y, z\} // \mathfrak{X}) &= \frac{1}{4} p(\{z\} // \{z\});
\end{aligned}$$

Итак,

$$\begin{aligned}
\mathcal{H}_{\mathfrak{X}|W} &= \sum_{X \subseteq \mathfrak{X}} p(X // \mathfrak{X}) \ln \frac{p(X // \mathfrak{X})}{p_W(X // \mathfrak{X})} = \\
&= \sum_{X \subseteq \mathfrak{X}} p(X // \mathfrak{X}) \ln p(X // \mathfrak{X}) + \ln 4 \sum_{X \subseteq \mathfrak{X}} p(X // \mathfrak{X}) - \\
&- p(\emptyset // \mathfrak{X}) \ln p(\emptyset // \{z\}) - \\
&- p(\{x\} // \mathfrak{X}) \ln p(\emptyset // \{z\}) - \\
&- p(\{y\} // \mathfrak{X}) \ln p(\emptyset // \{z\}) - \\
&- p(\{z\} // \mathfrak{X}) \ln p(\{z\} // \{z\}) - \\
&- p(\{x, y\} // \mathfrak{X}) \ln p(\emptyset // \{z\}) - \\
&- p(\{x, z\} // \mathfrak{X}) \ln p(\{z\} // \{z\}) - \\
&- p(\{y, z\} // \mathfrak{X}) \ln p(\{z\} // \{z\}) - \\
&- p(\{x, y, z\} // \mathfrak{X}) \ln p(\{z\} // \{z\}) = \\
&= \sum_{X \subseteq \mathfrak{X}} p(X // \mathfrak{X}) \ln p(X // \mathfrak{X}) + \ln 4 - \\
&- [p(\emptyset // \mathfrak{X}) + p(\{x\} // \mathfrak{X}) + \\
&+ p(\{y\} // \mathfrak{X}) + p(\{x, y\} // \mathfrak{X})] \ln p(\emptyset // \{z\}) - \\
&- [p(\{z\} // \mathfrak{X}) + p(\{x, z\} // \mathfrak{X}) + \\
&+ p(\{y, z\} // \mathfrak{X}) + p(\{x, y, z\} // \mathfrak{X})] \ln p(\{z\} // \{z\}) = \\
&= \sum_{X \subseteq \mathfrak{X}} p(X // \mathfrak{X}) \ln p(X // \mathfrak{X}) - \sum_{Y \subseteq W} p(Y // W) \ln p(Y // W) + \\
&+ \ln 4 = \ln 4 - H(p(X // \{x, y, z\})) + H(p(Y // \{z\})).
\end{aligned}$$

### 3.4.4 Пример для относительной энтропии вида $\mathcal{H}_{W|\mathfrak{X}}$

А теперь рассмотрим относительную энтропию вида  $\mathcal{H}_{W|\mathfrak{X}}$  для вероятностного распределения триплета событий  $\mathfrak{X} = \{x, y, z\}$  и равновероятной детализации  $p_W(X//\mathfrak{X})$  вероятностного распределения  $p(Y//\{z\})$ .

$$\begin{aligned}\mathcal{H}_{W|\mathfrak{X}} &= \sum_{X \subseteq \mathfrak{X}} p_W(X//\mathfrak{X}) \ln \frac{p_W(X//\mathfrak{X})}{p(X//\mathfrak{X})} = \\ &= -\ln 4 + \frac{1}{4} p(\emptyset//\{z\}) \times \\ &\quad \times \ln \frac{(p(\emptyset//\{z\}))^4}{p(\emptyset//\mathfrak{X}) p(\{x\}//\mathfrak{X}) p(\{y\}//\mathfrak{X}) p(\{x, y\}//\mathfrak{X})} + \\ &\quad + \frac{1}{4} p(\{z\}//\{z\}) \times \\ &\quad \times \ln \frac{(p(\{z\}//\{z\}))^4}{p(\{z\}//\mathfrak{X}) p(\{x, z\}//\mathfrak{X}) p(\{y, z\}//\mathfrak{X}) p(\{x, y, z\}//\mathfrak{X})}.\end{aligned}$$

Итак,

$$\begin{aligned}\mathcal{H}_{W|\mathfrak{X}} &= -\ln 4 - H(p(Y//\{z\})) + H(p(X//\{x, y, z\})) - \\ &\quad - p(\emptyset//\mathfrak{X}) \ln(p(\{x\}//\mathfrak{X}) \cdot p(\{y\}//\mathfrak{X}) \cdot p(\{x, y\}//\mathfrak{X})) - \\ &\quad - p(\{x\}//\mathfrak{X}) \ln(p(\emptyset//\mathfrak{X}) \cdot p(\{y\}//\mathfrak{X}) \cdot p(\{x, y\}//\mathfrak{X})) - \\ &\quad - p(\{y\}//\mathfrak{X}) \ln(p(\emptyset//\mathfrak{X}) \cdot p(\{x\}//\mathfrak{X}) \cdot p(\{x, y\}//\mathfrak{X})) - \\ &\quad - p(\{x, y\}//\mathfrak{X}) \ln(p(\emptyset//\mathfrak{X}) \cdot p(\{x\}//\mathfrak{X}) \cdot p(\{y\}//\mathfrak{X})) - \\ &\quad - p(\{z\}//\mathfrak{X}) \ln(p(\{x, z\}//\mathfrak{X}) \cdot p(\{y, z\}//\mathfrak{X}) \cdot p(\mathfrak{X}//\mathfrak{X})) - \\ &\quad - p(\{x, z\}//\mathfrak{X}) \ln(p(\{z\}//\mathfrak{X}) \cdot p(\{y, z\}//\mathfrak{X}) \cdot p(\mathfrak{X}//\mathfrak{X})) - \\ &\quad - p(\{y, z\}//\mathfrak{X}) \ln(p(\{z\}//\mathfrak{X}) \cdot p(\{x, z\}//\mathfrak{X}) \cdot p(\mathfrak{X}//\mathfrak{X})) - \\ &\quad - p(\mathfrak{X}//\mathfrak{X}) \ln(p(\{x, z\}//\mathfrak{X}) \cdot p(\{y, z\}//\mathfrak{X}) \cdot p(\{z\}//\mathfrak{X})) = \\ &= -\ln 4 - H(p(Y//\{z\})) + H(p(X//\{x, y, z\})) - \\ &\quad - \sum_{Y \subseteq \{z\}} \sum_{Z \subseteq \{x, y\}} p(Y + Z//\mathfrak{X}) \ln \prod_{K \subseteq \{x, y\} - Z} p(Y + K//\mathfrak{X}).\end{aligned}$$

### 3.4.5 Пример для неполного вероятностного распределения множества событий

Рассмотрим *неполное* вероятностное распределение триплета событий  $\mathfrak{X} = \{x, y, z\}$ :

$$\begin{aligned}p(\emptyset//\mathfrak{X}) &= \frac{1}{8}, \\ p(\{x\}//\mathfrak{X}) &= \frac{1}{8}, \\ p(\{y\}//\mathfrak{X}) &= 0, \\ p(\{x, y\}//\mathfrak{X}) &= \frac{3}{8}, \\ p(\{z\}//\mathfrak{X}) &= \frac{2}{8}, \\ p(\{x, z\}//\mathfrak{X}) &= 0, \\ p(\{y, z\}//\mathfrak{X}) &= 0, \\ p(\{x, y, z\}//\mathfrak{X}) &= \frac{1}{8}.\end{aligned}$$

и вероятностное распределение произвольного моноплета событий  $W = \{z\} \subseteq \mathfrak{X}$ ,

$$p(Y//W) = p(Y//\{z\}) = \{p(\emptyset//\{z\}), p(\{z\}//\{z\})\}$$

где

$$\begin{aligned}p(\emptyset//\{z\}) &= p(\emptyset//\mathfrak{X}) + p(\{x\}//\mathfrak{X}) + \\ &\quad + p(\{y\}//\mathfrak{X}) + p(\{x, y\}//\mathfrak{X}) = \frac{5}{8};\end{aligned}$$

$$\begin{aligned}p(\{z\}//\{z\}) &= p(\{z\}//\mathfrak{X}) + p(\{x, z\}//\mathfrak{X}) + \\ &\quad + p(\{y, z\}//\mathfrak{X}) + p(\{x, y, z\}//\mathfrak{X}) = \frac{3}{8}.\end{aligned}$$

Тогда согласно Замечанию 4, найдем для равновероятной детализации  $p_W(X//\mathfrak{X})$  вероятностного распределения  $p(Y//W)$  подмножества  $W$  числа  $N_{W^c}(Y)$  по формуле (13'):

$$N_{W^c}(\emptyset) = |\{Z \subseteq \{x, y\} : \text{ter}(\emptyset + Z//\mathfrak{X}) \neq 0\}| = 3,$$

$$N_{W^c}(\{z\}) = |\{Z \subseteq \{x, y\} : \text{ter}(\{z\} + Z//\mathfrak{X}) \neq 0\}| = 2.$$

Строим *неполную* равновероятную детализацию  $p_W(X//\mathfrak{X})$  вероятностного распределения  $p(Y//W)$ :

$$\begin{aligned}p_W(\emptyset//\mathfrak{X}) &= \frac{1}{3} \cdot p(\emptyset//\{z\}) = \frac{5}{24}; \\ p_W(\{x\}//\mathfrak{X}) &= \frac{1}{3} \cdot p(\emptyset//\{z\}) = \frac{5}{24}; \\ p_W(\{y\}//\mathfrak{X}) &= 0; \\ p_W(\{x, y\}//\mathfrak{X}) &= \frac{1}{3} \cdot p(\emptyset//\{z\}) = \frac{5}{24}; \\ p_W(\{z\}//\mathfrak{X}) &= \frac{1}{2} \cdot p(\{z\}//\{z\}) = \frac{3}{16}; \\ p_W(\{x, z\}//\mathfrak{X}) &= 0; \\ p_W(\{y, z\}//\mathfrak{X}) &= 0; \\ p_W(\{x, y, z\}//\mathfrak{X}) &= \frac{1}{2} \cdot p(\{z\}//\{z\}) = \frac{3}{16}.\end{aligned}$$

По формуле (17) получаем<sup>6</sup>:  $\mathcal{H}_{\mathfrak{X}|W} = 0.114$ , а делая расчет по формуле (18) имеем:  $\mathcal{H}_{W|\mathfrak{X}} = 0.112$ .

## Список литературы

- [1] О. Ю. Воробьев. *Эвентология*. Красноярск: СФУ, 2007, 435с.
- [2] Н. А. Лукьянова. Эвентологические подраспределения: энтропийно-информационный взгляд. *Труды X Международной ФАМЭБ'2011 конференции/ под ред. О.Ю. Воробьева*, Красноярск: НИИППБ, СФУ:206–216, 2011.
- [3] Н. А. Лукьянова. Мера неопределенности и меры расхождения эвентологических распределений множества случайных событий. *Труды X Международной ФАМЭБ'2011 конференции/ под ред. О.Ю. Воробьева*, Красноярск: НИИППБ, СФУ:217–223, 2011.
- [4] N. A. Lukyanova. Eventological subdistributions: matrix representation of the events-terraces for a set of events. *Proc. of the XV Int. EM Conf.*, Krasnoyarsk, Russia: SFU:20–26, 2011.
- [5] O. Yu. Vorobyev, N. A. Lukyanova. Properties of the entropy of multiplicative-truncated approximations of eventological distributions. *In Journal of Siberian Federal University. Mathematics & Physics*, 4(1), Krasnoyarsk, Russia:50–60, 2011.

<sup>6</sup>по определению  $0 \cdot \ln \frac{0}{0} = 0$ .

## Нахождение ассоциативных правил в статистике покупок продуктов

**Кристина Игоревна Максимова**

Сибирский Федеральный Университет  
Институт математики и фундаментальной информатики  
Красноярск  
Sabaka\_2007@mail.ru

**Ирина Владимировна Баранова**

Сибирский Федеральный Университет  
Институт математики и фундаментальной информатики  
Красноярск  
irinabar@yandex.ru

**Аннотация.** В работе рассматривается понятие интеллектуального анализа данных, его назначение и области его применения. Приводится подробное описание двух наиболее известных алгоритмов поиска ассоциативных правил: *Apriori* и *AprioriSome*. Решается практическая задача поиска ассоциативных правил в статистике покупок продуктов.

**Ключевые слова.** Интеллектуальный анализ данных, ассоциативные правила, рыночная корзина, закономерность, алгоритм *Apriori*, алгоритм *AprioriSome*

### 1 Введение

Интеллектуальный анализ данных (Data Mining) — мультидисциплинарная область, возникшая и развивающаяся на базе прикладной статистики, искусственного интеллекта, теории баз данных и других. Оригинальное англоязычное название Data Mining было предложено Григорием Пиатецким-Шапиро в 1989 [1]. Название происходит из двух понятий: поиска ценной информации в большой базе данных (Data) и добычи горной руды (Mining). Термин переводится как «добыча» или «раскопка» данных.

**Определение 65.** Интеллектуальный анализ данных — это процесс обнаружения в данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности.

Таким образом, Data Mining представляет собой технологию, предназначенную для поиска в больших объемах данных неочевидных, объективных и полученных на практике закономерностей.

К методам интеллектуального анализа данных относятся всевозможные методы классификации, моделирования и прогнозирования, основанные на применении деревьев решений, искусственных нейронных сетей, генетических алгоритмов, эволюционного программирования, ассоциативной памяти, нечеткой логики. Также методами Data Mining могут считаться статистические методы: дескриптивный анализ,

корреляционный и регрессионный анализ, факторный анализ, дисперсионный анализ, компонентный анализ, дискриминантный анализ, анализ временных рядов.

Один из самых востребованных методов интеллектуального анализа данных — метод поиска ассоциативных правил. Данный метод предназначен для выявления взаимосвязей между наборами данных из статистики. Поиск закономерностей осуществляется не на основе свойств анализируемого объекта, а между несколькими событиями, которые происходят одновременно. Наиболее известным алгоритмом решения задачи поиска ассоциативных правил является алгоритм *Apriori*.

Впервые задача поиска ассоциативных правил (association rule mining) была предложена для нахождения типичных шаблонов покупок, совершаемых в супермаркетах, поэтому иногда ее еще называют анализом рыночной корзины (market basket analysis). Часто покупатели приобретают не один товар, а некоторый набор товаров. В большинстве случаев между этими товарами существует взаимосвязь. С помощью поиска ассоциативных правил (association rule) мы найдем закономерности между связанными событиями в базах данных.

В работе приводится подробный алгоритм поиска ассоциативных правил. Решается практический пример нахождения ассоциативных правил на основе статистики покупок с помощью классического метода ассоциативных правил.

### 2 Основные задачи и применение интеллектуального анализа данных

#### 1. Классификация (Classification)

Краткое описание. Классификация представляет собой наиболее простую и распространенную задачу интеллектуального анализа данных. В результате решения задачи классификации обнаруживаются признаки, которые характеризуют группы объектов исследуемого набора данных — классы. По этим признакам новый объект можно отнести к тому или иному классу. Методы решения задачи. Для решения задачи классификации могут использоваться следу-

ющие методы: ближайшего соседа (Nearest Neighbor), k-ближайшего соседа (k-Nearest Neighbor), байесовские сети (Bayesian Networks), индукция деревьев решений, нейронные сети (neural networks).

## 2. Кластеризация (Clustering)

Краткое описание. Кластеризация является логическим продолжением идеи классификации. Это задача более сложная, чем предыдущая. Особенность кластеризации заключается в том, что классы объектов изначально не предопределены. Результатом кластеризации является разбиение объектов на группы.

Пример метода решения задачи кластеризации: обучение «без учителя» особого вида нейронных сетей — самоорганизующихся карт Кохонена.

## 3. Ассоциация (Associations)

Краткое описание. В ходе решения задачи поиска ассоциативных правил отыскиваются закономерности между связанными событиями в наборе данных. Отличие ассоциации от двух предыдущих задач заключается в том, что поиск закономерностей осуществляется не на основе свойств анализируемого объекта, а между несколькими событиями, которые происходят одновременно. Наиболее известным алгоритмом решения задачи поиска ассоциативных правил является алгоритм Apriori.

## 4. Прогнозирование (Forecasting)

Краткое описание. В результате решения задачи прогнозирования на основе особенностей исторических данных (предыстории) оцениваются пропущенные или же будущие значения целевых численных показателей.

Для решения таких задач широко применяются методы математической статистики, нейронные сети и др.

## 5. Оценивание (Estimation)

Задача оценивания сводится к предсказанию непрерывных значений признака.

## 6. Визуализация (Visualization, Graph Mining)

В результате визуализации создается графический образ анализируемых данных.

Для решения задачи визуализации используются графические методы, показывающие наличие закономерностей в данных. Примеры методов визуализации — представление данных в 2-D и 3-D измерениях.

## 7. Подведение итогов (Summarization)

Цель данной задачи — описание конкретных групп объектов из анализируемого набора данных.

Сфера применения Data Mining ничем не ограничена, он применим в любых сферах человеческой деятельности, в которых существуют статистические данные:

1. Розничная торговля
2. Банковское дело
3. Телекоммуникации
4. Страхование
5. Медицина
6. Молекулярная генетика и геновая инженерия
7. Прикладная химия

# 3 Области применения интеллектуального анализа данных

## Розничная торговля

Предприятия розничной торговли сегодня собирают подробную информацию о каждой отдельной покупке, используя кредитные карточки с маркой магазина и компьютеризованные системы контроля. Вот типичные задачи, которые можно решать с помощью интеллектуального анализа данных в сфере розничной торговли: анализ покупательской корзины (анализ сходства), исследование временных шаблонов, создание прогнозирующих моделей.

## Банковское дело

Достижения технологии интеллектуального анализа данных используются в банковском деле для решения следующих распространенных задач: выявление мошенничества с кредитными карточками, сегментация клиентов, прогнозирование изменений клиентуры.

## Телекоммуникации

В области телекоммуникаций характерен растущий уровень конкуренции. Здесь методы интеллектуального анализа данных помогают компаниям более энергично продвигать свои программы маркетинга и ценообразования, чтобы удержать существующих клиентов и привлечь новых. В число типичных мероприятий входят следующие: анализ записей о подробных характеристиках вызовов, выявление лояльности клиентов.

## Страхование

Страховые компании в течение многих лет накапливают большие объемы данных. Здесь большое поле деятельности для методов интеллектуального анализа данных: выявление мошенничества, разработка продуктов, анализ риска.

## Медицина

Известно много экспертных систем для постановки медицинских диагнозов. Они построены главным образом на основе правил, описывающих сочетания

различных симптомов различных заболеваний. С помощью таких правил узнают не только, чем болен пациент, но и как нужно его лечить. Правила помогают выбирать средства медикаментозного воздействия, определять показания — противопоказания, ориентироваться в лечебных процедурах, создавать условия наиболее эффективного лечения, предсказывать исходы назначенного курса лечения и т.п.

#### Молекулярная генетика и генная инженерия

Пожалуй, наиболее остро и вместе с тем четко задача обнаружения закономерностей в экспериментальных данных стоит в молекулярной генетике и геномной инженерии. Здесь она формулируется как определение так называемых маркеров, под которыми понимают генетические коды, контролирующие те или иные фенотипические признаки живого организма. Такие коды могут содержать сотни, тысячи и более связанных элементов. На развитие генетических исследований выделяются большие средства. В последнее время в данной области возник особый интерес к применению методов интеллектуального анализа данных. Известно несколько крупных фирм, специализирующихся на применении этих методов для расшифровки генома человека и растений.

#### Прикладная химия

Методы интеллектуального анализа данных находят широкое применение в прикладной химии (органической и неорганической). Особенно актуальна такая задача при анализе сложных химических соединений, описание которых включает сотни и тысячи структурных элементов и их связей.

## 4 Основы ассоциативных правил в интеллектуальном анализе данных

Как было сказано выше, в ходе решения задачи поиска ассоциативных правил отыскиваются закономерности между связанными событиями в наборе данных. В данной работе рассматриваются два наиболее известных алгоритма поиска ассоциативных правил в интеллектуальном анализе данных: Apriori и AprioriSome. Приведем основные понятия, связанные с данными методами.

Приведем основные понятия, связанные с данным методом.

**Определение 66.** *Транзакция* — это множество событий, которые произошли одновременно.

**Определение 67.** *Транзакционная (операционная) база данных* представляет собой двумерную таблицу, состоящую из номера транзакции (TID) и перечня событий, происходящих во время этой транзакции.

**Определение 68.** *Поддержка* — количество или процент транзакций, содержащих определенный набор данных.

## 5 Алгоритм Apriori

Рассмотрим классический ассоциативный метод Apriori [2].

Выявление часто встречающихся наборов элементов — достаточно сложная операция, требующая значительных вычислений. Простейшим способом решения данной задачи является простой перебор всех возможных наборов элементов. Чтобы сократить количество возможных кандидатов в методе Apriori используется так называемое **свойство антимонотонности**, которое позволяет понизить размерность пространства поиска. Данное правило формулируется следующим образом: «Поддержка любого набора элементов не может превышать минимальной поддержки любого из его подмножеств».

Приведем обозначения, используемые в алгоритме Apriori:

$D$  — множество транзакций,

$$D = \{T_1, T_2, \dots, T_m\},$$

где  $T_i$  — транзакции из транзакционной (операционной) базы данных.

$L_k$  — множество  $k$ -элементных наборов, чья поддержка не меньше заданной пользователем.

$C_k$  — множество потенциально частых  $k$ -элементных наборов.

Алгоритм поиска ассоциативных правил имеет следующий вид:

Шаг 1. Присвоить  $k = 1$  и выполнить отбор всех одноэлементных наборов, у которых поддержка больше минимально заданной пользователем  $\text{Suppmin}$ .

Шаг 2.  $k = k + 1$ .

Шаг 3. Если не удастся создавать  $k$ -элементные наборы, то завершить алгоритм, иначе выполнить следующий шаг.

Шаг 4. Создать множество  $k$ -элементных наборов кандидатов из частых наборов. Для этого необходимо объединить в  $k$ -элементные кандидаты  $(k-1)$ -элементные частые наборы. Каждый кандидат будет формироваться путем добавления к  $(k-1)$ -элементному частому набору —  $p$  элемента из другого  $(k-1)$ -элементного частого набора —  $q$ . Причем добавляется последний элемент набора  $q$ , который по порядку выше, чем последний элемент набора  $p$ .

При этом все  $k-2$  элемента обоих наборов одинаковы.

Шаг 5. Для каждой транзакции  $T$  из множества  $D$  выбрать кандидатов  $C_t$  из множества  $C_k$ , присутствующих в транзакции  $T$ . Для каждого набора из построенного множества  $C_k$  удалить набор, если хотя бы одно из его  $(k-1)$  подмножеств не является часто встречающимся т.е. отсутствует во множестве  $L_{k-1}$ .

Шаг 6. Для каждого кандидата из  $C_k$  увеличить значение поддержки на единицу.

Шаг 7. Выбрать только кандидатов  $L_k$  из множества  $C_k$ , у которых значение поддержки больше заданной



пользователем Suprmin. Вернуться к шагу 2. Результатом работы алгоритма является объединение всех множеств  $L_k$  для всех  $k$ .

Поиск закономерностей в классическом методе ассоциативных правил осуществляется между несколькими событиями, которые происходят одновременно. Результатом поиска ассоциативных правил является пара следующего вида: Из события А следует событие В. Таким образом, все полученные методом правила являются парными.

## 6 Алгоритм AprioriSome

Другим наиболее известным алгоритмом поиска ассоциативных правил является алгоритм AprioriSome.

Алгоритм Apriori формирует частые последовательности-кандидаты всех возможных длин. Однако, если из числа последовательностей определенной длины формируется мало частых последовательностей, то эту длину можно пропустить. Алгоритм AprioriSome использует как параметр длину последовательностей, анализируемых на предыдущем проходе, и возвращает длину последовательностей, которые будут анализироваться на следующем этапе. Иными словами, длина последовательностей, искомых на следующем проходе, определяется длиной последовательностей, найденных на предыдущем шаге.

Если обозначить номер прохода переменной  $t$ , то можно записать, что  $k_{t+1} = k_t + p$ . Это означает, что на следующем проходе будут анализироваться последовательности с длиной на  $p$  больше, чем на предыдущем. В случае если  $p = 1$ , то алгоритм идентичен Apriori, т.е. будут анализироваться все последовательности.

Задача заключается в том, чтобы определить, какие длины последовательностей могут быть пропущены.

Обозначим отношение числа частых  $k$ -последовательностей к числу всех  $k$ -последовательностей-кандидатов переменной  $h_k$ ,

$$h_k = \frac{F_k}{C_k}.$$

Для выбора длины последовательности воспользуемся алгоритмом, приведенным на рис. 1.

## 7 Практический пример нахождения ассоциативных правил

Решим практическую задачу нахождения ассоциативных правил в статистике покупок продуктов. Одним из продуктовых розничных магазинов города Красноярска была предоставлена статистика покупок продуктов за октябрь 2012 года и февраль 2013

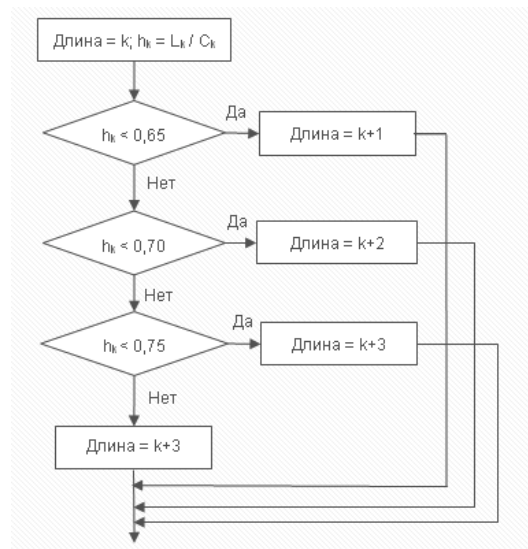


Рис. 1: Алгоритм выбора длины последовательности в AprioriSome

года. Всего статистика содержала около 1000 транзакций (чеков покупок). В статистике встречалось 50 различных наименований товаров: молоко, хлеб, сметана, творог, сыр, яйца, колбаса, сахар, различные виды мяса, полуфабрикаты и т.д. На основе статистики было необходимо найти закономерности между покупками в наборе данных.

Приведенные выше товары были обозначены переменными, в Табл. 1 приведены 14 самых часто встречающихся товаров, их обозначение и вычисленное значение поддержки.

Был определен уровень поддержки для данной задачи Suprmin=25. На первом этапе происходило формирование одноэлементных кандидатов.

Согласно алгоритму были подсчитаны поддержки одноэлементных наборов. Наборы с уровнем поддержки меньше установленного отсекались. В нашем примере это были все товары, которые имели поддержку менее 25. Оставшиеся наборы товаров считаются часто встречающимися одноэлементными наборами товаров: это наборы  $a, b, c, d, e, \dots, n$ .

На следующем этапе выполнялось формирование двухэлементных кандидатов, подсчет их поддержки и отсекались наборы с уровнем поддержки, меньшим 25. В результате данного процесса остались шесть двухэлементных наборов, принимающих участие в дальнейшей работе алгоритма. Это пары  $ac, ag, ak, bg, cg, gk$ .

Далее формировались трехэлементные наборы товаров:  $abf, adb, adf, bdf$ , подсчитывалась их поддержка и убирались наборы с уровнем поддержки, меньшим 25. В результате были получены два набора с поддержкой более 25:  $acg$  и  $agk$  (с поддержкой 25 и 26, соответственно).

Товар	Обозначение	Поддержка
Молоко	a	100
Сметана	b	31
Сыр твердый	c	44
Мясо куриное	d	26
Яйца	e	26
Полуфабрикаты	f	31
Хлеб	g	125
Конфеты	h	33
Газ.вода	i	27
Овощи	j	34
Фрукты	k	52
Консервы	l	28
Соусы	m	31
Крупа	n	32

Таблица 1: Таблица продуктов и их поддержки

Формирование четырехэлементных наборов не удалось, поэтому работа алгоритма была прекращена. Итак, в результате работы алгоритма Apriori было получено два ассоциативных правила: «Покупатель часто приобретает молоко, сыр и хлеб» и «Покупатель часто приобретает молоко, хлеб и фрукты».

Уровень поддержки первого правила равен 25, а второго — 26.

Теперь рассмотрим работу алгоритма AprioriSome на той же самой статистике покупок товаров.

При первом проходе алгоритма по базе данных было найдено множество частых последовательностей с единичной длиной. На втором шаге было определено множество кандидатов  $C_2$  для того, чтобы получить множество частых последовательностей  $F_2$ . На следующем шаге из  $F_2$  было получено  $C_3$ . На четвертом шаге сформированное множество  $C_4$  оказалось пустым.

После чего первая фаза была закончена, и была запущена обратная фаза алгоритма. Из  $F_3$  ничего не было исключено, поскольку более длинные последовательности отсутствуют в БД. В результате работы алгоритма были получены те же ассоциативные правила, что и при работе алгоритма Apriori.

## 8 Сравнение алгоритмов Apriori и AprioriSome

Основное преимущество алгоритма AprioriSome над Apriori заключается в том, что он позволяет избежать вычисления большого числа немаксимальных последовательностей.

Однако в случае нашей статистики данное преимущество сокращается. Причина этого заключается в том, что множество кандидатов  $C_k$  в алгоритме Apriori формируется с использованием множества  $L_{k-1}$ , а в алгоритме AprioriSome — с использованием множества  $C_{k-1}$ . Поскольку  $C_{k-1}$  является подмножеством  $L_{k-1}$ , число кандидатов, формируемых алгоритмом AprioriSome, может быть больше.

## 9 Итоги

В работе рассмотрены основные задачи и методы интеллектуального анализа данных. Особое внимание уделено методу ассоциативных правил в интеллектуальном анализе данных. Подробно изложены два наиболее часто используемых алгоритма поиска ассоциативных правил: алгоритмы Apriori и AprioriSome.

Также в работе решен практический пример нахождения ассоциативных правил на основе статистики покупок товаров с помощью обоих алгоритмов Apriori и AprioriSome.

## Список литературы

- [1] А.Ю. Афонин, П.П. Макарычев. *Оперативный и интеллектуальный анализ данных*. ПГУ, Пенза, 2010.
- [2] А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, И.И. Холод. *Технологии анализа данных. Data Mining, Visual Mining, Text Mining, OLAP*. БХВ-Петербург, СПб., 2004.

# Актuarные расчеты по добровольному медицинскому страхованию

**Алексей Николаевич Медницкий**

Самарский Государственный Университет  
Самара  
medik42@yandex.ru

**Алла Юрьевна Трусова**

Самарский Государственный Университет  
Самара  
alla.yurievna@gmail.com

**Аннотация.** В работе рассмотрены актуарные расчеты по добровольному медицинскому страхованию. Для расчета основной части нетто-ставки и рисковой надбавки применена формула Шуэтта-Несбитта. Показано три метода вычисления поправочных коэффициентов. Также составлены расчеты по вычислению франшизы для нескольких законов распределения. В среде программирования Borland Delphi написана программа, автоматизирующая актуарные расчеты. Все методы, описанные в теоретической части рассмотрены на реальных данных.

**Ключевые слова.** Страхование, актуарные расчеты, добровольное медицинское страхование, риск, вероятность, нетто-ставка, брутто-ставка, тарифы, поправочный коэффициент, убыточность, рисковая надбавка, франшиза.

## 1 Введение

Добровольное медицинское страхование - один из самых популярных видов страхования в настоящее время. Широкое его распространение обусловлено основным преимуществом перед другими видами страхования, а именно, активностью клиента в формировании страховой программы. Спектр рисков достаточно широк: риск смерти, инвалидности, опасных или инфекционных заболеваний, временной нетрудоспособности. Психологические аспекты в случае наступления рисков доминируют при выборе клиентом вида страхования.

В этой связи целью данной работы является изучение ключевых параметров добровольного медицинского страхования. Среди них важное место занимает вероятность наступления страхового случая, размер выплат при возникновении рисковой ситуации, брутто и нетто ставки, поправочные коэффициенты и величина страховой премии.

## 2 Основная часть

В настоящее время существуют различные способы расчета тарифных ставок по добровольному ме-

дицинскому страхованию. В работе использовались данные по самарской области за 2008-2012 год. Они содержат информацию по каждому договору (пол, возраст, сумма выплат по каждому виду страхования).

## 3 Расчет нетто-ставки по формуле Шуэтта-Несбитта

Для нахождения основной части нетто-ставки, по формуле Шуэтта-Несбитта, представленной в работе [1], вычисляется суммарный объем выплат:

$$H_0 = S = \sum_{j=1}^n S_{[j]}, \quad (1)$$

где  $S$  - суммарный объем выплат,  $n$  - количество классов болезней.

$$S_{[j]} = P_{[j]} = \sum_{i_1 \dots i_j} \frac{\prod_{k=1}^j q_{i_k}}{D_j} \left( \sum_{k=1}^j Y_{j_k} \right), \quad (2)$$

$$i_1 \neq i_2 \dots \neq i_j$$

где  $q(i_k)$  - вероятность наступления  $k$ -ой болезни,  $Y(j_k)$  - выплаты по этой болезни.

$$P_{[k]} = \sum_{j=1}^n (-1)^{j-k} \cdot C_k^j \cdot D_j, \quad (3)$$

$$D_j = \sum_{i_1 \dots i_j} \prod_{k=1}^j q_{i_k}. \quad (4)$$

$$i_1 \neq i_2 \dots \neq i_j$$

Дисперсия суммарного объема выплат, позволяющая оценить рисковую надбавку, находится по формуле (6):

$$\sigma_{[j]}^2 = P_{[j]} \sum_{i_1 \dots i_j} \frac{\prod_{k=1}^j q_{i_k}}{D_j} \cdot \left( \sum_{k=1}^j Y_{j_k} \right)^2 - S_{[j]}^2, \quad (5)$$

$$\sigma^2 = \sum_{i=1}^n \sigma_{[j]}^2. \quad (6)$$

Однако осуществить данные вычисления становится проблематично при большом количестве классов болезней. Для данных вычислений была составлена программа в среде Borland Delphi, которая позволяет автоматически рассчитать и ожидаемое значение выплат и дисперсию.

#### 4 Расчет франшизы

В работе представлены методы по вычислению безусловной франшизы. При страховании с учетом франшизы меняются вероятность наступления события и математическое ожидание убытка, поэтому для трех видов распределения: бета, гамма и экспоненциальное, были пересчитаны данные показатели. Расчет новой вероятности осуществлялся по формуле (7).

$$\hat{q} = q(1 - f(\omega)), \quad (7)$$

где  $\hat{q}$  – новая вероятность;  $q$  – исходная вероятность,  $f(\omega)$  – плотность вероятности выбранного закона распределения.

Новое математическое ожидание убытка рассчитывалось по формуле (8).

$$E(\tilde{x}) = \int_{\omega} \tilde{x} \cdot f(\tilde{x}) d\tilde{x}. \quad (8)$$

В итоге для бета-распределения получена вероятность, вычисляемая по формуле (9) и математическое ожидание, определяемое по формуле (10).

$$\hat{q} = q(1 - f(\omega, \alpha, \beta)), \quad (9)$$

$$E(\tilde{x}) = E(x)(1 - f(\omega, \alpha + 1, \beta)). \quad (10)$$

Для гамма-распределения получена вероятность, вычисляемая по формуле (11) и математическое ожидание, определяемое по формуле (12).

$$\hat{q} = q(1 - f(\omega, k, \theta)), \quad (11)$$

$$E(\tilde{x}) = E(x)(1 - f(\omega, k + 1, \theta)). \quad (12)$$

Для экспоненциального распределения получена вероятность, вычисляемая по формуле (13) и математическое ожидание, определяемое по формуле (14).

$$\hat{q} = q(1 - f(x, \lambda)), \quad (13)$$

$$E(\tilde{x}) = \exp(-\lambda \cdot \omega) \cdot \left( \omega + \frac{1}{\lambda} \right). \quad (14)$$

Из представленных видов распределения бета и гамма наиболее точно описали реальный процесс, поэтому для расчета франшизы можно опираться на любой из этих законов. Расчет параметров распределения можно проводить методом максимального правдоподобия, методом моментов или методом минимизации критерия Пирсона.

#### 5 Расчет поправочных коэффициентов

В работе представлено несколько видов расчета поправочных коэффициентов, из которых два были рассмотрены на реальных данных. Первый метод – составление эконометрической модели.

$$\hat{k} = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_n \cdot x_n, \quad (15)$$

где  $X$  – матрица факторов.

$$b = \begin{pmatrix} b_0 \\ b_1 \\ \dots \\ b_n \end{pmatrix} = (x^T x)^{-1} \cdot x^T k. \quad (16)$$

На основе выбранных данных была составлена следующая модель:

$$\hat{k} = 0.9 - 0.29x_1 + 0.0063x_2, \quad (17)$$

где  $\hat{k}$  – величина поправочного коэффициента,  $x_1$  – пол,  $x_2$  – возраст.

Второй метод, применяющийся в большинстве страховых компаний – статистический метод, основанный на вычислении процента больных в зависимости от различных факторов. После анализа исходной информации были получены следующие поправочные коэффициенты для мужчин:

Возраст	Коэффициент
0-20	0,5
21-35	0,62
Старше 35	1,05

Рис. 1: Зависимость коэффициента от возраста (м)

Для женщин:

Возраст	Коэффициент
0-22	0,8
23-30	1,95
31-40	1,3
41-55	1,4
Старше 55	1,05

Рис. 2: Зависимость коэффициента от возраста (ж)

Из рассмотренных методов, эконометрическая модель наиболее точно аппроксимирует реальный результат.

## 6 Расчет страховой премии

После расчета поправочных коэффициентов нетто-ставка перерасчитывается по формуле (18).

$$H = H \cdot \prod_{t=1}^i k_t, \quad (18)$$

где  $k_t$  – значение коэффициентов по отдельным факторам.

Брутто-ставка (сумма, которая взимается с клиента) вычисляется по формуле (19).

$$Brutto = \frac{H}{1 - f}, \quad (19)$$

где  $Brutto$  – брутто-ставка,  $H$  – нетто-ставка,  $f$  – нагрузка.

## 7 Заключение

В работе представлена методика расчета ключевых параметров, используемых в добровольном медицинском страховании.

## Список литературы

- [1] Н. Бауэрс, Х. Гербер, Д. Джанс, С. Несбитт, Д. Хикман. *Актuariальная математика. Перевод с английского под редакцией В. К. Малиновского*. Издательство “Янус-К”, Москва, 2001.
- [2] Н. Б. Грищенко, В. А. Клевно, В. В. Мищенко. *Добровольное медицинское страхование. Основы современной практики*. Издательство Алтайского государственного университета, Барнаул, 2001.
- [3] Т. Мак. *Математика рискованного страхования*. Издательство “Олимп-бизнес”, Москва, 2005.



# Расчет франшизы для бета, гамма и экспоненциального распределения

**Алексей Николаевич Медницкий**

Самарский Государственный Университет  
Самара  
medik42@yandex.ru

**Алла Юрьевна Трусова**

Самарский Государственный Университет  
Самара  
alla.yurievna@gmail.com

**Аннотация.** В работе рассмотрены принципы вычисления франшизы для бета, гамма, экспоненциального распределения. Были выведены вероятность и математическое ожидание для каждого из распределений и рассчитаны эти же показатели для условной франшизы. На основе реальных данных были выбраны распределения, наиболее точно аппроксимирующие реальный процесс поступления убытков.

**Ключевые слова.** Страхование, актуарные расчеты, франшиза, условная франшиза, безусловная франшиза, вероятность, математическое ожидание, сумма убытка, бета-распределение, гамма-распределение, экспоненциальное распределение, аппроксимация.

## 1 Введение

Франшиза - предусмотренное условиями страхования освобождение страховщика от возмещения убытков, не превышающих заданную величину. Франшиза бывает двух типов - условная и безусловная. Условная франшиза подразумевает выплату в полном размере, в случае, если убыток превысил размер франшизы. При безусловной франшизе убыток возмещается не полностью: за вычетом величины франшизы. Математически условную франшизу можно выразить формулой (1), а безусловную формулой (2).

$$Y_{\omega} = \begin{cases} 0, & \text{if } Y < \omega, \\ Y, & \text{if } Y \geq \omega, \end{cases} \quad (1)$$

$$Y_{\omega} = \begin{cases} 0, & \text{if } Y < \omega, \\ Y - \omega, & \text{if } Y \geq \omega, \end{cases} \quad (2)$$

где  $Y_{\omega}$  - размер выплаты при наличии франшизы;  $\omega$  - размер франшизы;  $Y$  - фактический убыток.

Иными словами, франшиза - способ деления риска между страховщиком и страхователем.

При заключении договора страхования часто принимают условную франшизу. Во-первых, это создает

программу ДМС более гибкой: клиент может сам выбрать, в каком соотношении будет поделен риск между ним и страховой компанией. Во-вторых, наличие франшизы избавит страховую компанию от незначительных убытков, при которых расходы на ведение дела будут на порядок больше чем сумма выплаты. В-третьих, франшиза позволяет уменьшить страховую премию, в случае если клиент не готов будет платить установленную сумму.

## 2 Основная часть

При наличии франшизы изменяется вероятность наступления страхового события и средняя выплата. Одним из способов нахождения величины франшизы - метод поправочных коэффициентов, однако в случае использования кластерного анализа для расчета будет получена дискретная модель, неудобная в использовании, а в случае составления эконометрической модели может пострадать ее адекватность, так как величина франшизы не принадлежит нормальному закону распределения. Поэтому следует сделать предположение, что франшиза имеет другое распределение, например бета, гамма, экспоненциальное. Тогда наша задача сведется к нахождению параметров выбранного распределения. Оценку параметров можно проводить разными методами: метод максимального правдоподобия, метод моментов, метод минимизации критерия Пирсона.

## 3 Расчет франшизы по бета-распределению

Бета-распределение наиболее часто используется при определении размера франшизы. Его плотность описывается формулой (3).

$$f(x, \alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}, \quad (3)$$

где  $B(\alpha, \beta)$  - бета-функция;  $\alpha, \beta$  - параметры распределения. Найдем математическое ожидание бета-

распределения по формуле (4).

$$E(x) = \int_{-\infty}^{\infty} xf(x, \alpha, \beta)dx = \int_{-\infty}^0 xf(x, \alpha, \beta)dx + \int_0^1 xf(x, \alpha, \beta)dx + \int_1^{\infty} xf(x, \alpha, \beta)dx. \quad (4)$$

Так как при бета-распределении  $x \in [0; 1]$ , то интегралы на остальных областях будут нулевыми и вычисления сведутся к формуле (5).

$$E(x) = \int_0^1 xf(x, \alpha, \beta)dx = \int_0^1 \frac{x^\alpha(1-x)^{\beta-1}}{B(\alpha, \beta)}dx. \quad (5)$$

Заменяем бета-функцию на гамма-функцию, используя формулу (6).

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}. \quad (6)$$

Подставим гамма-функцию в формулу (5) и получим формулу (7).

$$E(x) = \int_0^1 \frac{x^\alpha(1-x)^{\beta-1}\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}dx = \frac{\Gamma(\alpha + \beta)\Gamma(\alpha + 1)}{\Gamma(\alpha + \beta + 1)\Gamma(\alpha)} \times \quad (7)$$

$$\times \int_0^1 \frac{x^\alpha(1-x)^{\beta-1}\Gamma(\alpha + \beta + 1)}{\Gamma(\alpha + 1)\Gamma(\beta)}dx.$$

Подынтегральное выражение можно представить в виде формулы (8).

$$\int_0^1 \frac{x^\alpha(1-x)^{\beta-1}\Gamma(\alpha + \beta + 1)}{\Gamma(\alpha + 1)\Gamma(\beta)}dx = \quad (8)$$

$$= f(x, \alpha + 1, \beta) = 1.$$

Воспользуемся свойством гамма-функции по формуле (9).

$$\Gamma(\alpha + 1) = \alpha\Gamma(\alpha). \quad (9)$$

Тогда математическое ожидание будет выглядеть в виде формулы (10).

$$E(x) = \frac{\Gamma(\alpha + \beta)\Gamma(\alpha + 1)}{\Gamma(\alpha + \beta + 1)\Gamma(\alpha)} = \frac{\alpha}{\alpha + \beta}. \quad (10)$$

Так как при франшизе изменяется вероятность наступления страхового события и средний ущерб, пересчитаем эти показатели по формуле (11) и формуле (12).

$$\hat{q} = q(1 - f(\omega, \alpha, \beta)), \quad (11)$$

где  $\hat{q}$  – новая вероятность;  $\omega$  – размер франшизы. Новая убыточность будет вычисляться по формуле (12).

$$E(\tilde{x}) = \int_{\omega}^1 \tilde{x}f(\tilde{x}, \alpha, \beta)d\tilde{x} =$$

$$= \int_0^1 \tilde{x}f(\tilde{x}, \alpha, \beta)d\tilde{x} - \int_0^{\omega} \tilde{x}f(\tilde{x}, \alpha, \beta)d\tilde{x} = \quad (12)$$

$$= E(x) - \int_0^{\omega} \frac{\tilde{x}^\alpha(1-\tilde{x})^{\beta-1}}{B(\alpha, \beta)}d\tilde{x} =$$

$$= E(x) - E(x)f(\omega, \alpha + 1, \beta) =$$

$$= E(x)(1 - f(\omega, \alpha + 1, \beta)).$$

#### 4 Расчет франшизы по гамма-распределению

Плотность гамма-распределения задается формулой (13).

$$f(x, k, \theta) = \frac{x^{k-1}exp(\frac{-x}{\theta})}{\Gamma(k)\theta^k}, \quad (13)$$

где  $(k)$  – гамма-функция;  $k, \theta$  – параметры распределения.

Найдем математическое ожидание гамма распределения по формуле (14).

$$E(x) = \int_{-\infty}^{\infty} xf(x, k, \theta)dx = \int_{-\infty}^0 xf(x, k, \theta)dx + \quad (14)$$

$$+ \int_0^{\infty} xf(x, k, \theta)dx.$$

Так как при гамма-распределении  $x \in [0; \infty)$ , то интегралы на остальных областях будут нулевыми и вы-

числения сведутся к формуле (15).

$$\begin{aligned}
 E(x) &= \int_0^{\infty} x f(x, k, \theta) dx = \\
 &= \frac{1}{\Gamma(k)\theta^k} \int_0^{\infty} x^k \exp\left(\frac{-x}{\theta}\right) dx = \\
 &= \frac{\Gamma(k+1)\theta^{k+1}}{\Gamma(k)\theta^k} \int_0^{\infty} \frac{x^k}{\Gamma(k+1)\theta^{k+1}} \exp\left(\frac{-x}{\theta}\right) dx = \\
 &= k\theta \int_0^{\infty} f(x, k+1, \theta) dx = k\theta.
 \end{aligned} \tag{15}$$

Так как при франшизе изменяется вероятность наступления страхового события и средний ущерб, пересчитаем эти показатели по формулам (16) и (17)

$$\hat{q} = q(1 - f(\omega, k, \theta)), \tag{16}$$

где  $\hat{q}$  – новая вероятность;  $\omega$  – размер франшизы.

Новая убыточность будет вычисляться по формуле (17).

$$\begin{aligned}
 E(\tilde{x}) &= \int_{\omega}^{\infty} \tilde{x} f(\tilde{x}, k, \theta) d\tilde{x} = \\
 &= \int_0^{\infty} \tilde{x} f(\tilde{x}, k, \theta) d\tilde{x} - \int_0^{\omega} \tilde{x} f(\tilde{x}, k, \theta) d\tilde{x} = \\
 &= E(x) - \int_0^{\omega} \frac{\tilde{x}^k \exp\left(\frac{-\tilde{x}}{\theta}\right)}{\Gamma(k)\theta^k} d\tilde{x} = \\
 &= E(x) - E(x) f(\omega, k+1, \theta) = \\
 &= E(x)(1 - f(\omega, k+1, \theta)).
 \end{aligned} \tag{17}$$

## 5 Расчет франшизы по экспоненциальному распределению

Плотность экспоненциального распределения задается формулой (18).

$$f(x, \lambda) = \lambda \cdot \exp(-\lambda \cdot x). \tag{18}$$

Найдем математическое ожидание экспоненциально-

го распределения по формуле (19).

$$\begin{aligned}
 E(\tilde{x}) &= \int_{-\infty}^{\infty} \lambda \cdot \exp(-\lambda \cdot x) \cdot x dx = \\
 &= \int_{-\infty}^0 \lambda \cdot \exp(-\lambda \cdot x) \cdot x dx + \\
 &+ \int_0^{\infty} \lambda \cdot \exp(-\lambda \cdot x) \cdot x dx.
 \end{aligned} \tag{19}$$

Так как при экспоненциальном распределении  $x \in [0; \infty)$ , то интегралы на остальных областях будут нулевыми и вычисления сведутся к формуле (20).

$$E(\tilde{x}) = \int_0^{\infty} \lambda \cdot \exp(-\lambda \cdot x) \cdot x dx = \tag{20}$$

$$= -\exp(-\lambda \cdot x) \cdot \left(x + \frac{1}{\lambda}\right) \Big|_0^{\infty} = \frac{1}{\lambda}.$$

Так как при франшизе изменяется вероятность наступления страхового события и средний ущерб, пересчитаем эти показатели по формулам (21) и (22).

$$\hat{q} = q(1 - f(x, \lambda)), \tag{21}$$

где  $\hat{q}$  – новая вероятность;  $\omega$  – размер франшизы.

Новая убыточность будет вычисляться по формуле (22).

$$\begin{aligned}
 E(\tilde{x}) &= \int_{\omega}^{\infty} \tilde{x} f(\tilde{x}, \lambda) d\tilde{x} = \\
 &= \int_0^{\infty} \tilde{x} f(\tilde{x}, \lambda) d\tilde{x} - \int_0^{\omega} \tilde{x} f(\tilde{x}, \lambda) d\tilde{x} = \\
 &= E(x) - \int_0^{\omega} x \cdot \lambda \cdot \exp(-\lambda \cdot x) d\tilde{x} = \\
 &= E(x) + \exp(-\lambda \cdot x) \cdot \left(x + \frac{1}{\lambda}\right) \Big|_0^{\omega} = \\
 &= \exp(-\lambda \cdot \omega) \cdot \left(\omega + \frac{1}{\lambda}\right).
 \end{aligned} \tag{22}$$

## 6 Выбор наилучшего распределения

При выборе наилучшего распределения использовались несколько видов страхования: ДМС, ОСАГО,

КАСКО по разным регионам страны. При расчете поправочного коэффициента по франшизе была использована суммарная величина убытка. Задача состояла в прогнозировании этой величины для разных законов распределения. Параметры распределения рассчитывались путем минимизации критерия Пирсона, по формуле (23).

$$\sum_{\omega=1}^n \frac{(\hat{Y} - Y)^2}{\hat{Y}} \rightarrow \min, \quad (23)$$

где  $Y$  - реальный объем убытка;  $\hat{Y}$  - прогнозируемый объем убытка;  $\omega$  - размер франшизы.

Прогнозируемый объем убытка будет вычисляться по формуле (24).

$$\hat{Y} = q \cdot \bar{y} \cdot n, \quad (24)$$

где  $q$  - вероятность наступления события;  $\bar{y}$  - средний убыток;  $n$  - количество договоров.

На рисунке 1 представлены графики распределения суммарного убытка при разных размерах франшизы для ДМС, из которого видно, что гамма и бета распределение довольно хорошо описывают реальный процесс, при франшизе в 3000 рублей и более.

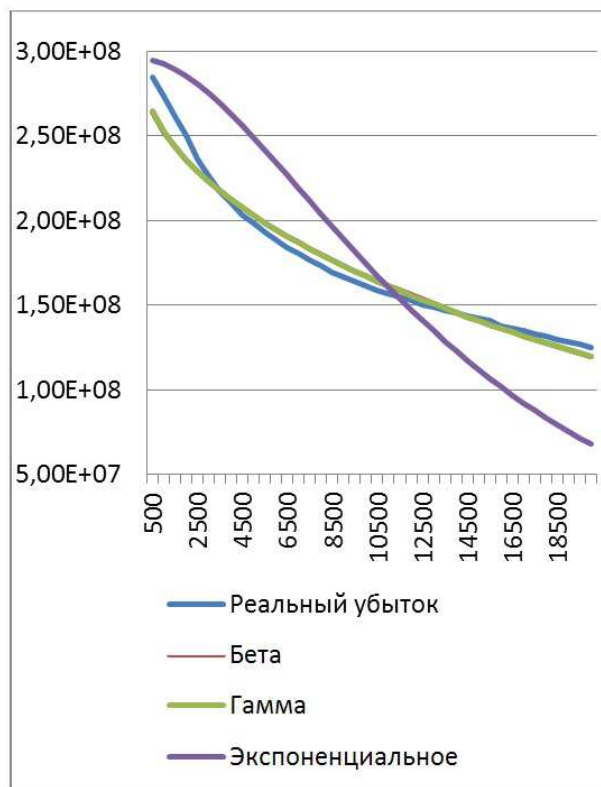


Рис. 1: Суммарный убыток

## 7 Заключение

В заключении хочется отметить, что расчет франшизы на данный момент является актуальной темой в сфере страхования. Причиной этого служат постоянно меняющиеся правила страхования, появление новых математических моделей в актуарных расчетах, социально-экономические изменения в различных сферах страхования.

## Список литературы

- [1] Н. Б. Грищенко, В. А. Клевно, В. В. Мищенко. *Добровольное медицинское страхование. Основы современной практики*. Издательство Алтайского государственного университета, Барнаул, 2001.
- [2] Н. Бауэрс, Х. Гербер, Д. Джанс, С. Несбитт, Д. Хикман. *Актуарная математика. Перевод с английского под редакцией В. К. Малиновского*. Издательство "Янус-К", Москва, 2001.
- [3] Т. Мак. *Математика рискованного страхования*. Издательство "Олимп-бизнес", Москва, 2005.

# Событийная визуализация распределения Пуассона и его средневероятного события

Нартов Яков Викторович

Сибирский федеральный университет  
Институт математики и фундаментальной информатики  
Красноярск  
yanartov@yandex.ru

**Аннотация.** Рассматривается визуализация действия средневероятных событий для схемы Бернулли и распределения Пуассона.

**Ключевые слова.** Средневероятное событие, вероятностное расстояние, схема Бернулли, распределение Пуассона.

## 1 Введение

Новый объект в эвентологии, именуемый средневероятным событием для конечного множества однородных событий, был введен в [1]. Как и любой математический объект он требует особого внимания для выявления его положительных качеств, которые смогли бы помочь нам в исследованиях и построениях математических моделей с одной стороны, и для выявления «подводных камней», которые следует иметь ввиду с другой стороны.

### 1.1 Основные определения

**Определение 1** (Средневероятное событие). Средневероятным событием для конечного множества однородных событий  $\mathfrak{X} \subset \mathcal{F}$  называется такое всеобщее событие  $\hat{x}_{\mathfrak{X}} \in \mathcal{F}$ , удовлетворяющее включению

$$\sum_{|X| > m} \text{ter}(X // \mathfrak{X}) \subseteq \hat{x}_{\mathfrak{X}} \subseteq \sum_{|X| \geq m} \text{ter}(X // \mathfrak{X}), \quad (1)$$

которое наступает с вероятностью

$$\mathbf{P}(\hat{x}_{\mathfrak{X}}) = \frac{1}{|\mathfrak{X}|} \sum_{x \in \mathfrak{X}} \mathbf{P}(x) \quad (2)$$

всякий раз, когда среди событий из  $\mathfrak{X}$  наступает не менее, чем  $m$  событий, где  $m \in \{0, 1, \dots, |\mathfrak{M}|\}$  удовлетворяет неравенствам

$$\sum_{|X| > m} p(X // \mathfrak{X}) \leq \mathbf{P}(\hat{x}_{\mathfrak{X}}) \leq \sum_{|X| \geq m} p(X // \mathfrak{X}). \quad (3)$$

В [2] было верно подмечено, что определение средневероятного события для множества однородных событий можно дать на языке террасных  $m$ -слоев и террасных  $m$ -пластов.

Итак, множественный  $m$ -слой  $C_{\mathfrak{X}}^m$  [3] определяется как совокупность подмножеств событий

$$C_{\mathfrak{X}}^m = \{X \subseteq \mathfrak{X} : |X| = m\} \subseteq 2^{\mathfrak{X}}. \quad (4)$$

Множественный  $m$ -пласт  $C_{\mathfrak{X}}^m$  [3] определяется как совокупность подмножеств событий

$$C_{\mathfrak{X}}^{m+} = \{X \subseteq \mathfrak{X} : |X| \geq m\} \subseteq 2^{\mathfrak{X}}. \quad (5)$$

Определим террасную операцию « $m$ -слой» следующим образом:

$$\text{terop}(C_{\mathfrak{X}}^m // 2^{\mathfrak{X}}) = \sum_{X \in C_{\mathfrak{X}}^m} \text{ter}(X // \mathfrak{X}) \subseteq \Omega \quad (6)$$

Аналогичным образом определим террасную операцию « $m$ -пласт»:

$$\text{terop}(C_{\mathfrak{X}}^{m+} // 2^{\mathfrak{X}}) = \sum_{X \in C_{\mathfrak{X}}^{m+}} \text{ter}(X // \mathfrak{X}) \subseteq \Omega \quad (7)$$

**Определение 2** (Средневероятное событие). Средневероятным событием для конечного однородного множества событий  $\mathfrak{X} \subset (\mathcal{F})$  называется любое такое всеобщее событие  $\hat{x}_{\mathfrak{X}} \in \mathcal{F}$ , которое содержится в террасном  $m$ -пласте, содержит террасный  $m+1$ -пласт:

$$\text{terop}(C_{\mathfrak{X}}^{(m+1)+} // 2^{\mathfrak{X}}) \subseteq \hat{x}_{\mathfrak{X}} \subseteq \text{terop}(C_{\mathfrak{X}}^{(m)+} // 2^{\mathfrak{X}}),$$

и наступает со средней вероятностью

$$\mathbf{P}(\hat{x}_{\mathfrak{X}}) = \frac{1}{|\mathfrak{M}|} \sum_{\mu \in \mathfrak{M}} \mathbf{P}(x_{\mu})$$

Всякий раз, когда среди событий из  $\mathfrak{X}$  наступает не менее, чем  $m$  событий, где  $m \in \{1, 2, \dots, |\mathfrak{X}|\}$  удовлетворяет неравенствам

$$\mathbf{P}(\text{terop}(C_{\mathfrak{X}}^{(m+1)+} // 2^{\mathfrak{X}})) \leq$$

$$\leq \mathbf{P}(\hat{x}_{\mathfrak{X}}) \leq$$

$$\leq \mathbf{P}(\text{terop}(C_{\mathfrak{X}}^{(m+1)+} // 2^{\mathfrak{X}}))$$



Если не вдаваться в подробности, то минимальное число  $m$  наступающих событий в результате наступления средневоятного события выбирается из соображения, что вероятность средневоятного события должна быть не менее вероятности террасного  $(m + 1)$ -пласта, но не более вероятности  $m$ -пласта.

Средневоятное событие обладает очень важным экстремальным свойством [1].

**Определение 3.** Вероятностное расстояние события  $\hat{x}_{\mathfrak{X}} \in \mathcal{F}$  до множества  $\mathfrak{X} \subset \mathcal{F}$  определяется формулой:

$$\begin{aligned} \rho(\hat{x}_{\mathfrak{X}}, \mathfrak{X}) &= \sum_{x \in \mathfrak{X}} \mathbf{P}(x \Delta \hat{x}_{\mathfrak{X}}) = \\ &= \sum_{x \in \mathfrak{X}} \mathbf{P}(\hat{x}_{\mathfrak{X}} \cap x^c) + \sum_{x \in \mathfrak{X}} \mathbf{P}((\hat{x}_{\mathfrak{X}})^c \cap x). \end{aligned} \quad (8)$$

**Теорема 20.** Средневоятное событие  $\hat{x}_{\mathfrak{X}}$  для множества событий  $\mathfrak{X}$  минимизирует вероятностное расстояние до  $\mathfrak{X}$ :

$$\rho(\hat{x}_{\mathfrak{X}}, \mathfrak{X}) = \min_{\alpha \in \mathcal{F}; \mathbf{P}(\alpha) = \hat{\mathbf{P}}_{\mathfrak{X}}} \rho(\alpha, \mathfrak{X}) \quad (9)$$

среди таких событий из алгебры  $\mathcal{F}$ , которые наступают с вероятностью, равной средней вероятности

$$\hat{\mathbf{P}}_{\mathfrak{X}} = \frac{1}{|\mathfrak{X}|} \sum_{x \in \mathfrak{X}} \mathbf{P}(x).$$

Совершенно очевидно, что средневоятное событие не единственно [4] и все попытки визуализации средневоятного события – в сущности попытки визуализировать семейство событий, каждое из которых в праве претендовать на средневоятное.

## 1.2 Событийная схема Бернулли

В [2] была дана событийная визуализация схемы Бернулли с параметрами  $(|\mathfrak{X}|, p)$ . Наряду с иными событийными средними (квантильные события, медианные события) была дана визуализация средневоятного события и его асимптотического поведения. Для визуализации использован следующий подход. Считаем, что пространство элементарных исходов  $\Omega$  – это  $[0, 1]^2$ . Разбиваем отрезок  $[0, 1]$  горизонтальной плоскости на  $N = |\mathfrak{X}|$  равных частей длины  $\frac{1}{N}$  точками  $\frac{m}{N}, m = 1, 2, \dots, N$ . Далее строим 2 графика: график биномиальных вероятностей:

$$b_n(|\mathfrak{X}|, p) = \mathbf{P}(\text{тероп}(C_{\mathfrak{X}}^n // \mathfrak{X}) = C_{|\mathfrak{X}|}^n * p^n * (1 - p)^{|\mathfrak{X}| - n},$$

а также график функции биномиального распределения:

$$B_n(|\mathfrak{X}|, p) = \sum_{m=0}^n b_m(|\mathfrak{X}|, p)$$

На рисунке 1 можно видеть как с ростом  $|\mathfrak{X}|$  семейство средневоятных событий сокращается, этот факт можно доказать [2].

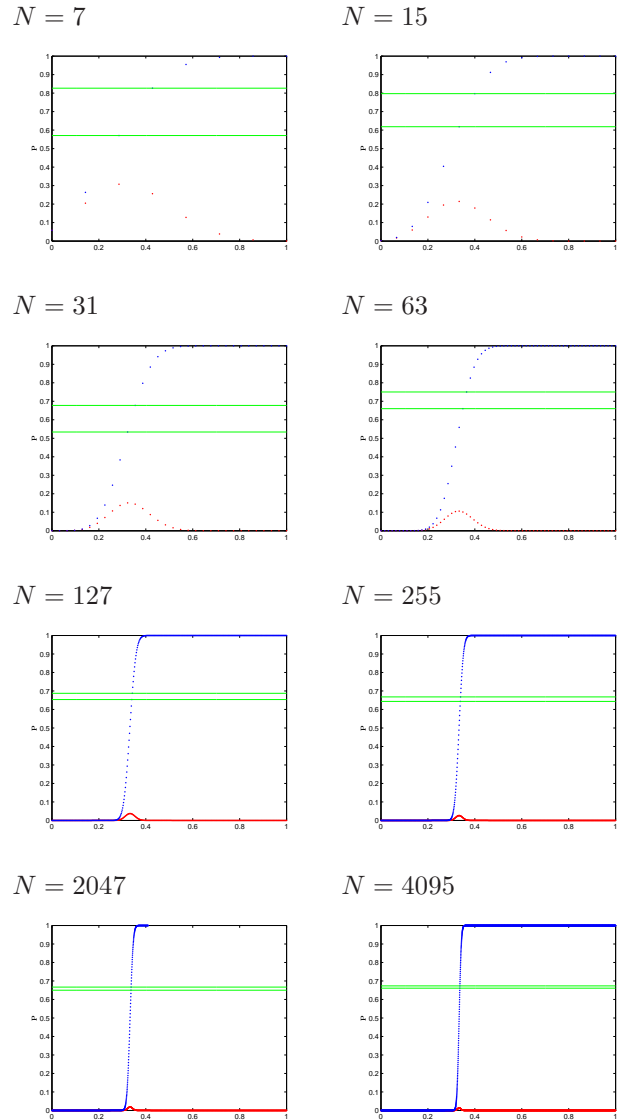


Рис. 1: Визуализация асимптотики семейства средневоятных событий в 8 схемах по  $N = 2^n - 1, n = 3, \dots, 12$  независимых испытаний, в каждом из которых вероятность события  $p = \frac{1}{3}$ . График биномиальных вероятностей  $b_n(|\mathfrak{X}|, p)$  изображен красным, график функции биномиального распределения  $B_n(|\mathfrak{X}|, p)$  синим, граница семейства средневоятных событий изображена зеленым.

## 2 Распределение Пуассона

Предлагается визуализировать асимптотику семейства средневоятных событий подходом, аналогичным подходу, описанным в [2] для распределения Пуассона [5].

Итак, рассматривается всеобщее вероятностное пространство  $(\Omega, \mathcal{F}, \mathbf{P})$  и схема серий Пуассона с параметром  $\lambda > 0$ . Каждая  $n$ -серия состоит из  $|\mathfrak{X}| = n$  событий, а весь эксперимент состоит из счетного числа серий. В каждой серии испытаний события равно-

вероятны, одноименны (однородны), различны и totally независимы.

Вероятность террасных слоев  $C_{\mathfrak{X}}^m, m = 1, \dots, N$  и террасных пластов  $C_{\mathfrak{X}}^{m+}, m = 1, \dots, N$  вычисляется следующим образом:

$$\mathbf{P}(\text{terop}(C_{\mathfrak{X}}^m/2^{\mathfrak{X}})) = e^{-\lambda} * \frac{\lambda^m}{m!}, m = 1, \dots, N; \quad (10)$$

$$\begin{aligned} \mathbf{P}(\text{terop}(C_{\mathfrak{X}}^{m+}/2^{\mathfrak{X}})) = \\ = \sum_{i=m}^N \mathbf{P}(\text{terop}(C_{\mathfrak{X}}^i/2^{\mathfrak{X}})), m = 1, \dots, N. \end{aligned} \quad (11)$$

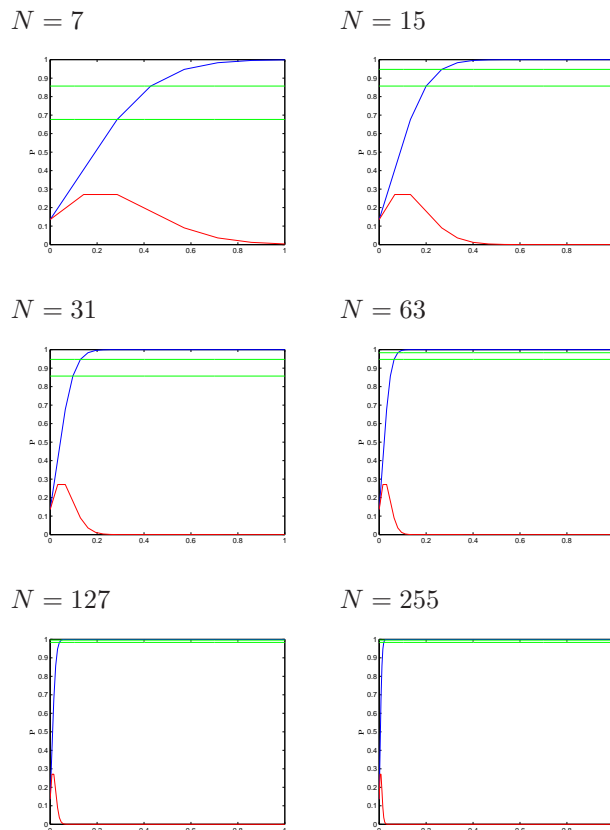


Рис. 2: Визуализация асимптотики семейства среднередких событий в 6 сериях по  $N = 2^n - 1, n = 3, \dots, 8$  независимых испытаний для распределения Пуассона с параметром  $\lambda = 2$ . График вероятностей  $b_n(|\mathfrak{X}|, p)$  изображен красным, график функции распределения  $B_n(|\mathfrak{X}|, p)$  синим, граница семейства среднередких событий изображена зеленым.

На рисунке 2 показаны серии испытаний распределения Пуассона при параметре  $\lambda = 2$  построенных указанным методом в каждой серии. Распределение Пуассона — дискретное распределение, и пусть никого не смущают «непрерывные графики» на рисунке 2, они созданы всего лишь для большей наглядности. Как и в случае со схемой Бернулли семейство среднередких событий сокращается с ростом числа событий  $|\mathfrak{X}|$ . Строгое доказательство этого автору еще предстоит завершить.

## Список литературы

- [1] О.Ю. Воробьев. Среднередкое событие для множества событий. *Труды XI международной конференции по финансово-актуарной математике и эвентологии безопасности*, -:139–147, 2012.
- [2] О.Ю. Воробьев. Событийные средние в эвентологии, их асимптотические свойства, интерпретация и визуализация. *Труды XVI международной ЭМ конференции по эвентологической математике и смежным вопросам*, 205:51–56, 2012.
- [3] О.Ю. Воробьев. *Эвентология*. Красноярск: Сибирский федеральный университет, 2007.
- [4] Д.В.Семенова Н.А. Лукьянова, Я.В. Нартов. Визуализация среднередкого события через матричное представление террасок-событий. *Труды XVI Международной ЭМ конференции по эвентологической математике и смежным вопросам*, 205:145–158, 2012.
- [5] З.А. Кузичева И.Г. Башмакова, Б.В. Гнеденко. *Математическая логика. Алгебра. Теория чисел. Теория вероятностей (Математика XIX века)*. Наука, 1978.

# Алгоритмы Монте-Карло генерирования распределений множеств событий и их применение в портфельном анализе валютных рынков и рынков драгоценных металлов

Александр Сергеевич Нифонтов

Сибирский федеральный университет  
Институт математики и фундаментальной информатики  
Красноярск  
nialeksander@gmail.com

**Аннотация.** Рассматривается прямая эвентологическая задача портфельного анализа событий — событийный аналог задачи Марковица из классического портфельного анализа. Рассматриваются обобщенные постановки прямой эвентологической задачи Марковица. Описывается метод равномерного выбора вероятностного распределения и метод равномерного выбора вероятностного распределения с заданными вероятностями событий. Решаются прямые обобщенные эвентологические задачи Марковица на основе статистических данных с валютных рынков и рынков драгоценных металлов.

**Ключевые слова.** Эвентология, событие, вероятность, портфельный анализ, портфельный анализ событий, валютный рынок, рынок драгметаллов.

## 1 Прямая эвентологическая задача Марковица

В [1, стр. 275–287] сформулирована прямая эвентологическая задачу Марковица, иными словами, *прямая эвентологическая портфельная задача*, в которой необходимо найти неизвестные доли событий в портфеле при известном вероятностном распределении множества этих портфельных событий.

Пусть  $\{\mathfrak{X}(\mathbf{p}), \mathbf{a}\}$  — портфель событий, где  $\mathfrak{X}$  — множество портфельных событий с известным вероятностным распределением

$$\mathbf{p} = \{p(X) : X \subseteq \mathfrak{X}\},$$

а

$$\mathbf{a} = \left\{ a_x : a_x \geq 0, x \in \mathfrak{X}, \sum_{x \in \mathfrak{X}} a_x = 1 \right\}$$

— множество неизвестных долей этих событий в портфеле.

Предполагается, что доходности портфельных событий  $x \in \mathfrak{X}$  определяются их индикаторами:

$$r_x = \mathbf{1}_x.$$

Тогда средняя доходность каждого портфельного события  $x \in \mathfrak{X}$  является вероятностью этого события:

$$\mathbf{E}r_x = \mathbf{E}\mathbf{1}_x = \mathbf{P}(x),$$

дисперсия доходности портфельного события равна

$$\mathbf{D}r_x = \mathbf{D}\mathbf{1}_x = \mathbf{P}(x)(1 - \mathbf{P}(x)) = \sigma_x^2,$$

где  $\sigma_x^2 = \mathbf{D}\mathbf{1}_x$  — дисперсия индикатора портфельного события  $x \in \mathfrak{X}$ .

Доходность всего портфеля событий полагается равной линейной комбинации доходностей его событий с долями из  $\mathbf{a}$ :

$$r_{\mathfrak{X}} = \sum_{x \in \mathfrak{X}} a_x \mathbf{1}_x.$$

Тогда средняя доходность портфеля событий определяется вероятностями его событий — это выпуклая комбинация вероятностей событий с коэффициентами, равными долям этих событий в портфеле:

$$\mathbf{E}(r_{\mathfrak{X}}) = \sum_{x \in \mathfrak{X}} a_x \mathbf{P}(x), \quad (1)$$

а дисперсия — это квадратичная форма от  $\mathbf{a}$ :

$$\mathbf{D}(r_{\mathfrak{X}}) = \sum_{x \in \mathfrak{X}} \sum_{y \in \mathfrak{X}} a_x a_y \text{Kov}_{xy}, \quad (2)$$

с коэффициентами, равными  $\text{Kov}_{xy} = \mathbf{P}(x \cap y) - \mathbf{P}(x)\mathbf{P}(y)$  — парной ковариации событий  $x$  и  $y$ .

Постановка прямой эвентологической портфельной задачи выглядит стандартным образом: *найти такое множество долей  $\mathbf{a}$  событий в данном портфеле с известным вероятностным распределением  $\mathbf{p}$ , которое обеспечивает данное фиксированное значение средней доходности портфеля при его минимальном риске (дисперсии портфеля):*

$$\mathbf{E}(r_{\mathfrak{X}}) = \langle r_{\mathfrak{X}} \rangle, \quad (3)$$

$$\mathbf{D}(r_{\mathfrak{X}}) \rightarrow \min_{\mathbf{a}}.$$

## 2 Обобщенные постановки: доходность события как линейная функция его индикатора

В [2] предложено линейное обобщение прямой и обратной эвентологических портфельных задач, когда доходность события определяется как линейная функция его индикатора.

Пусть  $(\mathfrak{X}, \mathbf{a})$  — портфель событий, где  $\mathfrak{X}$  — множество портфельных событий с вероятностным распределением

$$\mathbf{p} = \{p(X) : X \subseteq \mathfrak{X}\},$$

а

$$\mathbf{a} = \left\{ a_x : a_x \geq 0, x \in \mathfrak{X}, \sum_{x \in \mathfrak{X}} a_x = 1 \right\}$$

— множество долей этих событий в портфеле.

Предполагается, что доходности портфельных событий  $x \in \mathfrak{X}$  определяются линейными функциями их индикатора:

$$r_x = b_x \mathbf{1}_x + d_x.$$

Тогда средняя доходность каждого портфельного события  $x \in \mathfrak{X}$  является линейной функцией вероятности события:

$$\mathbf{E} r_x = b_x \mathbf{E} \mathbf{1}_x + d_x = b_x \mathbf{P}(x) + d_x,$$

дисперсия доходности портфельного события равна

$$\mathbf{D} r_x = b_x^2 \mathbf{D} \mathbf{1}_x = b_x^2 \mathbf{P}(x)(1 - \mathbf{P}(x)) = b_x^2 \sigma_x^2,$$

где  $\sigma_x^2 = \mathbf{D} \mathbf{1}_x$  — дисперсия индикатора портфельного события  $x \in \mathfrak{X}$ .

Доходность всего портфеля событий полагается равной выпуклой комбинации доходностей его событий с долями из  $\mathbf{a}$ :

$$r_{\mathfrak{X}} = \sum_{x \in \mathfrak{X}} a_x b_x \mathbf{1}_x + \sum_{x \in \mathfrak{X}} a_x d_x.$$

Тогда средняя доходность портфеля событий определяется вероятностями его событий по формуле:

$$\mathbf{E}(r_{\mathfrak{X}}) = \sum_{x \in \mathfrak{X}} a_x b_x \mathbf{P}(x) + \sum_{x \in \mathfrak{X}} a_x d_x, \quad (5)$$

а дисперсия — формулой:

$$\mathbf{D}(r_{\mathfrak{X}}) = \sum_{x \in \mathfrak{X}} \sum_{y \in \mathfrak{X}} a_x b_x a_y b_y \text{Kov}_{xy}, \quad (6)$$

где  $\text{Kov}_{xy} = \mathbf{P}(x \cap y) - \mathbf{P}(x)\mathbf{P}(y)$  — парная ковариация событий  $x$  и  $y$ .

Линейно обобщенная постановка прямой эвентологической задачи выглядит стандартным образом: *найти такое множество долей  $\mathbf{a}$  событий в данном портфеле с известным фиксированным вероятностным распределением  $\mathbf{p}$ , которое обеспечивает*

*данное фиксированное значение средней доходности портфеля при его минимальном риске (дисперсии портфеля):*

$$\begin{aligned} \mathbf{E}(r_{\mathfrak{X}}) &= \langle r_{\mathfrak{X}} \rangle, \\ \mathbf{D}(r_{\mathfrak{X}}) &\rightarrow \min_{\mathbf{a}}. \end{aligned} \quad (7)$$

### 2.1 Первая обобщенная постановка портфельной задачи

В первой постановке портфельной задачи оцениваются коэффициенты  $b_x$  и  $d_x$  по всему портфелю в целом.

$$d_x = d = \min_{x,m} S_{x,m},$$

$$b_x = b = \max_{x,m} S_{x,m} - d,$$

где  $S_{x,m}$  статистика портфельного актива  $x$  за период  $m$  из  $M$  наблюдаемых.

Получаем систему уравнений:

$$\begin{cases} b \sum_{x \in \mathfrak{X}} a_x \mathbf{P}(x) + d = \langle a \rangle \\ b^2 \sum_{x \in \mathfrak{X}} \sum_{y \in \mathfrak{X}} a_x a_y \text{Kov}_{xy} \rightarrow \min_{\mathbf{a}}. \end{cases}$$

### 2.2 Вторая обобщенная постановка портфельной задачи

Во второй постановке портфельной задачи оцениваются коэффициенты  $b_x$  и  $d_x$  для каждого портфельного актива по отдельности.

$$d_x = \min_m S_{x,m},$$

$$b_x = \max_m S_{x,m} - d_x,$$

где  $S_{x,m}$  статистика портфельного актива  $x$  за период  $m$  из  $M$  наблюдаемых.

Получаем систему уравнений:

$$\begin{cases} \sum_{x \in \mathfrak{X}} a_x b_x \mathbf{P}(x) + \sum_{x \in \mathfrak{X}} a_x d_x = \langle a \rangle \\ \sum_{x \in \mathfrak{X}} \sum_{y \in \mathfrak{X}} a_x b_x a_y b_y \text{Kov}_{xy} \rightarrow \min_{\mathbf{a}}. \end{cases}$$

## 3 Метод равномерного выбора вероятностного распределения

Давно известен алгоритм Монте-Карло генерации распределения множества событий  $\mathfrak{X}$ , когда оно состоит из  $|\mathfrak{X}| = N$  событий. Мощность полученного распределения  $|\mathfrak{X}| = 2^N$ , и ему соответствует точка в  $2^N$ -вершинном симплексе распределений. Для генерации равномерной в симплексе  $2^N$ -мерной точки методом Монте-Карло сначала генерируем псевдослучайные числа  $a_i^* \in [0, 1]$  при  $i = 1, \dots, 2^N - 1$ , равномерно распределенные на  $[0, 1]$ . Эти значения разбивают единичный отрезок на  $2^N$  отрезков (см. рис.

1), длины которых служат решением задачи. Чтобы провести вычисления, полученные значения  $a_i^*$  упорядочиваем по возрастанию

$$0 < a_1 < \dots < a_{2^N-1} < 1.$$

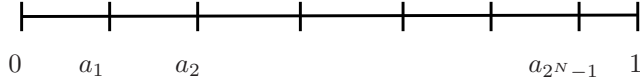


Рис. 1: Упорядоченные по возрастанию значения псевдослучайных чисел, равномерных на единичном отрезке.

После этого, вычитая из каждого значения  $a_i$  предыдущее, и считая, что  $a_0 = 0$ , получим значения (см. рис 10)

$$r_i = a_{i+1} - a_i$$

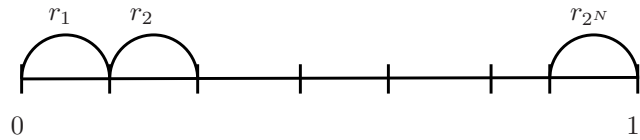


Рис. 2: Значения  $r_i = a_i - a_{i-1}$  координат  $2^N$ -мерной точки, равномерно распределенной в стандартном единичном  $2^N$ -вершинном симплексе.

для  $i = 0, 1, \dots, 2^N - 1$ , которые образуют координаты  $2^N$ -мерной точки, равномерно распределенной в стандартном единичном  $2^N$ -вершинном симплексе.

Затем устанавливаем произвольное взаимно однозначное отображение  $2^N$ -набора полученных значений

$$\{r_i, i = 0, \dots, 2^N - 1\}$$

на  $2^N$ -набор вероятностей террасных событий,

$$\{p(X), X \subseteq \mathfrak{X}\},$$

образующих распределение множества событий  $\mathfrak{X}$ .

#### 4 Равномерный выбор вероятностного распределения с заданными вероятностями событий

В [3] предложен алгоритм Монте-Карло генерирования распределения множества событий  $\mathfrak{X}$ , но с ограничением: вероятности событий  $x \in \mathfrak{X}$  фиксированы, т.е. справедливы  $N$  уравнений

$$\sum_{x \in X} p(X) = \mathbf{P}(x), \quad (1)$$

которым должно удовлетворять генерируемое распределение

$$\{p(X), X \subseteq \mathfrak{X}\}.$$

Проведем построение алгоритма по индукции по мощности  $N$  множества событий  $\mathfrak{X}$ .

При  $N = 1$  множество событий — это моноплет  $\mathfrak{X} = \{x\}$ , который разбивает  $\Omega$  на два террасных события, вероятности которых следуют из ограничений (1):

$$p(\emptyset // \{x\}) = 1 - \mathbf{P}(x),$$

$$p(\{x\} // \{x\}) = \mathbf{P}(x).$$

Таким образом, на этом шаге псевдослучайных чисел для генерации распределения не требуется, так как распределение определяется из ограничений.

При  $N = 2$  множество событий — это дуплет  $\mathfrak{X} = \{x, y\}$ , который разбивает  $\Omega$  на четыре террасных события. Эти четыре террасных события можно получить из двух, полученных на 1-м шаге, разбив каждое из них на два. Поступим следующим образом: разобьем отрезок  $[0, \mathbf{P}]$  на две части, генерацией случайного числа  $r_x$ , равномерного в отрезке

$$[\mathbf{P}(x) + \mathbf{P}(x) - 1, \mathbf{P}(x)], \quad (2)$$

а на отрезке  $[\mathbf{P}(x), 1]$  точка  $r_{x^c}$  определяется из ограничений:

$$r_{x^c} = \mathbf{P}(x) + \mathbf{P}(y) - r_x.$$

В результате получится распределение:

$$p(\{x, y\}) = r_x;$$

$$p(\{x^c, y\}) = r_{x^c} - \mathbf{P}(x);$$

$$p(\{x, y^c\}) = \mathbf{P}(x) - r_x;$$

$$p(\emptyset) = 1 - r_{x^c};$$

Этот принцип разбиения на два всех террасных событий, полученных на предыдущем шаге, может составить основу конструирования произвольного рекурсивного  $n$ -го шага алгоритма, на котором каждый из полученных на предыдущем шаге отрезков поочередно разбивается на два так, чтобы из полученных “половинок” удалось “собрать” очередную вероятность  $n$ -го события, которая задана. Для этого надо выбирать случайные числа не из целых предыдущих отрезков, а только из каких-то их частей так, как это сделано в (2).

##### 4.1 Метод равномерного выбора вероятностного распределения с заданными вероятностями событий

Следуя [3], предположим, что требуется построить распределение множества событий  $\mathfrak{X}$  из  $N = |\mathfrak{X}|$  событий, вероятности которых зафиксированы.

Допустим, что нам уже удалось решить эту проблему для некоторого  $n < N$ , т.е. генерировать распределение  $(n - 1)$ -подмножества  $\mathfrak{X}_{n-1} \subset \mathfrak{X}$ .

Тогда на очередном  $n$ -м шаге остается лишь “дополнить” его распределение еще одним  $n$ -м событием,



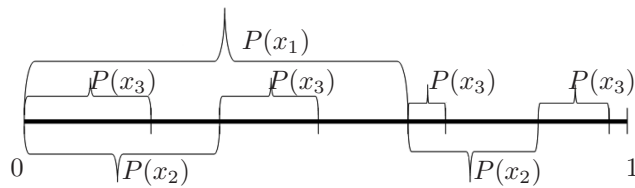


Рис. 3: Пример распределения множества событий для трех событий с заданными вероятностями

которое обозначим  $x \in \mathfrak{X} - \mathfrak{X}_{n-1}$ , и которое превратит это  $2^{n-1}$ -распределение в  $2^n$ -распределение, разбив каждое его террасное событие на две половины, одна из которых попадет в событие  $x$ , а другая в его дополнение  $x^c$ . И сделать это надо так, чтобы вероятность добавляемого события  $x$  оказалась равной заранее фиксированному числу  $\mathbf{P}(x)$ .

Итак, на очередном шаге требуется распределить (разбить) вероятностную массу  $\mathbf{P}(x)$  на  $2^{n-1}$  часть так, чтобы каждая уместилась в соответствующую вероятность террасного события из полученных на предыдущем шаге. Допустим, что вероятности этих террасных событий образуют набор

$$\pi_1, \dots, \pi_{2^{n-1}} \quad (1)$$

— вероятностное  $2^{n-1}$ -распределение, полученное на  $(n-1)$ -м шаге.

Представим пошаговое распределение вероятностной массы  $\mathbf{P}(x)$  на  $2^{n-1}$  частей по определенному правилу. И представим, что перед нами стоит задача распределения оставшейся нераспределенной массы  $0 < \pi_i \leq \mathbf{P}(x)$  на очередном  $i$ -м шаге ( $i = 1, \dots, 2^{n-1}$ ). Введем следующие обозначения

$$\sum_{j=1}^{i-1} p_j = S_{i-1}, \quad \sum_{j=i+1}^{2^{n-1}} p_j = D_{i+1},$$

в которых вероятностная нормировка распределения (1) принимает вид

$$S_{i-1} + p_i + D_{i+1} = 1.$$

Тогда на очередном  $i$ -м шаге нам надо решить задачу случайного распределения оставшейся вероятностной массы  $\pi_i$  на две половины:

$$\pi_i = \rho_i + \delta_i,$$

которые умещаются в величины

$$R = p_i \quad \text{и} \quad D = D_{i+1}$$

соответственно. Нам стоит, конечно, понять, что случайный выбор здесь определяется одним случайным числом  $r$ , например, определяющим величину  $\rho_i$ , а вторая величина определится из ограничения на их

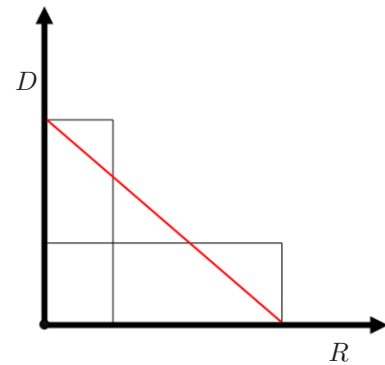


Рис. 4: Зависимость между R и D

сумму. Решение задачи на  $i$ -м шаге дается любым случайным числом  $r$  из интервала:

$$\max\{0, D - \pi\} \leq r \leq \min\{\pi, R\}. \quad (2)$$

На следующем  $(i+1)$ -м шаге решение для следующей доли  $\rho_{i+1}$ , записанное в виде (2), также подходит после соответствующего пересчета величин

$$\pi_{i+1} = \pi_i - \rho_i,$$

$$D = D - p_i,$$

$$R = p_{i+1}.$$

Завершение этого пошагового процесса дает  $2^n$ -распределение, а затем шаг за шагом по  $n$  даст  $2^N$ -распределение.

## 5 Применение на валютных рынках

Рассмотрим применение на валютных рынках. Статистика за период с 1.01.2013 по 1.04.2013 взята с сайта <http://www.finmarket.ru> Выбранный портфель состоит из:

- Японская Йена;
- Доллар США;
- ЕВРО.

Нормируем статистику: Вычтем из всех элементов статистики минимальную стоимость валюты и потом разделим на полученный максимум по цене. Получили статистику, элементы которой лежат в интервале  $[0, 1]$ . Для нахождения долей валют в портфеле, решим прямую обобщенную задачу Марковица в двух постановках.

### 5.1 Первая обобщенная постановка портфельной задачи

Для поиска функции средней доходности портфеля находим максимальную и минимальную стоимость валюты. В выбранном периоде получили: минимум

$= 0$ , а максимум  $= 1$ . Рассчитаем функцию средней доходности

$$Er_x = b_x \mathbf{E} \mathbf{1}_x + d_x = b_x \mathbf{P}(x) + d_x,$$

где

$$b_x = \max_{i,j} S_{i,j} - d_x,$$

а  $d_x = 0$  - минимум,  $b_x = 1$ . Рассчитаем обобщенную прямую задачу Марковица с использованием полученной средней доходности. Условием доходности будет ограничение

$$\langle a \rangle = 0.6$$

После решения поставленной задачи получили комбинацию долей:

$$a_1 = 0.4918, a_2 = 0.0357, a_3 = 0.4725.$$

Полученная комбинация имеет доходность 0.5903 и риск 0.0348.

## 5.2 Вторая обобщенная постановка портфельной задачи

Находим среднюю доходность портфельного события  $x \in \mathfrak{X}$  вероятность этого события записывается по формуле:

$$Er_x = b_x \mathbf{E} \mathbf{1}_x + d_x = b_x \mathbf{P}(x) + d_x,$$

$b_x$  и  $d_x$  рассчитываем отдельно по каждой валюте. Получаем:

- Японская Йена:  $b_1 = 0.2635, d_1 = 0.1805$ ;
- Доллар США:  $b_2 = 0.1059, d_2 = 0.0000$ ;
- ЕВРО:  $b_3 = 0.1123, d_3 = 0.8877$ .

Подставляем найденные коэффициенты в обобщенную прямую задачу Марковица с использованием полученной средней доходности. Условием доходности будет ограничение

$$\langle a \rangle = 0.6$$

После решения поставленной задачи получили комбинацию долей:

$$a_1 = 0.0062, a_2 = 0.3471, a_3 = 0.6467.$$

Полученная комбинация имеет доходность 0.5980 и риск 0.0012.

## 5.3 Сравнение методов

При использовании оценки портфеля в целом получили доходность 0.5903 и риск 0.0348, а при оценке валют по отдельности 0.5980 и риск 0.0012. В данном примере заметно выгоднее использовать оценку валют по отдельности потому, что получили доходность немного выше и в тоже время риск уменьшился в разы.

## 6 Применение на рынках драгоценных металлов

Рассмотрим анализ статистики на примере рынка драгоценных металлов. Статистика взята с сайта <http://mfd.ru/> в период с 9.01.2013 по 9.04.2013. Рассматриваемые металлы:

- Золото;
- Серебро;
- Платина;
- Палладий.

Нормируем статистику: вычтем из всех элементов статистики минимальную стоимость валюты и потом разделим на полученный максимум по цене. Получили статистику, элементы которой лежат в интервале  $[0, 1]$ . Для нахождения вкладываемых долей воспользуемся двумя обобщенными постановками. Для начала рассчитаем среднюю доходность по каждому металлу. Получим:

- Золото — 0.9527;
- Серебро — 0.0012;
- Платина — 0.9567;
- Палладий — 0.4271.

### 6.1 Первая обобщенная постановка портфельной задачи

Для поиска функции средней доходности портфеля находим максимальную и минимальную стоимость металла. В выбранном периоде получили: минимум  $= 0$ , а максимум  $= 1$ . Рассчитаем функцию средней доходности

$$Er_x = b_x \mathbf{E} \mathbf{1}_x + d_x = b_x \mathbf{P}(x) + d_x,$$

где

$$b_x = \max_i S_{i,j} - d_x,$$

а  $d_x = 0$  - минимум,  $b_x = 1$ . Рассчитаем обобщенную прямую задачу Марковица с использованием полученной средней доходности. Условием доходности будет ограничение:

$$\langle a \rangle = 0.6.$$

После решения поставленной задачи получили комбинацию долей:

$$a_1 = 0.1935, a_2 = 0.3374, a_3 = 0.2562, a_4 = 0.2128.$$

Полученная комбинация имеет доходность 0.6085 и риск 0.0073.

### 6.2 Вторая обобщенная постановка портфельной задачи

находим среднюю доходность портфельного события  $x \in \mathfrak{X}$  вероятность этого события записывается по формуле:

$$Er_x = b_x \mathbf{E} \mathbf{1}_x + d_x = b_x \mathbf{P}(x) + d_x,$$

$b_x$  и  $d_x$  рассчитываем отдельно по каждому металлу. Получаем

- Золото —  $b_1 = 0.0730$ ,  $d_1 = 0.9107$ ;
- Серебро —  $b_2 = 0.0024$ ,  $d_2 = 0.0000$ ;
- Платина —  $b_3 = 0.0789$ ,  $d_3 = 0.9211$ ;
- Палладий —  $b_4 = 0.0647$ ,  $d_4 = 0.3897$ .

Подставляем найденные коэффициенты в обобщенную прямую задачу Марковица с использованием полученной средней доходности. Условием доходности будет ограничение:

$$\langle a \rangle = 0.6.$$

После решения поставленной задачи получили комбинацию долей:

$$a_1 = 0.3200, a_2 = 0.1387, a_3 = 0.0959, a_4 = 0.4455.$$

Полученная комбинация имеет доходность 0.5911 и риск 0.0005.

### 6.3 Сравнение методов

При использовании первого способа получили риск 0.0073 и доходность 0.6085, а при использовании второго способа получили 0.0005 и 0.5911 соответственно. Таким образом при оценивании портфеля в целом получили чуть более доходный портфель, но риск получился во много раз больше.

### Список литературы

- [1] О. Ю. Воробьев. *Эвентология*. Сибирский федеральный университет, Красноярск, 2007, 435с., <http://eventology-theory.ru/0-books/1-VorobyevOleg~2007~Eventology~435p.pdf>.
- [2] О. Ю. Воробьев, А. С. Нифонтов. Портфельный анализ событий: расширенная версия. *Труды XII Международ. ФАМЭБ конференции по финансово-актуарной математике и эвентологии безопасности*, Красноярск: НИИППБ, СФУ (под ред. Олега Воробьева):141–144, 2013.
- [3] О. Ю. Воробьев. Измерение многомерных событийных зависимостей в эвентологическом портфельном анализе финансово-актуарных, гуманитарных и социо-экономических рынков. *Труды XII Международ. ФАМЭБ конференции по финансово-актуарной математике и эвентологии безопасности*, Красноярск: НИИППБ, СФУ (под ред. Олега Воробьева):133–140, 2013.

# Дележ риска между факторами и активами

Аркадий Арсеньевич Новоселов

Сибирский федеральный университет  
Институт математики и фундаментальной информатики  
Красноярск  
arcady@novosyolov.ru

**Аннотация.** В работе рассматривается дележ показателя рискованности системы между составляющими ее элементами, внешними факторами, оказывающими влияние на поведение системы в условиях риска, а также двумерные дележи между элементами и факторами. Изложение ведется на примере финансового портфеля и факторных риск-моделей финансового рынка, однако, результаты работы могут быть легко перенесены на поведение сложных систем в других областях риск-менеджмента.

**Ключевые слова.** Риск, фактор, актив, риск-модель, ковариационная матрица, дележ риска, теорема Эйлера.

## 1 Введение

Рискованность сложных систем описывается различными мерами (показателями) риска. На поведение таких систем в условиях риска оказывает влияние как состав самих систем, так и различные внешние факторы, которые включены в риск-модель, как показано на рис. 1.

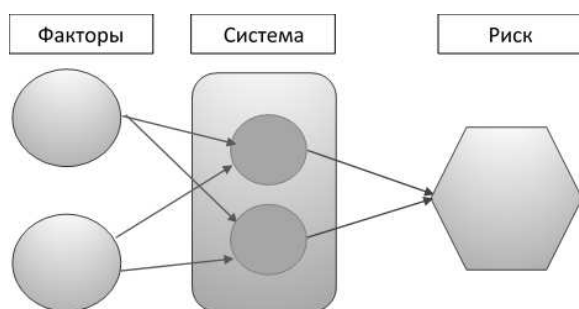


Рис. 1: Влияние состава системы и внешних факторов на риск системы.

Представляет интерес определение долей (вкладов) риска, которые имеет смысл приписать каждому элементу системы, или каждому фактору. Можно говорить также о доле риска, вкладываемой каждым фактором через посредство каждого элемента системы. Именно такое распределение вкладов в риск между элементами или факторами и называется в работе *дележом*.

Для определенности мы будем рассматривать риск-модели финансового портфеля, однако, принципиальные выводы справедливы и для других областей риск-менеджмента. В финансовых приложениях различают два типа риск-моделей: одноуровневые и двухуровневые. В одноуровневых моделях поведение системы описывается в терминах составляющих ее элементов, например, поведение финансового портфеля описывается в терминах активов, распределение которых задается непосредственно; факторы в модель не включаются. В двухуровневых моделях задается распределение факторов риска и функциональная зависимость элементов системы (активов портфеля) от факторов. Мы рассмотрим дележ риска в моделях этих двух типов отдельно.

Рискованность портфеля финансовых активов в финансовой индустрии принято измерять такими показателями, как tracking error (стандартное отклонение доходности портфеля), Value-at-Risk [1] (VaR, квантиль распределения доходности портфеля заданного уровня), Conditional Value-at-Risk [2] (CVaR, средние потери при условии превышения VaR; известен также под названием Expected tail loss, ETL) [3]. Более сложные меры риска ([4]), ([5]), ([6]), ([7]) пока используются чрезвычайно редко.

Для типичных распределений все эти показатели связаны со стандартным отклонением простыми масштабными преобразованиями. Например, для нормального распределения VaR уровня  $\alpha$ , обозначаемый  $v_\alpha$ , и стандартное отклонение  $\sigma$  связаны соотношением

$$v_\alpha = \sigma \Phi^{-1}(\alpha), \quad (1)$$

где  $\Phi$  — функция стандартного нормального распределения, а  $\Phi^{-1}$  — обратная к  $\Phi$  функция. Переход к значению  $c_\alpha$ , CVaR уровня  $\alpha$ , производится по одной из формул

$$c_\alpha = \frac{\sigma}{(1-\alpha)\sqrt{2\pi}} \exp\left(-\frac{1}{2}(\Phi^{-1}(\alpha))^2\right) \quad (2)$$

$$= \frac{\sigma}{1-\alpha} \varphi(\Phi^{-1}(\alpha)), \quad (3)$$

где  $\varphi$  — плотность стандартного нормального распределения.

Для других типичных распределений, например, Стюдента, эллиптических распределений, известны формулы перехода, аналогичные (1), (2), (3).

Поэтому в контексте работы нам достаточно изучать дележи стандартного отклонения портфеля; аналогичные дележи VaR и CVaR отличаются добавлением соответствующего масштабного коэффициента в каждый элемент разложения.

Работа построена следующим образом. В параграфе 2 описана теорема Эйлера, лежащая в основе всех дележей. Параграф 3 посвящен дележу риска в одноуровневых моделях, предложенному в [8]. В параграфе 4 представлены несколько новых дележей в двухуровневых линейных риск-моделях. В заключении подведены некоторые итоги.

## 2 Теорема Эйлера

Основой для дележей служит классическая теорема Эйлера о положительно однородных функциях [9]. Функция нескольких вещественных переменных  $f : R^n \rightarrow R$  называется положительно однородной, если  $f(\lambda x) = |\lambda|f(x)$  для произвольных  $x$  и области определения  $f$  и произвольного  $\lambda \in R$ . Теорема Эйлера утверждает, что если положительно однородная функция дифференцируема на  $R^n$ , то справедливо разложение (дележ)

$$f(x) = x' \cdot \nabla f(x) = \sum_{i=1}^n x_i \frac{\partial f}{\partial x_i}, \quad (4)$$

где  $x'$  обозначает вектор-строку, полученную транспонированием столбца  $x$ , а  $\nabla$  обозначает градиент функции.

## 3 Дележ стандартного отклонения в одноуровневых моделях

### 3.1 Дележ

Обозначим  $X = (X_1, X_2, \dots, X_n)'$  случайный вектор доходностей активов, имеющий нормальное распределение с нулевым средним и ковариационной матрицей  $C = EXX'$ . Далее, обозначим  $w = (w_1, \dots, w_n)'$  вектор весов, и составим из активов портфель

$$P = w'X.$$

Среднее значение доходности портфеля, очевидно, равно 0, а его дисперсия вычисляется по известной формуле

$$E(P^2) = E(w'XX'w) = w'Cw,$$

так что стандартное отклонение доходности портфеля, лежащее в основе вычисления всех показателей риска, равно

$$\sigma = f(w) = \sqrt{w'Cw}. \quad (5)$$

Дележ для (5) был предложен в [8].

Функция  $f$ , очевидно, является положительно однородной функцией  $w$ , поэтому к ней применима теорема Эйлера. Градиент  $f$  по  $w$  имеет вид

$$\nabla f = \frac{Cw}{\sqrt{w'Cw}},$$

а его  $j$ -я компонента, производная по  $w_j$ , равна

$$\frac{\partial f}{\partial w_j} = \frac{C_{j \cdot} w}{\sqrt{w'Cw}}, j = 1, \dots, n,$$

где  $C_{j \cdot}$  обозначает  $j$ -ю строку матрицы  $C$ .

Таким образом,  $j$ -я компонента дележа стандартного отклонения (ее еще называют маргинальным стандартным отклонением) равна

$$U_j = w_j \frac{\partial f}{\partial w_j} = \frac{w_j C_{j \cdot} w}{\sqrt{w'Cw}}, j = 1, \dots, n, \quad (6)$$

а сам дележ приобретает вид

$$\sigma = \sum_{i=1}^n U_i = \sum_{i=1}^n \frac{w_i C_{i \cdot} w}{\sqrt{w'Cw}}. \quad (7)$$

### 3.2 Простые случаи

Рассмотрим частные случаи. Пусть сначала только один из весов отличен от 0, например,

$$w_1 \neq 0, w_2 = 0, \dots, w_n = 0.$$

Тогда, очевидно,

$$U_1 = \sigma, U_2 = 0, \dots, U_n = 0.$$

Пусть теперь ковариационная матрица является диагональной, то есть  $c_{ij} = 0$  при  $i \neq j$ . Тогда

$$\sigma(w) = \sqrt{c_{11}w_1^2 + \dots + c_{nn}w_n^2},$$

$$\frac{\partial \sigma}{\partial w_i} = \frac{c_{ii}w_i}{\sqrt{\sum_{j=1}^n c_{jj}w_j^2}}, i = 1, \dots, n,$$

и маргинальные стандартные отклонения приобретают вид

$$w_i \frac{\partial \sigma}{\partial w_i} = \frac{c_{ii}w_i^2}{\sqrt{\sum_{j=1}^n c_{jj}w_j^2}}, i = 1, \dots, n.$$

Как видно, в рассмотренных примерах маргинальные стандартные отклонения неотрицательны. Это, однако, не является неизблемым правилом, в чем мы сейчас и убедимся.

### 3.3 Два актива

Для иллюстрации рассмотрим пример с двумя активами  $n = 2$ .



**Пример 1.** Пусть ковариационная матрица имеет вид

$$C = \begin{pmatrix} 1 & -1/2 \\ -1/2 & 1 \end{pmatrix}, \quad (8)$$

а вектор весов равен  $w = (1, 3)'$ . Тогда стандартное отклонение портфеля равно  $\sigma = \sqrt{w' C w} = \sqrt{7}$ , а его составляющие имеют вид

$$U_1 = -\frac{1}{2\sqrt{7}}, U_2 = \frac{15}{2\sqrt{7}}.$$

Как видим, несмотря на положительность весов обоих активов, маргинальный показатель риска для первого из них отрицателен. Содержательно это означает, что увеличение веса первого актива в портфеле должно приводить к снижению показателя риска портфеля. Прямое вычисление подтверждает это интуитивное предположение.

**Пример 2.** Рассмотрим более общую ситуацию с той же ковариационной матрицей (8), но с переменным весом первого актива,

$$w = (\beta, 1)', \beta \in [1, 3]. \quad (9)$$

Стандартное отклонение портфеля и маргинальные стандартные отклонения активов выражаются следующими формулами

$$\begin{aligned} \sigma(\beta) &= \sqrt{\beta^2 - 3\beta + 9}, \\ U_1(\beta) &= \frac{\beta^2 - \frac{3}{2}\beta}{\sqrt{\beta^2 - 3\beta + 9}}, \\ U_2(\beta) &= \frac{-\frac{3}{2}\beta + 9}{\sqrt{\beta^2 - 3\beta + 9}}, \end{aligned}$$

а графики этих функций приведены на рис. 2 – 4.

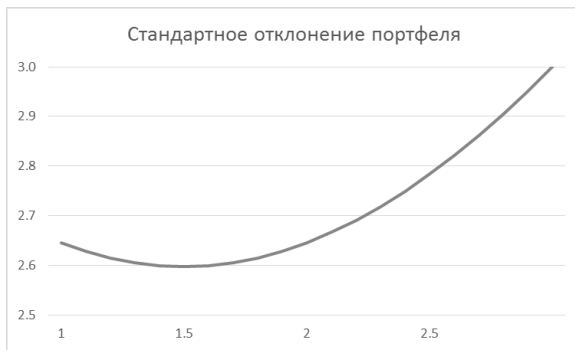


Рис. 2: Зависимость стандартного отклонения  $\sigma$  портфеля с параметрами (8), (9) от  $\beta$ .

При изменении значения  $\beta$  от 1 до 1.5 стандартное отклонение портфеля действительно уменьшается, достигая минимума  $3\sqrt{3}/2$  при  $\beta = 1.5$ . Вес второго актива оставался в данном примере неизменным. Поскольку второе маргинальное отклонение положительно, следует ожидать увеличения стандартного

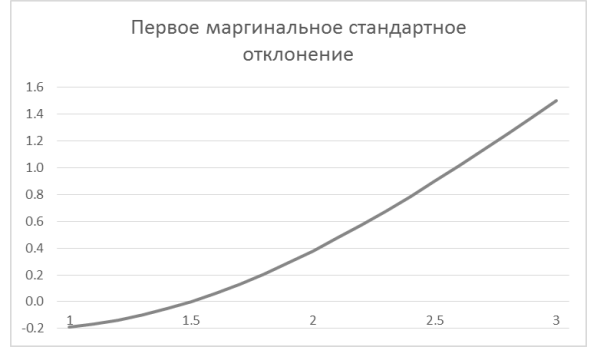


Рис. 3: Зависимость первой компоненты  $U_1$  стандартного отклонения портфеля с параметрами (8), (9) от  $\beta$ .

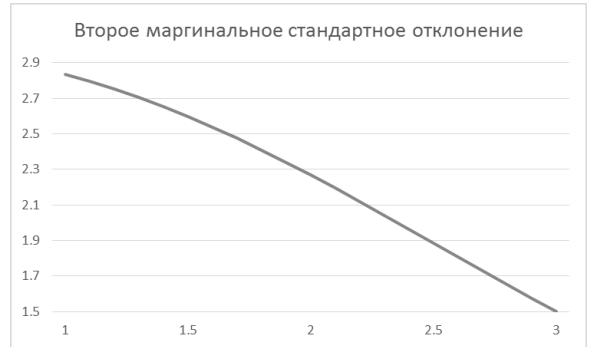


Рис. 4: Зависимость второй компоненты  $U_2$  стандартного отклонения портфеля с параметрами (8), (9) от  $\beta$ .

отклонения портфеля при увеличении веса второго актива.

**Пример 3.** На рис. 5 – 7 показаны аналогичные зависимости стандартного отклонения портфеля и его маргинальных составляющих в случае более общей ковариационной матрицы

$$C = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \quad (10)$$

при различных значениях  $\rho = -0.5, 0, 0.5$  и вектора весов

$$w = (\beta, 3)', \beta \in [-1, 1]. \quad (11)$$

## 4 Дележи стандартного отклонения в двухуровневых моделях

### 4.1 Описание модели

Рассмотрим двухуровневую факторную риск-модель следующего вида. Имеется  $n$  факторов риска, представляющих собой случайный вектор

$$X = (X_1, X_2, \dots, X_n) \quad (12)$$

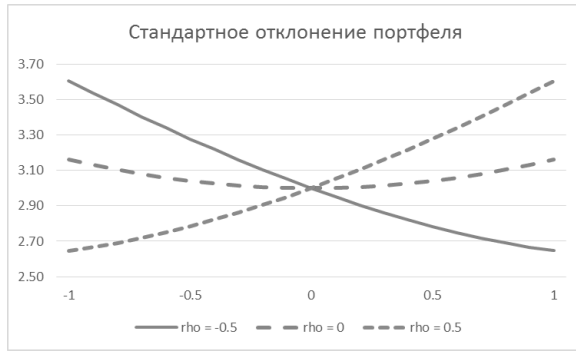


Рис. 5: Зависимость стандартного отклонения  $\sigma$  портфеля с параметрами (10), (11) от  $\beta$ .

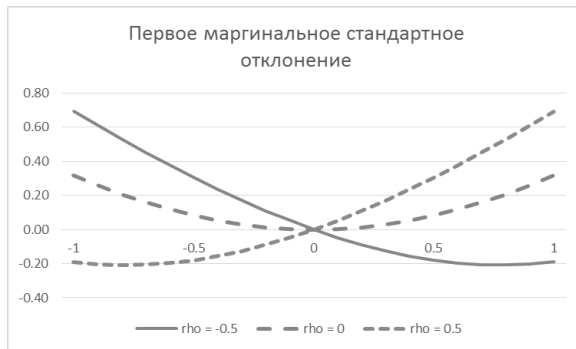


Рис. 6: Зависимость первого маргинального стандартного отклонения  $U_1$  портфеля с параметрами (10), (11) от  $\beta$ .

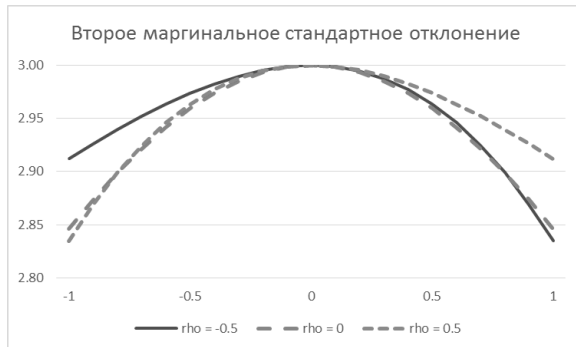


Рис. 7: Зависимость второго маргинального стандартного отклонения  $U_2$  портфеля с параметрами (10), (11) от  $\beta$ .

с нормальным распределением с нулевым средним и ковариационной матрицей

$$C = \mathbf{E}XX' = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{pmatrix}. \quad (13)$$

Рыночные активы  $A = (A_1, \dots, A_m)$  выражаются через факторы риска посредством линейных функций

$$A = HX + R, \quad (14)$$

где  $H$  — матрица нагрузок размера  $m \times n$ , а  $R$  — случайный вектор остатков, имеющий нормальное распределение с нулевым средним и диагональной ковариационной матрицей

$$G = \mathbf{E}RR' = \begin{pmatrix} g_{11} & 0 & \cdots & 0 \\ 0 & g_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & g_{mm} \end{pmatrix}. \quad (15)$$

Предполагается, что факторы  $X$  и остатки  $R$  независимы.

Активы, очевидно, имеют нормальное распределение с нулевым средним и ковариационной матрицей

$$D = \mathbf{E}AA' = \mathbf{E}((HX + R)(HX + R)') = HCH' + G.$$

Из активов составлен портфель

$$P = w'A \quad (16)$$

с весами  $w = (w_1, \dots, w_m)'$ . Вводя обозначение

$$W = H'w, \quad (17)$$

получаем выражение для портфеля через факторы

$$P = w'(HX + R) = W'X + w'R$$

Среднее значение портфеля равно 0, а его дисперсия вычисляется следующим образом

$$\mathbf{E}PP' = \mathbf{E}w'AA'w = w'(HCH' + G)w.$$

Таким образом, стандартное отклонение портфеля имеет вид

$$\sigma = \sqrt{w'(HCH' + G)w} \quad (18)$$

$$= \sqrt{W'CW + w'Gw} \quad (19)$$

$$= \sqrt{w'Dw}. \quad (20)$$

Здесь представляет интерес дележ стандартного отклонения портфеля по активам  $A_1, \dots, A_m$  и по факторам  $X_1, \dots, X_n$ . Кроме того, следует иметь в виду, что в данной модели кроме компонента  $X$  имеется также дополнительный фактор риска, представленный случайным вектором  $R$ ; его необходимо учитывать в дележе по факторам. Присвоим этому последнему фактору номер  $n + 1$ . Наша задача состоит в вычислении чисел  $U_{ij}$ ,  $i = 1, \dots, m$ ,  $j = 1, \dots, n + 1$  таких, что

$$\sigma = \sum_{i=1}^m \sum_{j=1}^{n+1} U_{ij}. \quad (21)$$

Кроме того, представляет интерес вывод дележей отдельно по факторам  $V_j$ ,  $j = 1, \dots, n + 1$  и активам  $T_i$ ,  $i = 1, \dots, m$  таких, что

$$\sigma = \sum_{j=1}^{n+1} V_j \quad (22)$$

и

$$\sigma = \sum_{i=1}^m T_i. \quad (23)$$

В идеальном варианте дележи (21) и (22), (23) должны быть согласованы в смысле

$$V_j = \sum_{i=1}^m U_{ij}, \quad j = 1, \dots, n+1 \quad (24)$$

и

$$T_i = \sum_{j=1}^{n+1} U_{ij}, \quad i = 1, \dots, m. \quad (25)$$

В настоящем параграфе рассмотрим несколько дележей, удовлетворяющих всем условиям, или только их части.

#### 4.2 Наивный дележ риска

Можно поступить довольно просто и назначить величины вкладов всех факторов через все активы одинаковыми:

$$U_{ij} = \frac{\sigma}{m(n+1)}, \quad i = 1, \dots, m, \quad j = 1, \dots, n+1. \quad (26)$$

Этот дележ удовлетворяет условиям согласования (21), (24), (25), однако, смысла в нем нет, и использование его для каких-либо целей проблематично. В последующих параграфах мы займемся поиском дележей риска, несущих в себе смысловую нагрузку вкладов факторов (активов) в риск портфеля.

#### 4.3 Первый дележ риска

Первый из дележей строится следующим образом. Сначала вычисляется дележ  $\sigma$  по активам в соответствии с теоремой Эйлера

$$T_i = \frac{w_i D_i w}{\sigma}, \quad i = 1, \dots, m, \quad (27)$$

так что выполнено (23). Затем вклады факторов в эти доли активов выделяются пропорционально нагрузкам и стандартным отклонениям факторов

$$U_{ij} = \frac{h_{ij} \sqrt{c_{jj}}}{k_i} T_i, \quad i = 1, \dots, m, \quad j = 1, \dots, n, \quad (28)$$

$$U_{i,n+1} = \frac{\sqrt{g_{ii}}}{k_i} T_i, \quad i = 1, \dots, m, \quad (29)$$

где нормирующие множители  $k_i, i = 1, \dots, n$  подбираются таким образом, чтобы выполнялось соотношение (25), то есть,

$$k_i = \sum_{j=1}^n h_{ij} \sqrt{c_{jj}} + \sqrt{g_{ii}}, \quad i = 1, \dots, m. \quad (30)$$

Далее вклады факторов можно определить посредством (24), при этом автоматически будет выполнено соотношение (22).

Недостатком такого подхода является тот факт, что на втором этапе, в формулах (28) — (30), не учитывается корреляционная структура распределения факторов, что в некоторых приложениях оказывается недопустимым.

**Пример 4.** Рассмотрим пример такого дележа риска. Пусть  $m = 3, n = 2$ ,

$$C = \begin{pmatrix} 1 & -0.5 \\ -0.5 & 1 \end{pmatrix}, \quad (31)$$

$$G = \begin{pmatrix} 0.7 & 0 & 0 \\ 0 & 0.4 & 0 \\ 0 & 0 & 0.9 \end{pmatrix}, \quad w = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}, \quad (32)$$

$$H = \begin{pmatrix} 0.2 & 0.5 \\ 0.2 & -0.7 \\ 0.5 & 0.4 \end{pmatrix}, \quad (33)$$

Стандартное отклонение портфеля  $\sigma$  равно 2.7749, а результаты его дележа, матрица  $U$  и векторы  $T, V$ , представлены в табл. 1.

	$X_1$	$X_2$	$R$	$T$
$A_1$	0.0680	0.1700	0.2845	0.5225
$A_2$	0.7890	-2.7615	2.4951	0.5225
$A_3$	0.4678	0.3743	0.8877	1.7298
$V$	1.3249	-2.2172	3.6673	2.7749

Таблица 1: Дележ риска по факторам и активам, первый метод.

В этом дележе вклады факторов определяются довольно произвольным образом. В следующем дележе этот недостаток будет устранен, и вклады факторов будут определяться в соответствии с чувствительностью  $\sigma$  к изменению веса фактора.

#### 4.4 Второй дележ риска

Дифференцируя (19) по компонентам  $W$ , получим

$$\frac{\partial \sigma}{\partial W_j} = \frac{C_j \cdot W}{\sigma},$$

поэтому справедливо разложение по факторам

$$\sigma = \sum_{j=1}^{n+1} V_j, \quad (34)$$

где

$$V_j = W_j \frac{\partial \sigma}{\partial W_j} = \frac{W_j C_j \cdot W}{\sigma}, \quad j = 1, \dots, n \quad (35)$$

и

$$V_{n+1} = \frac{w'Gw}{\sigma}. \quad (36)$$

Далее, используя (17) и (20), выразим компоненты дележа  $V_1, \dots, V_{n+1}$  через веса активов  $w$ :

$$\begin{aligned} V_j &= (H_j)'w \frac{C_j \cdot H'w}{\sqrt{w'Dw}} = w'H_j \frac{C_j \cdot H'w}{\sqrt{w'Dw}} \\ &= \frac{w'S^{(j)}w}{\sqrt{w'Dw}}, \quad j = 1, \dots, n \end{aligned} \quad (37)$$

где введено обозначение

$$S^{(j)} = H_j C_j H', \quad j = 1, \dots, n$$

и

$$V_{n+1} = \frac{w'Gw}{\sqrt{w'Dw}}. \quad (38)$$

Очевидно, все компоненты  $V_1, \dots, V_n, V_{n+1}$  являются положительно однородными функциями  $w$ , поэтому к каждой из них можно применить теорему Эйлера. В результате имеем

$$\begin{aligned} \frac{\partial V_j}{\partial w_i} &= \frac{2S_{i \cdot}^{(j)}w \sqrt{w'Dw} - w'S^{(j)}w \frac{D_{i \cdot}w}{\sqrt{w'Dw}}}{w'Dw} \\ &= \frac{2S_{i \cdot}^{(j)}w(w'Dw) - (w'S^{(j)}w)D_{i \cdot}w}{(w'Dw)^{3/2}} \end{aligned}$$

и

$$\begin{aligned} \frac{\partial V_{n+1}}{\partial w_i} &= \frac{2G_{i \cdot}w \sqrt{w'Dw} - w'Gw \frac{D_{i \cdot}w}{\sqrt{w'Dw}}}{w'Dw} \\ &= \frac{2(G_{i \cdot}w)(w'Dw) - (w'Gw)(D_{i \cdot}w)}{(w'Dw)^{3/2}} \end{aligned}$$

при всех  $j = 1, \dots, n$  и  $i = 1, \dots, m$ , так что

$$U_{ij} = w_i \frac{2(S_{i \cdot}^{(j)}w)(w'Dw) - (w'S^{(j)}w)(D_{i \cdot}w)}{(w'Dw)^{3/2}} \quad (39)$$

для  $i = 1, \dots, m$  и  $j = 1, \dots, n$ , и

$$U_{i,n+1} = w_i \frac{2(G_{i \cdot}w)(w'Dw) - (w'Gw)(D_{i \cdot}w)}{(w'Dw)^{3/2}} \quad (40)$$

для всех  $i = 1, \dots, m$ .

Отметим, что в этом дележе вклады факторов (35), (36) уже определяются на основании чувствительности риска портфеля к изменению соответствующих факторов.

**Пример 5.** Рассмотрим пример с параметрами (31) — (33) из примера 4. Здесь вклады факторов, приведенные в табл. 2, определяются в соответствии с чувствительностью  $\sigma$  к весу соответствующих факторов.

Однако, у этого способа распределения есть другой недостаток, который проявляется при рассмотрении немного отличного примера.

	$X_1$	$X_2$	$R$	$T$
$A_1$	0	0.3370	0.1856	0.5225
$A_2$	0	0.5532	-0.0307	0.5225
$A_3$	0	0.1910	1.5388	1.7298
$V$	0	1.0811	1.6938	2.7749

Таблица 2: Дележ риска по факторам и активам, второй метод.

**Пример 6.** Пусть параметры модели имеют те же значения (31) — (33) с единственным изменением элемента  $h_{32}$  в матрице нагрузок:

$$H = \begin{pmatrix} 0.2 & 0.5 \\ 0.2 & -0.7 \\ 0.5 & 0 \end{pmatrix}, \quad (41)$$

Дележ риска представлен в табл. 3.

	$X_1$	$X_2$	$R$	$T$
$A_1$	0.0344	0.2216	0.2076	0.4636
$A_2$	-0.0880	0.3551	0.0653	0.3323
$A_3$	0.2177	-0.2321	1.6556	1.6462
$V$	0.1641	0.3447	1.9284	2.4372

Таблица 3: Дележ риска по факторам и активам, второй метод, один из элементов  $H$  равен 0.

Здесь вклад второго фактора через третий актив оказывается отрицательным, хотя нагрузка третьего актива на второй фактор,  $h_{32}$ , равна 0. Это обстоятельство может показаться противоречащим интуиции. Поэтому имеет смысл ввести еще один метод дележа риска, в котором вклады факторов через активы, имеющие нулевые нагрузки на соответствующий фактор, равны нулю.

#### 4.5 Третий дележ риска

Для вывода третьего дележа введем переменные

$$z_{ij} = w_i h_{ij}, \quad i = 1, \dots, m, \quad j = 1, \dots, n. \quad (42)$$

Обозначим  $Q$  диагональную матрицу, на главной диагонали которой расположены веса  $w$ :

$$Q = \begin{pmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w_m \end{pmatrix}.$$

Очевидно, матрица  $Z$  с элементами (42) имеет вид

$$Z = QH.$$

Обозначим  $I = (1, 1, \dots, 1)'$  вектор размера  $m \times 1$ , состоящий из единиц. С учетом очевидного равенства  $w = Q'I$  из (18) получаем

$$\sigma = \sqrt{w'(HCH' + G)w} = \sqrt{I'QHCH'Q'I + w'Gw}.$$

Используя произвольную факторизацию  $C$  вида  $C = LL'$ , например, разложение Холецкого, получаем далее

$$\begin{aligned} \sigma &= \sqrt{I'QHLL'H'Q'I + w'Gw} \\ &= \sqrt{I'ZLL'Z'I + w'Gw}, \end{aligned} \quad (43)$$

$$= \sqrt{BB' + w'Gw}, \quad (44)$$

где использовано обозначение  $B = I'ZL$ . Из (43) видно, что  $\sigma$  является положительно однородной функцией набора переменных  $(Z, w)$ , поэтому можно применить теорему Эйлера.

Дифференцируя  $B$  по  $z_{ij}$  при фиксированных значениях индексов, получаем

$$\begin{aligned} \frac{\partial B}{\partial z_{ij}} &= \frac{\partial}{\partial z_{ij}} \left( \sum_{k=1}^m z_{k1}, \dots, \sum_{k=1}^m z_{kn} \right) L \\ &= e_j' L = L_j, \end{aligned}$$

где  $e_j$  обозначен  $j$ -й орт пространства  $\mathbf{R}^n$ . Отсюда

$$\begin{aligned} \frac{\partial BB'}{\partial z_{ij}} &= \frac{\partial B}{\partial z_{ij}} B' + B \frac{\partial B'}{\partial z_{ij}} \\ &= L_j B' + B (L_j)' = 2B(L_j)' \\ &= 2I'ZL(L_j)' = 2I'ZC_j = 2C_j Z'I \end{aligned}$$

для всех  $i = 1, \dots, m$ ,  $j = 1, \dots, n$ . Теперь из (44) вытекает

$$\frac{\partial \sigma}{\partial z_{ij}} = \frac{C_j Z'I}{\sigma}, \quad \frac{\partial \sigma}{\partial w_i} = \frac{G_i w}{\sigma}$$

при всех  $i = 1, \dots, m$ ,  $j = 1, \dots, n$ .

Для удобства записи обозначим  $S = CZ'I$  столбец с компонентами  $S = (s_1, \dots, s_n)'$ . Окончательно первые  $n$  столбцов матрицы дележа  $U$  вычисляются по формуле

$$U_{ij} = \frac{w_i h_{ij} s_j}{\sigma}, \quad i = 1, \dots, m, \quad j = 1, \dots, n,$$

а последний,  $(n+1)$ -й, столбец - по формуле

$$U_{i,n+1} = \frac{g_{ii} w_i^2}{\sigma}, \quad i = 1, \dots, m.$$

**Пример 7.** Рассмотрим дележ стандартного отклонения в модели с параметрами из примера 5. Этот дележ представлен в табл. 4.

	$X_1$	$X_2$	$R$	$T$
$A_1$	0	0.2703	0.2523	0.5225
$A_2$	0	0.3784	0.1441	0.5225
$A_3$	0	0.4324	1.2973	1.7298
$V$	0	1.0811	1.6938	2.7749

Таблица 4: Дележ риска по факторам и активам, третий метод.

	$X_1$	$X_2$	$R$	$T$
$A_1$	0.0328	0.1436	0.2872	0.4636
$A_2$	-0.0328	0.2010	0.1641	0.3323
$A_3$	0.1641	0	1.4771	1.6412
$V$	0.1641	0.3447	1.9284	2.4372

Таблица 5: Дележ риска по факторам и активам, третий метод, один из элементов  $H$  равен 0.

**Пример 8.** Дележ стандартного отклонения в модели с параметрами из примера 6 представлен в табл. 5. Как видим, вклад второго фактора через третий актив в этом дележе равен 0, согласуясь с нулевым значением соответствующей нагрузки.

Для удобства сравнения методов в табл. 6, 7 приведены дележи всеми методами для моделей примеров 4, 6.

Из табл. 6 видим, что первый дележ, использовавший распределение по факторам без учета чувствительности к изменению весов факторов, приводит к существенно отличным результатам.

Результаты второго и третьего дележей отличаются не очень значительно. Здесь обращает на себя внимание тот факт, что при втором дележе вклад фактора  $X_2$  осуществляется через актив  $A_3$  в наименьшей степени, а при третьем дележе, напротив, в наибольшей степени.

Из табл. 7 видим, что и в модели (31), (32), (41) первый дележ дает существенно отличные результаты, в частности, при дележе риска между факторами (вектор  $V$ ). Второй и третий дележ дают сопоставимые результаты, при выборе между этими дележами можно руководствоваться дополнительными соображениями.

## 5 Заключение

В работе предложены три способа двухуровневого дележа риска портфеля между факторами и активами. Первый из них частично опирается на теорему Эйлера о разложении положительно однородной функ-



	Первый дележ			
	$X_1$	$X_2$	$R$	$T$
$A_1$	0.0680	0.1700	0.2845	0.5225
$A_2$	0.7890	-2.7615	2.4951	0.5225
$A_3$	0.4678	0.3743	0.8877	1.7298
$V$	1.3249	-2.2172	3.6673	2.7749
	Второй дележ			
	$X_1$	$X_2$	$R$	$T$
$A_1$	0	0.3370	0.1856	0.5225
$A_2$	0	0.5532	-0.0307	0.5225
$A_3$	0	0.1910	1.5388	1.7298
$V$	0	1.0811	1.6938	2.7749
	Третий дележ			
	$X_1$	$X_2$	$R$	$T$
$A_1$	0	0.2703	0.2523	0.5225
$A_2$	0	0.3784	0.1441	0.5225
$A_3$	0	0.4324	1.2973	1.7298
$V$	0	1.0811	1.6938	2.7749

Таблица 6: Дележ риска по факторам и активам, модель (31) — (33)

ции, а второй и третий используют эту теорему в полной мере.

Выбор способа дележа в каждом конкретном случае, по-видимому, следует производить с учетом задач, решаемых при управлении портфелем, а также текущих условий функционирования. Немалую роль в выборе может сыграть и интуиция риск-менеджера, выработанная им при использовании различных способов дележа.

Стоит также отметить, что перечень способов дележа не исчерпывается приведенными методами. Среди решений уравнений (18), (24), (25) могут найтись и другие дележи, для которых векторы  $T$  и  $V$  сами являются дележами риска в соответствующих одноуровневых моделях, и обладающие различными дополнительными привлекательными свойствами.

## Благодарности

Автор выражает свою искреннюю признательность Д.С. Сачкову за многочисленные обсуждения методики дележа, результатом которых стали описанные в работе двухуровневые варианты дележа показателей риска.

	Первый дележ			
	$X_1$	$X_2$	$R$	$T$
$A_1$	0.0603	0.1505	0.2524	0.4636
$A_2$	0.5018	-1.7564	1.5869	0.3323
$A_3$	0.5665	0	1.0748	1.6412
$V$	1.1286	-1.6055	2.9141	2.4372
	Второй дележ			
	$X_1$	$X_2$	$R$	$T$
$A_1$	0.0344	0.2216	0.2076	0.4636
$A_2$	-0.0880	0.3551	0.0653	0.3323
$A_3$	0.2177	-0.2321	1.6556	1.6462
$V$	0.1641	0.3447	1.9284	2.4372
	Третий дележ			
	$X_1$	$X_2$	$R$	$T$
$A_1$	0.0328	0.1436	0.2872	0.4636
$A_2$	-0.0328	0.2010	0.1641	0.3323
$A_3$	0.1641	0	1.4771	1.6412
$V$	0.1641	0.3447	1.9284	2.4372

Таблица 7: Дележ риска по факторам и активам, модель (31), (32), (41)

## Список литературы

- [1] G. Holton. *Value-at-Risk: Theory and Practice*. Academic Press.
- [2] R.T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *The Journal of Risk*, 2:21–41, 2000.
- [3] C. Acerbi and D. Tasche. Expected shortfall: A natural coherent alternative to value at risk. *Economic Notes*, 31:379–388, 2002.
- [4] А.А. Новоселов. *Математическое моделирование финансового риска: теория измерения*. Наука, Новосибирск, 2001.
- [5] А.А. Новоселов. Обобщенные когерентные меры риска. *Труды IV Всероссийской конференции по финансово-актуарной математике и смежным вопросам*, 1:325–339, 2005.
- [6] Delbaen F. Eber J.-M. Artzner, P. and Heath D. Coherent measures of risk. *Mathematical Finance*, 9:203–228, 1999.
- [7] H. Follmer and A. Schied. Convex measures of risk and trading constraints. *Finance and Stochastics*, 6:429–447, 2002.
- [8] M. B. Garman. Taking var to pieces. *Risk Magazine*, 10:70–71, 1997.
- [9] А.Д. Мышкис. *Лекции по высшей математике*. Наука, М., 1973.

# Некоторые задачи управления пожарным риском

Аркадий Арсеньевич Новоселов

Сибирский федеральный университет  
Институт математики и фундаментальной информатики  
Красноярск  
arcady@novosyolov.ru

**Аннотация.** В работе рассматриваются некоторые задачи управления пожарным риском в условиях ограниченных ресурсов, приводятся методы их решения, производится сравнение методов, даны результаты численных экспериментов, а также рекомендации по применению методов.

**Ключевые слова.** Риск, пожар, управление, оптимизация, динамическое программирование.

## 1 Введение

Вопросам исследования пожарного риска посвящена обширная литература, см., напр., [1], а также библиографию в [2], [3]. В настоящей работе поставлены некоторые задачи оптимального управления риском возникновения пожара, а также предложены приближенные и точные методы решения этих задач. Работа построена следующим образом. Сначала в параграфе 2 вводятся необходимые понятия и описываются модели вероятности возникновения пожара, затем в параграфе 3 ставятся задачи управления вероятностью возникновения пожара и описываются методы их решения. В параграфе 4 приведены численные примеры решения задач управления.

## 2 Вероятность возникновения пожара

### 2.1 Общие соображения

Пусть  $(\Omega, \mathcal{A}, P)$  — вероятностное пространство. Для событий  $x, y \in \mathcal{A}$  обозначим

$$c(x, y) = P(x \cap y) - P(x)P(y)$$

ковариацию этих событий. Далее, обозначим  $F \in \mathcal{A}$  событие возникновения пожара, а  $\mathcal{X} \subseteq \mathcal{A}$  — множество событий, изображающих противопожарные мероприятия. Точнее, наступление события  $x \in \mathcal{X}$  будем интерпретировать, как осуществление некоторого противопожарного мероприятия.

От противопожарных мероприятий  $x \in \mathcal{X}$  естественно ожидать выполнения соотношений  $P(F|x) < P(F)$ , что эквивалентно отрицательности ковариации

событий  $F, x$ . Поэтому будем считать выполненным предположение

$$c(F, x) < 0, \quad x \in \mathcal{X}. \quad (1)$$

Далее, для подмножества мероприятий  $X \in \mathcal{X}$  обозначим

$$f(X) = P\left(F \mid \bigcap_{x \in X} x \bigcap_{x \in X^c} x^c\right) \quad (2)$$

вероятность возникновения пожара при условии, что выполнены мероприятия из подмножества  $X$ . Здесь  $x^c = \Omega \setminus x$  и  $X^c = \mathcal{X} \setminus X$ .

Поскольку мощность множества возможных противопожарных мероприятий  $\mathcal{X}$  обычно исчисляется сотнями, непараметрическое оценивание функции  $f$ , заданной на множестве  $2^{\mathcal{X}}$ , не представляется возможным. Поэтому будем искать функцию  $f$  в более или менее узких параметрических классах функций. Предварительно сделаем некоторые естественные предположения общего характера.

Функцию  $f$  будем предполагать монотонной по включению:

$$X \subseteq Y \subseteq \mathcal{X} \Rightarrow f(X) \geq f(Y). \quad (3)$$

Содержательно это означает, что расширение списка выполненных противопожарных мероприятий не повышает вероятность возникновения пожара.

Из (3) вытекает, что максимальное значение вероятности возникновения пожара соответствует пустому множеству мероприятий

$$q_M = \max_{X \subseteq \mathcal{X}} f(X) = f(\emptyset),$$

а минимальное значение этой вероятности — полному комплексу противопожарных мероприятий

$$q_m = \min_{X \subseteq \mathcal{X}} f(X) = f(\mathcal{X}).$$

Значения  $q_m, q_M$  также будем считать заданными.

### 2.2 Вероятность возникновения пожара; частные случаи

Мы будем искать функцию  $f$  в виде

$$f(X) = g(v(X)), \quad X \subseteq \mathcal{X}, \quad (4)$$

где  $g$  - невозрастающая функция эффективности  $v$  комплекса мероприятий:

$$g : [0, 1] \rightarrow [q_m, q_M],$$

$$g(u) > g(v), q_m \leq u < v \leq q_M$$

и

$$g(0) = q_M, g(1) = q_m. \quad (5)$$

Вид функции эффективности  $v : 2^{\mathcal{X}} \rightarrow [0, 1]$  уточним позже.

Рассмотрим некоторые параметрические формы модели зависимости  $g$  вероятности пожара от эффективности мероприятий  $v$ .

1. Линейная модель  $g(v) = a + bv$
2. Показательная модель  $g(v) = a + b \exp(cv)$

Линейная модель зависит от двух параметров, которые однозначно определяются граничными условиями (5). Она имеет вид

$$\begin{aligned} g(v) &= q_M + (q_m - q_M)v \\ &= q_m v + q_M(1 - v). \end{aligned} \quad (6)$$

В показательном классе имеется три параметра  $a, b, c$ , которые из условий (5) не могут быть определены однозначно. Некоторая свобода будет использована нами для вариации модели.

Показательная модель, с учетом граничных условий (5), имеет вид

$$g_a(v) = a + (q_M - a) \left( \frac{q_m - a}{q_M - a} \right)^v. \quad (7)$$

Эта модель корректна при  $a < q_m$ , и

$$\lim_{v \rightarrow \infty} g_a(v) = a.$$

Кроме того,

$$\lim_{a \rightarrow -\infty} g_a(v) = g(v), v \in [0, 1],$$

то есть, показательная модель (7) при  $a \rightarrow -\infty$  сходится к линейной модели (6), причем, сходимость равномерна на отрезке  $[0, 1]$ . В частном случае  $a = 0$  получаем

$$g_0(v) = q_M \left( \frac{q_m}{q_M} \right)^v. \quad (8)$$

Графики функций  $g$  и  $g_a$  при  $a = -0.00001$  и  $a = 0$  приведены на рис. 1.

Теперь введем функцию эффективности комплекса мероприятий  $v : 2^{\mathcal{X}} \rightarrow [0, 1]$ . Будем считать ее аддитивной, так что

$$v(X) = \sum_{x \in X} v(x),$$

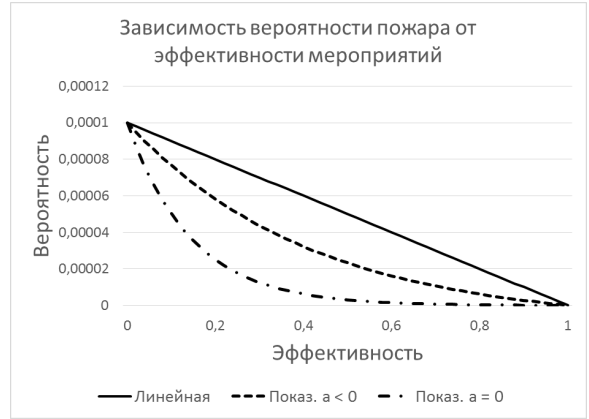


Рис. 1: Зависимость вероятности пожара от эффективности мероприятий  $v$ ; линейная (6) и показательная (7) модели с  $a = -0.00001$  и  $a = 0$

где для простоты используется сокращенное обозначение  $v(x) = v(\{x\})$ ,  $x \in \mathcal{X}$ , причем  $v(\emptyset) = 0$  и  $v(\{\mathcal{X}\}) = 1$ . Для  $x \in \mathcal{X}$  величину  $v(x)$  можно интерпретировать, как эффективность мероприятия  $x$ . Ее можно задать экспертным способом, а можно оценить по наблюдениям за возникновением пожаров на реальных объектах. Теперь зависимость вероятности возникновения пожара от эффективности комплекса  $X$  противопожарных мероприятий можно описать линейной

$$f(X) = q_M + (q_m - q_M)v(X) \quad (9)$$

$$= q_m v(X) + q_M(1 - v(X)). \quad (10)$$

и показательной

$$f_a(X) = a + (q_M - a) \left( \frac{q_m - a}{q_M - a} \right)^{v(X)} \quad (11)$$

моделями.

Показательная модель (11) представляет собой частный случай гиббсовской модели [Воробьев].

### 2.3 Затраты на проведение мероприятий

Противопожарные мероприятия требуют для своего осуществления некоторых затрат, что естественно учитывать при планировании комплекса противопожарных мероприятий. Обозначим  $u_x$  затраты на осуществление мероприятия  $x \in \mathcal{X}$ , и будем предполагать аддитивность затрат в составе комплекса, то есть затраты  $U_X$  на осуществление комплекса мероприятий  $X \in \mathcal{X}$  выражаются в виде

$$U_X = \sum_{x \in X} u_x.$$

## 3 Многокритериальное управление

При управлении пожарной безопасностью, целевыми могут оказаться многие показатели, в частности,

эффективность противопожарных мероприятий, выражающаяся в снижении условной вероятности возникновения пожара, а также затраты на осуществление мероприятий. В настоящем параграфе мы рассмотрим некоторые постановки задач с использованием стандартных способов сведения многокритериальных задач к однокритериальным задачам с ограничениями.

### 3.1 Максимизация эффекта

Предположим, что осуществлен некоторый набор противопожарных мероприятий  $Y \subseteq \mathcal{X}$ , и имеется бюджет  $Z$  для осуществления дополнительных мероприятий. Рассмотрим задачу выбора дополнительного мероприятия  $x \in Y^c$ , максимизирующего изменение вероятности возникновения пожара

$$\begin{cases} f(Y) - f(Y + \{x\}) \rightarrow \max_{x \in Y^c} \\ u_x \leq Z \end{cases}.$$

Ввиду (4), монотонности функции  $g$  и того факта, что  $Y$  фиксировано, эта задача эквивалентна следующей:

$$\begin{cases} v(Y + \{x\}) \rightarrow \max_{x \in Y^c} \\ u_x \leq Z \end{cases}. \quad (12)$$

Ввиду свойства аддитивности функции  $v$  такая задача легко преобразуется в форму

$$\begin{cases} v(x) \rightarrow \max_{x \in Y^c} \\ u_x \leq Z \end{cases}. \quad (13)$$

Решение этой задачи сводится к элементарному перебору элементов  $Y^c$ .

Более общая постановка заключается в добавлении целого комплекса мероприятий  $X \subseteq Y^c$  в рамках бюджета:

$$\begin{cases} f(Y) - f(Y + X) \rightarrow \max_{X \subseteq Y^c} \\ U_X \leq Z \end{cases}.$$

Последняя задача, в свою очередь, эквивалентна следующей:

$$\begin{cases} v(Y + X) \rightarrow \max_{X \subseteq Y^c} \\ U_X \leq Z \end{cases} \quad (14)$$

и, далее, ввиду аддитивности  $v$ ,

$$\begin{cases} v(X) \rightarrow \max_{X \subseteq Y^c} \\ U_X \leq Z \end{cases} \quad (15)$$

Задача (15) есть, по существу, задача о рюкзаке [4], ее решение перебором всех подмножеств  $X \subseteq Y^c$  по-

требует экспоненциальных временных затрат. Приближенное решение задачи (15) может быть получено последовательным применением решения простой задачи (13). Назовем такой алгоритм последовательным. Разновидность последовательного алгоритма, в котором мероприятия рассматриваются в порядке убывания эффективности, можно назвать "жадным". Оба алгоритма дают лишь приближенное решение задачи. Точное решение в случае целочисленных значений функций  $u, v$  дает метод динамического программирования (МДП). При относительно небольшом количестве различных значений метод динамического программирования дает выигрыш в эффективности по сравнению с методом полного перебора всех подмножеств. В случае вещественных значений  $u, v$  метод полного перебора является, по-видимому, единственным точным методом решения задачи (15).

### 3.2 Минимизация затрат

Рассмотрим в некотором смысле двойственную постановку задачи: минимизация затрат при условии достижения заданного уровня  $W$  эффективности мероприятий

$$\begin{cases} U_X \rightarrow \min_{X \subseteq Y^c} \\ v(X) \geq W \end{cases}. \quad (16)$$

Переходя к дополнению  $X$  в  $Y^c$ , то есть, к подмножеству  $T = Y^c \setminus X$ , можно переформулировать задачу (16) в форме

$$\begin{cases} U_T \rightarrow \min_{T \subseteq Y^c} \\ v(T) \leq W \end{cases}. \quad (17)$$

Для решения последней задачи уже напрямую применимы те же методы, что и в случае задачи (15).

## 4 Численные примеры

Рассмотрим примеры решения задач управления тремя методами: последовательным, жадным и методом динамического программирования. Всюду без ограничения общности будем считать, что  $Y = \emptyset$ . Мероприятия в примерах будем обозначать номерами.

**Пример 1.** Рассмотрим совокупность из пяти мероприятий, обладающих эффективностью  $v$  и требующих для своего осуществления затрат  $u$  в соответствии с табл. 1.

Решение задачи (15) при  $Z = 13$  тремя методами приведено в табл. 2.

Как видим, жадный алгоритм в данном примере нашел точное решение задачи. Вообще говоря, такой эффект достигается далеко не всегда, в следующем примере мы приллюстрируем это явление на более сложных задачах.

Номер	Затраты $u$	Эффективность $v$
1	3	1
2	4	6
3	5	4
4	8	7
5	9	6

Таблица 1: Характеристики мероприятий в примере 1.

Метод	Состав	Затраты $u$	Эфф. $v$
Посл.	{1,2,3}	12	11
Жадный	{2,4}	12	13
МДП	{2,4}	12	13

Таблица 2: Решения в примере 1.

### Пример 2.

Теперь рассмотрим решение серии однотипных задач. В каждой из задач серии затраты  $u$  и эффективность  $v$  представляют собой независимые реализации целочисленной случайной величины с равномерным распределением на сетке  $1, 2, \dots, 50$ . В таблице 3 приведены средние значения отношения эффективностей в решении задач приближенными и точным (МДП) методами, при различных ограничениях  $Z$  на имеющиеся ресурсы. Точнее, обозначим  $N$  количество задач в серии (в примере мы используем серию из  $N = 10000$  задач). Обозначим  $X_i$  множество добавляемых мероприятий в решении  $i$ -ой задачи методом динамического программирования, а  $X_{ia}$  - то же множество в решении задачи приближенным методом,  $i = 1, \dots, N$ . При этом относительная эффективность решения  $i$ -ой задачи приближенным методом выражается отношением

$$r_i = \frac{v(X_{ia})}{v(X_i)}, \quad i = 1, \dots, N,$$

а средняя относительная эффективность вычисляется в виде

$$r = \frac{1}{N} \sum_{i=1}^N r_i.$$

В табл. 4 приведена средняя мощность множества добавляемых мероприятий  $X$  в решении каждым из методов

$$m = \frac{1}{N} \sum_{i=1}^N |X_i|, \quad m_a = \frac{1}{N} \sum_{i=1}^N |X_{ia}|.$$

Ресурсы	Приблизж. метод	
	Посл.	Жадный
13	0.445	0.717
130	0.308	0.545
1300	0.656	0.945

Таблица 3: Относительная эффективность решений в примере 2.

Ресурсы	Мощность $X_a$		Мощность $X$ МДП
	Посл.	Жадный	
13	3.03	3.05	5.17
130	7.64	7.63	18.70
1300	53.29	53.25	63.12

Таблица 4: Средняя мощность решений в примере 2.

## 5 Заключение

В работе рассмотрены некоторые задачи управления риском возникновения пожара посредством выбора комплекса противопожарных средств в условиях ограниченных ресурсов. Предложены приближенные и точные методы решения этих задач, проведено сравнение методов, приведены численные иллюстрации.

Проведенные эксперименты позволяют заключить, что метод динамического программирования, дающий точное решение целочисленных задач управления, обеспечивает решение за приемлемое время, поэтому его можно использовать для решения практических задач посредством техники квантования.

## Благодарности

Автор выражает свою искреннюю признательность Олегу Юрьевичу Воробьеву и Сергею Петровичу Амелчугову за многочисленные обсуждения проблематики управления пожарной безопасностью.

## Список литературы

- [1] А.А. Новоселов. *Управление пожарным риском в естественных условиях и на технических объектах*, volume 1. 2012.
- [2] О.Ю. Воробьев. Применимая эвентология безопасности: неокончательные итоги. *Труды XII Международной конференции по финансово-актуарной математике и эвентологии безопасности*, 1:105–115, 2013.
- [3] А.А. Новоселов. *Математическое моделирование финансового риска: теория измерения*. Наука, Новосибирск, 2001.
- [4] Х. Таха. *Введение в исследование операций*. Вильямс, М., 2005.



## Алгоритм тематической классификации научных текстов

**Олейников Борис Васильевич**

Институт математики  
Сибирский федеральный университет  
Красноярск  
oleynik48@mail.ru

**Шалабай Андрей Игоревич**

Институт математики и фундаментальной информатики  
Сибирский федеральный университет  
Красноярск  
andrsh@gmail.com

**Аннотация.** В настоящей статье описан оригинальный алгоритм автоматической классификации научных текстов и проведен анализ его точности. Основными его преимуществами являются скорость работы при хранении данных в SQL базе данных.

**Ключевые слова.** тематическая классификация текстов, УДК, алгоритм, SQL

### 261 Введение

В [1],[2] описана концепция и разработка библиотечной Grid-системы, предназначенной для каталогизации Интернет-ресурсов в частично автоматическом режиме и объединения этого каталога с ресурсами традиционных библиотек.

С целью повышения релевантности поисковых запросов необходимо внедрять тематический поиск — подбор литературы строго определенной области знаний. Основной проблемой автоматизированного создания тематических коллекций является необходимость классификации огромного числа документов: если большинство бумажных книг классифицируется в процессе издания, то электронные документы, в силу свободы их распространения, чаще всего не имеют никаких формальных указаний на их тематику.

Таким образом, для полноценной реализации тематического поиска необходимо решить задачу классификации текстов, являющуюся одной из задач общего информационного поиска и заключающуюся в отношении документа к одной из нескольких категорий на основании содержания документа.

Существует три основных подхода к классификации:

1. Ручная. При поступлении новых книг в библиотеку, сотрудники самостоятельно классифицируют каждый документ.
2. Автоматическая, с использованием набора заранее созданных правил. Позволяет довольно точно классифицировать большинство текстов, но требует создания тысяч правил и их дальнейшего изменения грамотным специалистом.

3. Автоматическая, с использованием обучающих алгоритмов. Для реализации необходимо создать обучающую выборку: вручную классифицировать относительно небольшое число текстов и использовать один из алгоритмов классификации. Данный подход, как правило, является менее точным и менее быстрым в работе, чем предыдущий. Его главным преимуществом является минимизация человеческих ресурсов для обслуживания системы классификации.

Для дальнейшего практического использования важнейшим требованием к классификатору является скорость его работы, даже в ущерб точности: процесс классификации должен происходить в реальном времени при добавлении новых документов в электронную библиотеку.

### 262 Алгоритм классификации научных текстов

Для реализации возможности тематического поиска научных текстов можно значительно упростить задачу классификации. Предположим, что каждый научный текст содержит определенные термины, по которым можно сделать вывод о принадлежности информации к определенной тематике.

Для создания обучающей выборки, в каждый класс необходимо включать только термы  $t_k \in (X^k \setminus M)$ , где  $X^k$  — множество слов документа, включаемого в выборку,  $M$  — множество слов общей лексики, которые не являются научными терминами. На практике возможно включение в  $M$  всех слов, присутствующих в нескольких произведениях классической художественной литературы (в которых практически отсутствуют научные термины).

Для первоначального формирования обучающей выборки достаточно выбрать несколько научных текстов разной тематики и присвоить им код (например, УДК или ГРНТИ). В дальнейшем, система будет обучаться двумя способами:

1. При автоматической классификации система добавляет встречающиеся новые термины в определенный класс.
2. В случае, если подходящий класс не найден,

пользователю будет предложено самостоятельно выбрать код для документа из справочника, после чего произойдет создание нового класса и размещение в нем научных терминов классифицированного текста.

Основным требованием для алгоритма классификации является скорость работы, для его выполнения возможно даже понижение точности распознавания тематики. Это связано с тем, что реализация тематического поиска подразумевает работу с большим количеством объемных текстов (сотни тысяч документов, в каждом из которых может содержаться несколько тысяч терминов).

Исходя из вышеизложенного, предлагается классифицировать документ следующим образом:

1. При добавлении  $k$ -го документа неизвестной тематики сформировать для него множество  $X^k \setminus M$ .
2. Для каждого известного класса  $C_i$  вычислить  $R_i^k = |C_i \cap (D^k \setminus M)|$  - количество слов научной тематики  $k$ -го документа, имеющего отношения к классу  $C_i$
3. Принадлежность документа к классу  $C_i$  будет определяться некоторой неубывающей функцией  $f_i^k = f(R_i^k, |C_i|)$ , которую можно интерпретировать, как индекс принадлежности  $k$ -го документа  $i$ -му классу ( $|C_i|$  - количество уникальных слов научной тематики в  $C_i$ ).
4. Классы, с наибольшей величиной  $f_i^k$  и будут наиболее вероятными для  $k$ -го документа.
5. После принятия окончательного решения о классификации необходимо добавить  $(X^k \setminus M) \setminus C_i$  в соответствующее множество  $C_i$

Исследуем необходимость введения функции  $f_i^k$ . Пусть  $D^1$  и  $D^2$  - некоторые документы, а  $C_1$  и  $C_2$  - некоторые классы. Рассмотрим два предельных случая при классификации документа только по величине  $R_i^k$ :

1. Допустим, что  $D^1$  и  $D^2$  принадлежат классу  $C_1$ , но  $|D^1 \cap C_1| > |D^2|$  и  $D^1 \cap C_1 = C_1$ . В этом случае  $R_1^1 > R_1^2$ , а значит вероятность принадлежности документа  $D^1$  классу  $C_1$  выше. Однако, с другой стороны, так как всё множество слов документа  $D^2$  содержится в классе  $C_1$ , вероятность принадлежности  $D^2$  к  $C_1$  должна быть максимальной.
2. Допустим, что  $D^1$  принадлежит классу  $C_1$ , но содержит в себе некоторое количество слов (терминов) из класса  $C_2$ , то есть в документе присутствует множество терминов (слов)  $A \subset (C_1 \cap C_2)$ , соответствующее нескольким классам. В случае,

если  $|C_2 \cap A| > |C_1|$  - документ будет классифицирован неправильно.

Поэтому, экспериментально подобрана функция

$$f_i^k = \frac{R_i^k}{\min_j R_j^k} + \frac{R_i^k \min_j |C_j|}{C_i \min_j R_j^k},$$

служащая для того, чтобы исключить ошибку классификации в случае, когда количество слов в различных документах и классах отличается в десятки раз: при классификации объемного документа только по  $R_i^k$  преимуществом будут пользоваться классы с наибольшим количеством слов. Кроме того, при формировании и хранении множества  $X^k \setminus M$  становится возможным реализовать не только тематический, но и полнотекстовый поиск информации.

Для реализации алгоритма предлагается использовать SQL базу данных, которая позволит:

1. Хранить не только ключевые слова, но и дополнительные данные о литературе. Таким образом в одной базе можно сосредоточить всю информацию, необходимую для создания полноценной информационной системы.
2. Производить с помощью штатных средств SQL быстрые соединения таблиц и при минимальных трудовых затратах получать в результатах поискового запроса все необходимые для классификации данные.
3. Легко масштабировать систему хранения данных, выбирая СУБД в соответствии с реальными потребностями (зависит от числа классов, документов, рабочих мест пользователей): при реальном использовании словарь слов всех классов научной тематики может состоять из нескольких сотен тысяч записей (для литературы на разных языках), а количество книг может быть чрезвычайно большим.

Приведем простой пример практической реализации представленного алгоритма. Рассматривается SQL база данных:

1. Создается таблица Dict\_words (code, name), содержащая множество всех слов научной тематики.
2. Создается таблица Dict\_docs (code, name, author, class...), содержащая библиографическую информацию обо всех документах.
3. Создается таблица Link\_words\_docs (code\_dict\_words, code\_dict\_docs), содержащая записи, которые определяют принадлежность слов из Dict\_words документу из Dict\_docs.

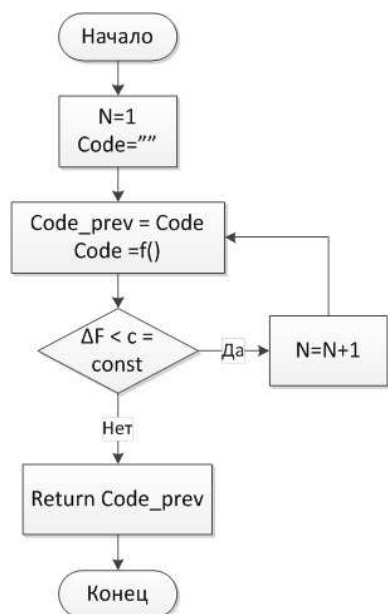


Рис. 1: Схема иерархической классификации

Тогда запрос по конкретному k-му документу, определяющий  $R_i^k$  будет выглядеть следующим образом:

```
SELECT count(*) FROM
```

```
(SELECT * FROM LINK_WORDS_DOCS
lwd1 INNER JOIN DICT_DOCS dd ON
lwd1.code_dict_docs = dd.code WHERE dd.class = Ci) t1 INNER JOIN
```

```
(SELECT * FROM LINK_WORDS_DOCS
lwd2 WHERE lwd2.code_dict_docs = d) t2 ON
t1.code_dict_words = t2.code_dict_words;
```

Главное преимущество рассматриваемого алгоритма заключается в том, что при подобной реализации возможно исключить использование более медленных соединений типа OUTER JOIN и тем самым значительно повысить скорость выполнения запроса. На стандартном офисном компьютере время выполнения такого запроса по базе данных с использованием таблицы LINK\_WORDS\_DOCS, содержащей 42 тысячи записей, составляет в среднем 0,045 секунды.

Пусть  $f_*^k = \max_j f_j^k$ ,  $f_{**}^k = \max_{j, f_j^k \neq f_*^k} f_j^k$ ,  $f_*^k - f_{**}^k = \Delta f$ .

Очевидно, что чем больше значение  $\Delta f$ , тем выше вероятность правильности классификации. Следовательно, если при некотором  $\Delta f$  считать, что классификация не удалась, то за счет сокращения общего числа классифицированных документов возможно повысить ее точность. Для классификации текстов по кодам УДК предлагается воспользоваться следующей схемой иерархической классификации, показанной на Рис. 1, где N - текущий уровень УДК, Code - текущий код УДК, определенный классификатором, f() - функция автоматической классификации. Следует пояснить, что классификация происхо-

Уровень УДК	Оценка по $R_i^k$	Оценка по $f_i^k$
1	84%	93,5%
2	67,5%	76%
3	42%	52%

Таблица 1: Точность классификации представленного алгоритма.

дит только среди классов (кодов УДК), являющихся дочерними для кода, определенного на предыдущем этапе.

## 263 Оценка точности классификации

Для оценки точности представленного алгоритма был проведен следующий эксперимент: для классификации использованы 200 научных статей из архивов электронных научных журналов (Табл. 1). В качестве множества M использовался набор слов из небольших отрывков произведений В.П. Астафьева, в результате количество слов в документах сократилось в среднем на 15%. Если при  $\Delta f < 0,2$  полагать, что документ не был классифицирован, то по первому уровню УДК точность классификации составит 100 %, но будет классифицировано всего 79% документов. Подбирать требуемое минимальное значение должен специалист, ответственный за классификацию литературы в соответствии с выбранным приоритетом (качества, либо количества).

Для сравнения, согласно [3], точность известного метода PrTFIDF (в его иерархическом варианте) для тематической классификации составляет 72% для второго уровня иерархии, и 66% для конечного уровня иерархии. Поэтому, можно заключить, что представленный оригинальный алгоритм целесообразно использовать для классификации по первому, либо второму уровню УДК. В случаях использования подобной SQL базы данных, представленный алгоритм значительно выигрывает в скорости, но при другой организации системы хранения данных сопоставление результатов требует дополнительных исследований.

## Список литературы

- [1] Олейников Б.В. Шалабай А.И. К вопросу размещения открытых Интернет-ресурсов в классических библиотечных системах. *XI международная ФАМЭБ'2012 конференция*, С.:293–296, 2012.
- [2] Олейников Б.В. Шалабай А.И. Программное обеспечение для построения Грид-системы консолидации ресурсов традиционных библиотек и Интернет-ресурсов. *XVI Международная ЭМ конференция по эвентологической математике и смежным вопросам*, 2012.
- [3] Шелестов А.А. Дунаев Е.В. Автоматическая рубрикация web-страниц в интернет-каталоге с иерархической структурой. *Интернет-математика 2005*, 2005.

## Восприятие рекламы в малых населённых пунктах на примере сотового оператора *Билайн*

Ольга Вячеславовна Пазенко

СФУ ИУБПиЭ, Красноярский филиал ОАО ВымпелКом  
opazenko@yandex.ru

Михаил Георгиевич Садовский

Институт вычислительного моделирования СО РАН  
Красноярск  
msad@icm.krasn.ru

**Аннотация.** Изучалось сравнительное восприятие рекламных имиджей в крупных городах и малых населённых пунктах на примере рекламы сотового оператора *Билайн*.

**Ключевые слова.** Частота, энтропия, условная энтропия, оптимальность.

### 1 Введение

В настоящее время рынок сотовой связи развивается в направлении малых городов и сельской местности, где операторы сотовой связи, используя все возможные средства пытаются привлечь потребителей. Однако потребности жителей крупных городов и потребности жителей малых городов разнятся в зависимости от географического местоположения; социально-экономических факторов; размера населённого пункта и т.д. Все это влияет на восприятие рекламы.

Уровень проникновения на рынке сотовой связи в настоящее время оценивается на уровне 161 %, общее число номинальных абонентов сотовой связи (по числу проданных SIM-карт) составило 229,8 млн. чел. (согласно данным АСaМ Consulting за III квартал 2012г.). В настоящее время проникновение увеличивается за счет молодежи и лиц пожилого возраста, а рост доходов идет за счет роста выручки на абонента. Средняя выручка на абонента, в свою очередь, растет за счет более активного использования мобильного Интернета и других сервисов, предоставляемых оператором, при этом рост голосового трафика и выручка от него падает. Именно поэтому компании вводят все больше и больше дополнительных услуг, а также весьма активно развивают услуги мобильного Интернета.

Всё вышесказанное приводит к тому, что большинство рекламных материалов от компании к компании, от тарифа к тарифу выполняются в определенном фирменном стиле. Это значит не только использование логотипа и макетирование согласно Brand Book, но и использование постоянно повторяющихся дизайнерских приемов, решений и образов: потребитель должен отличать рекламу одного опера-

тора сотовой связи от другого. На первый взгляд, это не сложно сделать целевой аудитории, т.к. предполагается, что абоненты сотовых операторов четко сегментированы по социально-экономическому статусу, ценностям, жизненным установкам и т.д. Однако почти полный охват жителей различных населённых пунктов приводит к размыванию границ целевой аудитории, что существенно осложняет задачу построения эффективного рекламного сообщения.

Существенная разница в укладе жизни, социально-экономическом и географическом положении, развитии городов и темпа жизни определяет восприятие действительности окружающего мира. В результате жители малых городов могут по другому воспринимать рекламные сообщения по сравнению с жителями крупных городов. Реклама, хорошо работающая в крупных городах, может иметь те же шансы на успех в городах других типов, однако гарантий этого нет. Потребности и отношения жителей небольших городов могут сильно отличаться от потребностей и отношений жителей крупных городов. В целом жители малых городов всё ещё активнее потребляют рекламу и уделяют ей больше внимания, по сравнению с жителями крупных городов. Для них образцами для подражания являются знаменитости, а в рекламе они хотят видеть не обычные сцены из повседневной жизни и отзывы рядовых потребителей, а образцы для подражания. Однако даже географическое местоположение населённого пункта накладывает свой отпечаток на восприятие рекламы.

Все выше сказанное позволило сформулировать гипотезу исследования: восприятие рекламы в крупных городах отличается от восприятия рекламы в малых городах, поэтому рекламные имиджи и считываемая информация могут по разному восприниматься в зависимости от величины населённого пункта, а эффективность рекламы требует её локализации. Цель настоящего исследования — выявить и описать различия в восприятии рекламы потребителями услуг сотового оператора, проживающими в Красноярске и в малых городах (сёлах) Красноярского края. Задачи исследования:

- определить метод исследования;
- разработать стимульный материал;
- провести исследования потребителей услуг со-



товых операторов связи;

- описать восприятие рекламных имиджей потребителя услуг сотовой связи.

## 2 Материалы и методы

Для тестирования восприятия рекламы был использован Модифицированный ТАТ (Тематический Апперцепционный Тест). Стимульный материал разработан на основе имиджей рекламных компаний сотового оператора "Билайн". Исследование проводилось в малых населенных пунктах Красноярского края среди пользователей услуг сотовой связи.

Наиболее подходящим методом для выявления внутренних, глубинных причин восприятия человеком материала являются проективные методики. Они позволяют раскрыть субъективные причины поведения человека. Проективные методики позволяют «обмануть» сознание человека и обойти психологическую защиту либо стереотипы поведения. С их помощью можно узнать реальное отношение потребителей к исследуемому предмету. Использование проективных методик в исследовании в рекламе позволяет исследователем ответить на такие вопросы как:

- Почему потребителям нравятся или не нравятся те или иные товары?
- Что влияет на их решение о выборе того или иного продукта?
- Какой имидж марки сложился в глазах потребителей?
- Какие моменты в рекламе могли бы оказать влияние на желание потребителей купить продукт?
- Какие герои/персонажи могут быть использованы в рекламе?

Для проведения данного исследования был выбран Модифицированный ТАТ (Тематический Апперцепционный Тест). Данная методика при помощи серии визуальных картинок позволяет выявить глубинные мотивы поведения потребителя, отношение к имиджу марки, на сколько потребитель соотносит себя с ситуацией потребления товара или услуги, или то, как воспринимает потребитель по отношению к себе людей, представленных в рекламных имиджах, воспринимает ли он их как равных себе или как людей из другого общества, страны и т. п.

Для проведения исследования был разработан стимульный материал на основе имиджей, используемых в рекламных компаниях сотового оператора *Билайн* при продвижении различных тарифных планов. Имиджи были взяты за последние половину 2012 г. В имиджах были убраны все фирменные элементы, позволяющие идентифицировать компанию, но в тоже время реципиенты могли запомнить рекламу и все же определить её как относящуюся к ре-



### Реклама нового оператора сотовой связи

Рис. 1: Пример стимульного материала.

кламе сотового оператора *Билайн*, что говорило бы о запоминаемости марки и ее рекламных компаний. Чтобы не исказить восприятие стимульного материала и не накладывать на восприятие позиционирование компании при проведении исследования имиджи представлялись как разработка рекламных макетов нового сотового оператора, которых готовится выйти на рынок Красноярского края.

## 3 Результаты и обсуждение

Опрос проводился среди жителей малых населенных пунктов где осуществляется продажа услуг и товаров компании *Билайн*. В опросе участвовали люди, которые пользуются сотовой связью, как настоящие, так и потенциальные потребители услуг компании *Билайн*.

### Реклама нового оператора сотовой связи



Рис. 2: Пример стимульного материала.

При проведении исследования реципиентам в определенной последовательности предлагается взглянуть на изображение и описать где, по его мнению, происходит действие, изображенное на картинке; к како-



му социальному статусу относятся люди, какой образ жизни ведут, род деятельности и т.д. Если реципиент узнает, что данная реклама относится к рекламе сотового оператора *Билайн* необходимо зафиксировать данный факт.



## Реклама нового оператора сотовой связи

Рис. 3: Пример стимульного материала.

Предварительные результаты показывают, что при восприятии имиджей рекламы потребители не отождествляют себя с персонажами как с равными себе ни по месту проживания, ни по социально-экономическому статусу. В основном персонажи ассоциируются с людьми живущими либо за границей, либо в европейской части России. Персонажи на имиджах работают в финансово-привлекательных отраслях экономики, такие как банковский сектор, Интернет бизнес или имеют собственное дело.

Также полученные результаты исследования сопоставимы с результатами полученными в г Красноярске. Что возможно говорит о том, что жители малых населенных пунктов больше считают себя жителями региональных центров, нежели чем жители Красноярска считают себя жителями столицы.

# Копулы сглаженных эмпирических распределений при наличии связок (совпадений) и их применение в имитационном моделировании

**Василий Васильевич Поддубный**

Национальный исследовательский Томский государственный университет  
Кафедра прикладной информатики  
Томск  
vvpoddubny@gmail.com

**Александр Сергеевич Пехтерев**

Национальный исследовательский Томский государственный университет  
Кафедра прикладной информатики  
Томск  
alexandr.pehterev@gmail.com

**Аннотация.** Рассмотрен подход к имитационному моделированию частотных признаков объектов классификации и распознавания, основанный на использовании копула-функций сглаженных эмпирических распределений обучающих выборок в условиях связок (совпадений значений) в обучающей выборке. На основе кусочно-линейного сглаживания маргинальных эмпирических функций распределения при наличии связок предложен алгоритм имитационного моделирования векторов, близких по статистическим свойствам (включая ранговые корреляции) к векторам обучающей выборки. Из условия равенства модельной эмпирической и гауссовой копул найдена многомерная функция распределения моделируемых векторов. В качестве примера проведено имитационное моделирование векторов со свойствами обучающей выборки относительных частот служебных слов в текстах русских прозаиков 19 века. С помощью непараметрических критериев согласия проведена проверка статистической адекватности законов распределения обучающей и модельной выборок. Это позволяет использовать имитационное моделирование для оценки качества алгоритмов классификации и распознавания в условиях отсутствия контрольных выборок реальных объектов.

**Ключевые слова.** Имитационное моделирование, обучающая выборка, связка, эмпирическое распределение, сглаживание, ранговая корреляция, копула, алгоритм моделирования, проверка адекватности.

## 1 Введение

В ряде приложений (квантитативная лингвистика, медицинская диагностика и др.) исследуемые объекты (тексты литературных произведений, пациенты клиник и т.д.) характеризуются частотными признаками – относительными частотами появления в текстах тех или иных слов (словосочетаний, буквосочетаний), относительными частотами проявления в анамнезе больных тех или иных симптомов болезней

и т.д. При этом объекты одного класса (определенного автора, жанра, типа патологии и т.п.) характеризуются своими более или менее устойчивыми распределениями относительных частот описывающих их признаков, тогда как объекты, принадлежащие разным классам, обычно значительно различаются по относительным частотам появления одних и тех же признаков. Относительно малый разброс значений относительных частот каждого из признаков в пределах одного класса и значительный их разброс между классами позволяет проводить классификацию объектов по выбранным частотным признакам.

Алгоритмы классификации (правила отнесения наблюдаемых объектов к тому или иному классу) обычно строятся так, чтобы объекты обучающей выборки, принадлежность которых к рассматриваемым классам точно известна, относились классификатором к своим классам с как можно меньшей вероятностью ошибки. Идеальный классификатор обеспечивает безошибочную классификацию объектов обучающей выборки.

А что произойдет, если предъявить классификатору объект известного класса, не участвовавший в обучении классификатора (не входивший в обучающую выборку)? Такой объект с определенной вероятностью будет правильно отнесен к своему классу. Но с некоторой вероятностью он может попасть в другой класс – произойдет ошибка классификации. Для достаточно надежной эмпирической оценки качества классификации необходимо предъявить классификатору достаточно большое число объектов разных классов (контрольную выборку достаточно большого объема). Однако на практике далеко не всегда может существовать такая выборка. Например, если обучающая выборка включает в себя все произведения определенного автора, то контрольную выборку сформировать просто не из чего. Выходом из этого затруднения может быть только имитационное моделирование контрольной выборки, обладающей (по возможности) теми же статистическими свойствами, что и обучающая выборка, представляющая данный класс объектов.

Общие особенности частотных признаков состоят в

том, что: 1) спектр значений относительных частот является подмножеством множества рациональных чисел из интервала  $[0, 1]$ ; 2) поскольку относительные частоты – рациональные числа, возможно совпадение их значений (связки); 3) в силу рациональности относительных частот их интегральная функция распределения не является непрерывной; 4) относительные частоты разных признаков в общем случае статистически зависимы, но не подчиняются многомерному (совместному) нормальному распределению.

Если функция распределения обучающей выборки неизвестна, информацию о ней можно извлечь только из самой выборки. Это можно сделать, построив по наблюдаемой обучающей выборке статистическую оценку многомерной функции распределения – многомерную эмпирическую функцию распределения. Из нее, используя те или иные методы сглаживания (например, ядерные оценки Парзена–Розенблатта [1, 2] для многомерной плотности распределения), можно построить сглаженную многомерную эмпирическую функцию распределения. Только на ее основе можно построить алгоритмы имитационного моделирования объектов, обладающих статистическими свойствами обучающей выборки.

Существует два подхода к имитационному моделированию многомерных объектов со статистически связанными компонентами. Один – на основе сглаженных условных и безусловных эмпирических функций распределения, построенных по обучающей выборке. Другой – на основе одномерных (частных, маргинальных) эмпирических функций распределения, тоже сглаженных, с использованием непараметрических статистических оценок корреляционной матрицы компонент моделируемого вектора (например, ранговых корреляционных матриц Спирмена, Кендалла [3, 4] или др.). Последний подход тесно связан с копула-функциями [5, 6, 7]. Первый подход в принципе позволяет извлечь из обучающей выборки наиболее полную информацию о статистический связи компонент многомерного случайного вектора, описывающего многомерный объект. Но для достаточно точного оценивания многомерной функции распределения требуется достаточно большой объем обучающей выборки. Кроме того, этот подход достаточно сложен в реализации. Второй подход, основанный на теории копула-функций, позволяет разделить информацию о маргинальных распределениях и информацию о корреляционных связях. Правда, при этом в общем случае может быть учтена только корреляционная (линейная) составляющая статистических связей между компонентами случайного вектора, описывающего объект. Зато алгоритм имитационного моделирования таких многомерных случайных векторов оказывается достаточно простым в реализации.

Рассмотрим оба подхода к имитационному моделированию статистически зависимых частотных призна-

ков, уделив особое внимание второму подходу. Этот подход рассмотрим в специальных условиях возможного наличия в обучающей выборке повторяющихся значений (связок). Особенно это касается нулевых значений относительных частот, свидетельствующих об отсутствии в ряде наблюдений некоторых признаков из исследуемого набора.

## 2 Постановка задачи

Пусть случайная величина  $\xi$  имеет непрерывную интегральную функцию распределения  $F(x)$ . Рассмотрим случайную величину  $\eta$ , являющуюся нелинейным преобразованием  $\eta = F(\xi)$  случайной величины  $\xi$ . Из теории вероятностей известно, что эта случайная величина имеет равномерное распределение  $U(y)$  в интервале  $[0, 1]$ . Действительно, если  $y = F(x)$ , то, в силу монотонности функций распределения,  $U(y) = P(\eta < y) = P(\xi < x) = F(x) = y$ , так что  $dU(y)/dy = 1$ , то есть плотность распределения величины  $\eta$  постоянна (равна 1) в интервале  $[0, 1]$ .

Пусть теперь  $\xi = (\xi_1, \xi_2, \dots, \xi_m)$  – многомерная (векторная) случайная величина со статистически зависимыми компонентами  $\{\xi_i, i = \overline{1, m}\}$ , маргинальные функции распределения которых есть  $\{F_i(x_i), i = \overline{1, m}\}$ . Тогда многомерная (векторная) случайная величина  $\eta = (\eta_1, \eta_2, \dots, \eta_m)$ , полученная покомпонентным нелинейным преобразованием  $\{\eta_i = F_i(\xi_i), i = \overline{1, m}\}$  многомерной случайной величины  $\xi$ , будет иметь многомерное в  $m$ -мерном гиперкубе  $[0, 1]^m$  распределение  $U(y_1, y_2, \dots, y_m)$ , называемое копулой [7]:

$$C(y_1, y_2, \dots, y_m) = F(F_1^{-1}(y_1), F_2^{-1}(y_2), \dots, F_m^{-1}(y_m)),$$

с равномерными маргинальными распределениями компонент. Очевидно (теорема Склара [8]),

$$C(F_1(x_1), F_2(x_2), \dots, F_m(x_m)) = F(x_1, x_2, \dots, x_m).$$

Если функция распределения  $F(x_1, x_2, \dots, x_m)$  неизвестна, ее можно оценить по обучающей матрице-выборке  $x^o$  размерности  $n \times m$ , полученной в результате  $n$  независимых наблюдений  $m$ -вектора  $\xi$ , так что  $x_{ij}^o$  – значение относительной частоты  $j$ -го признака в  $i$ -м наблюдении. Исходной оценкой функции распределения

$$F(x_1, x_2, \dots, x_m) = P(\xi_1 < x_1, \xi_2 < x_2, \dots, \xi_m < x_m)$$

является эмпирическая функция распределения  $F_n(x_1, x_2, \dots, x_m)$ . По определению,  $m$ -мерная эмпирическая функция распределения есть отношение числа точек наблюдения, попадающих в  $m$ -мерную область

$$B(x_1, x_2, \dots, x_m) = \{\xi_1 < x_1, \xi_2 < x_2, \dots, \xi_m < x_m\},$$

к общему числу  $n$  точек наблюдения:

$$F_n(x_1, x_2, \dots, x_m) = \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^m 1(x^o(i, j) < x_j), \quad (1)$$

где  $1(\cdot)$  – индикатор события, указанного в скобках. В частности, маргинальные безусловные  $F_n(x_j)$  и условные  $F_n(x_j|x_{j+1}, \dots, x_m)$  эмпирические функции распределения принимают вид:

$$F_n(x_j) = \frac{1}{n} \sum_{i=1}^n 1(x^o(i, j) < x_j), \quad j = \overline{1, m}, \quad (2)$$

$$F_n(x_j|x_{j+1}, \dots, x_m) = \frac{F_n(x_j, x_{j+1}, \dots, x_m)}{F_n(x_{j+1}, \dots, x_m)}, \quad j < m. \quad (3)$$

В соответствии с известной теоремой Гливленко–Кантелли [9], при  $n \rightarrow \infty$  эмпирическая функция распределения сходится почти наверное (п.н.) к генеральной функции распределения:

$$F_n(x_1, x_2, \dots, x_m) \xrightarrow[n \rightarrow \infty]{\text{п.н.}} F(x_1, x_2, \dots, x_m).$$

Однако построить копулу на основе эмпирической функции распределения невозможно, так как функция  $F_n^{-1}(y)$ , обратная к эмпирической функции распределения  $F_n(x)$ , не определена в областях постоянства функции  $F_n(x)$ . Поэтому для построения эмпирической копулы необходимо использовать сглаженные эмпирические оценки  $\tilde{F}_n(x)$  функции распределения  $F(x)$ . Сглаживание достигается, например, использованием непрерывных ядерных функций Парзена–Розенблатта, обеспечивающих непрерывность эмпирической функции распределения по всем переменным. На основе сглаженной непрерывной эмпирической функции распределения  $\tilde{F}_n(x_1, x_2, \dots, x_m)$  уже можно построить эмпирическую копулу. Простейшей является непрерывная кусочно-линейная аппроксимация эмпирической функции распределения  $\tilde{F}_n(x_1, x_2, \dots, x_m)$  по каждой переменной. Сглаженная эмпирическая функция распределения также сходится почти наверное к генеральной функции распределения. Поэтому и эмпирическая копула

$$\tilde{C}_n(y_1, \dots, y_m) = \tilde{F}_n(\tilde{F}_n^{-1}(y_1), \dots, \tilde{F}_n^{-1}(y_m)), \quad (4)$$

соответствующая сглаженному эмпирическому распределению, при  $n \rightarrow \infty$  сходится почти наверное к копуле  $C(y_1, \dots, y_m)$ , соответствующей генеральному распределению.

### 3 Алгоритмы имитационного моделирования

#### 3.1 Прямой метод

Если бы случайные величины  $\xi_1, \xi_2, \dots, \xi_m$  подчинялись сглаженному многомерному распределению

$\tilde{F}_n(x_1, x_2, \dots, x_m)$  со сглаженными безусловными и условными маргинальными распределениями

$$\tilde{F}_n(x_1), \tilde{F}_n(x_2|x_1), \tilde{F}_n(x_3|x_2, x_1), \dots, \tilde{F}_n(x_m|x_{m-1}, x_{m-2}, \dots, x_1), \quad (5)$$

их преобразования по формулам этих распределений  $\eta_1 = \tilde{F}_n(\xi_1)$ ,  $\eta_2 = \tilde{F}_n(\xi_2|\xi_1)$ ,  $\eta_3 = \tilde{F}_n(\xi_3|\xi_2, \xi_1)$ ,  $\dots$ ,  $\eta_m = \tilde{F}_n(\xi_m|\xi_{m-1}, \xi_{m-2}, \dots, \xi_1)$  дали бы  $m$ -вектор  $\eta$  равномерно распределенных в  $[0, 1]$  независимых случайных величин  $\eta_1, \eta_2, \dots, \eta_m$ . Следовательно, при обратном преобразовании независимые равномерно распределенные в  $[0, 1]$  случайные величины  $\eta_1, \eta_2, \dots, \eta_m$  дали бы  $m$ -вектор  $\xi$  зависимых случайных величин  $\xi_1, \xi_2, \dots, \xi_m$ , распределенных по закону (5). На этой основе в работах [10, 11, 12, 13, 14, 15, 16] и др. предложены и развиты различные формы алгоритма генерации зависимых случайных величин, подчиняющихся закону распределения такой же структуры, как (5). Этот алгоритм можно применить и к собственно эмпирическому закону (5). Для этого закона алгоритм можно записать следующим образом.

#### Алгоритм 1:

- 1) датчиком равномерно распределенных в  $[0, 1]$  случайных чисел сгенерировать независимые числа  $y_1, y_2, \dots, y_m$ ;
- 2) по  $y_1$  вычислить  $x_1 = \tilde{F}_n^{-1}(y_1)$ ;
- 3) по  $y_2$  и вычисленному  $x_1$  вычислить  $x_2 = \tilde{F}_n^{-1}(y_2|x_1)$ ;
- 4) по  $y_3$  и вычисленным  $x_1, x_2$  вычислить  $x_3 = \tilde{F}_n^{-1}(y_3|x_2, x_1)$ ;
- 5) продолжая этот процесс, вычислить, наконец,  $x_m = \tilde{F}_n^{-1}(y_m|x_{m-1}, x_{m-2}, \dots, x_1)$ .

Проблема состоит только в построении подходящим образом сглаженных эмпирических функций распределения  $\tilde{F}_n(x_1), \tilde{F}_n(x_1, x_2), \dots, \tilde{F}_n(x_1, x_2, \dots, x_m)$ , из которых по формулам вида (3) легко получить все маргинальные условные функции распределения (5), используемые алгоритмом. Однако, построение сглаженных многомерных эмпирических функций распределения представляется в вычислительном отношении достаточно сложной задачей.

#### 3.2 Метод копула-функции

Если бы случайные величины  $\xi_1, \xi_2, \dots, \xi_m$  подчинялись сглаженному многомерному распределению  $\tilde{F}_n(x_1, x_2, \dots, x_m)$  с безусловными маргинальными распределениями  $\tilde{F}_n(x_1), \tilde{F}_n(x_2), \dots, \tilde{F}_n(x_m)$ , то их преобразования по формулам этих функций распределений  $\eta_1 = \tilde{F}_n(\xi_1)$ ,  $\eta_2 = \tilde{F}_n(\xi_2)$ ,  $\dots$ ,  $\eta_m = \tilde{F}_n(\xi_m)$  дали бы  $m$ -вектор  $\eta$  статистически независимых случайных величин  $\eta_1, \eta_2, \dots, \eta_m$ , распределенных в  $[0, 1]^m$  по закону, выражаемому копулой (4). При обратном преобразовании эти зависимые случайные величины  $\eta_1, \eta_2, \dots, \eta_m$  дали бы



опять  $m$ -вектор  $\xi$  исходных зависимых случайных величин  $\xi_1, \xi_2, \dots, \xi_m$ . Однако, если обратное преобразование провести с использованием обратных маргинальных стандартных (с нулевым математическим ожиданием и единичной дисперсией) нормальных распределений  $\{\zeta_j = \Phi^{-1}(\eta_j), j = \overline{1, m}\}$ , получится гауссов  $m$ -вектор  $\zeta$  с зависимыми компонентами. Поскольку интегральные функции распределения – монотонно неубывающие функции, ранги одноименных компонент преобразуемых векторов  $\xi$ ,  $\eta$  и  $\zeta$  совпадают, так что матрицы ранговых коэффициентов корреляции этих векторов также совпадают. И хотя из нормальности частных (маргинальных) распределений не следует нормальность совместного распределения [17], построим алгоритм генерации случайных векторов  $\xi = (\xi_1, \dots, \xi_m)$  с такой же ранговой корреляционной матрицей и с такими же маргинальными функциями распределения, как у обучающей выборки, с использованием совместно-гауссовского распределения с той же ранговой корреляционной матрицей. Многомерная функция распределения  $\tilde{F}_{n(\text{мод})}(x_1, \dots, x_m)$  случайной величины, моделируемой таким образом, будет с некоторым приближением аппроксимировать неизвестную сглаженную эмпирическую функцию распределения  $\tilde{F}_n(x_1, \dots, x_m)$ , соответствующую обучающей выборке. Это дает возможность с определенной степенью приближения принять гауссову копулу

$$C_{\text{норм}}(y_1, \dots, y_m) = \Phi(\Phi^{-1}(y_1), \dots, \Phi^{-1}(y_m)), \quad (6)$$

точно совпадающую с “модельной” копулой

$$C_{\text{мод}}(y_1, \dots, y_m) = \tilde{F}_{n(\text{мод})}^{-1}(\tilde{F}_n^{-1}(y_1), \dots, \tilde{F}_n^{-1}(y_m)), \quad (7)$$

за сглаженную эмпирическую копулу (4):  $\tilde{C}_n(y_1, \dots, y_m) \approx C_{\text{норм}}(y_1, \dots, y_m) = C_{\text{мод}}(y_1, \dots, y_m)$ . На ее основе алгоритм имитационного моделирования случайных векторов  $x = (x_1, \dots, x_m)$  с маргинальными сглаженными эмпирическими функциями распределения  $\{\tilde{F}_n(x_j), j = \overline{1, m}\}$  и с эмпирической ранговой корреляционной матрицей Спирмена  $R_S$ , построенными по обучающей выборке, можно записать следующим образом.

#### Алгоритм 2:

- 1) стандартным датчиком нормально распределенных случайных чисел сгенерировать  $m$  независимых случайных чисел  $\{v_1, \dots, v_m\}$ ;
- 2) преобразовать вектор-столбец  $v$  с независимыми компонентами в вектор-столбец  $z = (z_1, \dots, z_m)^T$  с коррелированными компонентами по формуле:  $z = Av$ , где матрица коэффициентов преобразования  $A$  – нижняя треугольная матрица в разложении Холецкого  $R = AA^T$  (формально – результат матричной операции извлечения вещественного квадратного

корня из корреляционной матрицы  $R$ :  $A = \text{Re}(\sqrt{R})$ ),  $T$  – знак транспонирования,  $\text{Re}(\cdot)$  – знак вещественной части, а  $R = 2 \sin(\pi R_S/6)$  – корреляционная матрица Пирсона, соответствующая ранговой корреляционной матрице Спирмена  $R_S$  [18, 19];

- 3) преобразовать покомпонентно гауссов вектор  $z$  с коррелированными компонентами по формуле:  $\{y_j = \Phi(z_j), j = \overline{1, m}\}$  в вектор  $y$  со значениями в  $[0, 1]^m$ , распределение которого описывается копулой (6);
- 4) преобразовать покомпонентно вектор  $y$  в искомый вектор  $x$  с маргинальными распределениями  $\{\tilde{F}_n(x_j), j = \overline{1, m}\}$  и ранговой корреляционной матрицей Спирмена  $R_S$  по формуле:  $\{x_j = \tilde{F}_n^{-1}(y_j), j = \overline{1, m}\}$ .

Таким образом, алгоритм 2 преобразует случайный  $m$ -мерный гауссов вектор  $v$  с независимыми компонентами  $\{v_k, k = \overline{1, m}\}$ , подчиняющимся стандартному нормальному закону распределения (с нулевым математическим ожиданием и единичной дисперсией) в  $m$ -мерный вектор  $x$  с зависимыми компонентами  $\{x_j = \tilde{F}_n^{-1}(\Phi(\sum_{k=1}^m A_{jk}v_k)), j = \overline{1, m}\}$ , где  $A = \text{Re}(\sqrt{2 \sin(\pi R_S/6)})$ , а  $R_S$  – ранговая корреляционная матрица Спирмена, вычисленная по обучающей выборке. Компоненты вектора  $x$  подчиняются сглаженным маргинальным эмпирическим функциям распределения  $\{\tilde{F}_n(x_j), j = \overline{1, m}\}$ , построенным, как и обратные функции распределения  $\{\tilde{F}_n^{-1}(y_j), j = \overline{1, m}\}$ , по обучающей выборке.

Многомерная модельная функция распределения легко получается из равенства гауссовой и модельной копул  $C_{\text{норм}}(y_1, \dots, y_m) = C_{\text{мод}}(y_1, \dots, y_m)$  с учетом их выражений (6)–(7) после замены  $\{y_j = \tilde{F}_n(x_j), j = \overline{1, m}\}$ :

$$\begin{aligned} \tilde{F}_{n(\text{мод})}(x_1, \dots, x_m) &= \\ &= \Phi(\Phi^{-1}(\tilde{F}_n(x_1)), \dots, \Phi^{-1}(\tilde{F}_n(x_m))). \end{aligned} \quad (8)$$

Эта функция зависит только от маргинальных сглаженных эмпирических функций распределения, маргинальных стандартных гауссовых распределений и многомерной нормальной функции распределения, зависящей от эмпирической ранговой корреляционной матрицы, построенной по обучающей выборке. Именно этой функции распределения подчиняются случайные векторы  $x$ , получаемые алгоритмом 2 имитационного моделирования. Легко видеть, что соответствующие этому распределению маргинальные функции распределения совпадают со сглаженными маргинальными эмпирическими функциями распределения:  $\{\tilde{F}_{n(\text{мод})}(x_j) = \tilde{F}_n(x_j), j = \overline{1, m}\}$ .

Дифференцируя выражение (8), получим многомерную модельную плотность распределения случайных



векторов, генерируемых алгоритмом 2:

$$\begin{aligned} \tilde{f}_{n(\text{мод})}(x_1, \dots, x_m) &= \\ &= \frac{\partial^m \Phi(\Phi^{-1}(\tilde{F}_n(x_1)), \dots, \Phi^{-1}(\tilde{F}_n(x_m)))}{\partial x_1 \dots \partial x_m} = \\ &= f_{\text{норм}}(\Phi^{-1}(\tilde{F}_n(x_1)), \dots, \Phi^{-1}(\tilde{F}_n(x_m))) \times \\ &\quad \times \prod_{j=1}^m \frac{\tilde{f}_n(x_j)}{f_{\text{норм}}(\Phi^{-1}(\tilde{F}_n(x_j)))}, \end{aligned} \quad (9)$$

где  $\tilde{f}_n(x_j) = d\tilde{F}_n(x_j)/dx_j$  – плотность сглаженного маргинального эмпирического распределения  $\tilde{F}_n(x_j)$ ,  $f_{\text{норм}}(x) = d\Phi(x)/dx$  – плотность стандартного нормального распределения.

Построение сглаженных маргинальных эмпирических функций распределения, участвующих в алгоритме 2, является несравненно более простой в вычислительном отношении задачей, чем построение сглаженных многомерных функций распределения, используемых алгоритмом 1.

#### 4 Построение сглаженных маргинальных эмпирических функций распределений при наличии связок (совпадений) в наблюдениях

Поскольку алгоритм построения сглаженного маргинального эмпирического распределения по наблюдаемой выборке любого признака одинаков для всех признаков, опустим пока индекс признака  $j = \overline{1, m}$ , оставив только индексы наблюдений  $i = \overline{1, n}$ .

Пусть  $\{x_{(1)}, x_{(2)}, \dots, x_{(n)}\}$  – ранжированная в порядке возрастания значений выборка (вариационный ряд) наблюдений случайной величины  $\xi$ , интегральная функция распределения которой  $F(x)$  априорно неизвестна. Пусть возможные значения случайной величины  $\xi$  – рациональные числа из интервала  $[x_0, x_{fin}]$  (в случае, если  $\xi$  – относительная частота,  $x_0 = 0$ ,  $x_{fin} = 1$ ). Тогда возможно образование связок (совпадения значений) в наблюдениях.

Пусть в вариационном ряду наблюдений оказалось  $L \leq n$  различных значений наблюдаемой случайной величины  $\xi$ . Для общности будем называть связкой каждую группу одинаковых наблюдений, отличающихся от наблюдений других групп. Такое определение допускает наличие в связке одного и более одинаковых наблюдений. Обозначим через  $n_k \geq 1$  число наблюдений в  $k$ -й связке ( $k = \overline{1, L}$ ). В силу упорядоченности наблюдений в вариационном ряду совпадающие наблюдения в каждой связке являются подпоследовательностью соседних совпадающих значений вариационного ряда. В частности, при  $n_k = 1$  имеем вырожденную связку, состоящую из единственного значения.

Очевидно,  $\sum_{k=1}^L n_k = n$ . Обозначим число значений вариационного ряда в первых  $l$  связках через  $N_l = \sum_{k=1}^l n_k$ ,  $l = \overline{1, L-1}$ . Тогда эмпирическая функция распределения  $F_n(x)$  случайной величины  $\xi$  примет вид ступенчатой функции со скачками в точках связок  $\{x_{(N_l)}, l = \overline{1, L-1}\}$ , равными относительным частотам  $\{N_l/n, l = \overline{1, L-1}\}$  совпадающих значений наблюдений в связках:

$$F_n(x) = \begin{cases} 0, & \text{если } x \leq x_{(1)}, \\ \frac{N_l}{n}, & \text{если } x_{(N_l)} < x \leq x_{(N_{l+1})}, \\ & l = \overline{1, L-1}, \\ 1, & \text{если } x > x_{(n)}. \end{cases} \quad (10)$$

Сгладим функцию распределения  $F_n(x)$ , аппроксимировав ее простейшей непрерывной функцией – ломаной, проходящей через середины горизонтальных участков ступенек этой функции. Обозначив среднее расстояние между соседними связками (точнее, между значениями наблюдений, входящих в соседние связки) через  $h = (x_{(n)} - x_{(1)})/2(L-1)$ , установим нижнюю  $\bar{x}_1 = \max(x_{(1)} - h, x_0)$  и верхнюю  $\bar{x}_{L+1} = \min(x_{(n)} + h, x_{fin})$  границы интервала возможных значений случайной величины, подчиняющейся сглаженной эмпирической функции распределения  $\tilde{F}_n(x)$ , таким образом, чтобы эти границы отстояли от граничных значений наблюдаемой выборки не более, чем на половину среднего расстояния между соседними связками. Вычислим упорядоченное множество середин интервалов между соседними связками:  $\{\bar{x}_{l+1} = (x_{(N_l)} + x_{(N_{l+1})})/2, l = \overline{1, L-1}\}$ . Тогда сглаженная эмпирическая функция распределения примет вид:

$$\tilde{F}_n(x) = \begin{cases} 0, & \text{если } x \leq \bar{x}_1, \\ \frac{N_l}{n} + \frac{n_{l+1}}{n} \frac{x - \bar{x}_l}{\bar{x}_{l+1} - \bar{x}_l}, & \text{если } \bar{x}_l < x \leq \bar{x}_{l+1}, \\ & l = \overline{1, L-1}, \\ 1, & \text{если } x > \bar{x}_{L+1}. \end{cases} \quad (11)$$

Производная  $d\tilde{F}_n(x)/dx = \tilde{f}_n(x)$  дает ступенчатую функцию плотности этого распределения:

$$\tilde{f}_n(x) = \begin{cases} 0, & \text{если } x \leq \bar{x}_1, \\ \frac{n_{l+1}}{n} \frac{1}{\bar{x}_{l+1} - \bar{x}_l}, & \text{если } \bar{x}_l < x \leq \bar{x}_{l+1}, \\ & l = \overline{1, L-1}, \\ 0, & \text{если } x > \bar{x}_{L+1}. \end{cases} \quad (12)$$

Функция  $\tilde{F}_n^{-1}(y)$ , обратная к сглаженной маргинальной функции распределения  $\tilde{F}_n(x)$ , принимает вид:

$$\tilde{F}_n^{-1}(y) = \begin{cases} \bar{x}_1, & \text{если } y = 0, \\ \bar{x}_l + \frac{\bar{x}_{l+1} - \bar{x}_l}{n_{l+1}/n} (y - \frac{N_l}{n}), & \text{если } \frac{N_l}{n} < y \leq \frac{N_{l+1}}{n}, \\ & l = \overline{1, L-1}, \\ \bar{x}_{L+1}, & \text{если } y = 1. \end{cases} \quad (13)$$

Возвращая обратно индексы переменных, следует принимать во внимание, что в выражениях для сглаженной маргинальной эмпирической функции распределения (11) и ее плотности (12) для каждой  $j$ -й компоненты вектора  $\xi = (\xi_1, \dots, \xi_m)$  в качестве  $x$  следует подставлять  $x_j$ , в качестве  $\bar{x}_l$  – величину  $\bar{x}_{lj}$ , в качестве  $L$  – величину  $L_j$ , в качестве  $n_l$  и  $N_l$  – величины  $n_{lj}$  и  $N_{lj}$  соответственно. Таким образом, для каждого  $j = \overline{1, m}$  имеем:

$$\tilde{F}_n(x_j) = \begin{cases} 0, & \text{если } x_j \leq \bar{x}_{1,j}, \\ \frac{N_{lj}}{n} + \frac{(x_j - \bar{x}_{lj})n_{l+1,j}}{(\bar{x}_{l+1,j} - \bar{x}_{lj})n}, & \text{если } \bar{x}_{lj} < x_j \leq \bar{x}_{l+1,j}, \\ & l = \overline{1, L_j - 1}, \\ 1, & \text{если } x_j > \bar{x}_{L+1,j}. \end{cases} \quad (14)$$

Аналогично для плотности

$$\tilde{f}_n(x_j) = \begin{cases} 0, & \text{если } x_j \leq \bar{x}_{1,j}, \\ \frac{n_{l+1,j}/n}{\bar{x}_{l+1,j} - \bar{x}_{lj}}, & \text{если } \bar{x}_{lj} < x_j \leq \bar{x}_{l+1,j}, \\ & l = \overline{1, L_j - 1}, \\ 0, & \text{если } x_j > \bar{x}_{L+1,j}. \end{cases} \quad (15)$$

и обратной функции:

$$\tilde{F}_n^{-1}(y_j) = \begin{cases} \bar{x}_{1,j}, & \text{если } y_j = 0, \\ \bar{x}_{lj} + \frac{\bar{x}_{l+1,j} - \bar{x}_{lj}}{n_{l+1,j}/n} \times \\ \times (y_j - \frac{N_{lj}}{n}), & \text{если } \frac{N_{lj}}{n} < y_j \leq \frac{N_{l+1,j}}{n}, \\ & l = \overline{1, L_j - 1}, \\ \bar{x}_{L+1,j}, & \text{если } y_j = 1. \end{cases} \quad (16)$$

При этом

$$\begin{aligned} h_j &= (x_{(n),j} - x_{(1),j})/2(L_j - 1), \\ \bar{x}_{1,j} &= \max(x_{(1),j} - h_j, x_0), \\ \bar{x}_{L+1,j} &= \min(x_{(n),j} + h_j, x_{fin}), \\ \bar{x}_{l+1,j} &= (x_{(N_l),j} + x_{(N_{l+1}),j})/2, \\ & l = \overline{1, L_j - 1}, j = \overline{1, m}. \end{aligned}$$

Здесь  $x_{(k),j}$  –  $k$ -я порядковая статистика  $j$ -го вариационного ряда  $\{x_{(1),j}, \dots, x_{(n),j}\}$ , полученного ранжированием в порядке возрастания  $j$ -го столбца обучающей матрицы-выборки  $x^o = \{x_{ij}^o, i = \overline{1, n}, j = \overline{1, m}\}$ . Если  $n_{lj} \geq 1$  – число наблюдений в  $l$ -й связке  $j$ -го столбца этой матрицы, то, очевидно,  $\forall j \sum_{l=1}^{L_j} n_{lj} = n$  – общее число наблюдений векторной переменной  $\xi = (\xi_1, \dots, \xi_m)$  (объем выборки, число вариационных рядов, число строк обучающей матрицы), а  $N_{lj} = \sum_{k=1}^l n_{kj}$ ,  $l = \overline{1, L_j - 1}$ , – число наблюдений в первых  $l$  связках  $j$ -го вариационного ряда. Следовательно,  $N_{L_j} = n$ .

На рис. 1 пунктирной линией показано взаимно-однозначное отображение  $x_{(norm)} \leftrightarrow x_{(mod)}$  реализации  $x_{(norm)}$  нормально распределенной случайной величины в модельную  $x_{(mod)}$  и наоборот.

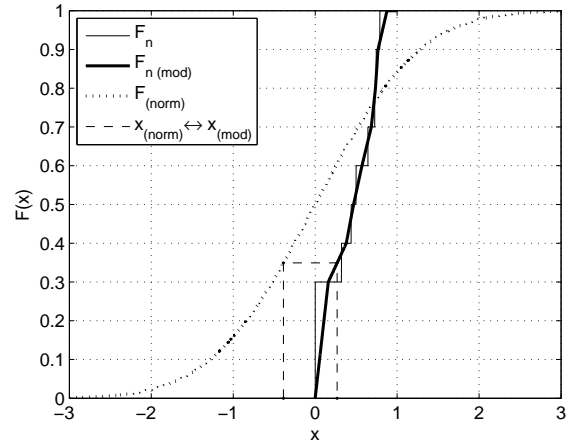


Рис. 1: Эмпирическая  $F_n(x)$ , сглаженная эмпирическая  $F_{n(mod)}(x) = \tilde{F}_n(x)$ , стандартная нормальная  $F_{(norm)}(x) = \Phi(x)$  функции распределения и линия взаимно-однозначного отображения  $x_{(norm)} \leftrightarrow x_{(mod)}$ .

## 5 Имитационное моделирование относительных частот признаков стилей текстов на примере произведений русской художественной прозы 19 века

Применим алгоритм 2 имитационного моделирования реализаций случайных векторов, подчиняющихся сглаженному эмпирическому распределению, к задаче имитационного моделирования векторов признаков стилей текстов.

$j$	Название	$в$	$на$	$с$	$за$	$по$	$из$
1	ВечераЧ1	611	502	272	165	147	109
2	ВечераЧ2	817	675	437	189	177	158
3	Вий	292	212	135	56	78	65
4	ЗапискСум	138	107	48	34	18	27
5	КакПоссор	303	207	186	79	58	47
6	МертДуш1	1945	1189	833	325	363	293
7	МертДуш2	1868	992	841	317	350	257
8	НевскПрос	269	203	144	41	43	34
9	Нос	207	115	89	35	41	22
10	Портрет	499	267	200	82	65	67
11	СтароПом	158	122	83	39	35	20
12	ТарБульб	859	737	386	202	204	190
13	Шинель	261	174	115	47	55	32

Таблица 1: Фрагмент таблицы абсолютных частот.

Используем для примера в качестве обучающей вы-

борки 13 текстов произведений Н.В. Гоголя, сокращенные названия которых приведены в табл. 1. Таблица содержит также фрагмент таблицы абсолютных частот употребления в текстах этого писателя первого набора служебных слов {в, на, с, за, по, из, при, лишь, уже} из табл. 2 минимальных тупиковых тестов (безизбыточных наборов признаков), различающих стили текстов художественных произведений 11 русских писателей 19 века по алгоритму дерева решений [20].

№	Признаки	Длина
1	в, на, с, за, по, из, при, лишь, уже	9
2	в, на, с, по, из, от, при, когда, лишь	9
3	в, на, с, по, из, во, при, когда, уже	9
4	в, на, с, по, из, при, если,нибудь	8

Таблица 2: Минимальные наборы признаков.

Для обучения алгоритма дерева решений использовались 154 (практически все) произведения этих писателей, в том числе 13 указанных в табл. 1 произведений Н.В. Гоголя. Эти произведения составляют обучающую выборку. Для статистической оценки качества алгоритма распознавания авторского стиля по дереву решений требуется достаточно большое количество текстов тех же авторских классов с теми же частотами употребления слов-признаков стилей текстов, что и тексты обучающей выборки. Поскольку таких контрольных текстов нет, остается использовать имитационное моделирование для построения таблиц частот признаков стилей текстов различных авторских классов, обладающих теми же статистическими свойствами, что и таблицы частот, полученных из обучающей выборки. В данной работе мы ограничимся иллюстрацией возможности такого имитационного моделирования на примере одного писателя – Н.В. Гоголя. Полную таблицу абсолютных частот можно представить  $13 \times 9$ -матрицей  $x^a$ , предпоследний столбец которой, соответствующий признаку  $j = 8$ , содержит связку из нулевых значений абсолютных частот:

$$x^a = \begin{pmatrix} 611 & 502 & 272 & 165 & 147 & 109 & 28 & 2 & 93 \\ 817 & 675 & 437 & 189 & 177 & 158 & 35 & 0 & 171 \\ 292 & 212 & 135 & 56 & 78 & 65 & 13 & 0 & 52 \\ 138 & 107 & 48 & 34 & 18 & 27 & 5 & 0 & 22 \\ 303 & 207 & 186 & 79 & 58 & 47 & 30 & 1 & 44 \\ 1945 & 1189 & 833 & 325 & 363 & 293 & 64 & 2 & 282 \\ 1868 & 992 & 841 & 317 & 350 & 257 & 66 & 1 & 250 \\ 269 & 203 & 144 & 41 & 43 & 34 & 24 & 0 & 34 \\ 207 & 115 & 89 & 35 & 471 & 22 & 13 & 1 & 25 \\ 499 & 267 & 200 & 82 & 65 & 67 & 16 & 0 & 101 \\ 158 & 122 & 83 & 39 & 35 & 20 & 4 & 0 & 34 \\ 859 & 737 & 386 & 202 & 204 & 190 & 29 & 1 & 177 \\ 261 & 174 & 115 & 47 & 55 & 32 & 12 & 0 & 46 \end{pmatrix}.$$

Связка нулевых абсолютных частот приведет и к связке нулевых значений относительных частот.

Матрица относительных частот  $x^o$  (обучающая матрица-выборка) получается построчным нормированием абсолютных частот:  $\{x_{ij}^o = x_{ij}^a / \sum_{j'=1}^m x_{ij'}^a, j = \overline{1, m}\}$ .

На рис. 2 представлены эмпирические функции распределения  $F_n(x)$ , полученные по обучающей выборке  $x^o$  (тонкие линии) при  $n = 13$  и путем имитационного моделирования по алгоритму 2 (полужирные линии) при  $n = 100$  для каждого из  $m = 9$  признаков  $j = \overline{1, m}$ .

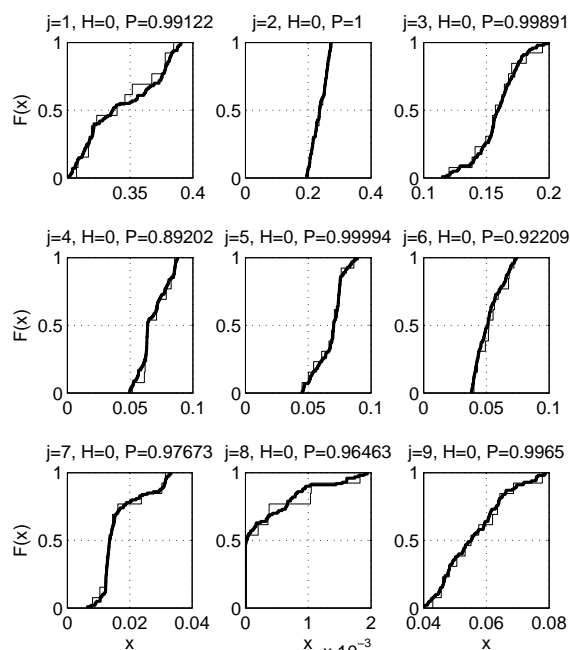


Рис. 2: Эмпирические функции распределения (ЭФР). Тонкие линии – исходные ЭФР при  $n = 13$ , полужирные – модельные ЭФР при  $n = 100$ .

С помощью двухвыборочного критерия согласия Колмогорова-Смирнова [19] были проверены гипотезы  $H_0$  о равенстве исходного (по обучающей выборке) и модельного (по алгоритму 2) эмпирических распределений для каждого признака.

На рис. 2 для каждого признака приведены результаты работы критерия (достигнутый уровень значимости  $P$  и принимаемая гипотеза  $H: H = 0$  соответствует  $H_0$ ). Видно, что во всех случаях нет оснований отвергнуть нулевую гипотезу, поскольку достигнутый уровень значимости существенно превышает критическое (пороговое) значение  $P \gg 0.05$ . Это означает, что генерация векторов-признаков по алгоритму 2 соответствует маргинальным статистическим характеристикам обучающей выборки.

На рис. 3 представлены столбиковые диаграммы эмпирических оценок математических ожиданий  $\{M_j = \bar{x}_j, j = \overline{1, m}\}$  и стандартных отклонений  $\{S_j = \hat{\sigma}_j, j = \overline{1, m}\}$  признаков по обучающей (“origin”)

выборке  $x^o$  объема  $n = 13$  и по модельной (“model”) выборке  $x$  объема  $n = 100$ . Видно, что математические ожидания практически совпадают, а различия стандартных отклонений пренебрежимо малы.

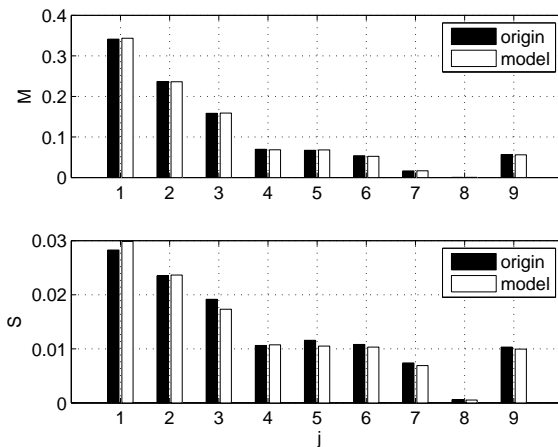


Рис. 3: Эмпирические оценки математических ожиданий и стандартных отклонений признаков по обучающей и модельной выборкам.

На рис. 4 представлены полутонные яркие изображения эмпирических оценок нормированных корреляционных матриц признаков.

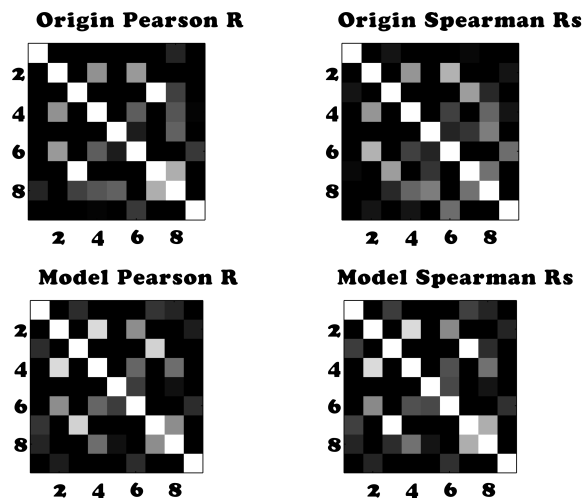


Рис. 4: Полутонные яркие изображения эмпирических оценок нормированных корреляционных матриц  $R$  и  $R_S$ .

Для проверки гипотез о равенстве корреляционных матриц элементы верхней наддиагональной треугольной матрицы каждой матрицы  $R$  последовательно объединялись в линейные массивы  $r = (R_{1,2}, \dots, R_{1,m}, R_{2,3}, \dots, R_{2,m}, \dots, R_{m-1,m})$  длины  $n_r = m(m-1)/2$  (в данном случае  $n_r = 36$  при  $m = 9$ ), представленные на рис. 5, для сравнения которых использовался непараметрический критерий ранговых сумм Вилкоксона-Манна-Уитни (Wilcoxon-Mann-Whitney rank sum test) [19].

Результаты этого сравнения сведены в табл. 3. Видно, что критерий Вилкоксона не обнаружил при критическом уровне значимости, равном 5%, статистически значимых различий между представленными на рис. 4 корреляционными матрицами: достигнутые уровни значимости  $P$  критерия при сравнении между собой всех корреляционных матриц (Пирсона  $R$  и Спирмена  $R_S$ ), вычисленных по обучающей выборке (помечены “Origin” на рис. 3–5 и “(исх)” в табл. 3) и по результатам имитационного моделирования (помечены “Model” и “(мод)”), составили величины, значительно превышающие критическое значение ( $P \gg 0.05$ ), что не дает оснований отвергнуть нулевую гипотезу  $H_0$  ( $H = 0$ ) о равенстве сравниваемых корреляционных матриц.

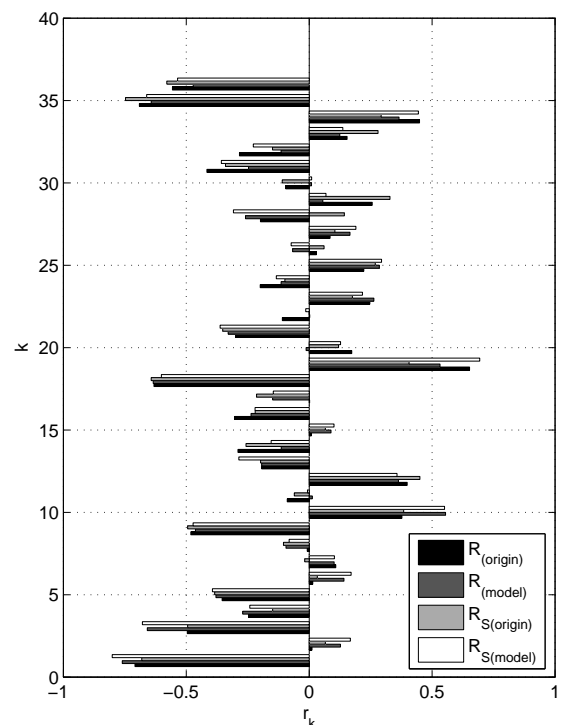


Рис. 5: Массивы элементов  $r$  верхних наддиагональных подматриц нормированных эмпирических корреляционных матриц  $R$  и  $R_S$ .

Множественное сравнение массивов элементов верхних наддиагональных подматриц нормированных эмпирических корреляционных матриц  $R_{(исх)}$ ,  $R_{(мод)}$ ,  $R_{S(исх)}$ ,  $R_{S(мод)}$  по непараметрическому медианному критерию Краскела-Уоллиса (Kruskal-Wallis one-way ANOVA test) [19] также дает достигнутый уровень значимости  $P = 0.9951 \gg 0.05$ , не позволяющий отвергнуть нулевую гипотезу  $H_0$  ( $H = 0$ ) об отсутствии различий в указанных корреляционных матрицах при критическом уровне значимости 5%.

Таким образом, во всех случаях нулевые гипотезы о равенстве исходных и модельных маргиналь-

Сравниваемые матрицы	$P$	$H$
$R_{(исх)} \& R_{(мод)}$	0.8350	0
$R_{(исх)} \& R_{S(мод)}$	0.9237	0
$R_{(исх)} \& R_{S(исх)}$	0.7913	0
$R_{S(исх)} \& R_{(мод)}$	0.9686	0
$R_{S(исх)} \& R_{S(мод)}$	0.9327	0
$R_{(мод)} \& R_{S(мод)}$	0.9596	0

Таблица 3: Парное сравнение корреляционных матриц по критерию ранговых сумм Вилкоксона.

ных эмпирических распределений, а также о равенстве эмпирических исходных и модельных корреляционных матриц не отвергаются, поскольку достигнутые уровни значимости нулевых гипотез значительно превышает критическое (пороговое) значение  $P \gg 0.05$ . Это означает, что генерация векторов-признаков по алгоритму 2 соответствует статистическим характеристикам обучающей выборки (по крайней мере на уровне сглаженных маргинальных распределений и ранговых корреляций между компонентами векторов-признаков объектов).

Используя выражения (8) и (9), можно вычислить модельные многомерные функции распределения. На рис. 6 в качестве примера приведена двумерная совместная модельная плотность распределения первых двух компонент  $\xi_1$  и  $\xi_2$  модельного вектора  $\xi$ , коррелированных с коэффициентом корреляции  $R_{(мод)}_{1,2} = -0.7603$ . Эти компоненты – модельные относительные частоты зависимых признаков “в” и “на”.

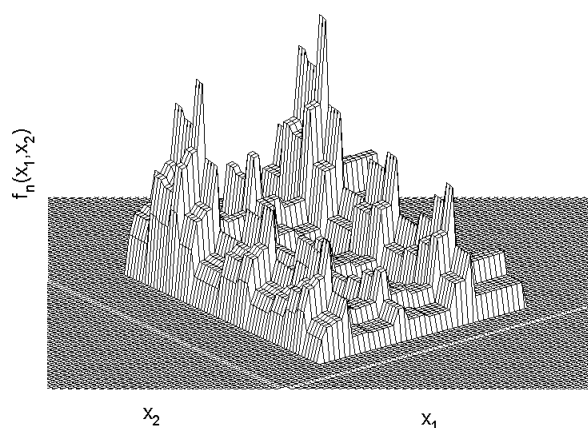


Рис. 6: Плотность совместной модельной функции распределения первых двух компонент вектора  $\xi$  – признаков “в” и “на” (табуляция с шагом 0.0001).

## 6 Заключение

Итак, рассмотренный в работе алгоритм 2 имитационного моделирования статистически зависимых случайных величин, подчиняющихся сглаженным маргинальным эмпирическим функциям распределения, построен полностью по обучающей выборке без использования каких-либо априорных сведений о функциях распределения наблюдаемых величин. В его основе лежит принцип равенства копулы сглаженного эмпирического распределения и гауссовой копулы, что предопределяет возможность учета при моделировании только корреляционных связей между случайными величинами.

Общая структура такого алгоритма при заданных маргинальных распределениях и заданной ранговой корреляционной матрице хорошо известна, реализована программно в математических пакетах (например, в системе программирования Matlab [21, 22]) и широко используется в различных приложениях (например, [23, 24, 25, 26, 27, 28] и др.).

Особенность алгоритма 2, представленного в данной работе, состоит в том, что он может работать в условиях связей (совпадений значений) в обучающей выборке, в том числе связей нулевых значений. Это позволяет использовать его для имитационного моделирования случайных векторов, близких по статистическим характеристикам к обучающей выборке, содержащей связи. При этом нулевые связи будут автоматически учитывать отсутствие в некоторых объектах (например, текстах) того или иного признака.

Стохастические объекты, генерируемые алгоритмом 2, могут быть использованы в качестве контрольной (тестовой) выборки в разнообразных задачах статистической оценки качества алгоритмов классификации и распознавания при дефиците или отсутствии реальных объектов для тестирования.

## Благодарности

Работа поддержана грантами РФФИ №№ 11-07-00776-а, 12-07-31109-мол-а, 13-07-00373-а.

## Список литературы

- [1] E. Parzen. On the estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33:1065–1076, 1962.
- [2] M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, 27 (3):832–837, 1956.
- [3] C. Spearman. The proof and measurement of association between two things. *American Journal of Psychology*, 15 (1):72–101, 1904.
- [4] M. G. Kendall. *Rank Correlation Methods (4th ed.)*. Charles Griffin & Co., London, 1970. Русский перевод: М. Кендэл. *Ранговые корреляции*, Статистика, Москва, 1975.



- [5] D. G. Clayton. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65:141–151, 1978.
- [6] E. W. Frees, E. A. Valdez. Understanding relationships using copulas. *North American Actuarial Journal*, 2:1–25, 1998.
- [7] R. B. Nelsen. *An Introduction to Copulas (Second Edition)*. Springer, Berlin, 2006.
- [8] A. Sklar. Fonctions de repartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de L'Université de Paris*, 8:229–231, 1959.
- [9] А. А. Боровков. *Математическая статистика*. Наука, Гл. ред. физ.-мат. лит., Москва, 1984.
- [10] K. Aas, C. Czado, A. Frigessi and H. Bakken. Pair copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44 (2):182–198, 2009.
- [11] Д. Фантаццини. Моделирование многомерных распределений с использованием копула-функций. ii. *Прикладная экономика*, 3 (23):98–132, 2011.
- [12] M. Soto, Ya. González-Fernández and A. Ochoa. *Modeling with Copulas and Vines in Estimation of Distribution Algorithms. 2012*. <http://arxiv.org/pdf/1210.5500.pdf>.
- [13] G. Salvadori, C. De Michele, N. T. Kottegoda and R. Rosso. *Extremes in Nature. An Approach Using Copulas*. Springer, Netherlands, 2007.
- [14] C. Czado. *Pair-Copula Constructions of Multivariate Copulas*. In: *Copula Theory and Its Applications: Proceedings of the Workshop Held in Warsaw, 25–26 September 2009*, pages 93–109. Springer, Heidelberg Dordrecht London New York, 2009.
- [15] R. Doman. *Modeling the Dependence Structure of the WIG20 Portfolio Using a Pair-copula Construction. Dynamic Econometric Models*, 10:31–42, Nicolaus Copernicus University, Torun, 2010.
- [16] V. Erhardt and C. Czado. Modeling dependent yearly claim totals including zero claims in private health insurance. *Scandinavian Actuarial Journal*, 2:106–129, 2012.
- [17] М. Кендалл, А. Стьюарт. *Теория распределений*. Наука, Гл. ред. физ.-мат. лит., Москва, 1966.
- [18] Б. Л. Ван дер Варден. *Математическая статистика*. Иностранная Литература, Москва, 1960.
- [19] Л. Закс. *Статистическое оценивание*. Статистика, Москва, 1976.
- [20] В. В. Поддубный, А. С. Пехтерев. Минимизация числа информативных признаков методом тестового распознавания в задаче классификации текстов по дереву решений. *Труды XVI международной ЭМ'2012 конференции по эвентологической математике и смежным вопросам*, Красноярск: СФУ, НИИППБ, 1 (под ред. О. Ю. Воробьева):180–184, 2012.
- [21] *Simulating Dependent Random Variables Using Copulas. 2013*. <http://www.mathworks.com/help/stats/examples/simulating-dependent-random-variables-using-copulas.html>.
- [22] J. C. Strelen. *Tools for Dependent Simulation Input with Copulas. 2009*. <http://web.informatik.uni-bonn.de/IV/strelen/Forschung/Publikationen/093Simutools.pdf>.
- [23] С. В. Стрелков. Стохастическое моделирование операционных рисков кредитных организаций. *Аудит и финансовый анализ*, 2:172–177, 2010.
- [24] U. Cherubini, E. Luciano, and W. Vecchiato. *Copula Methods in Finance*. John Wiley & Sons Ltd., Chichester, 2004.
- [25] D. Brigo, A. Pallavicini and R. Torresetti. *Credit Models and the Crisis. A Journey into CDOs, Copulas, Correlations and Dynamic Models*. John Wiley & Sons Ltd., Chichester, 2010.
- [26] J.-F. Jouanin, G. Rapuch, G. Riboulet and T. Roncalli. *Modelling dependence for credit derivatives with copulas. 2001*. <http://www.thierry-roncalli.com/download/copula-intensity.pdf>.
- [27] D. J. Erdman and A. Sinko. *Using Copulas to Model Dependency Structures in Econometrics. SAS Global Forum 2008, Paper 321-2008, pages 1–13*, <http://www2.sas.com/proceedings/forum2008/321-2008.pdf>.
- [28] G. Papaefthymiou and D. Kurowicka. Using copulas for modeling stochastic dependence in power system uncertainty analysis. *IEEE Transactions on Power Systems*, 24 (1):40–49, 2009.

# Информационно-поисковые системы для электронных библиотек

**Попов Александр Викторович**

Сибирский Федеральный Университет  
Институт математики и фундаментальной информатики  
Базовая кафедра вычислительных и информационных технологий  
Красноярск  
a.v.popov89@gmail.com

**Шабанов Владислав Саидович**

Сибирский Федеральный Университет  
Институт математики и фундаментальной информатики  
Базовая кафедра вычислительных и информационных технологий  
Красноярск  
vshabanov88@gmail.com

**Аннотация.** В статье рассмотрены вопросы построения полнотекстовый информационно-поисковых систем для электронных библиотек. Произведен обзор существующих ИПС. Представлена модель описания документа. Предложены способы индексации и поиска. Затронут аспект представления результатов поиска. В заключении отмечены дальнейшие варианты развития ИПС.

**Ключевые слова.** Информационно-поисковые системы, электронные библиотеки, индексация, поиск, релевантность, ранжирование.

## 1 Введение

Широкое распространение персональных компьютеров и быстрое развитие телекоммуникационных сетей привели к интенсивному росту электронных библиотек. В настоящее время многие статьи и книги публикуются непосредственно в электронном формате, а бумажные источники информации активно оцифровываются. Одной из наиболее актуальных потребностей, которые возникают при работе с большим количеством электронных источников, очевидно, является быстрый поиск с предоставлением релевантных и точных результатов, т.е. необходимо найти не только сам документ но и конкретную страницу или абзац в этом документе.

Особую ценность представляет собой контекстный поиск, который учитывает тематику текста при выдаче результатов на поисковый запрос. Это в большей степени важно при поиске научной литературы так как многие термины имеют различное значение в различных науках и релевантность результатов существенно зависит от определения контекста, в котором необходимо найти информацию.

На текущий момент электронная литература представлена во множестве различных форматах, что усложняет процесс обработки. Наибольшую популярность приобрели форматы PDF и DjVu. Особенностью данных форматов является их инвариантность при изменении средств просмотра и операционной системы у пользователя. Помимо текстовой информации PDF и DjVu могут содержать и графиче-

ские элементы, что очень удобно, в частности для технической литературы.

В наше время разработано несколько систем управления электронными библиотеками.

## 2 Существующие системы

Из этого множества библиотечных систем по функциональности, известности, использованию и русской локализации, выделяются три: DSpace, EPrints и Greenstone. Все эти системы широко применяются по всему миру и в России в частности. Так например библиотека Уральского государственного университета использует DSpace, библиотека правительства республики Марий Эл - Greenstone.

Каждая из вышеперечисленных систем разрабатывается уже не один год, для их усовершенствования работают десятки программистов. Информационно поисковые системы вошли в нашу повседневную жизнь и сложно представить использование персонального компьютера без их помощи. Однако каждая из перечисленных систем имеет как свои достоинства, так и недостатки. Ни в одной из рассмотренных АБИС не предусмотрена поддержка формата DjVu, который очень популярен в настоящее время и число книг данного формата (особенно технической литературы) продолжает постоянно увеличиваться. АБИС EPrints и DSpace не являются кроссплатформенными, для их установки нужна Unix-подобная ОС, и доступ к библиотеке со стороны конечного пользователя осуществляется по сети либо через интернет по средствам интернет браузера. Для домашней библиотеки и небольшой общественной библиотеки это является существенным недостатком, поскольку около 90 процентов компьютеров в мире находится под управление ОС Windows. [1]

Отдельный интерес представляла бы система для работы с электронными библиотеками, которая позволяла бы обеспечить совместную работу в некотором научном сообществе, где у участников имеются общая литература.

### 3 Информационный поиск

В настоящее время работу большинства поисковых систем можно упрощенно представить в виде следующих шагов: [2]:

1. Пользователь формирует поисковый запрос
2. Запрос передается поисковой системе
3. Система производит поиск ресурса в базе данных на основе этого запроса
4. Отображение списка результатов, наиболее подходящего данному поисковому запросу

Существует три основных подхода к организации поиска, каждый из которых обладает своими преимуществами и недостатками:

1. Булева модель. Относительно проста в реализации, позволяет обрабатывать большие объемы данных. Отсутствие контекстных операторов. Невысокая эффективность поиска.
2. Векторно-пространственная модель. В классическом виде малоприспособлена для обработки больших объемов данных в связи с применением массивов высокой размерности.
3. Вероятностная модель. Характеризуется низкой вычислительной масштабируемостью и необходимостью обучения системы.

Помимо переведенных выше, существуют и другие подходы, но их эффективность пока не достигла того уровня, который бы способствовал их распространению. [3]

#### 3.1 Модель

Содержание документов и запросов в информационно-поисковых системах обычно описывается некоторыми наборами терминов, представляющих собой отдельные слова или словосочетания.

Часто для характеристики терминов используются их веса, которые отражают предполагаемую важность каждого из терминов.

Документ  $F$  можно представить как покрытие:

$$F = \bigcup_i F_i$$

Где  $F_i$  - фрагменты документа. Каждому фрагменту  $F_i$  сопоставляется набор  $K_i = \{(K_l, z_l) | l = 1, \dots, n_i\}$ , где  $K_l$  - ключевое слово и  $z_l$  - его значимость. Значения  $z_l$  принадлежат некоторому упорядоченному множеству. Например  $\{1, 2, \dots, 5\}$ .

Далее будем предполагать, что выполняются следующие соотношения: если некоторый фрагмент  $F_i$  содержится во фрагменте  $F_j$ , то набор ключевых  $K_j$  содержится в  $K_i$ . Более того, будем считать, что фрагменту  $F_i \cap F_j$  соответствует набор ключевых слов  $K_i \cap K_j$ . Таким образом, мы определили правило на следования ключевых слов.

Решение о выдаче того или иного документа принимается в результате сравнения наборов терминов, относящихся соответственно к документам и запросам. Вниманию пользователя предлагаются те документы, наборы терминов которых совпадают с наборами терминов запросов. [4]

#### 3.2 Индексация

Для осуществления поиска среди множества электронных документов в настоящее время наиболее эффективной и распространенной является схема с применением индекса документов. Это значит, что перед осуществлением поиска необходимо произвести индексацию документов. При добавлении нового электронного документа информация о нем добавляется в индекс.

**Определение 4.** Индексация — это процесс построения поискового индекса.

При индексации необходимо получить доступ к информации в текстовом виде. Особую сложность на данном шаге представляют собой случаи когда информация представлена в виде изображения. Для ее преобразования в текстовый формат можно использовать метод оптического распознавания, однако результаты будут не всегда точны.

Принято выделять три способа индексации:

1. Ручной, при котором человек (индексатор) выделяет наиболее значимые слова и указывает область их действия. Достоинством такого способа является точность созданного индекса, а недостатком - трудоемкость.
2. Автоматический. Программа обрабатывает текст, выделяя ключевые слова и определяя их область действия. В настоящее время существует множество способов выделения ключевых слов. Например, можно использовать частотный анализ, учитывая количество использования слова в тексте. Для исключения попаданий малозначимых слов в индекс примется стоп-словарь, куда заносятся слова, которые следует пропускать во время индексирования. В качестве области действия слов можно использовать типографические области текста, такие как параграф или абзац.
3. Смешанный. Объединяет перечисленные выше способы, позволяя человеку корректировать ре-

зультаты автоматической индексации. Данный способ является наиболее эффективным т.к. дает возможность охватить хоть сколько много информации и точно выделить основные ключевые слова.

### 3.3 Представление результатов

Часто в ответ на поисковый запрос система представляет достаточно большое множество документов, которые в разной степени подходят пользователю. Способ их отображения во многом связан с удобством пользования системой. Обычно результаты отображаются в виде списка.

**Определение 5.** *Ранжирование — процесс, при котором поисковая система выстраивает результаты поиска в определенном порядке по принципу наибольшего соответствия конкретному запросу.*

Ранжирование доступно не для всех моделей поиска (например недоступно для булевой). [3] Ранжирование в текстовых и гипертекстовых документах существенно различается. В модели, представленной выше, для ранжирования текстовых документов можно использовать веса ключевых слов. Ранжирование гипертекстовых документов возможно также по свойствам обуславливаемым сетевой структурой.

Поскольку в одной книге может быть несколько страниц, дающих ответ на запрос пользователя, из них необходимо выбрать наиболее релевантную. Чтобы показать наибольшее количество книг, в поисковой выдаче для отдельной книги показывается как правило только одна страница, в редких случаях в поисковой выдаче могут быть показаны несколько страниц из одной книги.

Так же важен способ отображения. Для пользователя было бы удобно видеть результат поиска непосредственно в документе.

## 4 Заключение

Развитие контекстного поиска в полнотекстовых электронных библиотеках помогает уменьшить затрачиваемое время на поиск нужной информации, что является несомненным плюсом для потребностей современного общества. Так же необходимо расширить возможность поиска в большем количестве форматов поскольку с появлением новых устройств увеличивается и количество форматов электронных книг. Для улучшения качества поиска может быть полезным разработка предметных тезаурусов и применение поиска по словарю синонимов. [5] Еще одной перспективной ветвью развития системы является разработка приложения для web, тем самым открыв доступ к своей библиотеке из любого места где есть доступ к интернету. [6]

Все это поможет реализовать более быстрый контекстный поиск и сделает использование поисковых систем более удобным для пользователя.

## Список литературы

- [1] К. А. Кудим и В. А. Резниченко. Сравнение систем электронных библиотек eprints 3.0 и dspace 1.4.1. *Электронные библиотеки: перспективные методы и технологии, электронные коллекции. Труды Девятой Всероссийской научной конференции RCDL'2007*, pages 241–252, 2007.
- [2] В. Н. Романенко и Г. В. Никитина. *Сетевой информационный поиск: Практическое пособие*. кПрофессия, Санкт-Петербург, 2005.
- [3] Д. В. Ландэ и А. А. Снарский. *Интернетика: Навигация в сложных сетях: модели и алгоритмы*. Книжный дом кЛИБРОКОМ, Москва, 2009.
- [4] Б. С. Добронев и А. А. Мамедов. Разработка информационно-поисковых систем для полнотекстовых библиотек. *Повышение качества высшего профессионального образования: Материалы Всероссийской научно-методической конференции с международным участием: в 2ч, 1:252–256*, 2007.
- [5] М. Р. Коголовский. *Перспективные технологии информационных систем*. ДМК Пресс, Москва, 2003.
- [6] П. Храпцев. *Информационно-поисковые системы internet. Открытые системы*, 3, 1996.

## К истории закона больших чисел

Попова Елена Александровна

Сибирский федеральный университет  
Торгово-экономический институт  
Красноярск  
popovaelena15@mail.ru

**Аннотация.** В 1713 году было опубликовано сочинение Я. Бернулли “*Ars Conjectandi*”, содержащее доказательство теоремы, положившей начало закону больших чисел. Позже С. Д. Пуассон доказал более общее предложение, назвав его законом больших чисел. Важное дальнейшее обобщение этого закона было дано в П. Л. Чебышёвым.

**Ключевые слова.** Закон больших чисел, история закона больших чисел, теорема Бернулли, 300-летие закона больших чисел.

### 1 Введение

В 2013 году исполняется 300 лет со дня издания сочинения Я. Бернулли “*Ars Conjectandi*” (“Искусство предположений”), содержащего доказательство теоремы, которая получила название “закона больших чисел”. Историк математики В. В. Бобынин отмечает, что закон больших чисел “с достаточной ясностью был выражен в XVI столетии Джероламо Кардано в его статье “*De ludo aleae*”. Совершенно новым и вполне принадлежащим Якобу Бернулли было данное им ... впервые математическое доказательство закона больших чисел, попыток найти которые до него, по-видимому, совсем не делалось” [1]. В “математическом оформлении” он появляется в “*Ars Conjectandi*” в 1713 г. — посмертной работе Я. Бернулли. Позже С. Д. Пуассон доказал более общее предложение, назвав его законом больших чисел, причем С. Д. Пуассон “в различных местах своего сочинения (1836) обозначает одним и тем же термином два совершенно различных утверждения” [2]. Ю. В. Прохоров в энциклопедии “Вероятность и математическая статистика” определяет закон больших чисел, как “общий принцип, в силу которого совместное действие случайных факторов приводит при некоторых весьма общих условиях к результату почти не зависящему от случая. Математическая формулировка общей идеи всех теорем, известных под названием З.Б.Ч., предложена в 1933 году А. Н. Колмогоровым. Сближение частоты наступления случайного события с его вероятностью при возрастании числа испытаний (подмеченный сначала, по-видимому, на азартных играх) может служить примером действия этого принципа”

[3]. Для читателя, далекого от математики и ее приложений, Б. В. Гнеденко, выдающийся математик, ученик и соратник А. Я. Хинчина и А. Н. Колмогорова, дал следующую трактовку закона больших чисел: “среднее арифметическое очень большого числа случайных величин, принимающих свои значения независимо от друг от друга, с практической достоверностью равно постоянной величине” [4].

### 2 История написания и издания сочинения Якоба Бернулли “*Ars Conjectandi*”

Труд был напечатан в Базеле на латинском языке. Это сочинение Я. Бернулли готовил около 20 лет почти до самой смерти, а вышло оно из печати в августе 1713 года. “Согласно опубликованным ныне дневникам, Бернулли доказал свою теорему еще в 1687-1689 годах” [5]. Принято считать, что издателем этого труда был Николай I Бернулли, так как в предисловие, помещенном в самом начале к “*Ars Conjectandi*”, написано: “Николай Бернулли приветствует читателя! Наконец ... выходит в свет долгожданный трактат моего дяди об искусстве предположений...” [5]. Это неверно. Наследники Я. Бернулли, его вдова и сын передали хранившуюся у них рукопись издателям братьям Турнэйзен летом 1712 г, когда Н. Бернулли находился в Лондоне. Издатели хотели бы, чтобы незавершенность работы восполнил брат покойного — И. Бернулли, как человек в наибольшей степени для этого подходящий; “однако они не сочли возможным обременять его такой просьбой из-за чрезвычайной занятости” [там же]. Весной 1713 г. Н. Бернулли вернулся в Базель, но в это время книга была уже почти полностью напечатана, и “он смог только, по специальной просьбе сына покойного, написать краткое предисловие и составить список опечаток, пропущенных неквалифицированными корректорами” [5].

Полный перевод надписи на титульном листе первого издания “*Ars Conjectandi*” (1713) таков:

ЯКОБА БЕРНУЛЛИ,  
профессора Базельского университета и члена обоих  
Королевских научных обществ Франции и Пруссии,  
знаменитого математика,  
ИСКУССТВО ПРЕДПОЛОЖЕНИЙ,  
посмертное сочинение



С добавлением  
Трактата  
о бесконечных рядах  
и написанного по-французски послания  
Об игре в мяч  
Ракеткой.  
В Базеле  
на средства братьев Турнэйзен  
1713  
На гербе (издательской марке) надпись  
“Торопись медленно”.

Я. Бернулли свое сочинение разделил на четыре части. Перевод их названий [5]: “Искусства предположений часть первая, содержащая трактата Гюйгенса о подсчете шансов при игре в кости с примечаниями Якоба Бернулли”; “Искусства предположений часть вторая, содержащая учение о перестановках и сочетаниях”; “Искусства предположений часть третья, объясняющая применение вышеизложенного учения при различных жеребьевках и различных играх в кости”; “Искусства предположений часть четвертая, излагающая использование и применение предшествующего учения в гражданских, моральных и экономических делах”.

Четвертую часть сочинения, в которой автор намеревался рассказать об использовании всего изложенного выше в гражданских, моральных и экономических вопросах, он из-за продолжительной болезни и безвременной наступившей смерти оставил незавершенной. Эта часть содержит классическое и статистическое определение вероятности, элементы исчисления предположений и основную теорему — закон больших чисел в форме Бернулли. Я. Бернулли доказал теорему средствами “чистой математики”, предпослав ей пять лемм.

Главное предположение (здесь под этим названием содержится собственно теорема Бернулли). Пусть число благоприятных случаев относится к числу неблагоприятных точно или приближенно, как  $r$  к  $s$ , или к числу всех случаев — как  $r$  к  $r+s$  или  $r$  к  $t$ , каковое отношение заключается в пределах  $\frac{r+1}{t}$  и  $\frac{r-1}{t}$ . Требуется доказать, что можно взять столько опытов, чтобы в какое угодно данное число раз (с раз) было вероятнее, что число благоприятных наблюдений попадает в эти пределы, а не вне их, т.е. что отношение числа благоприятных наблюдений к числу всех будет не более, чем на  $\frac{r+1}{t}$ , и не менее, чем  $\frac{r-1}{t}$  [5]. Современная формулировка теоремы Бернулли: если в каждом из  $n$  независимых испытаний вероятность появления события постоянна, то как угодно близка к единице вероятность того, что отклонение относительной частоты от вероятности по абсолютной величине будет сколь угодно малым, если число испытаний достаточно велико. Другими словами, если  $\varepsilon$  — сколь угодно малое положительное число, то при соблюдении условий теоремы имеет место

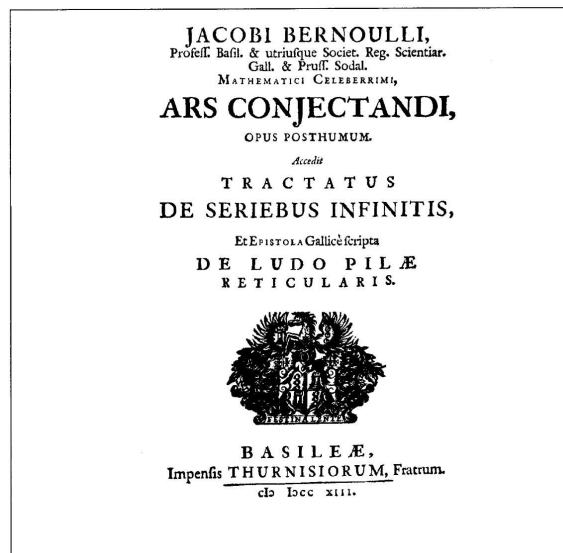


Рис. 1: Титульный лист “Ars Conjectandi” (изображение с сайта [http://en.wikipedia.org/wiki/Ars\\_Conjectandi](http://en.wikipedia.org/wiki/Ars_Conjectandi), Public Domain).

равенство  $\lim_{n \rightarrow \infty} P(|\frac{r_n}{n} - p| < \varepsilon) = 1$ . Я. Бернулли заканчивает книгу словами: “... если бы наблюдения над всеми событиями продолжать всю вечность (при чем вероятность, наконец, перешла бы в полную достоверность), то было бы замечено, что все в мире управляется точными отношениями и постоянным законом изменений, так что даже в вещах, в высшей степени случайных, мы принуждены были бы признать как бы некоторую необходимость и, скажу я, рок. Не знаю, не это ли имел в виду уже сам Платон в своем учении о восстановлении всех вещей, согласно которому все по истечении несметного числа веков возвратится в прежнее состояние” [5]. На этом “Ars Conjectandi” обрывается. Л. Е. Майстров предполагает: “работа осталась незаконченной, потому, что Бернулли не видел серьезных применений теории вероятностей к упомянутым вопросам” [6].

### 3 История подготовки юбилейного заседания Императорской Академии наук, посвященного 200-летию закона больших чисел

В декабре 1912 года на заседании Физико-математического отделения Императорской Академии наук предстояло избрать комиссию для участия в праздновании 300-летия дома Романовых. В протоколе заседания от 5 декабря 1912 г. записано: “Академик А. А. Марков заявил, что он не находит возможным участвовать в юбилее, а потому от баллотировки устраняется” [7]. Более того, в противовес, он организует научный юбилей — празднование 200-летия математического закона

больших чисел.

12 января 1913 г. А. А. Марков выступил с заявлением: “Считаю своим долгом обратить внимание Общего собрания на то, что в текущем году исполняется двухсотлетний юбилей закона больших чисел. Начало этому закону положено знаменитой теоремой Якоба Бернулли, которая опубликована в его сочинении *Ars Conjectandi*, помеченном 1713 годом. Полагаю, что Академии наук следует так или иначе ознаменовать этот юбилей, который непосредственно касается не только I-го Отделения, но и III-го, ибо закон больших чисел играет важную роль в статистике” [5].

Его идею поддержали академики К. Г. Залеман, И. И. Янжул, А. М. Ляпунов и В. А. Стеклов, которые и образовали вместе с А. А. Марковым юбилейную комиссию. 22 января 1913 года газета “Речь” поместила следующее обращение: “Представителям науки и ее почитателям. В 1713 году появилось в свет посмертное сочинение Якоба Бернулли “*Ars Conjectandi*”, где впервые установлена его знаменитая теорема, развившаяся впоследствии в закон больших чисел. Теорема эта относится непосредственно к математике и специально к теории вероятностей; но она встречает и должна встречать, как заметил сам Якоб Бернулли, многочисленные приложения во всех науках и вопросах практики, где приходится пользоваться статистическими приемами. Полагая, что следует, так или иначе, торжественно ознаменовать этот 200-летний юбилей, обращаюсь ко всем сочувствующим этой идее с предложением помочь мне в осуществлении ее. Академик А. Марков”

9 февраля 1913 года был зачитан доклад Комиссии по вопросу об ознаменовании 200-летнего юбилея закона больших чисел. Обсудив вопрос, Комиссия пришла к заключению, что ознаменовать юбилей можно следующим образом:

“Во-первых, устроить особо торжественное заседание Академии, при чем к деятельному участию в этом заседании желательно из посторонних ученых привлечь члена государственного Совета А. В. Васильева и профессора С.-Петербургского Политехнического института А. А. Чупрова.

Во-вторых, издать перевод на русский язык четвертой части *Ars Conjectandi*; такой перевод может выполнить приват-доцент С. Петербургского Университета Я. В. Успенский.

Наконец, академик А.А. Марков предлагает сделать Академическим изданием приготовленный им перевод на французский язык дополнительных статей третьего издания его “Исчисление вероятностей”, которое печатается и должно появиться в текущем году” [5].

В мае 1913 года Андрей Андреевич Марков написал предисловие к изданному Академией наук русскому

переводу сочинения Якоба Бернулли, опубликованного в 1713 г., в котором была доказана его знаменитая теорема, положившая начало закону больших чисел. К переводу, исполненному Я.В. Успенским, приложен портрет Якоба Бернулли, воспроизведенный по портрету масляными красками, который находится в актовом зале Базельского музея. Эта фотография была прислана А. А. Маркову библиотекой Базельского Университета.

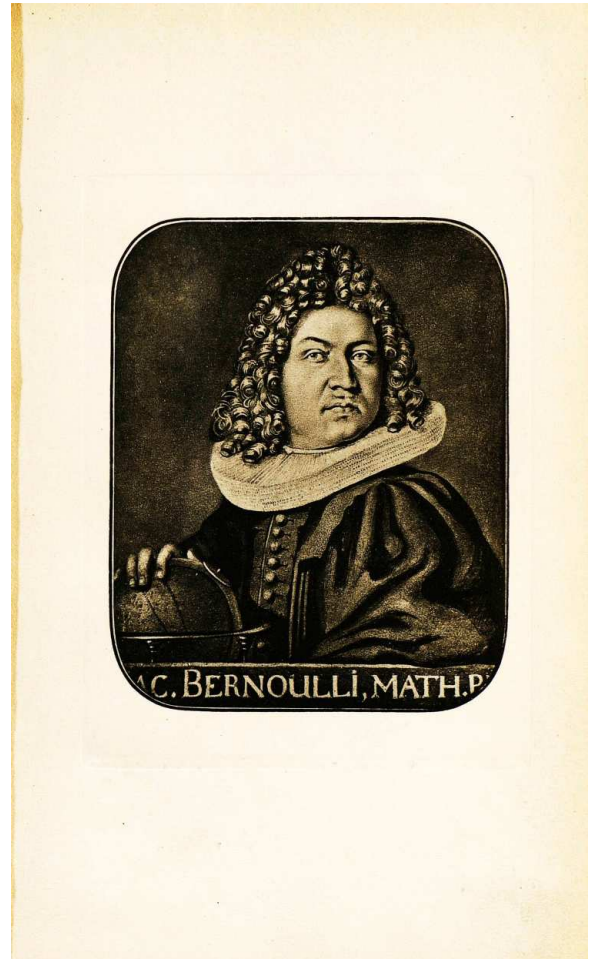


Рис. 2: Якоб Бернулли (Изображение с сайта Wikimedia Commons, Public Domain).

На заседании Академии наук 20 октября 1913 года А. А. Марковым представлено 2 экземпляра 3-го издания, один из которых веленевый, в кожаном переплете (1913 г.) труда “Исчисление вероятностей”, выпущенного в свет к 200-летию закона больших чисел. Издание было снабжено портретом Якоба Бернулли [8].

1 декабря 1913 года в большом конференц-зале Императорской Академии наук состоялось торжественное собрание, посвященное 200-летию закона больших чисел. Первым выступил А. В. Васильев с докладом “Вопросы теории вероятностей до теоремы Якоба Бернулли”. Затем состоялось выступление А. А. Маркова. Его доклад “Очерк развития за-

кона больших чисел как совокупность математических теорем” привлек всеобщее внимание. “Желающих попасть на выступление А. А. Маркова оказалось столько, что мест в зале не хватило и многим пришлось слушать ученого, находясь в соседнем помещении” [7]. Заседание закрылось сообщением А. А. Чупрова “Закон больших чисел в современной науке”.

Речь А. А. Маркова, произнесенная на заседании Академии наук 1 декабря 1913 года, посвященном 200-летию закона больших чисел, опубликована впервые на страницах журнала “Вестник Опытной физики и Элементарной математики”, с 603 [9]. Выступая 1 декабря 1913 года на торжественном заседании Российской Императорской Академии наук, посвященном празднованию двухсотлетия закона больших чисел, выдающийся русский математик, академик Андрей Андреевич Марков так охарактеризовал величие Якоба Бернулли: “В биографиях его упоминается, что следуя примеру Архимеда, он завещал начертить на его памятнике логарифмическую спираль и сделать надпись: “Eadem mutata resurgo” (измененная я возрождаюсь прежней). Надпись эта, конечно, указывает на найденные им свойства кривой. Но она имеет и другой смысл. В ней выражается надежда Бернулли на воскресение и вечную жизнь. Мы можем сказать, что надежда его осуществляется. Со времени смерти Бернулли прошло более 200 лет, однако он живет и будет жить в своей теореме” [5].

## 4 Итоги и выводы

Работа Я. Бернулли всегда ценилась очень высоко. В этой работе “впервые была опубликована и доказана знаменитая ... теорема, положившая начало закону больших чисел ... Свою теорему Я. Бернулли высказал точно и доказал с полной строгостью” [5]. В заключении отметим, четвертая часть “Ars Conjectandi” впервые была переведена на русский язык в 1913 г. [10], второе издание с комментариями 1986 г. [8]. Первые три части “Искусства предположений” в России недостаточно известны. В 2006 году в Берлине появился русский перевод книги “Искусство предположений” частей 1-3 О. Б. Шейнина [11], автора комментария “Якоб I Бернулли и начало теории вероятностей” [5].

## Список литературы

- [1] В. В. Бобынин. Якоб I Бернулли и теория вероятностей (с портретом). По поводу исполнения в 1913 году 200-летия со дня появления в свет книги “Ars Conjectandi”. *Математическое образование*, 4:161–168, 1914.
- [2] Р. Мизес. *Вероятность и статистика*. КомКнига, М., 2007.
- [3] Ю. В. Прохоров. *Вероятность и математическая статистика: Энциклопедия*. Большая Российская энциклопедия, М., 2003.
- [4] Б. В. Гнеденко. *Очерки по истории математики в России*. КомКнига, М., 2005.

- [5] Я. Бернулли. *О законе больших чисел*. Наука. Гл. ред. физ.-мат.лит., М., 1986.
- [6] Л. Е. Майстров. *Теория вероятностей: исторический очерк*. Наука, М., 1967.
- [7] С. Я. Гродзенский. *Андрей Андреевич Марков. 1856-1922*. Наука, М., 1987.
- [8] А. А. Марков. *Исчисление вероятностей: с портретом Якоба Бернулли. К 200-летию закона больших чисел*. Тип. Имп. Акад. наук, Санкт-Петербург, 1913.
- [9] А. А. Марков. Двухсотлетие закона больших чисел. *Вестник Опытной физики и Элементарной математики*, 603:59–64, 1914.
- [10] Я. Бернулли. *Четвертая часть сочинения Якоба Бернулли “Ars Conjectandi”: С портр. Якоба Бернулли. 200-летнему юбилею закона больших чисел*. Тип. Имп. Акад. наук, Санкт-Петербург, 1913.
- [11] Я. Бернулли. *Искусство предположений. Ч. 1-3*. N-G Verlag, Berlin, 2006.

## О подходах к постановкам оптимизационных задач с неопределенными данными

Ольга Аркадьевна Попова

Институт космических и информационных технологий  
Сибирский федеральный университет  
Красноярск  
OlgaArc@yandex.ru

**Аннотация.** В статье изучается вопрос необходимости учета неопределенных факторов при решении практических задач оптимизации и принятия решений. В рамках неопределенного программирования приводится классификация неопределенностей и описываются общие подходы к постановке и решению оптимизационных задач в условиях стохастической и интервальной неопределенности.

**Ключевые слова.** Неопределенное программирование, стохастическая оптимизация, оператор математического ожидания, численный вероятностный анализ.

### 1 Введение

Многие практические задачи, в том числе задачи принятия решений, требуют для своего изучения применения оптимизационного подхода. Эффективность полученных решений определяется рядом факторов. К таким факторам в первую очередь можно отнести данные, необходимые для описания и решения поставленной проблемы.

Необходимость учитывать характер и особенности данных показал исторический опыт применения методов линейного программирования, а именно кризис линейного программирования в шестидесятых и семидесятых годах двадцатого века. Стремительное развитие теории линейного программирования при решении практических задач планирования национальной экономики столкнулось с проблемой несоответствия теоретических предположений и реальных результатов при решении конкретных экономических задач. Результаты решения задач линейного программирования часто не соответствовали ожиданиям составителей планов. Одним из существенных факторов, которые не учитывались в предлагаемых моделях и методах, была неопределенность входных данных – явление, присущее многим практическим задачам экономики.

На актуальность учета неопределенных факторов в данных указывает Технический Актуарный Стандарт D Данные (ТАС) (Великобритания, ноябрь

2009), в котором отмечается, что при работе с данными должен быть установлен определенный уровень их надежности. Надежность данных заключается в том, что пользователи, для которых была создана некая актуарная информация, должны иметь возможность в значительной степени доверять релевантности информации, прозрачности допущений, полноте и понятности, включая сообщения о любой неопределенности, присущей информации. В нем также отмечается, что любые действия, предпринятые по причине неточности или неполноты данных, увеличивают надежность полученной актуарной информации.

В случаях, когда требуемые данные являются существенно неполными или неадекватными, проводится оценка для определения того, может ли надежность данных быть улучшена путем их изменения или дополнения.

Примером способа, которым данные могут быть скорректированы, является замена средними значениями неверных или отсутствующих записей.

Данные могут быть дополнены с использованием дополнительных источников данных, косвенных методов или методов выборки. Любые дополнительные источники данных должны быть проверены на надежность.

В некоторых случаях корректировки могут быть внесены в допущения, используемые в моделях или расчетах, или в их результаты вместо корректировки неполных или неточных данных или в дополнение к ней. Любые такие изменения должны документироваться, а их влияние быть выявлено.

Понятие “неопределенные данные” трактуется учеными, исследователями и практиками достаточно широко и неоднозначно. Различают, например, элиторную и эпистемистическую неопределенность. Для решения экономических некоторыми исследователями используется понятие экономическая неопределенность [1].

В рамках понятия “неопределенные данные” можно выделить три типа неопределенности: стохастическую, нечеткую и неточную.

Природа неопределенных данных может быть слу-



чайной, связана с ошибками измерений, или определяться неполнотой информации. Тогда говорят о случайных, неточных, неполных данных.

Можно найти множество аргументов, в поддержку, какую сторону или тип неопределенности подвергнуть исследованию, однако, понятно, что результаты полученных решений зависят от количества и качества соответствующей информации, доступных и ограниченных познавательных способностей лиц, принимающих решения, а также численных методов, выбранных для расчета.

## 2 Неопределенное программирование

Присутствие неопределенной информации в данных при решении оптимизационных задач требует от исследователя более сложных методов ее представления и более сложных методов расчета показателей. Чем более точно будет представлена неопределенность и чем лучше она будет учтена методом, тем эффективнее будет полученное решение.

Важно заметить, что после представления полученных неопределенностей формируется проблема выбора метода, который позволит осуществить последующие расчеты таким образом, чтобы получить реальные результаты, с тем чтобы не получить дополнительные неопределенности [1].

В настоящее время с этой целью развивается математический аппарат неопределенного программирования. Неопределенное программирование представляет собой теоретические основы решения оптимизационных задач в условиях различных видов неопределенности.

В работах [2] известный специалист в области теории и практики решения оптимизационных задач в условиях неопределенности B.Lju при описании теории неопределенного программирования выделяет три основных вида неопределенности: случайность, нечеткость и неточность.

При этом случайная величина представляет собой измеримую функцию из вероятностного пространства на вещественную прямую. Нечеткая переменная есть функция из возможностного пространства на вещественную прямую. Неточная величина — это измеримая функция из пространства приближений на вещественную прямую.

В работе [2] отмечается, что все большее и большее внимание привлекает многократная неопределенность. Например, нечетко-случайная величина представляет собой измеримую функцию из вероятностного пространства в набор нечетких величин, в то время как случайно-нечеткая величина есть некоторая функция из возможностного пространства в набор случайных величин. Вводятся также

следующие виды величин с двукратной неопределенностью: случайно-неточная величина, неточно-случайная величина, нечетко-неточная величина, неточно-нечеткая величина, бислучайная (случайно-случайная) величина, бинечеткая (нечетко-нечеткая) величина, бинеточная (неточно-неточная) величина, а также даются определения величин с многократной неопределенностью более общего вида, таких как трислучайная (trirandom) величина, тринечеткая (trifuzzy) величина и тринеточная (trirough) величина.

Каждому виду неопределенной величины соответствует определенный вид математического программирования. Такой подход представляет собой важный шаг в решении проблемы учета неопределенности в задачах оптимизации.

## 3 Оптимизационные задачи неопределенного программирования

Неопределенность данных существенным образом сказывается на постановках и решениях оптимизационных проблем. И в первую очередь это связано с тем, что задачи оптимизации в конечномерных пространствах могут быть охарактеризованы некоторым числом фиксированных входных параметров, которые определяют ее структуру и методы решения. В задачах линейного и нелинейного программирования такими фиксированными параметрами являются коэффициенты целевой функции, матрицы ограничений и правых частей ограничений. Решение таких оптимизационных задач состоит в поиске оптимального решения для заданных входных фиксированных параметров.

Входные данные моделей математического программирования часто точно неизвестны или (и) носят случайный характер. В этих условиях для целей моделирования часто берутся “средние” величины коэффициентов, а далее получается некоторое решение, оптимальное в данной модели, но которое не всегда является оптимальным в исходной задаче.

Одним из подходов, который работает с неточными (случайными) коэффициентами в задачах линейного и нелинейного программирования и “пытается встроить” влияние неточности коэффициентов в модель, является стохастическое программирование. Прогресс в этой области был получен в шестидесятые и семидесятые годы XX века и связан с именами Дж. Уэтса, А. Прекопа и К. Кала.

Среди других подходов к решению оптимизационных задач в условиях неопределенности данных являются нечеткое и неточное программирование,

В основе нечеткого программирования лежит теория нечетких множеств и нечеткая арифметика. Неточ-



ное программирование опирается на теорию неточных множеств и неточную арифметику.

Теория неточных множеств (rough set theory), начало которой было положено Павлаком, представляет собой “математическое средство для работы с неясными описаниями объектов”. Основными рабочими понятиями теории неточных множеств являются: пространство приближений, неточная величина, неточная арифметика, мера доверия, оператор ожидаемого значения.

Под неточным программированием понимается теория оптимизации в неточной среде.

Анализ публикаций показал, что основной подход к решению оптимизационных задач нечеткого программирования содержит идею аппроксимировать нечеткое множество двумя другими четкими множествами, которые порождаются отношением эквивалентности (рефлексивным, симметричным, транзитивным) и определяют верхнее и нижнее приближения. На основе теории неточных множеств строятся модели неточного ожидаемого значения (неточная EVM-модель) и модели с ограничениями на шансы (неточная ССР-модель).

Поскольку интервальное число можно рассматривать как специальный вид неточной величины, то в раздел неточного программирования попадает интервальный анализ, интервальная арифметика и, соответственно, интервальное программирование.

Подходы и методы неопределенного программирования имеют ряд достоинств и определенные недостатки. Среди достоинств, следует указать на прикладной аспект данной теории, предлагающей методы решения различных оптимизационных задач с широким спектром неопределенностей. К числу таких задач можно отнести, например, задачи массового обслуживания, производственно - финансовые проблемы, задачи, присущие энергетическим и экологическим системам, широкий круг задач из области принятия оптимальных (эффективных) решений. Для каждого вида неопределенностей разрабатывается своя арифметика неопределенностей. Развитие теории неопределенного программирования способствует разработке нового унифицированного подхода к решению оптимизационных задач в неопределенных средах.

В данном направлении следует отметить разработку гибридных алгоритмов совмещающих в себе идеи и подходы статистического моделирования, нейронные сети, генетические алгоритмы, имитационный отжиг и табу-поиск [2].

С вычислительной точки зрения необходимо создание более эффективных и мощных численных алгоритмов. Это касается прежде всего статистического моделирования. В тех случаях, когда извест-

ны функции плотности вероятности для стохастических данных, можно с успехом использовать метод Монте-Карло. При всех его положительных качествах этот метод обладает рядом недостатков. Один из самых существенных — низкая скорость сходимости и большой объем вычислений. В некоторых ситуациях, альтернативой методу Монте-Карло является Численный вероятностный анализ, который позволяет в условиях элиторной и эпистемистической неопределенности решать различные практические задачи, в том числе задачи, содержащие оптимизационные постановки неопределенного программирования. Отметим также проблемы, связанные с численным преобразованием задачи стохастического линейного программирования в детерминированную задачу нелинейного программирования. Хорошо известен тот факт, что алгоритмы нелинейного программирования на практике применимы к задачам с относительно небольшой размерностью.

В большинстве алгоритмах неопределенного программирования применяется оператор математического ожидания и проводятся процедуры усреднения.

#### 4 О постановках и методах решений задач стохастического программирования

В формулировке задач кнеопределенной оптимизации, обычно пытаются найти хороший компромисс между реалистичностью модели оптимизации и ее позвольительной способностью в выборе и дальнейшем использовании соответствующего численного метода решения изучаемой проблемы. Эти две составляющие в совокупности обычно влияют на полезность и качество получаемых решений. В результате этих соображений существует большое количество различных подходов к постановке и решению задач оптимизации в условиях неопределенности. Невозможно дать полный обзор всех таких моделей и методов в одной статье. Поэтому остановимся лишь на стохастическом подходе к решению оптимизационных задач.

Рассмотрим общую постановку задачи стохастического программирования (Stochastic Programming Problem (SPP)) [3]

$$\begin{aligned} \max f(x, \xi), \\ g_i(x, \xi) \leq 0, \quad i = 1, \dots, p, \end{aligned}$$

где  $x$  — вектор решения,  $\xi$  — случайный вектор,  $f(x, \xi)$  — целевая функция,  $g_i(x, \xi)$  — случайные функции ограничений.

С целью применения соответствующих подходов к решению оптимизационных задач в условиях стохастической неопределенности общая задача СП может быть сформулирована в М и Р постановках по

отношению к записи целевой функции и ограничений. М-постановка означает оптимизацию математического ожидания целевой функции и представляет собой первый тип постановки задачи стохастического программирования (Stochastic Programming Problem (SPP)) [3]. В зарубежной литературе такие задачи называются моделями ожидаемого значения или expected model value (EMV) [2]. М-постановка имеет вид

$$\max M[f(x, \xi)] \quad (1)$$

$$M[g_i(x, \xi)] \leq 0, \quad i = 1, \dots, p. \quad (2)$$

Во многих случаях задача стохастической оптимизации может ставиться как многокритериальная. В этом случае имеет место многокритериальное стохастическое программирование.

Наряду с EMV часто пользуются модели стохастического программирования с регрессом, которая имеет вид

$$\min_x M[f(x, \xi) + Q(x, \xi)].$$

$Q$  — регрессная функция и может быть выбрана как

$$Q(x, \xi) = \min_y \{q(y) | G_j(y|x, \xi) \leq 0, j = 1, \dots, p\}.$$

$G_j(y|x, \xi)$  — функции ограничений от  $y$  при заданных значениях решения  $x$  и реализованного значения вектора  $\xi$  для  $j = 1, \dots, p$ .

Суть Р-постановки заключается в том, что необходимо найти такие значения  $x_j$ , при которых максимизируется вероятность того, что целевая функция будет не хуже предельно допустимого значения. Ограничения задачи, которые должны выполняться при всех реализациях параметров условий задачи, называются жесткими ограничениями.

Часто возникают ситуации, в которых постановка задачи позволяет заменить жесткие ограничения их усреднением по распределению случайных параметров. Такие ограничения называют статистическими: т.е. вероятность выполнения каждого заданного ограничения должна быть не менее назначенной величины  $\alpha$ .

Р-постановка задачи СП имеет вид

$$\min P\{f(x, \xi) \geq \alpha\} \quad (3)$$

при ограничениях

$$P\{g_i(x, \xi) \leq 0\} \geq P_i, \quad i = 1, \dots, p. \quad (4)$$

где  $\alpha$   $P_i$  — некоторые числа,  $P$  — вероятность.

Возможные некоторые комбинации задач (1), (2) и (3), (4).

Например, найти минимум (1) при условиях (4) или минимум (3) при условиях (2). Несмотря на кажущееся различие в постановках задач (1), (2) и (3), (4),

они могут быть сведены к некоторой общей формулировке, например вида (1), (2). Для этого необходимо ввести характеристические функции:

$$h_0(x, \xi) = \begin{cases} 1, & \text{если } f(x, \xi) \geq \alpha; \\ 0, & \text{в противном случае;} \end{cases}$$

$$h_i(x, \xi) = \begin{cases} 1, & \text{если } g_i(x, \xi) \leq 0; \\ 0, & \text{в противном случае;} \end{cases}$$

для которых

$$M[h_0(x, \xi)] = P\{f(x, \xi) \geq \alpha\};$$

$$M[h_i(x, \xi)] = P\{g_i(x, \xi) \leq 0\}.$$

Задача (3), (4) тогда приводится к виду

$$\min M[h_0(x, \xi)]$$

при условии

$$M[h_i(x, \xi)] \geq P_i, \quad i = 1, \dots, p\}.$$

Существует два основных подхода к решению задач стохастического программирования:

1) не прямые методы, которые заключаются в нахождении функций  $F(x)$ ,  $G_i$  и решении эквивалентной задачи НП вида (1), (2);

2) прямые методы стохастического программирования, основанные на информации о значении функций  $f(x, \xi)$ ,  $g_i(x, \xi)$ , получаемой в результате проведения экспериментов.

Рассмотрим некоторые варианты моделей задач стохастического программирования

1. Пусть задана задача линейного стохастического программирования с вероятностными ограничениями типа

$$\max M(c, x) \quad (5)$$

при условиях

$$P\left\{\sum_{j=1}^n a_{ij}x_j \leq b_i\right\} \geq \alpha_i, \quad i = 1, \dots, m, \quad (6)$$

$$x_j \geq 0, \quad i = 1, \dots, n. \quad (7)$$

При детерминированной матрице  $A$  и случайном векторе  $b$  задача (5)–(7) сводится к эквивалентной детерминированной задаче ЛП следующим образом.

Пусть  $p(b_1, b_2, \dots, b_m)$  — совместная плотность распределения составляющих  $b_i$  случайного вектора  $b$ . Находим плотность распределения  $b_i$ :

$$P(b_i) = \int \dots \int p(b_1, b_2, \dots, b_m) db_1 db_2 \dots db_{i-1} db_{i+1} \dots db_m.$$

Вычислим  $\tilde{b}_i$  из уравнения

$$\int_{b_i}^{\infty} P(b_i) db_i = \alpha_i, i = 1, \dots, m, \quad (8)$$

Очевидно, что условия (6) при этом эквивалентны неравенствам

$$\sum_{j=1}^n a_{ij} x_j \leq \tilde{b}_i,$$

Отсюда следует, что задаче стохастического программирования (5)–(7) будет эквивалентна следующая детерминированная задача ЛП:

$$(\bar{c}, x) \rightarrow \max$$

при условиях

$$\begin{aligned} \sum_{j=1}^n a_{ij} x_j &\leq \tilde{b}_i, \\ x_j &\geq 0, \end{aligned}$$

где  $\bar{c} = M[c]$ .

Рассмотренные подходы на основе стохастического программирования имеют следующие два практических недостатка. Первый связан с численным преобразованием задачи стохастического линейного программирования в детерминированную задачу нелинейного программирования. Хорошо известен тот факт, что алгоритмы нелинейного программирования на практике применимы к задачам с относительно небольшой размерностью. Во-вторых, основное предположение в задачах стохастического линейного программирования состоит в том, что вероятностные распределения (их функции распределения или функции плотности распределения) известны заранее. Это требование обычно не выполняется. Коэффициенты являются неточными, но дополнительная информация не имеет вероятностной природы. Гораздо чаще эти коэффициенты оцениваются экспертами и получаются, в конечном, счете по совокупности уровней неточности или неопределенности проблемы.

Следует указать на относительно новые постановки оптимизационных задач в условиях интервальной неопределенности. Например, задача линейного программирования с интервальными данными формулируется следующим образом [4]:

$$(c, x) \rightarrow \min, \quad (9)$$

$$Ax = b, x \geq 0. \quad (10)$$

$$A \in \mathbf{A}, b \in \mathbf{b}, c \in \mathbf{c}, \quad (11)$$

где  $\mathbf{A}$  — интервальная матрица,  $\mathbf{b}$ ,  $\mathbf{c}$  — интервальные векторы размерности  $n$ .

Аналогично задаче с интервальными данными в рамках численного вероятностного анализа можно поставить задачу (1)–(10) линейного программирования со случайными данными.

$$A \in \mathbf{A}, b \in \mathbf{b}, c \in \mathbf{c}, \quad (12)$$

где  $\mathbf{A}$  — случайная матрица,  $\mathbf{b}$ ,  $\mathbf{c}$  — случайные векторы размерности  $n$ . Точка  $x_*$  — решение задачи (1) – (2), если

$$(c, x^*) = \inf_U (c, x),$$

где

$$U = \{x | Ax = b, x \geq 0\}$$

Множество решений (1), (10), (12) определяются следующим образом

$$\mathcal{X} = \{x | (c, x) \rightarrow \min, Ax = b, x \geq 0, A \in \mathbf{A}, b \in \mathbf{b}, c \in \mathbf{c}\}$$

Заметим, что  $x^*$  — случайный вектор, поэтому в отличие от детерминированной задачи, для  $x^*$  необходимо определять функции плотности вероятности для каждой компоненты  $x_i^*$  как совместную плотность вероятности.

## 5 Заключение

Постановки оптимизационных задач в условиях неопределенности дают новые направления развития теории и практики неопределенного программирования. В настоящее время в этом контексте отметим разработку гибридных алгоритмов, успешно соединяющих в себе статистическое моделирование, генетические методы и нейронные сети [2], новые подходы и методы линейной оптимизации с неточными данными [4], численные вероятностный анализ для решения оптимизационных задач с случайными данными [5].

## Список литературы

- [1] Hans Schjaer-Jacobsen. Representation and calculation of economic uncertainties: Intervals, fuzzy numbers, and probabilities. *Int. J. Production Economics*, 78:91–98, 2002.
- [2] B. Liu. *Theory and Practice of Uncertain Programming (2nd Edition)*. Springer-Verlag, Berlin, 2009.
- [3] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM, Philadelphia, 2009.
- [4] М. Фидлер, Й. Недома, Я. Рамик, И. Рон и К. Циммерманн. *Задачи линейной оптимизации с неточными данными*. НИЦ “РХД”, М.– Ижевск, 2008.
- [5] Б. С. Добронев и О. А. Попова. Численный вероятностный анализ для исследования систем в условиях неопределенности. *Вестник Томского государственного университета. Управление, вычислительная техника и информатика*, 21(4):39–46, 2012.

# Об обобщении результатов Никольского С.М. для погрешности усложненных квадратурных формул на случай класса функций, имеющих дробные производные

Севастьянова Нэлли Александровна

Сибирский федеральный университет  
Торгово-экономический институт  
Красноярск  
nelliseva@mail.ru

**Аннотация.** Целью данной работы является исследование погрешности приближенного интегрирования с помощью усложненных квадратурных формул функций, имеющих дробную производную порядка  $\alpha$ , суммируемую на отрезке  $[a, b]$  в степени  $p$ .

**Ключевые слова.** Квадратурная формула, погрешность, функционал ошибок, дробные производные.

## 1 Основные тезисы

В теории приближенного интегрирования хорошо известен алгоритм, позволяющий по фиксированной квадратурной формуле строить другие формулы сходного строения с большим количеством узлов. В общем случае данный процесс выглядит так:

— выбираем некоторую «базовую» квадратурную формулу:

$$\int_0^1 f(x)dx \approx \sum_{j=0}^{m-1} c_j f(x_j) \quad (1)$$

— отрезок интегрирования  $[a, b]$  делим на  $n$  равных частей шагом

$$h = \frac{b-a}{n}$$

и на каждом интервале  $[x_i, x_{i+1}]$ ,  $i = 0, 1, 2, \dots, n$  применяем квадратурную формулу. подобную формуле (1)

$$\int_{x_i}^{x_{i+1}} f(x)dx \approx \sum_{j=0}^{m-1} hc_j f(hx_j + hi + a)$$

— тогда на интервале интегрирования  $[a, b]$  получаем

$$\int_a^b f(x)dx \approx \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} hc_j f(hx_j + hi + a). \quad (2)$$

Для удобства будем рассматривать не сами квадратурные формулы (2), а их функционалы ошибок

$$(l^h, f) = \int_a^b f(x)dx - \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} hc_j f(hx_j + hi + a). \quad (3)$$

В работе Никольского С.М. рассматривались классы функций  $W_{L_p}^{(r)}(M; a, b)$ , имеющих абсолютно непрерывную производную до порядка  $(r-1)$  включительно и производную порядка  $r$ , суммируемую в степени  $p$

$$\left( \int_a^b |f^{(r)}(x)|^p dx \right)^{\frac{1}{p}} \leq M.$$

Для таких функций доказана оценка

$$|(l^h, f)| \leq (b-a)^{\frac{1}{q}} h^r A_p^r M, \quad (4)$$

где постоянная  $A_p^r$  выписывается в явном виде и зависит только от  $r$  и  $p$ .

В данной работе решен вопрос о возможности обобщения результата Никольского С.М. (4) на дробный показатель  $r$ .

Использован метод получения оценок погрешности с применением неравенства типа

$$|(l^h, f)| \leq \|l\|_{B^*} \|f\|_B,$$

где  $B$  — линейное нормированное пространство функций, метрика которого индуцирована на функциях из некоторого класса функций полунормой  $\|\cdot\|_B$ , вложенное в пространство непрерывных функций  $C$ , а функционал  $l$  принадлежит сопряженному к  $B$  пространству  $B^*$ . Норма  $\|l\|_{B^*}$  характеризует качество квадратурной формулы.

Классы интегрируемых функций из данной работы сходны с классами  $W_{L_p}^{(r)}(M; a, b)$  из книги Никольского С.М. Однако существенным отличием является то, что в рассматриваемом случае порядок производных у интегрируемых функций — дробный ( $\alpha$ ).

Рассматриваются пространства  $L_p^\alpha(a, b)$ . Данные пространства индуцируются на линейных многообразиях функций, имеющих дробные производные Римана-Лиувилля порядка  $\alpha$ , суммируемые в степени  $p$  полунормами

$$\|f\|_{L_p^\alpha(a, b)} = \|f^{(\alpha)}\|_{L_p(a, b)} = \left( \int_a^b |f^{(\alpha)}(x)|^p dx \right)^{\frac{1}{p}},$$

где  $f^{(\alpha)}$  — дробная производная Римана-Лиувилля.

Данне функции представимы на отрезке  $[a, b]$  в виде

$$f(x) = f(a) + f'(a)(x-a) + \dots + \frac{f^{([\alpha])}(a)}{[\alpha]!}(x-a)^{[\alpha]} + \\ + \frac{1}{\Gamma(\alpha)} \int_a^x (x-t)^{\alpha-1} g(t) dt,$$

где  $x > a$ , функция  $g(x) \in L_p(a, b)$ , т.е.

$$\int_a^b |g(x)|^p dx < \infty,$$

$\Gamma(\alpha)$  — гамма функция Эйлера. Функция  $g(x)$  называется левосторонней дробной производной Римана-Лиувилля и обозначается  $f^{(\alpha)}(x)$ .

Доказано утверждение: пусть  $\{l^h\}$  — последовательность функционалов вида (3) и выполнены условия

$$(l^h(x), x^s) = 0, s = 0, 1, 2, \dots, [\alpha].$$

Тогда при  $h \rightarrow 0$

$$\|l^h\|_{L_p^{\alpha*}(a,b)} = h^\alpha (b-a)^{\frac{1}{p}} B_p^\alpha (1 + o(1)),$$

где константа

$$B_p^\alpha = \|Q(\lambda)\|_{L_q(0,1)} \quad (5)$$

зависит только от  $\alpha$  и  $p$ .

При разложении функций

$$\left(1 + \frac{1}{i-\lambda}\right)^\alpha, \left(1 + \frac{x_j}{i-\lambda}\right)^{\alpha-1},$$

используемых в выведении функции  $Q(\lambda)$  из (5), в степенные ряды при натуральном  $\alpha$  будет лишь конечное число членов, не равных нулю, и функция  $Q(\lambda)$  в данном случае совпадает с функцией  $F_r(t)$  из работы Никольского С.М. [1].

## Список литературы

- [1] Никольский С.М. *Квадратурные формулы*. Наука, М., 1979.
- [2] Соболев С.Л. *Введение в теорию кубатурных формул*. Наука, М., 1974.
- [3] Севастьянова Н.А. *Усложненные квадратурные формулы в пространствах функций с дробными производными*. Деп в ВИНТИ 12.11.93 №2803-В93, Красноярск, 1993.
- [4] Половинкин В.И., Севастьянова Н.А. *Усложненные квадратурные формулы в пространствах функций, имеющих дробные производные. Оптимальные методы вычислений и их применение к обработке информации: Межвузовский сборник научных трудов. — Пенза: изд. Пензенского политехнического института, вып. II: 86–95, 1992.*



# Эвентологические распределения случайных множеств событий на основе ассоциативных функций

Семенова Дарья Владиславовна

Сибирский федеральный университет  
Институт математики и фундаментальной информатики  
Красноярск  
dariasdv@gmail.com

Кочанова Юлия Сергеевна

Сибирский федеральный университет  
Институт математики и фундаментальной информатики  
Красноярск  
kochanova\_y@mail.ru

**Аннотация.** В работе предлагается новый подход описания эвентологических распределений случайного множества событий с помощью аппарата ассоциативных функций. Особенность этого подхода заключается в том, что для определения эвентологического распределения случайного множества событий достаточно знать  $N$  вероятностей событий и вид ассоциативной функции.

**Ключевые слова.** Случайное множество событий, эвентологическое распределение, ассоциативная функция.

## Введение

Статистическую систему можно определить как случайное множество событий, образующих своеобразную структуру статистических взаимосвязей случайных событий друг с другом. Чем больше событий содержит данное множество событий, тем более сложными структурами вероятностных зависимостей событий оно обладает [1, 2, 3, 4]. В структуру вероятностных зависимостей входят виды зависимости между событиями, которые составляют пары событий, тройки событий и более мощные подмножества событий данного множества. Если мощность рассматриваемого множества событий равна  $N$ , то имеется  $2^N$  видов вероятностных зависимостей между событиями этого множества, т.е. ровно столько, сколько у этого множества подмножеств. Таким образом, каждое множество из  $N$  событий характеризуется набором из  $2^N$  вероятностей, которые играют для конечного множества событий ту же роль, что и вероятностное распределение — для случайной величины с конечным множеством значений. В эвентологии набор из  $2^N$  вероятностей, определяющий  $N$ -множество событий, называется эвентологическим распределением этого множества событий. Таким образом, эвентологические распределения — это класс специальных вероятностных распределений, характеризующих множества событий. В эвентологическом распределении множества событий содержится всеобъемлющая информация о всех видах вероятностных зависимостей событий из данного

множества. Следовательно, изучение структур статистических взаимосвязей случайных событий означает, в итоге, изучение вероятностных распределений соответствующих случайных множеств событий.

В работе предлагается новый подход описания эвентологических распределений случайных множеств событий с помощью аппарата ассоциативных функций. Особенность этого подхода заключается в том, что для определения эвентологического распределения случайного множества событий достаточно знать  $N$  вероятностей событий и вид ассоциативной функции.

## 1 Предварительные сведения

### 1.1 Случайное множество событий

Рассмотрим всеобщее вероятностное пространство  $(\Omega, \mathcal{F}, \mathbf{P})$ . Пусть  $\mathfrak{X} \subset \mathcal{F}$  — конечное множество событий, выбранное из алгебры  $\mathcal{F}$  этого пространства.

Эвентологическое распределение (Э-распределение) случайного множества событий  $\mathfrak{X} \subseteq \mathcal{F}$  можно представить несколькими эквивалентными распределениями вероятностей, порожденных этим множеством событий [3]:

- Э-распределение I-го рода — набор из  $2^{|\mathfrak{X}|}$  вероятностей вида<sup>1</sup>  $\{p(X), X \subseteq \mathfrak{X}\}$ , где

$$p(X) = \mathbf{P}(K = X) = \mathbf{P}\left(\bigcap_{x \in X} x \bigcap_{x \in X^c} x^c\right);$$

- Э-распределение II-го рода — набор из  $2^{|\mathfrak{X}|}$  вероятностей вида  $\{p_X, X \subseteq \mathfrak{X}\}$ , где

$$p_X = \mathbf{P}(X \subseteq K) = \mathbf{P}\left(\bigcap_{x \in X} x\right).$$

Каждое из этих распределений вероятностей замечательно тем, что его одного достаточно для полного

<sup>1</sup>Заметим, что под  $\bigcap_{x \in X} x \bigcap_{x \in X^c} x^c$  имеется ввиду

$$\left(\bigcap_{x \in X} x\right) \cap \left(\bigcap_{x \in X^c} x^c\right).$$

определения эвентологического распределения множества случайных событий  $\mathfrak{X}$ ,  $|\mathfrak{X}| = N$ . Эвентологические распределения I и II рода связаны взаимнообратными формулами обращения Мёбиуса [3].

$$p_X = \sum_{X \subseteq Y} p(Y), \quad (1)$$

$$p(X) = \sum_{X \subseteq Y} (-1)^{|Y|-|X|} p_Y. \quad (2)$$

Следует заметить, что только *распределение I-го рода*  $p(X)$ ,  $X \subseteq \mathfrak{X}$ , обладает привычным для распределения вероятности свойством нормировки,

$$\sum_{X \subseteq \mathfrak{X}} p(X) = 1,$$

которое очевидно следует из того, что соответствующие<sup>2</sup> события  $\text{ter}(X) = \bigcap_{x \in X} x \bigcap_{x \in X^c} x^c$ ,  $X \subseteq \mathfrak{X}$ , образуют разбиение пространства элементарных событий  $\Omega$ . А события  $\text{ter}_X = \bigcap_{x \in X} x$ ,  $X \subseteq \mathfrak{X}$  образуют не разбиение, а всего лишь покрытие  $\Omega$ , то соотношение нормировки для соответствующих им распределений вероятностей, разумеется, не выполняется. Поэтому суммы вероятностей этих событий всегда больше единицы.

Будем говорить, что  $\mathfrak{E}$ -распределение II рода является легитимным, если при переходе по формулам обращения Мёбиуса к  $\mathfrak{E}$ -распределению I рода мы будем получать легитимное распределение, то есть  $0 \leq p(X) \leq 1$ ,  $X \subseteq \mathfrak{X}$ , и  $\sum_{X \subseteq \mathfrak{X}} p(X) = 1$ .

Зависимость событий, отсутствие независимости событий в пределах данного множества событий — относительное понятие, которое определяется эвентологическим распределением данного множества событий; иначе говоря, определяется вероятностью  $\mathbf{P}$  и конечным множеством событий  $\mathfrak{X} \subseteq \mathcal{F}$ , выбранным из алгебры событий  $\mathcal{F}$  вероятностного пространства  $(\Omega, \mathcal{F}, \mathbf{P})$  [3, 5]. В работе [3] был предложен удобный инструмент анализа структур эвентологической зависимости — ковариация. Ковариация двух событий  $x, y \in \mathfrak{X}$  определяется как величина

$$\text{Kov}_{xy} = \mathbf{P}(x \cap y) - \mathbf{P}(x)\mathbf{P}(y),$$

которая обращается в нуль, когда эти события независимы; больше нуля, когда они наступают вместе чаще (статистически притягиваются), и меньше нуля когда реже (статистически отталкиваются), чем в независимой ситуации. Ковариация событий служит мерой аддитивного отклонения событий от независимой ситуации. Для множества событий  $X \subseteq \mathfrak{X}$  арная

<sup>2</sup>террасные события [3]

ковариация определяется по формуле [3]:

$$\text{Kov}_X = \mathbf{P}\left(\bigcap_{x \in X} x\right) - \prod_{x \in X} \mathbf{P}(x), \quad X \subseteq \mathfrak{X}. \quad (3)$$

В эвентологии, как и в классической теории вероятностей для вероятностей пересечения множеств событий получены границы Фреше [3]

$$\max \left\{ \sum_{x \in X} p_x - |X| + 1, 0 \right\} \leq \mathbf{P}\left(\bigcap_{x \in X} x\right) \leq \min_{x \in X} p_x.$$

Заметим, что любой набор из  $2^N$  вероятностей  $\mathbf{P}(X)$ ,  $X \subseteq \mathfrak{X}$ , удовлетворяющих границам Фреше будет определять некоторую структуру зависимости случайного множества событий (легитимное  $\mathfrak{E}$ -распределение).

## 1.2 Треугольные нормы

Треугольные нормы ( $t$ -нормы) играют важную роль в различных математических дисциплинах, таких как вероятностные метрические пространства, многозначные логики и, в частности, нечеткая логика, нечеткие множества и др. Своему названию треугольные нормы обязаны К.Менгеру<sup>3</sup> [6, 7], который в середине прошлого века ввел специальный класс двухместных функций на единичном квадрате, позволяющих расширить неравенство треугольника от классических метрических пространств для более общего случая вероятностных метрических пространств. В середине 1950-ых А. Скляр и Б. Швейцер начали совместную работу над теорией вероятностных метрических пространств. В 1958 ими был предложен окончательный набор аксиом для  $t$ -норм [6, 8, 7].

## 2 Ассоциативные случайные множества событий

В работах О.Ю. Воробьева [9, 3] был предложен подход применения классического понятия копулы к эвентологическим распределениям множеств событий. Суть подхода заключается в следующем. Пусть  $\mathfrak{X} \subseteq \mathcal{F}$  — множество избранных событий. И пусть известны вероятности событий  $p_x$ ,  $x \in \mathfrak{X}$ . Тогда копула связывает вероятность пересечения множества событий  $\mathfrak{X}$  с их вероятностями

$$\mathbf{P}\left(\bigcap_{x \in \mathfrak{X}} x\right) = \text{copula}_X(\mathbf{P}(x), x \in X; 1, x \in X^c),$$

<sup>3</sup>В 1942 в статье «Статистические метрики» Карл Менгер ввел понятие пространства, в котором расстояния определены функциями распределения вероятности. Чтобы расширить неравенство треугольника в этом случае, он использовал специальные функции, которые назвал треугольными нормами.

где  $\text{copula}_X$  — эвентологическая копула [9]. В [3] были определены независимые эвентологические копулы.

В работе предлагается использовать подход О.Ю. Воробьева и в качестве копулы использовать ассоциативные функции, которые будут связывать вероятности пересечения множества событий  $\mathbf{P}\left(\bigcap_{x \in X} x\right)$  с вероятностями самих событий  $\mathbf{P}(x)$ ,  $x \in X$ ,  $X \subseteq \mathfrak{X}$ .

## 2.1 Ассоциативные функции в эвентологии

В данной работе будем использовать следующее определение ассоциативной функции.

**Определение 1.** Ассоциативная функция в эвентологии (*Associative Eventological Distribution — AED*)

$$\text{AED} : [0, 1]^2 \rightarrow [0, 1]$$

определяется как двуместная функция, удовлетворяющая следующим аксиомам:

*A1. граничные условия*

$$\text{AED}(a, 0) = \text{AED}(0, a) = 0, \quad (4)$$

$$\text{AED}(a, 1) = \text{AED}(1, a) = a,$$

$$a \in [0, 1]$$

*A2. монотонность*

$$\text{AED}(a_1, b_1) \leq \text{AED}(a_2, b_2), \quad (5)$$

$$\text{когда } a_1 \leq a_2, b_1 \leq b_2$$

*A3. коммутативность, т.е. для всех  $a, b$  на  $[0, 1]$*

$$\text{AED}(a, b) = \text{AED}(b, a);$$

*A4. ассоциативность, т.е. для всех  $a, b, c \in [0, 1]$*

$$\text{AED}(\text{AED}(a, b), c) = \text{AED}(a, \text{AED}(b, c)) \quad (6)$$

*A5. условие Липшиц-непрерывности*

$$\text{AED}(c, b) - \text{AED}(a, b) \leq c - a, \quad a \leq c. \quad (7)$$

Заметим, что аксиомы A1-A4 соответствует определению  $t$ -нормы. Условие A5 соответствует теореме [7], утверждающей, что Липшиц-непрерывная  $t$ -норма является копулой. Таким образом, в эвентологии под ассоциативной функцией мы будем понимать непрерывную  $t$ -норму, удовлетворяющую условию Липшица, или, что эквивалентно, ассоциативную, коммутативную копулу.

Из свойств треугольных норм и копул [8, 10, 7] непосредственно вытекает следующее утверждение

**Утверждение 1.** Ассоциативная функция  $\text{AED}(a, b)$  удовлетворяет границам Фреше - Хёвдинга

$$\max\{a + b - 1, 0\} \leq \text{AED}(a, b) \leq \min\{a, b\}, \quad (8)$$

$$\forall a, b \in [0, 1].$$

Свойство A4 позволяет получить рекуррентную формулу [6, 8] для  $n$ -местной ассоциативной функции  $\text{AED} : [0, 1]^n \rightarrow [0, 1]$

$$\text{AED}(a_1, \dots, a_n) = \text{AED}(a_1, \text{AED}(a_2, \dots, a_n)), \quad (9)$$

$$a_1, \dots, a_n \in [0, 1].$$

## 2.2 Ассоциативное случайное множество событий

Поскольку функции AED являются коммутативными ассоциативными копулами, то мы смело можем применить данный математический аппарат для описания новых классов эвентологических распределений.

**Определение 2.** Ассоциативным случайным множеством событий называется случайное множество событий  $K$  под конечным множеством избранных событий  $\mathfrak{X}$  с  $\mathfrak{A}$ -распределением II-го рода  $\{p_X, X \subseteq \mathfrak{X}\}$ , где для всех  $X \subseteq \mathfrak{X}$ ,  $|X| > 1$  вероятности пересечения множеств событий  $p_X$  определяются рекуррентным соотношением при известных вероятностях событий  $p_x = \mathbf{P}(x)$ ,  $x \in \mathfrak{X}$

$$p_{xy} = \mathbf{P}(x \cap y) = \text{AED}(p_x, p_y)$$

$$\begin{aligned} p_{xyz} &= \mathbf{P}(x \cap y \cap z) = \text{AED}(p_x, \text{AED}(p_y, p_z)) = \\ &= \text{AED}(p_x, \mathbf{P}(y \cap z)), \end{aligned}$$

$$p_X = \mathbf{P}\left(\bigcap_{x \in X} x\right) = \text{AED}\left(p_x, \mathbf{P}\left(\bigcap_{y \in X \setminus x} y\right)\right),$$

при условии, что соответствующее  $\mathfrak{A}$  - распределение I-го рода будет легитимным.

$\mathfrak{A}$ -распределения, определяемые функциями AED будем называть ассоциативными  $\mathfrak{A}$  - распределениями. Таким образом, ассоциативное случайное множество событий полностью определяется своим ассоциативным  $\mathfrak{A}$  - распределением.

## 2.3 Семейства ассоциативных $\mathfrak{A}$ -распределений

В эвентологии под ассоциативной функцией мы понимаем непрерывную  $t$ -норму, удовлетворяющую условию Липшица, или, что эквивалентно, ассоциативную, коммутативную копулу.

В [6, 8, 10, 7] рассмотрены различные семейства  $t$ -норм и указаны условия, при которых они являются копулами. Рассмотрим некоторые из этих семейств в качестве функции AED и исследуем свойства полученных ассоциативных случайных множеств событий.

Рассмотрим всеобщее вероятностное пространство  $(\Omega, \mathcal{F}, \mathbf{P})$ . Пусть  $\mathfrak{X} \subset \mathcal{F}$  – конечное множество событий, выбранное из алгебры  $\mathcal{F}$  этого пространства. И пусть известны вероятности событий  $p_x = \mathbf{P}(x)$ ,  $x \in \mathfrak{X}$ .

Будем рассматривать ассоциативные случайные множества событий, определяемые различными функциями AED.

### 2.3.1 Независимые (независимо-точечные) случайные множества событий

Независимым (независимо точечным) случайным множеством событий называется случайное множество событий  $K$  под конечным множеством событий  $\mathfrak{X}$ , независимых в совокупности, его распределение вероятностей имеет вид

$$p(X) = \mathbf{P}(K = X) = \prod_{x \in X} \mathbf{P}(x) \prod_{x \in X^c} \mathbf{P}(x^c), \quad X \in 2^{\mathfrak{X}}.$$

Значениями независимого случайного множества  $K$  служат подмножества  $X \subseteq \mathfrak{X}$  наступивших событий, независимых в совокупности  $\mathfrak{X}$ . Дополнение независимого случайного множества  $K^c = \mathfrak{X} - K$  – независимое случайное множество.

**Теорема 1.** Ассоциативное случайное множество событий с функцией вида  $\text{AED}(a, b) = a \cdot b$  является независимо точечным случайным множеством событий.

**Доказательство.** По определению 2

$$p_{xy} = \text{AED}(p_x, p_y) = p_x \cdot p_y,$$

$$p_{xyz} = \text{AED}(p_x, \text{AED}(p_y, p_z)) = p_x \cdot p_y \cdot p_z.$$

В общем случае для  $X \subseteq \mathfrak{X}$

$$p_X = \text{AED} \left( p_x, \mathbf{P} \left( \bigcap_{y \in X \setminus x} y \right) \right) = \prod_{x \in X} p_x.$$

В [3] показано, что Э-распределение вероятностей II-го рода независимо точечного случайного множества событий имеет вид

$$p_X = \mathbf{P} \left( \bigcap_{x \in X} x \right) = \prod_{x \in X} p_x, \quad X \subseteq \mathfrak{X}.$$

◇

### 2.3.2 Вложенные случайные множества событий

Два случайных события  $x, y \in \mathfrak{X}$ ,  $x \neq \emptyset, y \neq \emptyset$  называются вложенными, если между ними возможны только два отношения

$$x \cap y = \begin{cases} x, \\ y, \end{cases}$$

то есть одно из этих событий вложено в другое:  $x \subseteq y$  или  $y \subseteq x$ .

Если каждые два события из  $X$  вложенные, то  $X$  называется множеством вложенных событий, или множеством случайных событий с вложенной эвентологической структурой.

**Теорема 2.** Ассоциативное случайное множество событий с функцией вида  $\text{AED} = \min\{a, b\}$  является случайным множеством вложенных событий.

**Доказательство.** По определению 2

$$p_{xy} = \text{AED}(p_x, p_y) = \min\{p_x, p_y\},$$

$$p_{xyz} = \text{AED}(p_x, \text{AED}(p_y, p_z)) = \min\{p_x, p_y, p_z\}.$$

В общем случае для  $X \subseteq \mathfrak{X}$

$$p_X = \text{AED} \left( p_x, \mathbf{P} \left( \bigcap_{y \in X \setminus x} y \right) \right) = \min_{x \in X} p_x.$$

В [3] показано, что Э-распределение вероятностей II-го рода случайного множества вложенных событий имеет вид

$$p_X = \mathbf{P} \left( \bigcap_{x \in X} x \right) = \min_{x \in X} p_x, \quad X \subseteq \mathfrak{X}.$$

◇

### 2.3.3 Фреше-граничное случайное множество событий

Фреше-граничным случайным множеством событий называется случайное множество событий  $K$  под конечным множеством событий  $\mathfrak{X}$ , распределение вероятностей которого имеет вид

$$p_X = \mathbf{P} \left( \bigcap_{x \in X} x \right) = \max \left\{ \sum_{x \in X} p_x - |X| + 1, 0 \right\},$$

$X \subseteq \mathfrak{X}$ .

**Теорема 3.** Ассоциативная функция  $\text{AED}(a, b) = \max\{a + b - 1, 0\}$  определяет

1. случайное множество с непересекающейся структурой зависимостей, если вероятности

событий  $p_x = \mathbf{P}(x)$ ,  $x \in \mathfrak{X}$  удовлетворяют системе из  $2^{|\mathfrak{X}|} - |\mathfrak{X}| - 1$  неравенств

$$\sum_{x \in X} p_x \leq |X| - 1, \quad X \subseteq \mathfrak{X}, \quad |X| > 1. \quad (10)$$

2. Фреше-граничное случайное множество событий, которые пересекаются с наименьшими возможными вероятностями, если вероятности событий  $p_x = \mathbf{P}(x)$ ,  $x \in \mathfrak{X}$  удовлетворяют системе из  $2^{|\mathfrak{X}|} - 1$  неравенств

$$|X| - 1 \leq \sum_{x \in X} p_x \leq |X|, \quad X \subseteq \mathfrak{X}. \quad (11)$$

**Доказательство.** Рассмотрим ассоциативную функцию вида

$$\text{AED}(a, b) = \max\{a + b - 1, 0\}.$$

В [6, 8] получен вид  $n$ -местной функции

$$\text{AED}(a_1, \dots, a_n) = \max \left\{ \sum_{i=1}^n a_i - n + 1, 0 \right\}.$$

Рассмотрим пересечение множества событий  $X \subseteq \mathfrak{X}$ ,  $|X| > 0$ . Тогда по определению 2

$$\begin{aligned} p_X &= \text{AED} \left( p_x, \mathbf{P} \left( \bigcap_{y \in X \setminus x} y \right) \right) = \\ &= \max \left\{ \sum_{x \in X} p_x - |X| + 1, 0 \right\}. \end{aligned} \quad (12)$$

Формула (12) есть левая граница Фреше [3] для вероятностей пересечения множества событий  $X$ . Возможны следующие ситуации

1.  $p_X = 0$ , т.е. события не пересекаются. Тогда из (12) следует

$$\sum_{x \in X} p_x \leq |X| - 1.$$

2.  $p_X \neq 0$ , т.е. события пересекаются с наименьшей возможной вероятностью. Тогда из (12) и условия  $0 \leq p_X \leq 1$  следует

$$0 < \sum_{x \in X} p_x - |X| + 1 \leq 1$$

или, что эквивалентно,

$$|X| - 1 < \sum_{x \in X} p_x \leq |X|.$$

◇

### 2.3.4 Э-распределение $\mathcal{R}$

**Теорема 4.** Пусть вероятности событий  $p_x = \mathbf{P}(x) > 0$ ,  $x \in \mathfrak{X}$ . Тогда ассоциативная функция

$$\text{AED}(a, b) = \frac{ab}{a + b - ab}$$

определяет ассоциативное случайное множество с Э-распределением II рода:

$$p_X = \frac{\prod_{x \in X} p_x}{\sum_{Y \subseteq C_X^{|X|-1}} \left[ \prod_{x \in Y} p_x \right] - (|X| - 1) \cdot \prod_{x \in X} p_x}, \quad (13)$$

где  $|X| > 1$ ,  $X \in \mathfrak{X}$ ,

$$C_X^{|X|-1} = \{Y : Y \subseteq X, |Y| = |X| - 1\}.$$

Доказательство теоремы проводится индукцией по мощности множества  $\mathfrak{X}$ .

Будем обозначать Э-распределения такого вида  $\mathcal{R}$ .

**Теорема 5.** Для ассоциативного случайного множества, определяемого функцией

$$\text{AED}(a, b) = \frac{ab}{a + b - ab},$$

все арные ковариации  $\text{Kov}_X$  неотрицательны и имеют вид

$$\begin{aligned} \text{Kov}_X &= \\ &= \prod_{x \in X} p_x \cdot \left[ \frac{1}{\sum_{Y \subseteq C_X^{|X|-1}} \left[ \prod_{x \in Y} p_x \right] - (|X| - 1) \cdot \prod_{x \in X} p_x} - 1 \right], \end{aligned}$$

для всех  $X \subseteq \mathfrak{X}$ ,  $|X| > 1$ .

**Доказательство.** Выразим арную ковариацию для  $|X| > 1$

$$\begin{aligned} \text{Kov}_X &= \mathbf{P} \left( \bigcap_{x \in X} x \right) - \prod_{x \in X} p_x = \\ &= \prod_{x \in X} p_x \cdot \left[ \frac{1}{\sum_{Y \subseteq C_X^{|X|-1}} \left[ \prod_{x \in Y} p_x \right] - (|X| - 1) \cdot \prod_{x \in X} p_x} - 1 \right]. \end{aligned}$$

Пусть  $w = \max_{x \in \mathfrak{X}} p_x$ . Рассмотрим знаменатель

$$\sum_{Y \subseteq C_X^{|X|-1}} \left[ \prod_{x \in Y} p_x \right] - (|X| - 1) \cdot \prod_{x \in X} p_x \leq$$



$$\begin{aligned}
&\leq \sum_{Y \subseteq C_X^{|X|-1}} \left[ \prod_{x \in Y} w \right] - (|X| - 1) \cdot \prod_{x \in X} w = \\
&= |X| w^{|X|-1} - (|X| - 1) w^{|X|} \leq \\
&\leq w^{|X|-1} (w + |X|(1 - w)) \leq 1.
\end{aligned}$$

Следовательно,

$$\frac{1}{\sum_{Y \subseteq C_X^{|X|-1}} \left[ \prod_{x \in Y} p_x \right] - (|X| - 1) \cdot \prod_{x \in X} p_x} \geq 1,$$

и  $\text{Kov}_X \geq 0$ .  $\diamond$

### 2.3.5 Семейство Али-Михаэля-Хака ( $R_{AMH}$ )

**Теорема 6.** Пусть вероятности событий  $p_x = \mathbf{P}(x) > 0$ ,  $x \in \mathfrak{X}$ . Тогда ассоциативная функция

$$\text{AED}(a, b) = \frac{ab}{1 - \alpha(1 - a)(1 - b)}, \quad \alpha \in [-1, 1]$$

определяет семейство  $\mathcal{R}_{AMH}$  ассоциативных случайных множеств Али-Михаэля-Хака с  $\mathcal{A}$ -распределением  $\Pi$  рода для всех  $X \subseteq \mathfrak{X}$ ,  $|X| > 1$

$$p_X = \frac{\prod_{x \in X} p_x}{(1 - \alpha)^{|X|-1} + \sum_{k=1}^{|X|-1} \alpha^k \cdot S_k}, \quad (14)$$

где

$$S_k = \left[ \sum_{Y \subseteq C_X^k} \left[ (-1)^{k-|Y|} \cdot \delta_k(Y) \cdot \prod_{x \in Y} p_x \right] - \prod_{x \in X} p_x \right],$$

здесь  $C_X^k = \{Y : Y \subseteq X, |Y| = k\}$ ,

$$\delta_k(Y) = \begin{cases} 1, & k = 1, \text{ или } k = |X| - 1; \\ k - |Y| + 1, & 1 < k < |X| - 1. \end{cases}$$

Доказательство теоремы проводится индукцией по мощности множества  $\mathfrak{X}$ .

**Следствие.** Для ассоциативного  $\mathcal{A}$ -распределения Али-Михаэля-Хака

- при  $\alpha = 1$  получаем случайное множество  $\mathcal{R}$ ;
- при  $\alpha = 0$  получаем независимо-точечное случайное множество.

## Благодарности

Авторы выражают искреннюю признательность своим друзьям и коллегам Н.А. Лукьяновой, И.В. Барановой, Е.Е. Голденку и А.А. Новоселову, плодотворное сотрудничество с которыми явилось причиной

появления основных результатов этой работы. Особую признательность хотелось бы выразить своему учителю проф. О.Ю. Воробьеву, который стимулировал в течении многих лет исследования авторов в эвентологии.

## Список литературы

- [1] О.Ю. Воробьев и Е.Е. Голденко. *Структурный сет-анализ зависимостей случайных событий*. КрасГУ, ИВМ СО РАН, Красноярск, 2003.
- [2] О.Ю. Воробьев и А.Ю. Фомин. *Сет-регрессионный анализ зависимостей событий в статистических системах: Учеб.пособие*. Краснояр. гос. ун-т, Красноярск, 2004.
- [3] О. Ю. Воробьев. *Эвентология*. Сиб. фед. ун-т, Красноярск, 2007.
- [4] О. Ю. Воробьев. Широкая зависимость событий и аппроксимация эвентологических распределений широко-мультипликативными сет-функциями. *Труды VIII Международ. ФАМ конференции по финансово-актуарной математике*, Красноярск: СФУ (под ред. Олега Воробьева):101–122, 2009.
- [5] О. Ю. Воробьев. Зависимость событий. *Труды VII Международ. ФАМ конференции по финансово-актуарной математике*, Красноярск: СФУ (Ч.1 под ред. Олега Воробьева):65–66, 2008.
- [6] E.P. Klement and R. Mesiar and E. Pap. *Triangular Norms*. Springer, 2000.
- [7] S. Alsina and M. Frank and B. Schweizer. *ASSOCIATIVE FUNCTIONS Triangular Norms and Copulas*. World Scientific Publishing Co. Pte. Ltd., Singapore, 2006.
- [8] E.P. Klement and R. Mesiar and E. Pap. *Logical, Algebraic, Analytic, and Probabilistic Aspects of Triangular Norms*. Dordrecht: Kluwer., 2005.
- [9] О. Ю. Воробьев. Эвентологический скоринг в экономике, социологии, праве и бизнесе. *Труды III Всеросс. ФАМ конференции по финансово-актуарной математике*, Красноярск: СФУ (Ч.1 под ред. Олега Воробьева):113–128, 2006.
- [10] Roger B. Nelsen. *An Introduction to Copulas (Second Edition)*. Springer Science+Business Media, Inc., New York, 2006.
- [11] Д.В. Семенова и Ю.С.Кочанова. Моделирование эвентологических распределений ассоциативными функциями. *Труды XVI Международ. ЭМ конференции по эвентологической математике*, Красноярск: СФУ (под ред. Олега Воробьева):185–190, 2012.

## Примеры семейств ассоциативных случайных множеств

Семенова Дарья Владиславовна

Сибирский федеральный университет  
Институт математики и фундаментальной информатики  
Красноярск  
dariasdv@gmail.com

Кочанова Юлия Сергеевна

Сибирский федеральный университет  
Институт математики и фундаментальной информатики  
Красноярск  
kochanova\_y@mail.ru

**Аннотация.** В работе на основе известных параметрических семейств треугольных норм введены семейства ассоциативных эвентологических распределений случайных множеств событий, рассмотрены примеры для множеств событий малой мощности.

**Ключевые слова.** Ассоциативное случайное множество событий, эвентологическое распределение.

### Введение

В [1, 2] было введено понятие ассоциативного случайного множества событий. В этой работе на основе известных параметрических семейств треугольных норм рассмотрены примеры для ассоциативных множеств событий малой мощности.

Рассмотрим вообще вероятностное пространство  $(\Omega, \mathcal{F}, \mathbf{P})$ . Пусть  $\mathfrak{X} \subset \mathcal{F}$  – конечное множество событий, выбранное из алгебры  $\mathcal{F}$  этого пространства.

**Определение 1.** Ассоциативная функция в эвентологии (Associative Eventological Distribution – AED)

$$\text{AED} : [0, 1]^2 \rightarrow [0, 1]$$

определяется как двухместная функция, удовлетворяющая следующим аксиомам:

A1. граничные условия

$$\text{AED}(a, 0) = \text{AED}(0, a) = 0, \quad (1)$$

$$\text{AED}(a, 1) = \text{AED}(1, a) = a,$$

$$a \in [0, 1]$$

A2. монотонность

$$\text{AED}(a_1, b_1) \leq \text{AED}(a_2, b_2), \quad (2)$$

$$\text{когда } a_1 \leq a_2, b_1 \leq b_2$$

A3. коммутативность, т.е. для всех  $a, b$  на  $[0, 1]$

$$\text{AED}(a, b) = \text{AED}(b, a);$$

A4. ассоциативность, т.е. для всех  $a, b, c \in [0, 1]$

$$\text{AED}(\text{AED}(a, b), c) = \text{AED}(a, \text{AED}(b, c)) \quad (3)$$

A5. условие Липшиц-непрерывности

$$\text{AED}(c, b) - \text{AED}(a, b) \leq c - a, \quad a \leq c. \quad (4)$$

**Определение 2.** Ассоциативным случайным множеством событий называется случайное множество событий  $K$  под конечным множеством избранных событий  $\mathfrak{X}$  с  $\mathcal{E}$ -распределением II-го рода  $\{p_X, X \subseteq \mathfrak{X}\}$ , где для всех  $X \subseteq \mathfrak{X}$ ,  $|X| > 1$  вероятности пересечения множеств событий  $p_X$  определяются рекуррентным соотношением при известных вероятностях событий  $p_x = \mathbf{P}(x)$ ,  $x \in \mathfrak{X}$

$$p_{xy} = \mathbf{P}(x \cap y) = \text{AED}(p_x, p_y)$$

$$\begin{aligned} p_{xyz} &= \mathbf{P}(x \cap y \cap z) = \text{AED}(p_x, \text{AED}(p_y, p_z)) = \\ &= \text{AED}(p_x, \mathbf{P}(y \cap z)), \end{aligned}$$

$$p_X = \mathbf{P}\left(\bigcap_{x \in X} x\right) = \text{AED}\left(p_x, \mathbf{P}\left(\bigcap_{y \in X \setminus x} y\right)\right),$$

при условии, что соответствующее  $\mathcal{E}$  - распределение I-го рода будет легитимным.

$\mathcal{E}$ -распределения, определяемые функциями AED будем называть ассоциативными  $\mathcal{E}$  - распределениями. Таким образом, ассоциативное случайное множество событий полностью определяется своим ассоциативным  $\mathcal{E}$  - распределением.

Будем рассматривать ассоциативные случайные множества событий, определяемые различными функциями AED.

### 1 Примеры ассоциативных случайных множеств

#### 1.1 Семейство Али-Михаэля-Хака ( $R_{AMH}$ )

Рассмотрим вообще вероятностное пространство  $(\Omega, \mathcal{F}, \mathbf{P})$ . Пусть  $\mathfrak{X} \subset \mathcal{F}$  – конечное множество событий, выбранное из алгебры  $\mathcal{F}$  этого пространства.

И пусть известны вероятности событий  $p_x = \mathbf{P}(x)$ ,  $x \in \mathfrak{X}$ .

Ассоциативная функция [3, 4, 5]

$$\text{AED}(a, b) = \frac{ab}{1 - \alpha(1 - a)(1 - b)}, \quad \alpha \in [-1, 1] \quad (5)$$

определяет семейство  $\mathcal{R}_{AMH}$  ассоциативных случайных множеств Али-Михаэля-Хака

**Пример 1.** Пусть  $\mathfrak{X} = \{x, y\}$  и пусть известны вероятности событий  $p_x = \mathbf{P}(x)$  и  $p_y = \mathbf{P}(y)$ . Тогда согласно (5)

$$p_{xy} = \mathbf{P}\left(x \cap y\right) = \frac{p_x p_y}{1 - \alpha(1 - p_x)(1 - p_y)}. \quad (6)$$

По формулам обращения Мёбиуса перейдем к  $\mathfrak{E}$ -распределению I-го рода

$$\begin{aligned} p(xy) &= p_{xy} = \frac{p_x p_y}{1 - \alpha(1 - p_x)(1 - p_y)}, \\ p(x) &= p_x - p_{xy} = \frac{p_x(1 - p_y)(1 - \alpha(1 - p_x))}{1 - \alpha(1 - p_x)(1 - p_y)}, \\ p(y) &= p_y - p_{xy} = \frac{p_y(1 - p_x)(1 - \alpha(1 - p_y))}{1 - \alpha(1 - p_x)(1 - p_y)}, \end{aligned} \quad (7)$$

$$\begin{aligned} p(\emptyset) &= 1 - p_x - p_y + p_{xy} = \\ &= \frac{(1 - p_x)(1 - p_y)[1 + \alpha(p_x + p_y - 1)]}{1 - \alpha(1 - p_x)(1 - p_y)}. \end{aligned}$$

Рассмотрим  $\mathfrak{E}$ -распределения при различных значениях параметра  $\alpha$ .

- При  $\alpha = 0$  получаем независимо точечное  $\mathfrak{E}$ -распределение [6, 7]:  $p_{xy} = \mathbf{P}\left(x \cap y\right) = p_x \cdot p_y$ .
- При  $\alpha = 1$  получаем  $\mathfrak{E}$ -распределение  $\mathcal{R}$  [2]:  $p_{xy} = \mathbf{P}\left(x \cap y\right) = \frac{p_x p_y}{p_x + p_y - p_x p_y}$ .

На рис. 1 дана геометрическая интерпретация семейства Али-Михаэля-Хака при различных параметрах  $\alpha \in [-1, 1]$  в симплексе распределений для дуплета событий.

В работе [7] был предложен удобный инструмент анализа структур эвентологической зависимости — ковариация. Ковариация двух событий  $x, y \in \mathfrak{X}$  определяется как величина

$$\text{Kov}_{xy} = \mathbf{P}(x \cap y) - \mathbf{P}(x)\mathbf{P}(y), \quad (8)$$

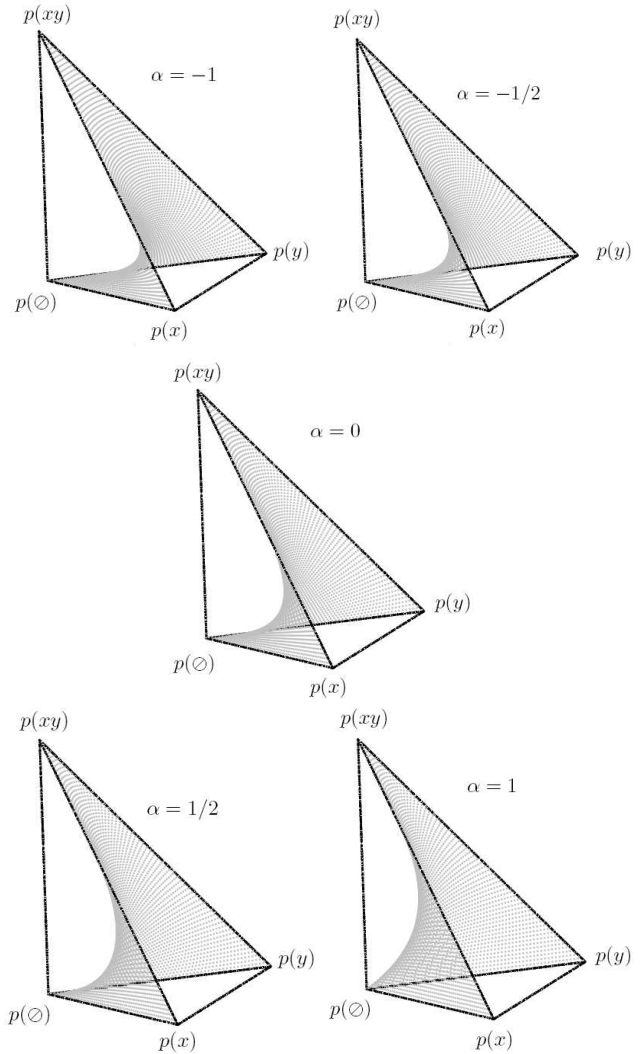


Рис. 1: Семейство  $\mathfrak{E}$ -распределений Али-Михаэля-Хака  $\mathcal{R}_{AMH}$  при различных параметрах  $\alpha$  в симплексе распределений для дуплета событий.

которая обращается в нуль, когда эти события независимы; больше нуля, когда они наступают вместе чаще (статистически притягиваются), и меньше нуля когда реже (статистически отталкиваются), чем в независимой ситуации. Ковариация событий служит мерой аддитивного отклонения событий от независимой ситуации.

Для семейства Али-Михаэля-Хака парная ковариация имеет вид

$$\text{Kov}_{xy} = \frac{\alpha p_x p_y (1 - p_x)(1 - p_y)}{1 - \alpha(1 - p_x)(1 - p_y)}. \quad (9)$$

Из (9) видно, что знак парной ковариации зависит только от знака параметра  $\alpha$ . Парная ковариация будет иметь вид  $\text{Kov}_{xy} = \frac{3\alpha}{8(8 - \alpha)}$  (рис. 3).

Пусть  $p_x = \frac{1}{2}$  и  $p_y = \frac{3}{4}$ . Тогда согласно (5) ве-

роятность пересечения событий  $x$  и  $y$  будет  $p_{xy} = \frac{3}{8-\alpha}$ . Из (7) получаем (рис. 2)  $p(x) = \frac{2-\alpha}{2(8-\alpha)}$ ,  $p(y) = \frac{3(4-\alpha)}{4(8-\alpha)}$ ,  $p(xy) = \frac{3}{8-\alpha}$ ,  $p(\odot) = \frac{4+\alpha}{4(8-\alpha)}$ .

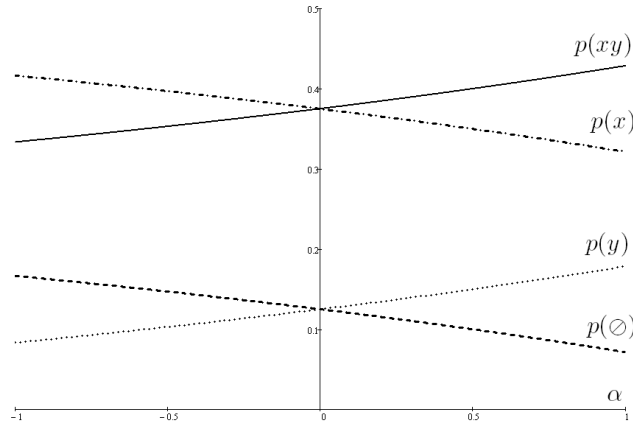


Рис. 2: Зависимость вероятностей I-рода Э-распределения Али-Михаэля-Хака от параметра  $\alpha \in [-1; 1]$  при  $p_x = \frac{1}{2}$  и  $p_y = \frac{3}{4}$ .

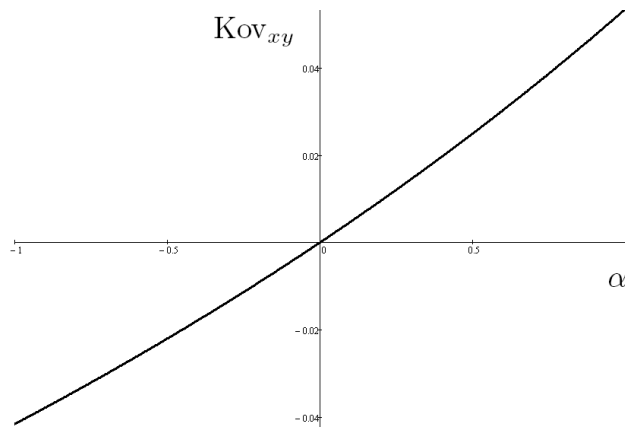


Рис. 3: Зависимость парной ковариации семейства Э-распределений Али-Михаэля-Хака от параметра  $\alpha \in [-1; 1]$  при  $p_x = \frac{1}{2}$  и  $p_y = \frac{3}{4}$ .

**Пример 2.** Пусть  $\mathfrak{X} = \{x, y, z\}$  и пусть известны вероятности событий  $p_x = \mathbf{P}(x)$ ,  $p_y = \mathbf{P}(y)$  и  $p_z = \mathbf{P}(z)$ .

Тогда согласно (5) и определению 2

$$\begin{aligned} p_{xy} &= \frac{p_x p_y}{(1-\alpha) + \alpha(p_x + p_y - p_x p_y)}, \\ p_{xz} &= \frac{p_x p_z}{(1-\alpha) + \alpha(p_x + p_z - p_x p_z)}, \\ p_{yz} &= \frac{p_y p_z}{(1-\alpha) + \alpha(p_y + p_z - p_y p_z)}, \\ p_{xyz} &= \frac{p_x p_y p_z}{(1-\alpha)^2 - \alpha^2 k_2 + \alpha k_1}, \end{aligned} \quad (10)$$

где

$$\begin{aligned} k_1 &= p_x + p_y + p_z - p_x p_y p_z, \\ k_2 &= p_x + p_y + p_z - \\ &\quad - p_x p_y - p_x p_z - p_y p_z + p_x p_y p_z. \end{aligned} \quad (11)$$

Арные ковариации будут иметь вид

$$\begin{aligned} \text{Kov}_{xy} &= \frac{\alpha p_x p_y (1-p_x)(1-p_y)}{(1-\alpha) + \alpha(p_x + p_y - p_x p_y)}, \\ \text{Kov}_{xz} &= \frac{\alpha p_x p_z (1-p_x)(1-p_z)}{(1-\alpha) + \alpha(p_x + p_z - p_x p_z)}, \\ \text{Kov}_{yz} &= \frac{\alpha p_y p_z (1-p_y)(1-p_z)}{(1-\alpha) + \alpha(p_y + p_z - p_y p_z)}, \end{aligned} \quad (12)$$

$$\text{Kov}_{xyz} = \frac{p_x p_y p_z (1 - (1-\alpha)^2 + \alpha^2 k_2 - \alpha k_1)}{(1-\alpha)^2 - \alpha^2 k_2 + \alpha k_1},$$

где  $k_1$  и  $k_2$  определяются по (11). Из (12) видно, что знак арной ковариации зависит только от знака параметра  $\alpha$ .

Пусть  $p_x = \frac{3}{8}$ ,  $p_y = \frac{1}{4}$  и  $p_z = \frac{1}{2}$ . Тогда

$$\begin{aligned} p_{xy} &= \frac{3}{32-15\alpha}, \quad p_{xz} = \frac{3}{16-5\alpha}, \\ p_{yz} &= \frac{1}{8-3\alpha}, \quad p_{xyz} = \frac{3}{15\alpha^2 - 59\alpha + 64}. \end{aligned}$$

По формулам обращения Мёбиуса получим зависимость вероятностей I-рода Э-распределения Али-Михаэля-Хака от параметра  $\alpha \in [-1; 1]$  (рис. 4). Из (12) получаем зависимость арных ковариаций от значения параметра  $\alpha$  (рис. 5):

$$\text{Kov}_{xy} = \frac{45\alpha}{1024 - 480\alpha}, \quad \text{Kov}_{xz} = \frac{15\alpha}{16(16 - 5\alpha)},$$

$$\text{Kov}_{yz} = \frac{3\alpha}{8(8 - 3\alpha)}, \quad \text{Kov}_{xyz} = \frac{3\alpha(59 - 15\alpha)}{64(15\alpha^2 - 59\alpha + 64)}.$$

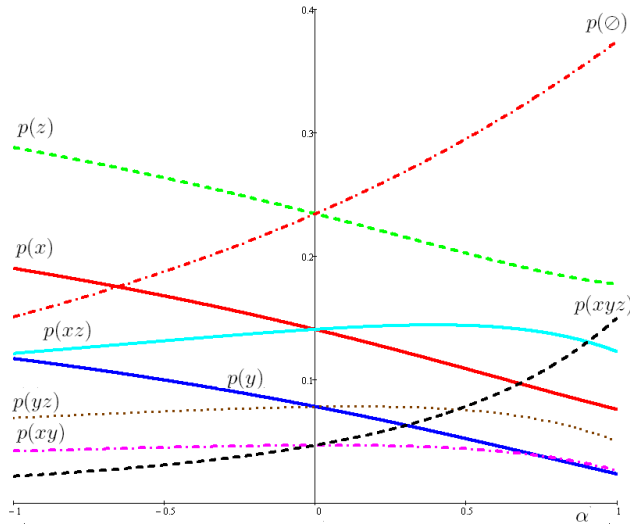


Рис. 4: Зависимость вероятностей I-рода Э-распределения Али-Михаэля-Хака от параметра  $\alpha \in [-1; 1]$  при  $p_x = \frac{3}{8}$ ,  $p_y = \frac{1}{4}$  и  $p_z = \frac{1}{2}$ .

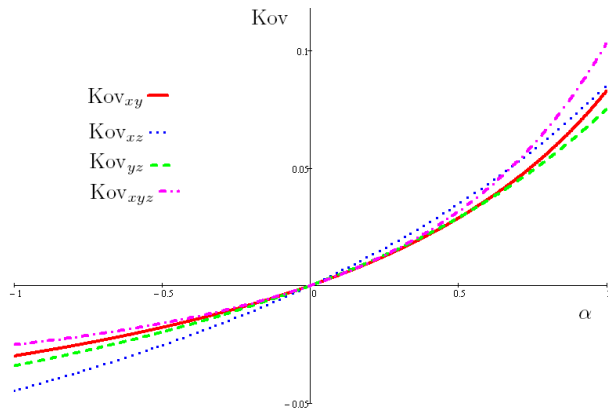


Рис. 5: Зависимость арных ковариаций семейства Э-распределений Али-Михаэля-Хака от параметра  $\alpha \in [-1; 1]$  при  $p_x = \frac{3}{8}$ ,  $p_y = \frac{1}{4}$  и  $p_z = \frac{1}{2}$ .

## 1.2 Семейство Гумбеля

Ассоциативная функция [3, 4, 5] при  $\alpha \in [1; \infty]$ ,  $a, b \in (0; 1)$

$$\text{AED}(a, b) = e^{-[(-\ln a)^\alpha + (-\ln b)^\alpha]^{\frac{1}{\alpha}}}, \quad (13)$$

определяет семейство  $\mathcal{R}_G$  ассоциативных случайных множеств Гумбеля.

**Пример 3.** Пусть  $\mathfrak{X} = \{x, y\}$  и пусть известны вероятности событий  $p_x = \mathbf{P}(x)$  и  $p_y = \mathbf{P}(y)$ . Тогда согласно (13)

$$p_{xy} = \mathbf{P}(x \cap y) = e^{-[(-\ln p_x)^\alpha + (-\ln p_y)^\alpha]^{\frac{1}{\alpha}}}. \quad (14)$$

На рис. 6 дана геометрическая интерпретация семейства Гумбеля при различных параметрах  $\alpha \in [1; \infty]$  в симплексе распределений для дуплета событий.

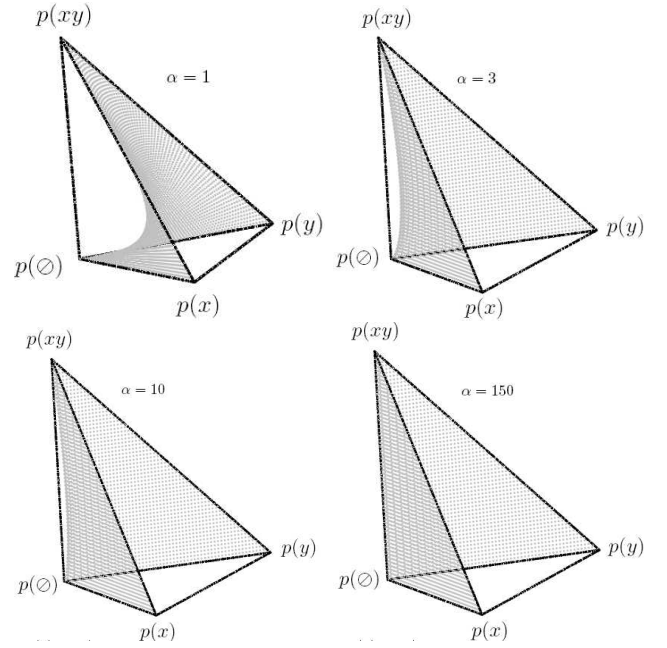


Рис. 6: Семейство Э-распределений Гумбеля  $\mathcal{R}_G$  при различных параметрах  $\alpha$  в симплексе распределений для дуплета событий.

Рассмотрим Э-распределения при различных значениях параметра  $\alpha$ .

- При  $\alpha = 1$  получаем независимо точечное Э-распределение [7]:  $p_{xy} = \mathbf{P}(x \cap y) = p_x \cdot p_y$ .
- При  $\alpha = \infty$  получаем Э-распределение вложенных событий [7]:  $p_{xy} = \mathbf{P}(x \cap y) = \lim_{\alpha \rightarrow \infty} e^{-[(-\ln p_x)^\alpha + (-\ln p_y)^\alpha]^{\frac{1}{\alpha}}} = \min\{p_x, p_y\}$ .

**Пример 4.** Пусть  $\mathfrak{X} = \{x, y, z\}$  и пусть известны вероятности событий  $p_x = \mathbf{P}(x) = \frac{3}{8}$ ,  $p_y = \mathbf{P}(y) = \frac{1}{4}$  и  $p_z = \mathbf{P}(z) = \frac{1}{2}$ . На рис.7 представлена зависимость вероятностей I-рода Э-распределения Гумбеля от параметра  $\alpha \in [1; 7]$ .

На рис.8 представлена зависимость арных ковариаций семейства Э-распределений Гумбеля от параметра  $\alpha \in [1; 7]$ . Как видно из рис.8 ассоциативное случайное множество Гумбеля состоит из событий, которые наступают вместе чаще (статистически притягиваются).

## 1.3 Семейство Франка

Рассмотрим эвентологическое распределение, где вероятность пересечения событий задается функцией

$$\text{AED}(p_x, p_y) = -\frac{1}{\alpha} \ln \left( 1 + \frac{(e^{-\alpha p_x} - 1)(e^{-\alpha p_y} - 1)}{e^{-\alpha} - 1} \right),$$



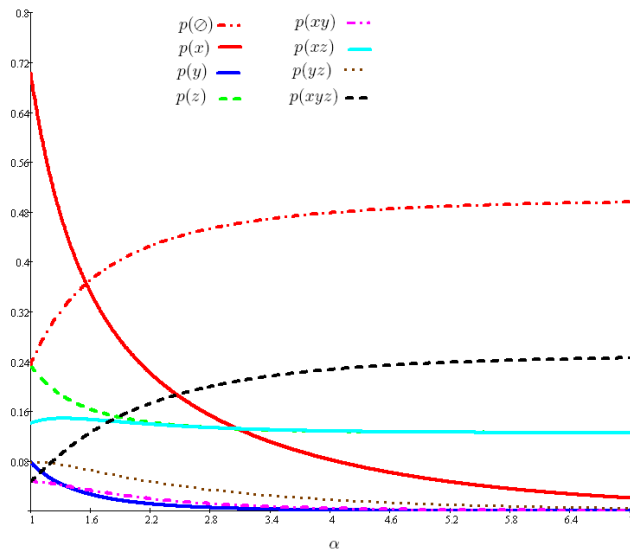


Рис. 7: Зависимость вероятностей I-рода Э-распределения Гумбеля от параметра  $\alpha \in [1; 7]$  при  $p_x = \frac{3}{8}$ ,  $p_y = \frac{1}{4}$  и  $p_z = \frac{1}{2}$ .

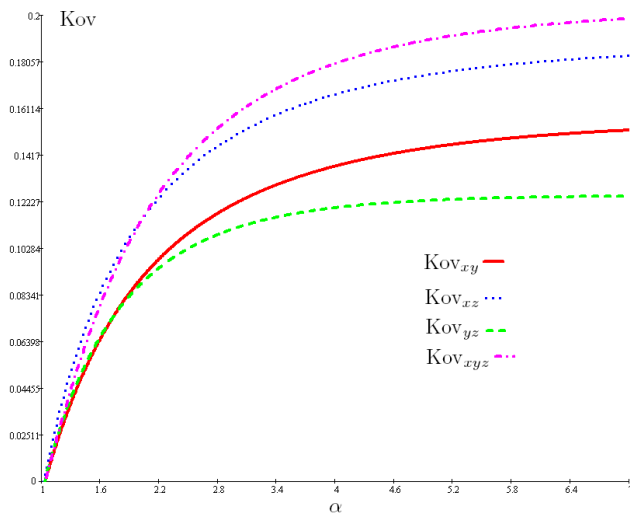


Рис. 8: Зависимость арных ковариаций семейства Э-распределений Гумбеля от параметра  $\alpha \in [1; 7]$  при  $p_x = \frac{3}{8}$ ,  $p_y = \frac{1}{4}$  и  $p_z = \frac{1}{2}$ .

данная функция является строгой  $t$ -нормой Франка [3, 4, 5], которая порождает семейство ассоциативных Э-распределений при  $\alpha \in (-\infty; +\infty) \setminus \{0\}$ .

На рис. 9 дана геометрическая интерпретация семейства Франка при различных параметрах  $\alpha \in [-\infty, \infty] \setminus \{0\}$  в симплексе распределений для дуплета событий.

На рис.10 представлена зависимость вероятностей I-рода Э-распределения Франка от параметра  $\alpha \in [-10; 10] \setminus \{0\}$ . На рис.11 представлена зависимость арных ковариаций семейства Э-распределений Фран-

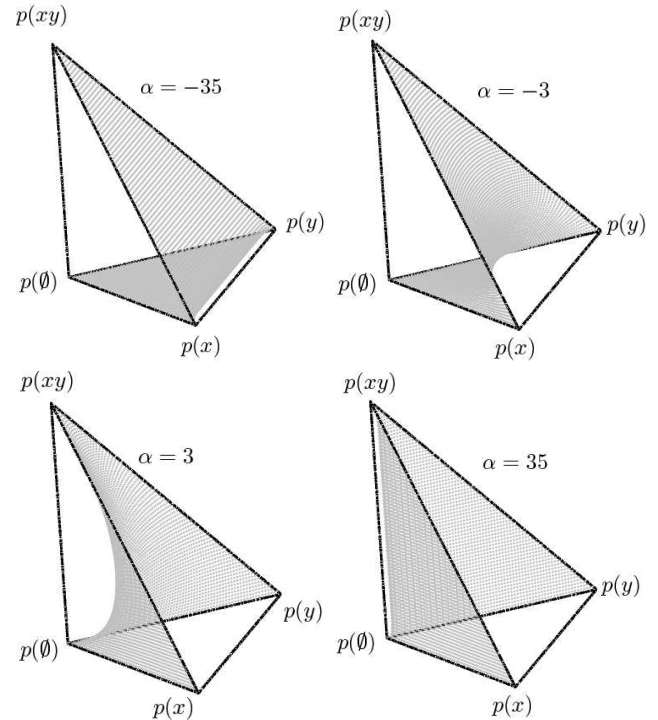


Рис. 9: Семейство Э-распределений Франка  $\mathcal{R}_F$  при различных параметрах  $\alpha$  в симплексе распределений для дуплета событий.

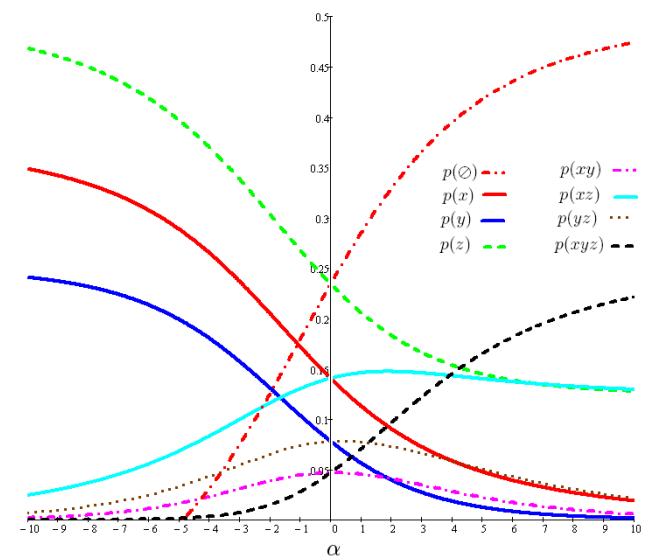


Рис. 10: Зависимость вероятностей I-рода Э-распределения Франка от параметра  $\alpha \in [-10; 10] \setminus \{0\}$  при  $p_x = \frac{3}{8}$ ,  $p_y = \frac{1}{4}$  и  $p_z = \frac{1}{2}$ .

ка от параметра  $\alpha \in [-10; 10] \setminus \{0\}$ .

#### 1.4 Семейство Швайцера-Скляра

Рассмотрим семейство ассоциативных эвентологических распределений, где вероятность пересечения со-

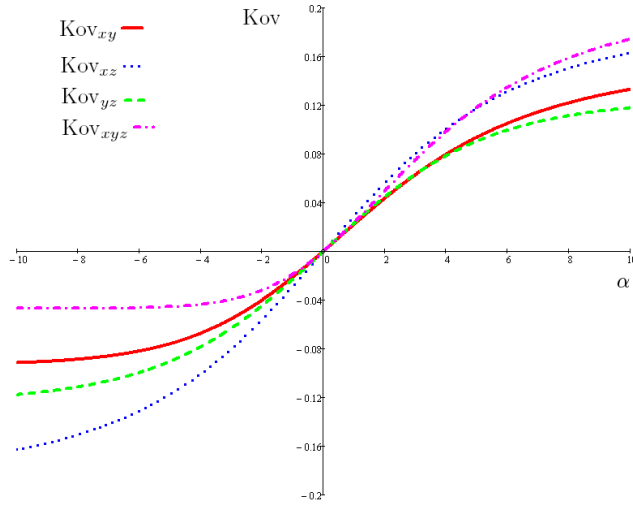


Рис. 11: Зависимость арных ковариаций семейства Э-распределений Франка от параметра  $\alpha \in [-10; 10] \setminus \{0\}$  при  $p_x = \frac{3}{8}$ ,  $p_y = \frac{1}{4}$  и  $p_z = \frac{1}{2}$ .

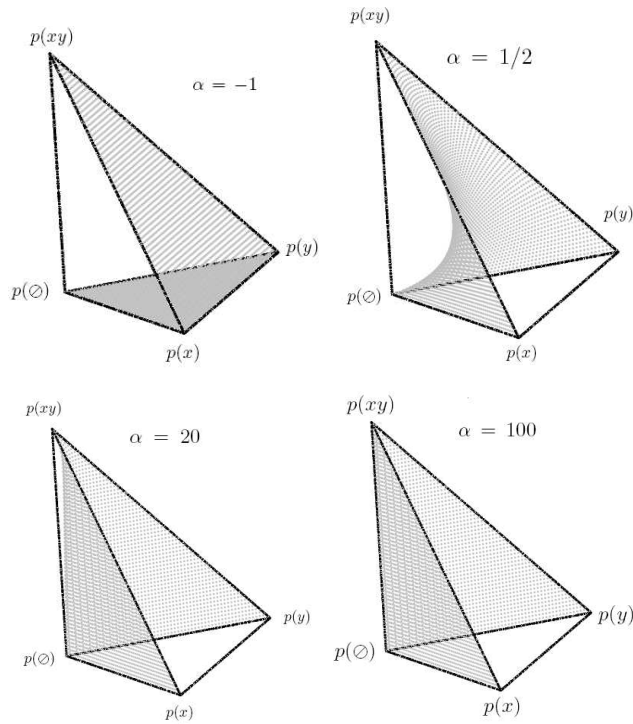


Рис. 12: Семейство Э-распределений Швайцера-Скляра  $\mathcal{R}_{SS}$  при различных параметрах  $\alpha$  в симплексе распределений для дуплета событий.

бытий задается функцией

$$\text{AED}(a, b) = \max\{a^{-\alpha} + b^{-\alpha} - 1, 0\}^{-1/\alpha},$$

данная функция является  $t$ -нормой Швайцера-Скляра [3, 4, 5], где параметр  $\alpha \in (-\infty, 0) \cup (0, +\infty)$ . При  $\alpha \geq -1$   $t$ -норма является копулой [8]. Тогда ассоциативное эвентологическое распределение будет вы-

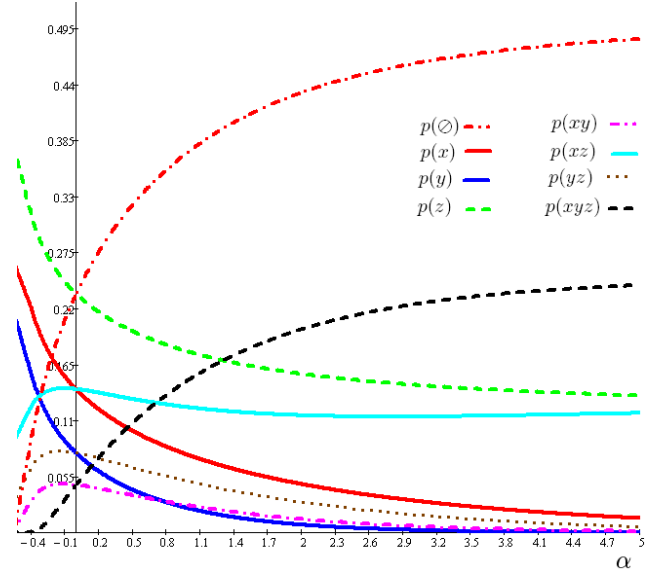


Рис. 13: Зависимость вероятностей I-рода Э-распределения Швайцера-Скляра от параметра  $\alpha \in [-0.4; 5] \setminus \{0\}$  при  $p_x = \frac{3}{8}$ ,  $p_y = \frac{1}{4}$  и  $p_z = \frac{1}{2}$ .

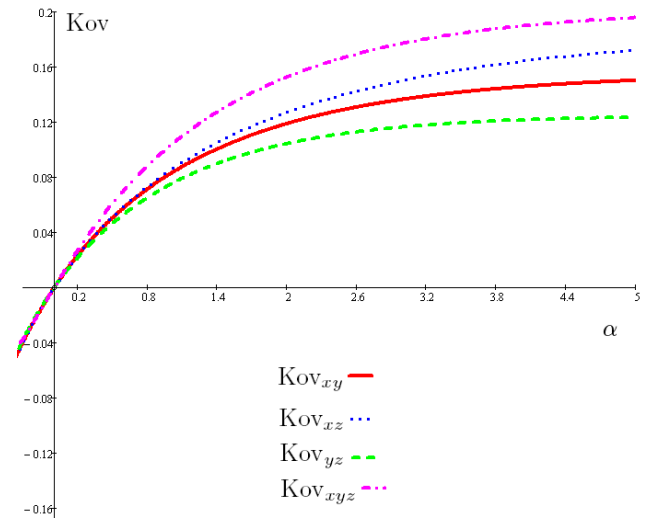


Рис. 14: Зависимость арных ковариаций семейства Э-распределений Швайцера-Скляра от параметра  $\alpha \in [-0.4; 5] \setminus \{0\}$  при  $p_x = \frac{3}{8}$ ,  $p_y = \frac{1}{4}$  и  $p_z = \frac{1}{2}$ .

глядеть следующим образом:

$$\text{AED}(p_x, p_y) = SS_\alpha(p_x, p_y) = \max\{a^{-\alpha} + b^{-\alpha} - 1, 0\}^{-1/\alpha},$$

где  $\alpha \in [-1; \infty] \setminus \{0\}$ .

- При  $\alpha = -1$  мы получаем Фреше-граничное случайное множество событий. В этом случае AED описывает левую границу Фреше для вероятности пересечения событий:  $\text{AED}(p_x, p_y)_{-1} = \max\{p_x + p_y - 1, 0\}$ .
- При  $\alpha = 1$  мы получаем распределение вида

$$\text{AED}(p_x, p_y)_1 = \frac{p_x p_y}{p_x + p_y - p_x p_y}.$$

Семейство  $\Theta$ -распределений Швайцера-Скляра при различных параметрах  $\alpha$  в симплексе распределений для дуплета событий представлено на Рисунке 12.

На рис.14 представлена зависимость арных ковариаций семейства  $\Theta$ -распределений Швайцера-Скляра от параметра  $\alpha \in [-0.4; 5] \setminus \{0\}$ . На рис.13 представлена зависимость вероятностей I-рода  $\Theta$ -распределения Швайцера-Скляра от параметра  $\alpha \in [-0.4, 5] \setminus \{0\}$ .

## Список литературы

- [1] Д.В. Семенова и Ю.С.Кочанова. Моделирование эвентологических распределений ассоциативными функциями. *Труды XVI Международ. ЭМ конференции по эвентологической математике*, Красноярск: СФУ (под ред. Олега Воробьева):185–190, 2012.
- [2] Д.В. Семенова и Ю.С.Кочанова. Эвентологические распределения случайных множеств событий на основе ассоциативных функций. *Труды XII Международ. ФАМ конференции по финансово-актуарной математике и эвентологии безопасности*, Красноярск: СФУ (под ред. Олега Воробьева):(в этой книге), 2013.
- [3] E.P. Klement and R. Mesiar and E. Pap. *Triangular Norms*. Springer, 2000.
- [4] E.P. Klement and R. Mesiar and E. Pap. *Logical, Algebraic, Analytic, and Probabilistic Aspects of Triangular Norms*. Dordrecht: Kluwer., 2005.
- [5] S. Alsina and M. Frank and B. Schweizer. *ASSOCIATIVE FUNCTIONS Triangular Norms and Copulas*. World Scientific Publishing Co. Pte. Ltd., Singapore, 2006.
- [6] О. Ю. Воробьев. Эвентологический скоринг в экономике, социологии, праве и бизнесе. *Труды III Всеросс. ФАМ конференции по финансово-актуарной математике*, Красноярск: СФУ (Ч.1 под ред. Олега Воробьева):113–128, 2006.
- [7] О. Ю. Воробьев. *Эвентология*. Сиб. фед. ун-т., Красноярск, 2007.
- [8] Roger B. Nelsen. *An Introduction to Copulas (Second Edition)*. Springer Science+Business Media, Inc., New York, 2006.

## Восстановление модели бизнес-процессов дистанционного обучения на основе данных информационной системы. Постановка задачи

Федоров Евгений Викторович

Сибирский Государственный Технологический Университет  
Факультет Автоматизации и Информационных Технологий  
Красноярск  
evgeniyfedorov1990@gmail.com

**Аннотация.** Системы дистанционного обучения в процессе своей работы сохраняют все данные о прохождении курсов обучаемыми. Благодаря этому появляется возможность восстановить модель процессов дистанционного обучения и провести ее анализ. Система Prom позволяет проводить анализ процессов, представленных в виде журнала событий, необходимо лишь привести данные системы дистанционного обучения к требуемому виду.

**Ключевые слова.** Система дистанционного обучения, moodle, манифест анализа процессов, process mining, база данных, журнал событий.

### 1 Постановка задачи

В настоящее время широко распространена практика так называемого «Дистанционного обучения». Преимуществом такого метода является то, что обучаемому не требуется непосредственное присутствие на занятиях – он может проходить обучение в удобное время и в удобном для него месте.

Наиболее популярной системой дистанционного обучения в нашей стране является Moodle. По уровню предоставляемых возможностей Moodle выдерживает сравнение с известными коммерческими СДО, в то же время выгодно отличается от них тем, что распространяется в открытом исходном коде – это дает возможность перестроить систему под особенности конкретного образовательного проекта, а при необходимости и встроить в нее новые модули.

С точки зрения исследования бизнес-процессов СДО Moodle хороша тем, что все данные о прохождении обучаемыми курса фиксируются и сохраняются, поэтому появляется возможность восстановления модели процессов дистанционного обучения и ее последующий анализ – Process mining.

Согласно «Манифеста анализа процессов», выпущенного рабочей группой IEEE по Анализу процессов (IEEE Task Force on Process Mining) 7 октября 2011 года, идея Process mining заключается в том, чтобы

выявить, отследить и произвести улучшение реальных процессов путем извлечения знаний из журналов событий, легко доступных в современных информационных системах.

Данные о прохождении обучения, хранящиеся в базе данных системы Moodle, как нельзя лучше подходят для осуществления анализа процессов, необходимо лишь на их основе сформировать журнал событий (Event Log).

Исходя из этого, была сформулирована цель исследования: Восстановление модели бизнес-процессов дистанционного обучения на основе данных информационной системы.

Для достижения данной цели были поставлены следующие задачи:

- анализ системы дистанционного обучения Moodle.
- получение данных о процессах обучения в системе Moodle – формирование event log.
- анализ полученного event log в системе Prom и генерация модели EPC по итогам анализа.

В результате анализа структуры данных системы дистанционного обучения Moodle было выявлено, что для составления журнала событий необходимы данные, содержащиеся в следующих таблицах: «mdl lesson», хранящей информацию об имеющихся в курсе обучения уроках, «mdl lesson timer», хранящей информацию о выполнении этих уроков каждым обучаемым, «mdl quiz», содержащей информацию обо всех имеющихся в курсе тестах и «mdl quiz attempts», которая содержит информацию о попытках прохождения имеющихся тестов. Для получения данных о процессах обучения в системе Moodle использовалось прямое подключение к рабочей базе данных системы посредством утилиты «MySQL Workbench», так как средствами самой системы Moodle это не представляется возможным. В результате был сформирован event log (журнал событий) содержащий все необходимые данные для дальнейшей работы. Следующим этапом исследования является экспорт полученного журнала событий в систему Prom и, непосредственно, анализ, результатом которого станет модель бизнес-процессов дистанционного обучения в системе Moodle.

## Распознавание образов с помощью многомерных интервалов

Гурами Шалвович Цициашвили

Институт прикладной математики ДВО РАН  
Владивосток  
guram@iam.dvo.ru

**Аннотация.** В работе рассматриваются алгоритмы интервального распознавания образов с помощью многомерных интервалов.

**Ключевые слова.** Алгоритм, многомерный интервал, распознавание образов.

В работе [1] изложен алгоритм интервального распознавания образов. В случае единственного признака, характеризующего объекты первого и второго классов, речь идет о построении минимального отрезка, который включает в себя все объекты первого класса. В случае, когда каждый объект характеризуется несколькими признаками, строится многомерный отрезок, являющийся прямым произведением одномерных отрезков указанного типа. В качестве правила, в силу которого новый объект распознается как объект первого класса, полагается принадлежность признака (в многомерном случае — набора признаков), характеризующего этот объект к уже построенному одномерному (многомерному) отрезку. Преимуществом алгоритма интервального распознавания образов перед известными алгоритмами является линейная по числу признаков и по числу объектов вычислительная сложность в реализации данного алгоритма. Этот алгоритм успешно применялся во многих задачах медицинской географии и экологии, в задачах метеорологии и рыболовства [2, 3, 4, 5, 6, 7, 8, 9].

Алгоритм интервального распознавания хорошо сработал, когда общее число объектов в обучающей выборке было порядка 20–30, а число признаков — более 3. Однако применительно к задачам горного дела возникла ситуация, когда единственный признак (несколько признаков), характеризующих объекты первого класса — проявления горного давления, не могут прогнозироваться с помощью единственного отрезка. Иными словами, есть проявления, имеющие предвестников, а есть проявления, которые не имеют предвестников. В этой ситуации единственный отрезок на множестве признаков уже не характеризует все проявления, пропуская те из них, накануне которых возникает явление так называемого молчания. В настоящей работе метод интервального распознавания развивается в направлении рассмотрения подобной ситуации. Он базируется на построении не одного отрезка или интервала, а нескольких непересекающихся интервалов, внутри которых на множе-

стве признаков содержатся точки, характеризующие объекты первого класса. Тем самым объекты первого класса разбиваются на подклассы, для каждого из которых распознавание производится отдельно. Если исходная выборка на каждом шаге пополняется новым объектом, то тогда возникает последовательность классификаций объектов первого класса, которую можно характеризовать как иерархическую классификацию.

Пусть первый класс объектов характеризуется набором вещественных чисел  $B = \{b_j, 1 \leq j \leq m\}$ , а второй класс объектов — набором вещественных чисел  $A = \{a_i, 1 \leq i \leq n\}$ ,  $-\infty \in A$ ,  $\infty \in A$ . Предположим, что  $m$  много меньше  $n$ . Пусть вещественные числа  $c, d$  удовлетворяют неравенству  $c \leq d$ . Определим интервал  $(c, d)$  условием  $(c, d) = \{f : c < f < d\}$ , если  $c < d$ . Если же  $c = d$ , то полагает интервал  $(c, d)$  состоящим из единственной точки  $c = d$ . Построим следующее правило распознавания объекта  $b$  набора  $B$ . Каждому числу  $b \in B$  сопоставим два числа

$$k(b) = \max\{a \in A : a \leq b\}, \quad r(b) = \min\{a \in A : a \geq b\}.$$

В результате вокруг каждого числа  $b \in B$  построен интервал  $(k(b), r(b))$ .

**Лемма 2.** Если  $b_i, b_j \in B$ , то интервалы  $(k(b_i), r(b_i))$ ,  $(k(b_j), r(b_j))$  либо совпадают, либо не пересекаются.

**Доказательство.** Пусть между точками  $b_i, b_j$  на вещественной оси нет точек набора  $A$ , тогда по построению интервалы  $(k(b_i), r(b_i))$ ,  $(k(b_j), r(b_j))$  совпадают. Наоборот, если между точками  $b_i, b_j$  на вещественной оси имеются точки набора  $A$ , тогда по построению интервалы  $(k(b_i), r(b_i))$ ,  $(k(b_j), r(b_j))$  не пересекаются. Таким образом точки набора  $b$  разбиваются на классы (эквивалентности) по их принадлежности совпадающим интервалам.  $\square$

Предположим теперь, что множество  $A$  состоит из  $n$  объектов, причем каждый объект  $i$  характеризуется  $l$ -мерным вектором  $a_i = (a_i^1, \dots, a_i^l)$ . Аналогично считаем, что множество  $B$  состоит из  $m$  объектов, причем каждый объект  $j$  характеризуется  $l$ -мерным вектором  $b_j = (b_j^1, \dots, b_j^l)$ . Определим интервал  $(k(b_j^t), r(b_j^t))$  равенствами:

$$k(b_j^t) = \max\{a_i^t : a_i^t \leq b_j^t, 1 \leq i \leq n\},$$

$$r(b_j^t) = \min\{a_i^t : a_i^t \geq b_j^t, 1 \leq i \leq n\}.$$



По этим интервалам построим  $l$ -мерный интервал, являющийся их прямым произведением  $\bigotimes_{t=1}^l (k(b_i^t), r(b_i^t))$ .

**Лемма 3.** Если  $1 \leq i \neq j \leq m$ , то  $l$ -мерные интервалы  $\bigotimes_{t=1}^l (k(b_i^t), r(b_i^t))$ ,  $\bigotimes_{t=1}^l (k(b_j^t), r(b_j^t))$  либо совпадают, либо не пересекаются.

**Доказательство.** Действительно, по построению для любого  $t$ ,  $1 \leq t \leq l$ , одномерные интервалы  $(k(b_i^t), r(b_i^t))$ ,  $(k(b_j^t), r(b_j^t))$  либо совпадают, либо не пересекаются. Если эти одномерные интервалы при всех  $t$ ,  $1 \leq t \leq l$  совпадают, то совпадают и их прямые произведения  $\bigotimes_{t=1}^l (k(b_i^t), r(b_i^t))$ ,  $\bigotimes_{t=1}^l (k(b_j^t), r(b_j^t))$ . В противном случае хотя бы при одном  $t$  эти интервалы не пересекаются и значит не пересекаются их прямые произведения. Таким образом, вектора набора  $B$  разбиваются на подмножества (классы эквивалентности) по их принадлежности к совпадающим  $l$ -мерным интервалам.  $\square$

Предположим теперь, что на вход нашей распознающей системы поступают  $(l+1)$ -мерные вектора. Первая компонента равна нулю, если этот вектор принадлежит множеству  $A$  и равна единице, если вектор принадлежит множеству  $B$ . Пусть на шаге 0 в систему введено два  $(l+1)$ -мерных вектора  $(0, +\infty, \dots, +\infty)$ ,  $(0, -\infty, \dots, -\infty)$ . Далее на шаге  $n > 0$  поступает один вектор  $\alpha_n, c_{n,1}, \dots, c_{n,l}$ . Пусть  $n_0$ -первый вектор, в котором  $\alpha_n = 1$ . Тогда строится первый многомерный интервал, содержащий вектор  $(c_{n_0,1}, \dots, c_{n_0,l})$ .

Далее пусть на шаге  $n > n_0$  поступает вектор  $\alpha_n, c_{n,1}, \dots, c_{n,l}$ . Предположим, что  $\alpha_n = 0$ . Тогда, если вектор  $(c_{n,1}, \dots, c_{n,l})$  не принадлежит ни одному из уже построенных интервалов, то система этих интервалов сохраняется. Если же  $(c_{n,1}, \dots, c_{n,l})$  принадлежит одному из уже существующих интервалов, то тогда этот интервал разбивается на подинтервалы по известному правилу. Предположим, что  $\alpha_n = 1$ . Тогда, если вектор  $(c_{n,1}, \dots, c_{n,l})$  не принадлежит ни одному из уже построенных интервалов, то строится новый интервал, содержащий этот вектор. Если же  $(c_{n,1}, \dots, c_{n,l})$  принадлежит одному из уже существующих интервалов, то тогда система интервалов на данном шаге сохраняется.

**Замечание 2.** В результате такого построения формируется последовательность разбиений векторов с первой компонентой, равной единице. При этом на шаге  $n > n_0$ , когда поступает новый вектор  $(\alpha_n, c_{n,1}, \dots, c_{n,l})$  образуется либо новый интервал, содержащий  $(c_{n,1}, \dots, c_{n,l})$ , либо вектор  $(c_{n,1}, \dots, c_{n,l})$  попадает в одну из компонент разбиения. Тем самым данные вектора подчиняются описанной выше иерархической классификации.

**Благодарности.** Автор выражает свою искреннюю

признательность Д. В. Семенову за переформатирование текста работы.

## Список литературы

- [1] Цициашвили Г.Ш. и Болотин Е.И. Разработка быстрого алгоритма распознавания образов в приложении к задачам прогнозирования. *Информатика и системы управления*, 2:25–27, 2010.
- [2] Болотин Е.И., Цициашвили Г.Ш. и Голычева И.В. Некоторые аспекты и перспективы факторного прогнозирования эпидемического проявления очагов клещевого энцефалита на основе многомерного анализа временных рядов. *Паразитология*, 36(2):89–95, 2002.
- [3] Болотин Е.И., Цициашвили Г.Ш., Бурухина И.Г. и Голычева И.В. Возможности факторного прогнозирования заболеваемости клещевым энцефалитом в Приморском крае. *Паразитология*, 36(4):280–285, 2002.
- [4] Болотин Е.И. и Цициашвили Г.Ш. Пространственно-временное прогнозирование функционирования очагов клещевого энцефалита. *Вестник ДВО РАН*, 1:5–19, 2003.
- [5] Шатилина Т.А., Цициашвили Г.Ш. и Радченкова Т.В. Опыт использования метода интервального распознавания для прогноза экстремальной ледовитости Татарского пролива (Японское море). *Метеорология и гидрология*, 10:65–72, 2006.
- [6] Горяинов, А.А., Шатилина Т.А., Цициашвили Г.Ш., Лысенко А.В. и Радченкова Т.В. Климатические причины снижения запасов амурских лососей в 20-м столетии. *Дальневосточный регион — Рыбное хозяйство*, 1,2(6,7):94–114, 2007.
- [7] Болотин Е.И., Цициашвили Г.Ш., Федорова С.Ю. и Радченкова Т.В. Факторное временное прогнозирование критических уровней инфекционной заболеваемости. *Экология человека*, 10:23–29, 2009.
- [8] Болотин Е.И., Цициашвили Г.Ш. и Федорова С.Ю. Теоретические и практические аспекты факторного прогнозирования инфекционной заболеваемости. *Экология человека*, 7:42–47, 2010.
- [9] Болотин Е.И., Ананьев В.Ю. и Цициашвили Г.Ш. Прогнозирование инфекционной заболеваемости: новые подходы. *Здоровье населения и среда обитания*, 5:15–19, 2010.

# О распределении статистики отношения правдоподобия, применяемой для тестирования нормальной выборки на один выброс

Людмила Константиновна Ширяева

Самарский государственный экономический университет  
Кафедра математической статистики и эконометрики  
Самара  
Shiryeva\_LK@mail.ru

**Аннотация.** Рассмотрена задача тестирования нормальной выборки на наличие в ней одного выброса. Для проверки значимости подозрительного наблюдения был использован критерий наибольшего по абсолютной величине нормированного отклонения. Найдена интегральная функция распределения статистики критерия. Определена область критических значений, в которой уровень значимости этого критерия равен удвоенному уровню значимости критерия Граббса.

**Ключевые слова.** Выброс, нормальная выборка, критерий наибольшего по абсолютной величине нормированного отклонения, интегральная функция распределения.

## 1 Введение

Пусть  $X_1, X_2, \dots, X_{n-1}, X_n$  — случайная выборка из  $n$  значений нормально распределенной случайной величины  $X$ ;  $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n-1)} \leq X_{(n)}$  — построенный по ней упорядоченный вариационный ряд;  $X_{(j)}$  —  $j$ -ая порядковая статистика ( $j = \overline{1, n}$ ).

Проверяемая нулевая гипотеза  $H_0$  состоит в том, что наблюдения  $X_1, X_2, \dots, X_{n-1}, X_n$  являются независимыми случайными величинами с нормальным распределением  $N(a, \sigma^2)$ . В качестве конкурирующей гипотезы  $H_1$  рассмотрим случай, когда какие-либо  $(n-1)$  из  $n$  наблюдений имеют одинаковое  $N(a, \sigma^2)$  распределение, а одно из них — выброс  $X_{out}$  — имеет распределение  $N(a + \lambda\sigma, \sigma^2)$ . Параметр сдвига  $\lambda \neq 0$  характеризует среднее (ожидаемое) смещение выброса вправо или влево.

Если есть основания предполагать, что аномальное наблюдение может иметь сдвиг математического ожидания вправо ( $\lambda > 0$ ), то для проверки на один верхний выброс используется односторонняя статистика Граббса:

$$T_n^{(1)} = (X_{(n)} - \bar{X}) / S,$$

где  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  и  $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ .

Если аномальное наблюдение может иметь сдвиг математического ожидания влево ( $\lambda < 0$ ), то для проверки на один нижний выброс используется статистика Граббса следующего вида:

$$T_{n,(1)} = (\bar{X} - X_{(1)}) / S.$$

Статистики  $T_n^{(1)}$  и  $T_{n,(1)}$  являются статистиками отношения правдоподобия для справедливых альтернатив  $H_1 : \lambda > 0$  и  $H_1 : \lambda < 0$  соответственно [1].

Легко проверить справедливость равенства:

$$P(T_n^{(1)} < t | H_0) = P(T_{n,(1)} < t | H_0), \forall t \in R. \quad (1)$$

Следовательно, нулевые распределения статистик  $T_n^{(1)}$  и  $T_{n,(1)}$  совпадают. Таблица критических точек распределения статистики  $T_n^{(1)}$  впервые была построена Ф. Граббсом [2], однако ее интегральная функция распределения  $F_n^{(1)}(t) = P(T_n^{(1)} < t | H_0)$  была найдена позже в работах [1], [3]:

$$F_n^{(1)}(t) = \begin{cases} 0, & t \leq \frac{1}{\sqrt{n}}, n \geq 2; \\ n \int_{\frac{1}{\sqrt{n}}}^t F_{n-1}^{(1)}(g_n(x)) dT_n(x), & \frac{1}{\sqrt{n}} < t \leq \frac{n-1}{\sqrt{n}}, n \geq 3; \\ 1, & t > \frac{n-1}{\sqrt{n}}, n \geq 2; \end{cases} \quad (2)$$

где  $dT_n(x) = f_{T_n}(x)dx$ ;

$$f_{T_n}(x) = \begin{cases} C_n \left(1 - \frac{n}{(n-1)^2} x^2\right)^{\frac{n-4}{2}}, & |x| < \frac{n-1}{\sqrt{n}}; \\ 0, & |x| \geq \frac{n-1}{\sqrt{n}}, \end{cases} \quad (3)$$

$$C_n = \frac{1}{n-1} \sqrt{\frac{n}{\pi}} \Gamma\left(\frac{n-1}{2}\right) / \Gamma\left(\frac{n-2}{2}\right)$$

$$g_n(x) = \frac{n}{n-1} x / \sqrt{\frac{n-1}{n-2} \left(1 - \frac{n}{(n-1)^2} x^2\right)}. \quad (4)$$

Если есть основания предполагать, что аномальное наблюдение может иметь сдвиг математического ожидания в разных направлениях, то для тестирования выборки на одно резко выделяющееся наблюдение применяют критерий наибольшего абсолютного

отклонения. Он основан на статистике:

$$T_n^{(1,ts)} = \max_{1 \leq i \leq n} \{|X_i - \bar{X}|\} / S.$$

Статистика  $T_n^{(1,ts)}$  является статистикой отношения правдоподобия, если в выборке имеется один выброс с параметрами  $\lambda \neq 0$  [1].

Гипотеза  $H_0$  отвергается, если на данном уровне значимости  $\alpha$  наблюдаемое значение статистики  $T_n^{(1,ts)}$  превысит критическое  $T_{cr}^{(1,ts)}(n; \alpha)$ , определяемое из условия  $P(T_n^{(1,ts)} > T_{cr}^{(1,ts)}(n; \alpha) | H_0) = \alpha$ .

Критические точки распределения статистики  $T_n^{(1,ts)}$ , как правило, находили либо по приближенным формулам [4], либо методом Монте-Карло [5], так как нулевое ее распределение до сих пор оставалось неизвестным. В данной работе найдено нулевое распределение статистики  $T_n^{(1,ts)}$ .

## 2 Закон распределения статистики $T_n^{(1,ts)}$ для случая справедливой нулевой гипотезы

Предположим, что гипотеза  $H_0$  справедлива. Обозначим

$$T_i = \frac{X_i - \bar{X}}{S}, \quad (i = 1, 2, \dots, n).$$

Очевидно, что случайные величины  $T_1, T_2, \dots, T_n$  распределены одинаково. Для плотности распределения вероятностей  $f_{i;n}(x)$  любой из них справедливо равенство [3]:

$$f_{i;n}(x) = f_{T_n}(x), \quad (i = 1, 2, \dots, n),$$

где  $f_{T_n}(x)$  вычисляется по формуле (3).

Каждая из случайных величин  $T_i$  представляет собой студентизированное отклонение наблюдения  $X_i$  от среднего по выборке,  $(i = \overline{1, n})$ . Для статистик  $T_n^{(1,ts)}, T_n^{(1)}$  и  $T_{n,(1)}$  справедливы следующие равенства:

$$T_n^{(1,ts)} = \max_{1 \leq i \leq n} \{|T_i|\}, \quad (5)$$

$$T_n^{(1)} = \max_{1 \leq i \leq n} \{T_i\}, \quad (6)$$

$$T_{n,(1)} = -\min_{1 \leq i \leq n} \{T_i\}. \quad (7)$$

Заметим, что из соотношений (6) и (2) вытекает

$$\frac{1}{\sqrt{n}} \leq \max_{1 \leq i \leq n} \{T_i\} \leq \frac{n-1}{\sqrt{n}}, \quad (8)$$

а из соотношений (7), (1) и (2) следует, что

$$-\frac{n-1}{\sqrt{n}} \leq \min_{1 \leq i \leq n} \{T_i\} \leq -\frac{1}{\sqrt{n}}. \quad (9)$$

Найдем функцию распределения статистики  $T_n^{(1,ts)}$ :

$$F_n^{(1,ts)}(t) = P(T_n^{(1,ts)} < t).$$

Очевидно, что в случае  $n = 2$  имеем  $T_2^{(1,ts)} = \frac{1}{\sqrt{2}}$ . Следовательно, функция распределения случайной величины  $T_2^{(1,ts)}$  примет вид:

$$F_2^{(1,ts)}(t) = \begin{cases} 0, & t \leq \frac{1}{\sqrt{2}}; \\ 1, & t > \frac{1}{\sqrt{2}}. \end{cases} \quad (10)$$

Найдем функцию распределения случайной величины  $T_n^{(1,ts)}$  для  $n \geq 3$ .

С учетом выражения (5) для  $t \leq 0$  можно записать

$$F_n^{(1,ts)}(t) = P(T_n^{(1,ts)} < t) = P(\max_{1 \leq i \leq n} |T_i| < t) = 0.$$

Для  $t \geq (n-1)/\sqrt{n}$  с учетом (3) и (5) получим

$$F_n^{(1,ts)}(t) = P\{\cap_{i=1}^n (|T_i| < t)\} = 1.$$

Далее для  $n \geq 3$  и  $0 < t < (n-1)/\sqrt{n}$  имеем

$$F_n^{(1,ts)}(t) = P\{\cap_{i=1}^n (|T_i| < t)\} = P\{\cap_{i=1}^n (-t < T_i < t)\}.$$

Заметим, что для  $0 < t < 1/\sqrt{n}$  в силу справедливости неравенства (8) событие  $\{\cap_{i=1}^n (-t < T_i < t)\}$  является невозможным. Поэтому

$$F_n^{(1,ts)}(t) = 0 \quad (0 < t \leq 1/\sqrt{n}).$$

Найдем  $F_n^{(1,ts)}(t)$  для  $1/\sqrt{n} < t < (n-1)/\sqrt{n}$ . Обозначим

$$\Psi_n(t_1, t_2) = P\{\cap_{i=1}^n (t_1 < T_i < t_2)\}. \quad (11)$$

С учетом последнего обозначения для  $n \geq 3$  можно записать:

$$F_n^{(1,ts)}(t) = \begin{cases} 0, & t \leq \frac{1}{\sqrt{n}}; \\ \Psi_n(-t, t), & \frac{1}{\sqrt{n}} < t < \frac{n-1}{\sqrt{n}}; \\ 1, & t \geq \frac{n-1}{\sqrt{n}}. \end{cases} \quad (12)$$

Таким образом, для вычисления значений функции  $F_n^{(1,ts)}(t)$  нужно уметь вычислять вероятность  $\Psi_n(t_1, t_2)$ .

**Теорема 21.** Пусть  $X_1, X_2, \dots, X_n$  независимые случайные величины, имеющие одинаковое нормальное  $N(a, \sigma^2)$  распределение. Тогда для  $n = 2$  вероятность

$$\Psi_2(t_1, t_2) = \begin{cases} 1, & (t_1, t_2) \in \Delta_2; \\ 0, & (t_1, t_2) \notin \Delta_2, \end{cases} \quad (13)$$

где  $\Delta_2 = [-\infty < t_1 < -\frac{\sqrt{2}}{2}; \frac{\sqrt{2}}{2} < t_2 < \infty]$   
для  $n = 3$  вероятность

$$\Psi_3(t_1, t_2) = \begin{cases} F_3^{(1)}(t_2), & t_1 \leq -\frac{2}{\sqrt{3}}; \\ F_3^{(1)}(-t_1), & t_2 \geq \frac{2}{\sqrt{3}}; \\ F_3^{(1)}(t_2) - F_3^{(1)}(\theta(t_1)), & (t_1, t_2) \in \tilde{\Delta}_3; \\ 0, & (t_1, t_2) \notin \tilde{\Delta}_3, \quad t_1 > -\frac{2}{\sqrt{3}}, \quad t_2 < \frac{2}{\sqrt{3}}, \end{cases} \quad (14)$$

где

$$F_3^{(1)}(t) = \begin{cases} 0, & t \leq \frac{1}{\sqrt{3}}; \\ \frac{3}{\pi} \arcsin\left(\frac{\sqrt{3}}{2}t\right) - \frac{1}{2}, & \frac{1}{\sqrt{3}} < t \leq \frac{2}{\sqrt{3}}; \\ 1, & t > \frac{2}{\sqrt{3}}, \end{cases}$$

$$\tilde{\Delta}_3 = \left[-\frac{2}{\sqrt{3}} < t_1 < -\frac{1}{\sqrt{3}}; \theta(t_1) < t_2 < \frac{2}{\sqrt{3}}\right];$$

$$\theta(t_1) = -\frac{t_1}{2} + \sqrt{1 - \frac{3}{4}t_1^2},$$

а для  $n \geq 4$  вероятность

$$\Psi_n(t_1, t_2) = \begin{cases} F_n^{(1)}(t_2), & t_1 \leq -\frac{n-1}{\sqrt{n}}; \\ F_n^{(1)}(-t_1), & t_2 \geq \frac{n-1}{\sqrt{n}}; \\ A_n^2 \int_{\frac{1}{\sqrt{n}}}^{t_2} I_n(t_1, x) dT_n(x), & (t_1, t_2) \in \Delta_n; \\ 0, & (t_1, t_2) \notin \Delta_n, \quad t_1 > -\frac{n-1}{\sqrt{n}}, \quad t_2 < \frac{n-1}{\sqrt{n}}, \end{cases} \quad (15)$$

где  $F_n^{(1)}(t)$  вычисляется по формуле (2);

$$A_n^2 = n(n-1);$$

$$I_n(t_1, x) = \int_{\rho_n(t_1, x)}^{\rho_n(-1/\sqrt{n}, x)} \Psi_{n-2}(g_{n-1}(y), \rho_{n-1}(g_n(x), y)) dT_{n-1}(y)$$

$$\rho_n(u, v) = \frac{u + \frac{v}{n-1}}{\sqrt{\frac{n-1}{n-2} \left(1 - \frac{n}{(n-1)^2} v^2\right)}}, \quad |v| < \frac{n-1}{\sqrt{n}}; \quad (16)$$

$g_n(x)$  вычисляется по формуле (4); функции  $f_{T_j}(\xi)$  вычисляются по формуле (3) для  $j = n-1, n$ ;

$$\Delta_n = \left[-\frac{n-1}{\sqrt{n}} < t_1 < -\frac{1}{\sqrt{n}}; \frac{1}{\sqrt{n}} < t_2 < \frac{n-1}{\sqrt{n}}\right].$$

*Доказательство.* Пусть  $n = 2$ . Легко проверить, что в этом случае

$$T_1 = \frac{\text{sign}(X_1 - X_2)}{\sqrt{2}} = -T_2.$$

Следовательно, вероятность

$$\Psi_2(t_1, t_2) = P\left\{\left(t_1 < -\frac{1}{\sqrt{2}} < t_2\right) \cap \left(t_1 < \frac{1}{\sqrt{2}} < t_2\right)\right\},$$

откуда и вытекает формула (13).

Рассмотрим случай  $n > 3$ . Найдем вероятность  $\Psi_n(t_1, t_2)$  для  $t_1 \leq -\frac{n-1}{\sqrt{n}}$ . С учетом (3) имеем

$$\Psi_n(t_1, t_2) = P\{\cap_{i=1}^n (T_i < t_2)\} = P\left(\max_{1 \leq i \leq n} \{T_i\} < t_2\right).$$

Далее с учетом (6) можно записать

$$P\left(\max_{1 \leq i \leq n} \{T_i\} < t_2\right) = P(T_n^{(1)} < t_2) = F_n^{(1)}(t_2),$$

где  $F_n^{(1)}(t)$  вычисляется по формуле (2).

Объединяя два последних равенства, получим

$$\Psi_n(t_1, t_2) = F_n^{(1)}(t_2), \quad \left(t_1 \leq -\frac{n-1}{\sqrt{n}}\right). \quad (17)$$

Пусть  $t_2 \geq (n-1)/\sqrt{n}$ . Тогда с учетом (3) получим

$$\Psi_n(t_1, t_2) = P\{\cap_{i=1}^n (T_i > t_1)\} = P\left(\min_{1 \leq i \leq n} \{T_i\} > t_1\right).$$

Теперь с учетом (9) можно записать

$$\Psi_n(t_1, t_2) = P\{T_{n,(1)} < -t_1\} = P\{T_n^{(1)} < -t_1\}.$$

Здесь также учтено, что нулевые распределения статистик  $T_{n,(1)}$  и  $T_n^{(1)}$  совпадают (см. равенство (1)).

Поэтому

$$\Psi_n(t_1, t_2) = F_n^{(1)}(-t_1), \quad t_2 \geq \frac{n-1}{\sqrt{n}}. \quad (18)$$

Пусть  $\Delta_n = [-\frac{n-1}{\sqrt{n}} < t_1 < -\frac{1}{\sqrt{n}}; \frac{1}{\sqrt{n}} < t_2 < \frac{n-1}{\sqrt{n}}]$ . С учетом (8) и (9) для  $(t_1, t_2) \in \Delta_n$  можно записать

$$\begin{aligned} \Psi_n(t_1, t_2) = A_n^2 P \{ \cap_{i=1}^{n-2} (T_{n-1} < T_i < T_n) \cap \\ \cap \left(\frac{1}{\sqrt{n}} < T_n < t_2\right) \cap \\ \cap \left(t_1 < T_{n-1} < -\frac{1}{\sqrt{n}}\right) \}, \end{aligned} \quad (19)$$

где  $A_n^2 = n(n-1)$ .

Обозначим

$$T_j^* = \frac{X_j - \bar{X}^*}{S^*}, \quad (j = 1, 2, \dots, n-1),$$

где  $\bar{X}^* = \frac{1}{n-1} \sum_{k \neq n} X_k$ ,  $S^{*2} = \frac{1}{n-2} \sum_{k \neq n} (X_k - \bar{X}^*)^2$ .

Справедливо соотношение для  $j = \overline{1, n-1}$  [3]

$$T_j = T_j^* \sqrt{\frac{n-1}{n-2} \left(1 - \frac{n}{(n-1)^2} T_n^2\right)} - \frac{1}{n-1} T_n. \quad (20)$$

Поэтому неравенство:  $T_{n-1} < T_i < T_n$  равносильно следующему:

$$T_{n-1}^* < T_i^* < g_n(T_n), \quad (i = \overline{1, n-2}),$$

где функция  $g_n(x)$  вычисляется по формуле (4).

Аналогично рассуждая, неравенство

$$t_1 < T_{n-1} < -1/\sqrt{n}$$

можно привести к виду:

$$\rho_n(t_1, T_n) < T_{n-1}^* < \rho_n(-1/\sqrt{n}, T_n),$$

где  $\rho_n(u, v)$  вычисляется согласно (16).

Поэтому соотношение (19) примет вид:

$$\begin{aligned} \Psi_n(t_1, t_2) = A_n^2 P \left\{ \bigcap_{i=1}^{n-2} (T_{i-1}^* < T_i^* < g_n(T_n)) \cap \right. \\ \left. \cap (\rho_n(t_1, T_n) < T_{n-1}^* < \rho_n(-\frac{1}{\sqrt{n}}, T_n)) \cap \right. \\ \left. \cap \left( \frac{1}{\sqrt{n}} < T_n < t_2 \right) \right\}. \end{aligned}$$

Обозначим:

$$T_i^{**} = \frac{X_i - \bar{X}^{**}}{S^{**}}, \quad (i = \overline{1, n-2}),$$

$$\text{где } \bar{X}^{**} = \frac{1}{n-2} \sum_{\substack{j \neq n \\ j \neq n-1}} X_j, \quad S^{**2} = \frac{1}{n-3} \sum_{\substack{j \neq n \\ j \neq n-1}} (X_j - \bar{X}^{**})^2.$$

Применим соотношение (20) для выборки объема  $(n-1)$ . В результате получим:

$$\begin{aligned} \Psi_n(t_1, t_2) = A_n^2 P \left\{ \bigcap_{i=1}^{n-2} B_n \left( \frac{1}{\sqrt{n}} < T_n < t_2 \right) \cap \right. \\ \left. \cap \left( \rho_n(t_1, T_n) < T_{n-1}^* < \rho_n(-\frac{1}{\sqrt{n}}, T_n) \right) \right\}. \end{aligned} \quad (21)$$

где  $B_n = \{g_{n-1}(T_{n-1}^*) < T_i^{**} < \rho_{n-1}(g_n(T_n), T_{n-1}^*)\}$ . Легко убедиться, что

$$T_n = \frac{n-1}{n} \cdot \frac{X_n - \bar{X}^*}{\sqrt{\frac{n-2}{n-1} S^{*2} + \frac{(X_n - \bar{X}^*)^2}{n}}},$$

т.е. случайная величина  $T_n$  является функцией случайных величин  $\bar{X}^*$ ,  $S^*$  и  $X_n$ .

Известно также, что случайные величины  $\{T_1^*, T_2^*, \dots, T_{n-1}^*\}$  распределены независимо от случайных величин  $\bar{X}^*$  и  $S^*$ . Следовательно, случайная величина  $T_n$  распределена независимо от случайных величин  $\{T_1^*, T_2^*, \dots, T_{n-1}^*\}$  [3]. Поэтому и случайные величины  $\{T_1^{**}, T_2^{**}, \dots, T_{n-2}^{**}\}$  распределены независимо от случайной величины  $T_{n-1}^*$ . Следовательно, соотношение (21) можно представить в виде

$$\begin{aligned} \Psi_n(t_1, t_2) = A_n^2 \int_{1/\sqrt{n}}^{t_2} dT_n(x) \times \\ \rho_n(-1/\sqrt{n}, x) \times \int P \left\{ \bigcap_{i=1}^{n-2} \{g_{n-1}(y) < T_i^{**} < \rho_{n-1}(g_n(x), y)\} \right\} dT_{n-1}(y). \end{aligned}$$

Откуда следует, что для  $(t_1, t_2) \in \Delta_n$  можно записать

$$\begin{aligned} \Psi_n(t_1, t_2) = A_n^2 \int_{1/\sqrt{n}}^{t_2} dT_n(x) \times \\ \rho_n(-1/\sqrt{n}, x) \times \int \Psi_{n-2}(g_{n-1}(y), \rho_{n-1}(g_n(x), y)) dT_{n-1}(y). \end{aligned} \quad (22)$$

Пусть  $(t_1, t_2) \notin \Delta_n$ ,  $t_1 > -\frac{n-1}{\sqrt{n}}$ ,  $t_2 < \frac{n-1}{\sqrt{n}}$ . Тогда в силу выполнения условий (8) и (9) событие

$$\{\bigcap_{i=1}^n (t_1 < T_i < t_2)\}$$

становится невозможным.

Поэтому для  $t_1 > -\frac{n-1}{\sqrt{n}}$  и  $t_2 < \frac{n-1}{\sqrt{n}}$ ,  $(t_1, t_2) \notin \Delta_n$  получим

$$\Psi_n(t_1, t_2) = 0. \quad (23)$$

Объединяя соотношения (17), (18), (22) и (23), получим (15), что и требовалось доказать.

Найдем рекурсивную формулу для вычисления вероятности  $\Psi_n(t_1, t_2)$  в случае  $n=3$ . Для этого заметим, что для  $n=3$  соотношение (2) примет вид:

$$F_3^{(1)}(t) = \begin{cases} 0, & t \leq \frac{1}{\sqrt{3}}; \\ 3 \int_{1/\sqrt{3}}^t dT_3(x), & \frac{1}{\sqrt{3}} < t \leq \frac{2}{\sqrt{3}}; \\ 1, & t > \frac{2}{\sqrt{3}}, \end{cases} \quad (24)$$

где  $f_{T_3}(x)$  согласно (3) вычисляется по формуле

$$f_{T_3}(x) = \begin{cases} \frac{\sqrt{3}}{2\pi} (1 - \frac{3}{4}x^2)^{-\frac{1}{2}}, & |x| < \frac{2}{\sqrt{3}}; \\ 0, & |x| \geq \frac{2}{\sqrt{3}}. \end{cases}$$

Откуда получим

$$F_3^{(1)}(t) = \begin{cases} 0, & t \leq \frac{1}{\sqrt{3}}; \\ \frac{3}{\pi} \arcsin\left(\frac{\sqrt{3}}{2}t\right) - \frac{1}{2}, & \frac{1}{\sqrt{3}} < t \leq \frac{2}{\sqrt{3}}; \\ 1, & t > \frac{2}{\sqrt{3}}. \end{cases} \quad (25)$$

Пусть  $t_1 \leq -2/\sqrt{3}$  и  $t_2 \geq t_1$ . Найдем вероятность  $\Psi_3(t_1, t_2) = P\{\bigcap_{i=1}^3 (t_1 < T_i < t_2)\}$ .

С учетом (3) и (6) имеем  $\Psi_3(t_1, t_2) = P\{\bigcap_{i=1}^3 (T_i < t_2)\} = P\{T_3^{(1)} < t_2\}$ . Поэтому

$$\Psi_3(t_1, t_2) = F_3^{(1)}(t_2), \quad t_1 \leq -2/\sqrt{3}, \quad t_2 \geq t_1. \quad (26)$$

Пусть  $t_2 \geq \frac{2}{\sqrt{3}}$  и  $t_1 \leq t_2$ . С учетом (3) и (7) можно записать

$$\Psi_3(t_1, t_2) = P\{\bigcap_{i=1}^3 (T_i > t_1)\} = P(\min_{1 \leq i \leq 3} \{T_i\} > t_1)$$

$$\Psi_3(t_1, t_2) = P\{T_{3,(1)} < -t_1\}.$$

Из (1) вытекает, что

$$P\{T_{3,(1)} < -t_1\} = P\{T_3^{(1)} < -t_1\}.$$

Поэтому

$$\Psi_3(t_1, t_2) = F_3^{(1)}(-t_1), \quad t_2 \geq \frac{2}{\sqrt{3}}, \quad t_1 \leq t_2. \quad (27)$$



Пусть  $|t_1| < \frac{2}{\sqrt{3}}$ ,  $|t_2| < \frac{2}{\sqrt{3}}$  и  $t_1 \leq t_2$ . Тогда

$$\Psi_3(t_1, t_2) = P \left[ \bigcap_{i=1}^3 \left( \min_{1 \leq i \leq 3} \{T_i\} \leq T_i \leq \max_{1 \leq i \leq 3} \{T_i\} \right) \cap \right. \\ \left. \cap \left( t_1 < \min_{1 \leq i \leq 3} \{T_i\} < -1/\sqrt{3} \right) \cap \right. \\ \left. \cap \left( 1/\sqrt{3} < \max_{1 \leq i \leq 3} \{T_i\} < t_2 \right) \right].$$

Обозначим  $G = \left[ -\frac{2}{\sqrt{3}} < t_1 < -\frac{1}{\sqrt{3}}; \frac{1}{\sqrt{3}} < t_2 < \frac{2}{\sqrt{3}} \right]$ . Из (8) и (9) следует, что

$$\Psi_3(t_1, t_2) = 0, (t_1, t_2) \notin G, t_1 > -2/\sqrt{3}, t_2 < 2/\sqrt{3}.$$

Найдем вероятность  $\Psi_3(t_1, t_2)$  для  $(t_1, t_2) \in G$ . Для этого заметим, что

$$\Psi_3(t_1, t_2) = P \{ \bigcap_{i=1}^3 (t_1 < T_i < t_2) \}; \\ \Psi_3(t_1, t_2) = P \{ \bigcap_{i=1}^3 (t_1 < T_i \leq \max_{1 \leq i \leq 3} \{T_i\}) \cap \\ \cap (\frac{1}{\sqrt{3}} \leq \max_{1 \leq i \leq 3} \{T_i\} < t_2) \};$$

$$\Psi_3(t_1, t_2) = 3P \{ \bigcap_{i=1}^2 (t_1 < T_i < T_3) \cap (1/\sqrt{3} < T_3 < t_2) \}$$

Используя соотношение (20), неравенство:

$$t_1 < T_i < T_3$$

можно привести к виду:

$$\rho_3(t_1, T_3) < T_i^* < g_3(T_3), \quad (i = 1, 2),$$

где  $\rho_3(t_1, T_3) = \frac{t_1 + T_3/2}{\sqrt{2(1-3T_3^2/4)}}$ ;  $g_3(x) = \frac{3T_3/2}{\sqrt{2(1-3T_3^2/4)}}$  (см. формулы (4) и (16) для  $n = 3$ ).

Поэтому

$$\Psi_3(t_1, t_2) = 3P \{ \bigcap_{i=1}^2 (\rho_3(t_1, T_3) < T_i^* < g_3(T_3)) \cap \\ \cap (1/\sqrt{3} < T_3 < t_2) \}.$$

Теперь заметим, что для выборки, "урезанной" до двух наблюдений, имеем

$$T_1^* = \frac{\text{sign}(X_1 - X_2)}{\sqrt{2}} = -T_2^*.$$

Поэтому

$$\Psi_3(t_1, t_2) = 3P \left\{ \left( \rho_3(t_1, T_3) < -\frac{1}{\sqrt{2}} \right) \cap \right. \\ \left. \cap \left( \frac{1}{\sqrt{2}} < g_3(T_3) \right) \cap \left( \frac{1}{\sqrt{3}} < T_3 < t_2 \right) \right\}.$$

Легко убедиться, что для  $T_3 > 1/\sqrt{3}$  выполняется неравенство:  $g_3(T_3) > 1/\sqrt{2}$ , следовательно,

$$\Psi_3(t_1, t_2) = 3P \{ (\rho_3(t_1, T_3) < -1/\sqrt{2}) \cap (1/\sqrt{3} < T_3 < t_2) \}.$$

Нетрудно проверить, что для  $(t_1, t_2) \in G$  неравенство  $\rho_3(t_1, T_3) < -1/\sqrt{2}$  выполняется, только если

$$\theta(t_1) < T_3 < t_2,$$

где  $\theta(t_1) = -\frac{t_1}{2} + \sqrt{1 - \frac{3}{4}t_1^2}$ .

Обозначим

$$\tilde{\Delta}_3 = \left[ -\frac{2}{\sqrt{3}} < t_1 < -\frac{1}{\sqrt{3}}; \theta(t_1) < t_2 < \frac{2}{\sqrt{3}} \right].$$

Тогда

$$\Psi_3(t_1, t_2) = 0, (t_1, t_2) \notin \tilde{\Delta}_3, t_1 > -\frac{2}{\sqrt{3}}, t_2 < \frac{2}{\sqrt{3}}. \quad (28)$$

Для  $(t_1, t_2) \in \tilde{\Delta}_3$  имеем

$$\Psi_3(t_1, t_2) = 3P \{ (\theta(t_1) < T_3 < t_2) \cap (1/\sqrt{3} < T_3 < t_2) \}.$$

Легко проверить, что при  $-2/\sqrt{3} < t_1 < -1/\sqrt{3}$  имеет место условие  $1/\sqrt{3} < \theta(t_1) < 2/\sqrt{3}$ . Поэтому для  $(t_1, t_2) \in \tilde{\Delta}_3$  можно записать

$$\Psi_3(t_1, t_2) = 3P \{ \theta(t_1) < T_3 < t_2 \} = 3 \int_{\theta(t_1)}^{t_2} dT_3(x).$$

Откуда с учетом (24) получим для  $(t_1, t_2) \in \tilde{\Delta}_3$

$$\Psi_3(t_1, t_2) = 3 \left[ F_3^{(1)}(t_2) - F_3^{(1)}(\theta(t_1)) \right], \quad (29)$$

где функция  $F_3^{(1)}(t)$  вычисляется по формуле (25).

Объединяя (26), (27), (28) и (29), получим соотношение (14), что и требовалось доказать.

Найдем функцию распределения  $F_n^{(1,ts)}(t)$  в случае  $n = 3$ . Используя формулы (12) и (14), нетрудно убедиться, что

$$F_3^{(1,ts)}(t) = \begin{cases} 0, & t \leq 1; \\ \frac{3}{\pi} \left[ \arcsin \left( \frac{\sqrt{3}}{2}t \right) - \arcsin(v(t)) \right], & 1 < t < \frac{2}{\sqrt{3}}; \\ 1, & t \geq \frac{2}{\sqrt{3}}, \end{cases} \quad (30)$$

где  $v(t) = \frac{\sqrt{3}}{2}\theta(-t)$ . В случае  $n > 3$  с учетом (15) функция распределения  $F_n^{(1,ts)}(t)$  примет вид

$$F_n^{(1,ts)}(t) = \begin{cases} 1, & t \geq \frac{n-1}{\sqrt{n}}; \\ A_n^2 \int_{\frac{1}{\sqrt{n}}}^t I_n(-t, x) dT_n(x), & t \in \left( \frac{1}{\sqrt{n}}; \frac{n-1}{\sqrt{n}} \right); \\ 0, & 0 < t < \frac{1}{\sqrt{n}}. \end{cases} \quad (31)$$

где  $I_n(-t, x) = \int_{\rho_n(-t, x)}^{\rho_n(-1/\sqrt{n}, x)} \Psi_{n-2}(g_{n-1}(y), \rho_{n-1}(g_n(x), y)) dT_{n-1}(y)$ . Найдем вероятность  $\varphi_n(t)$ . С учетом (6) и (7) для  $\frac{1}{\sqrt{n}} < t < \frac{n-1}{\sqrt{n}}$  можно записать

Теперь найдем правый хвост распределения случайной величины  $T_n^{(1,ts)}$ . Обозначим:

$$\overline{F_n^{(1,ts)}} = P(T_n^{(1,ts)} > t).$$

Справедлива следующая теорема о правом хвосте распределения статистики  $T_n^{(1,ts)}$ .

**Теорема 22.** Пусть  $X_1, X_2, \dots, X_n$  независимые случайные величины, имеющие одинаковое нормальное  $N(a, \sigma^2)$  распределение. Тогда для  $n > 3$  справедливо равенство

$$\overline{F_n^{(1,ts)}}(t) = \begin{cases} 1, & t \leq \frac{1}{\sqrt{n}}; \\ 2\overline{F_n^{(1)}}(t) - \varphi_n(t), & \frac{1}{\sqrt{n}} < t < t_n^*; \\ 2\overline{F_n^{(1)}}(t), & t_n^* \leq t < \frac{n-1}{\sqrt{n}}; \\ 0, & t \geq \frac{n-1}{\sqrt{n}}, \end{cases} \quad (32)$$

где  $\overline{F_n^{(1)}}(t) = 1 - F_n^{(1)}(t)$ ;  $F_n^{(1)}(t)$  вычисляется по формуле (2);  $t_n^* = \sqrt{(n-1)/2}$ ;

$$\varphi_n(t) = A_n^2 \int_{\frac{1}{\sqrt{n}}}^{\frac{n-1}{\sqrt{n}}} J_n(t, x) dT_n(x);$$

$$J_n(t, x) = \int_{\rho_n(-\frac{n-1}{\sqrt{n}}, x)}^{\rho_n(-t, x)} \Psi_{n-2}(g_{n-1}(y), \rho_{n-1}(g_n(x), y)) dT_{n-1}(y) \quad (33)$$

$dT_j(\xi) = f_{T_j}(\xi) d\xi$ ,  $j = n, n-1$ , плотность  $f_{T_j}(\xi)$  вычисляется по формуле (3).

**Доказательство.** Поскольку  $\overline{F_n^{(1,ts)}}(t) = 1 - F_n^{(1,ts)}(t)$ , то из (12) вытекает, что для  $n > 3$

$$\overline{F_n^{(1,ts)}}(t) = \begin{cases} 1, & t \leq \frac{1}{\sqrt{n}}; \\ 0, & t \geq \frac{n-1}{\sqrt{n}}. \end{cases} \quad (34)$$

Найдем  $\overline{F_n^{(1,ts)}}(t)$  для  $\left(\frac{1}{\sqrt{n}} < t < \frac{n-1}{\sqrt{n}}\right)$ . Из равенств (5)–(7) следует, что  $T_n^{(1,ts)} = \max\{T_n^{(1)}; T_{n,(1)}\}$ . Тогда  $\overline{F_n^{(1)}}(t) = P(\{T_n^{(1)} > t\} \cup \{T_{n,(1)} > t\})$ . Откуда получим

$$\overline{F_n^{(1,ts)}}(t) = P(\{T_n^{(1)} > t\}) + P(\{T_{n,(1)} > t\}) - P(\{T_n^{(1)} > t\} \cap \{T_{n,(1)} > t\}). \quad (35)$$

Используя формулу (1), получим для  $t \in \left(\frac{1}{\sqrt{n}}, \frac{n-1}{\sqrt{n}}\right)$

$$\overline{F_n^{(1,ts)}}(t) = 2\overline{F_n^{(1)}}(t) - \varphi_n(t), \quad (36)$$

где  $\overline{F_n^{(1)}}(t) = 1 - F_n^{(1)}(t)$ ;  
 $\varphi_n(t) = P(\{T_n^{(1)} > t\} \cap \{T_{n,(1)} > t\})$ .

$$\varphi_n(t) = A_n^2 P(\cap_{i=1}^{n-2} \{T_{n-1} < T_i < T_n\} \cap \left\{ -\frac{n-1}{\sqrt{n}} < T_{n-1} < -t \right\} \cap \left\{ t < T_n < \frac{n-1}{\sqrt{n}} \right\}).$$

Применим соотношение (20) сначала для выборки объема  $n$ , а затем для выборки объема  $n-1$ . В результате получим

$$\varphi_n(t) = A_n^2 P[\cap \{T_n > t\} \cap \{\rho_n(-(n-1)/\sqrt{n}, T_n) < T_{n-1}^* < \rho_n(-t, T_n)\} \cap \cap_{i=1}^{n-2} \{g_{n-1}(T_{n-1}^*) < T_i^{**} < \rho_{n-1}(g_n(T_n), T_{n-1}^*)\}].$$

Учтем, что случайные величины  $\{T_1^{**}, T_2^{**}, \dots, T_{n-2}^{**}\}$  распределены независимо от случайной величины  $T_{n-1}^*$ , а случайная величина  $T_n$  распределена независимо от случайной величины  $T_{n-1}^*$ . Следовательно, с учетом (11) можно записать

$$\varphi_n(t) = A_n^2 \int_{\frac{1}{\sqrt{n}}}^{\frac{n-1}{\sqrt{n}}} J_n(t, x) dT_n(x),$$

где интеграл  $J_n(t, x)$  вычисляется согласно (33).

Обозначим

$$S(t) = \left[ t < x < \frac{n-1}{\sqrt{n}}; \rho_n\left(-\frac{n-1}{\sqrt{n}}, x\right) < y < \rho_n(-t, x) \right].$$

Тогда для  $\frac{1}{\sqrt{n}} < t < \frac{n-1}{\sqrt{n}}$  можно записать

$$\varphi_n(t) = A_n^2 \int_{S(t)} \int \Psi_{n-2}(g_{n-1}(y), \rho_{n-1}(g_n(x), y)) \times \\ \times dT_{n-1}(y) dT_n(x).$$

Пусть  $t_n^* = \sqrt{(n-1)/2}$ . Легко проверить, что:

$$\rho_n(-t_n^*, t_n^*) = -(n-2)/\sqrt{n-1}.$$

Нетрудно убедиться, что для  $t_n^* < t < (n-1)/\sqrt{n}$  имеем

$$\rho_n(-t_n^*, x) < y \leq -(n-2)/\sqrt{n-1}, \quad (x, y) \in S(t).$$

Используя формулу (3), для  $t_n^* < t < (n-1)/\sqrt{n}$  можно записать

$$f_{T_{n-1}}(y) = 0, \quad (x, y) \in S(t).$$

Откуда, для  $t_n^* < t < (n-1)/\sqrt{n}$  имеем

$$\varphi_n(t) = 0, \quad \overline{F_n^{(1,ts)}}(t) = 2\overline{F_n^{(1)}}(t). \quad (37)$$

Объединяя (34), (36) и (37), получим (32), что и требовалось доказать.

## Список литературы

- [1] V.Barnett and T. Lewis. *Outliers in statistical data (second edition)*. Wiley Sons, Chichester, 1984.
- [2] F. Grubbs. Sample criteria for testing outlying observations. *Annals of mathematical statistics*, 21 (1):27–58, 1950.
- [3] J.Zhang and Y. Keming. The null distribution of the likelihood-ratio test for one or two outliers in a normal sample. *An Official journal of the Spanish society of statistics and operations research*, 15 (1):141–150, 2006.
- [4] Л.Н. Большев и Н.В. Смирнов. *Таблицы математической статистики (третье издание)*. Наука, Москва, 1983.
- [5] S.P. Verma and A. Quiroz-Ruiz. Critical values for 22 discordancy test variants for outliers in normal samples up to size 100, and application in science and engineering. *Revista Mexicana de Ciencias Geolgyicas*, 23 (3):302–311, 2006.

## Пример обработки нечетких правил вывода в экспертных системах

Рейхан Юсиф гызы Шихлинская

Институт прикладной математики Бакинского государственного университета  
Баку, Азербайджан  
reyhan\_sh@rambler.ru

**Аннотация.** В статье предложено логическое правило вывода с применением дефаззификационного метода WABL.

**Ключевые слова.** Нечеткое моделирование, логическое правило вывода.

В последнее время в области управления и принятия решений одним из интенсивных направлений исследований является нечеткое моделирование. Нечеткое моделирование, т.е. моделирование, основанное на нечеткой теории, позволяет получить более адекватные результаты в области управления технических систем. В статье предложено логическое правило вывода с применением дефаззификационного метода WABL. Сначала дадим краткое описание метода

**Определение.** Усредненным представителем нечеткого числа или WABL дефаззификацией называется следующая величина:

$$I_W(A) = c_L \int_0^1 L_A(\xi) p(\xi) d\xi + c_R \int_0^1 R_A(\xi) p(\xi) d\xi, \quad (1)$$

где коэффициенты  $c_L$  и  $c_R$  отражают степень существенности, соответственно, левых и правых сторон при проведении операций над нечетким числом и удовлетворяют следующим условиям:

$$c_L \geq 0, c_R \geq 0, c_L + c_R = 1.$$

$p(\xi)$  – функция важности степеней:

$$p: [0, 1] \rightarrow E_+ \equiv [0, +\infty]$$

и

$$\int_0^1 p(\xi) d\xi = 1.$$

Функцию  $p(\xi)$  можно выбрать следующим образом:  $p(\xi) = (k+1)\xi^k$ ,  $k \geq 0$ .  $L_A(\xi)$  и  $R_A(\xi)$  соответственно левая и правая часть нечеткого множества  $A$ :  $L_A(\xi) = \mu_{\uparrow}^{-1}(\xi)$ ,  $R_A(\xi) = \mu_{\downarrow}^{-1}(\xi)$ .  $\mu_{\uparrow}^{-1}$  и  $\mu_{\downarrow}^{-1}$  обратные функции  $\mu$ , соответствующие возрастающей и убывающей части функции принадлежности  $\mu$ .

Теперь дадим логическое правило вывода, опираясь на метод WABL. Рассмотрим этапы логического правила вывода:

1. Формирование базы правил систем нечеткого вывода.
2. Фаззификация входных и выходных переменных, т.е. формирование функций принадлежности  $\mu_{k_i}$ ,  $i = \overline{1, n}$  и  $\mu_{l_j}$ ,  $j = \overline{1, m}$  соответствующие нечетким входам  $k_i$  и нечетким выходам  $l_j$ .
3. Вычисление  $\overline{k_i}$  – степени истинности принадлежности точного входного значения  $x = \overline{x}$  в множество значений нечетких входов  $\mu_{k_i}$ ,  $i = \overline{1, n}$ , т.е.  $\overline{k_i} = \mu_{k_i}(\overline{x})$ ,  $i = \overline{1, n}$ .
4. Дефаззификация методом WABL выходных нечетких подмножеств  $l_j$ , т.е. нахождение  $I_W(l_j)$ ,  $j = \overline{1, m}$ .
5. Суммирование результатов дефаззификации с коэффициентами, равными степени истинности точных входных значений:

$$I_W = \sum_{i=1}^n \sum_{j=1}^m \overline{k_i} I_W(l_j). \quad (2)$$

В статье рассмотрен пример обработки нечетких правил вывода в ЭС, использующий работу вентилятора комнатного кондиционера. Разработан пакет прикладных программ на основе предложенных алгоритмов. Полученные в виде визуально - интерактивных графиков результаты сравнены с существующими в литературе результатами.

### Список литературы

- [1] Базы данных. Интеллектуальная обработка информации. / Под ред. В.В. Корнеева.
- [2] Р. Беллман, Л. Заде. Принятия решений в расплывчатых условиях. В сб.: Вопросы анализа и процедуры принятия решений, М.:172–215, 1976.
- [3] А. Леоненков. Нечеткое моделирование в среде MATLAB и fuzzyTECH. СПб.: БХВ, Петербург, 2003.
- [4] Э.Н. Насибов. Методы обработки нечеткой информации в задачах принятия решений. Баку: Элм, 2000.

# Функциональные моментные стратегии управления портфелем ценных бумаг

Дим Тахирович Юмагулов

Уфимский государственный авиационный технический университет  
yumagulov@bk.ru

**Аннотация.** В данной работе предложен один из вариантов сохранения денежных средств на российском фондовом рынке – функциональные моментные стратегии, при котором портфель переформируется с определенной периодичностью по формальным правилам.

**Ключевые слова.** Рынок ценных бумаг.

## 1 Введение

На сегодняшний день в России наблюдается целый ряд негативных макроэкономических факторов, препятствующих благоприятному экономическому росту и возможности сохранения и накопления средств. С этими негативными факторами сталкиваются как хозяйствующие субъекты страны, так и физические лица. На этом фоне проблема сохранения и накопления средств встает особенно остро и становится актуальной разработка средств, позволяющих снизить риск потерь от чрезмерной инфляции и прочих негативных факторов.

Одним из наиболее популярных и прибыльных инструментов сохранения денежных средств инвесторов является рынок ценных бумаг. Рынок ценных бумаг предоставляет инвесторам возможности по применению различных стратегий для получения эффективных портфелей ценных бумаг [1].

Одним из классов таких стратегий являются моментные стратегии. Суть их заключается в том, что покупка акций, которые имели высокие доходности в течение прошлого периода ранжирования (6–12 месяцев) (назовем их “победители”) и продажи акций, которые имели низкие доходности в течение этого прошлого периода ранжирования (назовем их “проигравшие”), позволяет получать прибыль [2], [3], [4]. Природа этой прибыли остается неизвестной и необъясненной [5].

Таким образом, теория, объясняющая прибыльность моментных стратегий несовершенна, что обуславливает необходимость дальнейших исследований в данной области.

Целью данной работы является разработка эффективного средства сохранения и накопления денеж-

ных средств – функциональных моментных стратегий, а также проверка их эффективности в условиях российского фондового рынка.

## 2 Функциональные моментные стратегии

Сущность предлагаемых нами функциональных моментных стратегий управления портфелем заключается в следующем. В некоторый момент времени инвестор обладает либо суммой средств  $S$ , либо портфелем акций, цена которого в этот момент времени по рыночной стоимости равна  $S$ . Далее портфель, аналогично обыкновенным моментным стратегиям, переформируется через временной интервал  $T$ , который выбирает для себя инвестор в каждом конкретном случае. Временные интервалы  $T$  в данном исследовании принимались равными неделе, месяцу, кварталу, полугодию, году. Отбор акций в новый портфель (“победителей” и “проигравших”) осуществляется с помощью моментной функции  $f(x)$ . Также, в портфель попадут не все акции с рассматриваемого сегмента, а только  $k$ . Продажа проигравших и покупка победителей осуществляется с учетом транзакционных расходов  $r$  за операции купли-продажи.

## 3 Экспериментальная проверка эффективности функциональных моментных стратегий

В ходе работы был проведен ряд исследований относительно эффективности применения функциональных моментных стратегий. Для исследования применялись современные ЭВМ, позволяющие обрабатывать большие объемы данных. Были специально разработаны программные продукты, позволяющие смоделировать процесс игры на рынке ценных бумаг.

В качестве исходных данных для анализа использовались дневные котировки на бирже ММВБ таких акций как: AFLT (ОАО “Аэрофлот”), AVAZ (“Автоваз”), MSNG (ОАО “Мосэнерго”), IRGZ (ОАО “Иркутск-энерго”), LKOH (ОАО “Лукойл”), RTKM (ОАО “Ростелеком”), SBER (Сбербанк России), SNGS (ОАО “Сургутнефтегаз”), TATN (ОАО “Татнефть”), URSI (ОАО “Уралсвязьинформ”) и др.



По результатам эксперимента можно представить следующие заключения:

1. функциональные моментные стратегии являются эффективным средством портфельного инвестирования;
2. на сроке инвестирования 5 лет наиболее прибыльными являются малые (2–3 акции) и средние (8–12 акций) сегменты рынка. Причем, чаще всего в данном случае наиболее эффективной является моментная функция  $f(x) = \frac{1}{x}$ , небольшие размеры портфеля (3–4 акции) и период переформирования 6 месяцев;
3. при долгосрочном инвестировании денежных средств функциональные моментные стратегии позволяют получить доход в два раза больше, чем средний по рынку;
4. при моделировании процесса игры на российской бирже ММВБ с 2001 по 2010 год выявлено, что функциональные моментные стратегии за этот период времени могли дать доход от 6180,92% до 7374,66%, наряду с тем как инвестиции всех средств только в одного эмитента могли дать в среднем 3664,01%. Уровень инфляции за этот период составил 340,08% [6];
5. наиболее эффективными при формировании портфеля является применение моментной функции  $f(x) = \frac{1}{x}$ , небольшие размеры портфеля (3–4 акции) и период переформирования 6 месяцев.

## 4 Заключение

Как показывают наблюдения, российская экономика еще далека от совершенства. Об этом, например, свидетельствует такой фактор как чрезмерная инфляция. Последнее в свою очередь препятствует благоприятному экономическому развитию и возможности стабильного сохранения и накопления денежных средств.

И хозяйствующие субъекты страны, и физические лица, нуждаются в методах снижения риска потери от неблагоприятных экономических воздействий. С проблемой сохранения денежных средств сталкиваются также и различные государственные финансовые фонды, управляющие средствами, предназначенными для удовлетворения социально значимых потребностей. В конечном итоге, обеспеченность граждан страны является одним из ключевых факторов, направленных на поддержание национального благосостояния.

Разработка таких методов поможет дать более широкое представление о механизмах финансового рынка.

Исследование несет также важное научное значение, позволяя найти и открыть новые закономерности существования и развития финансового рынка в частности и экономической системы в целом.

В данной работе предложен один из вариантов сохранения денежных средств на российском фондовом рынке - функциональные моментные стратегии, при котором портфель переформируется с определенной периодичностью по формальным правилам. Стратегии такого типа называются моментными. Проведено эмпирическое исследование подобных стратегий, которое продемонстрировало их высокую эффективность при соответствующем выборе параметров.

Эмпирическое исследование проводилось с применением современных ЭВМ, позволяющих обрабатывать большие объемы данных. Были специально разработаны программные продукты, позволяющие смоделировать процесс игры на рынке ценных бумаг. Результатом исследования стало доказательство эффективности применения функциональных моментных стратегий. Исследование показало, что функциональные моментные стратегии дают доход гораздо больший, чем средний по рынку.

## Список литературы

- [1] Sergio Ortobelli, Svetlozar Rachev, Isabella Huber, Almira Biglova. Optimal portfolio selection and risk management: A comparison between the stable paretian approach and the gaussian one. *Handbook of Numerical Methods in Finance*, Birkhauser Boston:197–252, 2004.
- [2] N. Jegadeesh and S. Titman. Profitability of momentum strategies: An evaluation of alternative explanations. *Journal of Finance*, 56:699–720, 2001.
- [3] K. G. Rouwenhorst. International momentum strategies. *Journal of Finance*, 53:267–284, 1998.
- [4] Griffin, J.N, Ji, X. and J.S. Martin. Momentum investing and business cycle risk: Evidence from pole to pole. *Journal of Finance*, 58:2515–2547, 2003.
- [5] А.Ф. Биглова. Моментные стратегии и их применение в условиях российского фондового рынка. *Финансы и кредит*, 9:74–79, 2005.
- [6] *Инвестиционно-аналитический портал*.  
[http : // www.fundshub.ru / finances /](http://www.fundshub.ru/finances/) .

## С п и с о к а в т о р о в

Абасов Рагиб Зейнал оглы .....	50
Абдушукуров Абдурахим Ахмедович .....	20, 55, 15, 57
Азимов Ж.Б. ....	60
Баранова Ирина Владимировна .....	62, 73, 220, 272
Белосохова Анна Владимировна .....	77
Бескорсый Никита Сергеевич .....	80
Боровиков Игорь Анатольевич .....	83
Веретнова Ксения Юрьевна .....	92
Воробьев Олег .....	33,40,97,104,115,133,142,150,158,162
Глик Людмила Андреевна .....	165
Гнездилова Маргарита Андреевна .....	169
Гордеева Ксения Владимировна .....	175
Горохов Александр Андреевич .....	180
Григорьев Юрий Дмитриевич .....	183,195
Демочкин Валерий .....	22
Добронец Борис Станиславович .....	199
Доррер Георгий Алексеевич .....	244
Доррер Михаил Георгиевич .....	203
Елина Мария Вадимовна .....	208
Есин Роман Витальевич .....	210
Зырянов Антон Александрович .....	203, 215
Жаткина Кристина Владимировна .....	220
Имомов Аъзам Абдурахимович .....	224
Какаджанова Лейла Рэшитовна .....	55
Киреев Виталий Александрович .....	227
Кирилук Анна Алексеевна .....	231
Клокова Ксения Андреевна .....	237
Коморовский Витольд Станиславович .....	244
Кочанова Юлия Сергеевна .....	335, 341
Кустицкая Татьяна Алексеевна .....	255
Кучеров Матвей Николаевич .....	249
Ланцев Евгений Алексеевич .....	203, 261
Логинова Екатерина Борисовна .....	208
Лукьянова Наталья Александровна .....	264
Максимова Кристина Игоревна .....	272
Масаев Сергей Николаевич .....	29
Медницкий Алексей Николаевич .....	277, 280
Мурадов Рустамжон .....	15
Мусурмонова Шахло Гулямовна .....	224
Нартов Яков Викторович .....	284
Нифонтов Александр Сергеевич .....	162, 287
Новоселов Аркадий Арсеньевич .....	293, 301
Нурмухамедова Наргиза Сайдиллаевна .....	20
Олейников Борис Васильевич .....	80, 305
Пазенко Ольга Вячеславовна .....	308
Поддубный Василий Васильевич .....	311
Попов Александр Викторович .....	321
Попова Елена Александровна .....	324
Попова Ольга Аркадьевна .....	199, 328
Распопов Виталий Евгеньевич .....	77, 180
Садовский Михаил Георгиевич .....	83, 169, 208, 237, 308
Сагидуллаев Калмурза Сапарбаевич .....	57
Сафарова Зохра Аббасали кызы .....	50
Севастьянова Нэлли Александровна .....	333
Семенова Дарья Владиславовна .....	335, 341
Семенов Михаил Евгеньевич .....	165
Темерова Елена Николаевна .....	73
Трусова Алла Юрьевна .....	277, 280
Федоров Евгений Викторович .....	348
Цициашвили Гурами Шалвович .....	349
Шабанов Владислав Саидович .....	321
Шалабай Андрей Игоревич .....	305
Ширияева Людмила Константиновна .....	351
Шихлинская Рейхан Юсиф гызы .....	358
Юмагулов Дим Тахирович .....	359
Янковская Анна Ефимовна .....	??

## L i s t o f a u t h o r s

Abasov Rahib .....	50
Abdushukurov Abdurakhim .....	20, 55, 15, 57
Azimov Jakhongir .....	60
Baranova Irina .....	62, 73, 220, 272
Belosokhova Anna .....	77
Beskorsyi Nikita .....	80
Borovikov Igor .....	83
Demotchkin Val .....	22
Dobronets Boris .....	199
Dorrer Georgy .....	244
Dorrer Michael .....	203
Elina Maria .....	208
Esin Roman .....	210
Fedorov Eugeny .....	348
Glik Ludmila .....	165
Gnezdilova Margaret .....	169
Gordeeva Ksenia .....	175
Gorokhov Alexandr .....	180
Grigoriev Yuri .....	183,195
Imomov Azam .....	224
Kakadganova Leyla .....	55
Kireev Vitaliy .....	227
Kirilyuk Anna .....	231
Kloкова Xenia .....	237
Kochanova Yulia .....	335, 341
Komarovski Vitold .....	244
Kuchеров Matvey .....	249
Kustitskaya Tatyana .....	255
Lantsev Eugeny .....	203, 261
Loginova Ekaterina .....	208
Lukyanova Natalia .....	264
Maksimova Kristina .....	272
Masaev Sergey .....	29
Mednitskiy Aleksey .....	277, 280
Muradov Rustamjon .....	15
Musurmonova Shahlo .....	224
Nartov Jakov .....	284
Nifontov Alexander .....	162, 287
Novosyolov Arcady .....	293, 301
Nurmuhamedova Nargiza .....	20
Oleynikov Boris .....	80, 305
Pazenko Olga .....	308
Poddubny Basil .....	311
Popov Alexander .....	321
Popova Elena .....	324
Popova Olga .....	199, 328
Raspopov Vitali .....	77, 180
Sadovsky Michael .....	83, 169, 208, 237, 308
Safarova Zohra .....	50
Sagidullaev Khalmurza .....	57
Semenova Daria .....	335, 341
Semenov Mikhail .....	165
Sevastianova Nelli .....	333
Shabanov Vladislav .....	321
Shalabay Andrey .....	305
Shiryaeva Ludmila .....	351
Shikhlin'skaya Reyhan Yusufqizi .....	358
Temerova Elena .....	73
Trusova Alla .....	277, 280
Tsitsiashvili Gurami .....	349
Veretnova Kseniya .....	92
Vorobyev Oleg .....	33,40,97,104,115,133,142,150,158,162
Yankovskaja Anna .....	??
Yumagulov Dim .....	359
Zhatkina Kristina .....	220
Zyryanov Anton .....	203, 215

## A b s t r a c t s

★ Abasov Rahib and Zohra Safarova (Baku, Azerbaijan) 50

**Is it time to thing about universal methods for solving olympiad problems**

**Abstract.** In this paper we propose a universal method of a constructive nature for solving effectively with some Olympiad class of problems on the theme: «Triangles». The proposed material can be used at the lessons of mathematical circles, as well as for the preparation various competitions.

**Keywords.** Triangle, universal method, constructive method, olympic tasks, efficient solution.

★ Abdushukurov Abdurahim and Rustamjon Muradov (Tashkent, Uzbekistan) ..... 20

**Dependent random censoring from the right and the survival function copula estimates**

**Abstract.** In this article a new estimators for survival functions with using Archimedean copulas under onedimensional and multidimensional random censoring from the right are constructed. The property of uniform consistency of estimators is proved.

**Keywords.** Random censorship, survival function, Archimedean copulas, relative-risk power estimator.

★ Абдушукуров Абдурахим Ахмедович и Нурмухамедова Наргиза Сайдиллаевна (Национальный Университет Узбекистана, Ташкент) ..... 20

**Об асимптотической минимаксной эффективности оценок максимального правдоподобия по неполным данным**

**Abstract.** Показана асимптотическая минимаксная эффективность оценок максимального правдоподобия для неизвестного параметра в модели конкурирующих рисков при случайном цензурировании с двух сторон.

**Keywords.** Оценка максимального правдоподобия, модель конкурирующих рисков.

★ Abdushukurov Abdurahim and Nargiza Nurmuhamedova (Tashkent, Uzbekistan) ..... 20

**On asymptotic minimax efficiency of maximum likelihood estimates by incomplete data**

**Abstract.** We demonstrate asymptotic minimax efficiency of maximum likelihood estimate of unknown parameter in competing risk model under random censoring from both sides.

**Keywords.** Maximum likelihood estimate, competing risk model.

★ Abdushukurov Abdurahim and Leyla Kakadjanova (Tashkent, Uzbekistan) ..... 55

**About some properties of a class of empirical processes**

**Abstract.** In this paper, we discuss the asymptotic properties of generalized empirical processes for certain classes of events and integrable functions, and also considered weak convergence to a Gaussian process.

**Keywords.** Empirical processes, Glivenko-Cantelli classes, metric entropy.

★ Abdushukurov Abdurakhim and Khalmurza Sagidullaev (Tashkent, Uzbekistan) ..... 57

**Estimating the derivatives of the distribution function in the proportional hazards model**

**Abstract.** In this paper we introduce the semiparametric kernel estimates of derivatives of the distribution function in the proportional hazards model of random censorship on the right. We establish the strong consistency of estimates.

**Keywords.** random censoring proportional to the intensity, the nuclear assessment

★ Azimov Jakhongir (Tashkent, Uzbekistan) ..... 60

**About the asymptotic behavior of branching processes with inhomogeneous immigration depending on the state**

**Abstract.** Asymptotic behavior of branching processes with inhomogeneous immigration depending on the state is considered.

**Keywords.** Branching processes.

★ Baranova Irina (Krasnoyarsk, Siberian Federal University) ..... 62

**Regression analysis of bipartite sets of events**

**Abstract.** In work are considered the notion of bipartite set of events, consisting of the sets of events, whose first part corresponds to the random variables, and second part - to the sets, and also the set of bipartite sets. In work are given the classical approach to regression (including pair regression, plural regression and others) and the eventological approach to regression developed by professor O.Ju. Vorob'ov. We study the relationship between bipartite sets of events, and construct eventological regression of these sets.

**Keywords.** Regression, multiple regression, conditional event, eventological regression, bipartite set of events.

★ Baranova Irina and Temerova Elena ..... 73

**The application of entropy to study the distribution of bipartite sets of events**

**Abstract.** In work are considered various definitions of concept of entropy, including eventological approach. Particular attention is paid to the notion of bipartite set of events, consisting of the sets of events, whose first part corresponds to the random variables, and second part - to the sets, and the entropy of bipartite sets of events.

**Keywords.** Thermodynamic entropy, statistical entropy, information entropy, the information, eventological entropy, relative entropy, random event, bipartite set.

★ Belosokhova Anna (Krasnoyarsk) Raspopov Vitaly (Krasnoyarsk) ..... 77

**Numerical research distribution of polluting substances from the Yenisei ferroalloy factory in ecologically significant zone of the city Krasnoyarsk**

**Abstract.** In this work by means of two mathematical models numerically investigated the alleged distribution of pollutants released into the atmosphere by the Yenisei ferroalloy factory (YFF). The software product with the friendly interface is developed, computing experiments are made.

**Keywords.** Gauss model, the convection diffusion equation, the Yenisei ferroalloy factory, the scheme predictor -proofreader, computing experiment, concentration of polluting substance.

★ Beskorsyi Nikita (Krasnoyarsk, Russia) and Oleynikov Boris Vasilevich (Krasnoyarsk, Russia) ..... 80

**Solution of inverse problem of multiple-criteria decision in the methodology of the analytic hierarchy process**

**Abstract.** The this work we formulate inverse problem in the methodology of analytic hierarchy process and offers two methods. Described method based on solution of matrix equation. Described the software developed by the authors and provided results of practical application in the processing of medical data.

**Keywords.** Inverse problem, multiple-criteria decision, analytic hierarchy process, Saaty, pseudoinverse

- ★ Borovikov Igor (California), Sadovsky Michael (Krasnoyarsk)..... 83

#### N-Grams Analysis for Financial Time Series

**Abstract.** By exploring n-grams frequency dictionaries derived from financial time series, we obtain a new view at some aspects of technical analysis on market events study. (TENTATIVE)

**Keywords.** Frequency, money, leisure, more money, more and more and more money.

- ★ Demotchkin Val (London, United Kingdom) ..... 22

#### Hedge Fund Strategies, Financial Instability, and Stock Prices Stabilization

**Аннотация.** The paper discusses hedge funds in connection to the latest financial crisis.

**Ключевые слова.** Hedge Funds, Performance fees, Hedge Fund Redemptions, Capital Raising, traders, speculators, risk, volatility, financial markets, Currency speculations, Bank Run, Global Financial Crisis, eventology, Financial Instability, Nash Equilibrium, Government intervention.

- ★ Dobronets Boris (Krasnoyarsk, Russia), Popova Olga (Krasnoyarsk, Russia) ..... 199

#### Numerical optimization with random data

**Abstract.** The article discusses a new approach to optimization problems with random input parameters, which is defined as a random programming. This approach uses a numerical probability analysis and allows us to construct the set of solutions of the optimization problem based on the joint probability density function.

**Keywords.** Numerical probabilistic analysis, random programming, mathematical programming.

- ★ Dorner Mikhail (Krasnoyarsk, Russia), Zyryanov Anton (Krasnoyarsk, Russia), and Lancev Evgenij (Krasnoyarsk, Russia) ..... 203

#### Identification and prediction of functioning business process models of e-learning

**Abstract.** The article describes the technology of modeling, analysis and simulation of business processes in social systems, eliminates the factor of subjectivity of expert in the construction of a business-model and reduce the complexity of development of a simulation models and evaluate simulations. The technology includes the Process Mining, construction of agent-based simulation model of AnyLogic and forecast the statistical characteristics of the process by GERT-networks. The efficiency of technology tested in the study of the process of e-learning with using the information system Moodle.

**Keywords.** Business-process, Process Mining, GERT-network, agent-based simulation model, e-learning

- ★ Elina Maria (Krasnoyarsk, Russia), Loginova Ekaterina Borisovna (Krasnoyarsk, Russia), and Sadovsky Michael Georgievich (Krasnoyarsk, Russia)..... 208

#### One approach to identify the attitude of voters to a candidate (case study based on mayor elections at Krasnoyarsk on June, 2012)

**Abstract.** New method to identify and reveal some fine peculiarities of the attitude of voters to a politician is presented. The method is based on the analysis of the attribution of a candidate with some proverbs and sayings. Some constraints and difficulties of the provided work are discussed.

**Keywords.** Psychology, linguistics, sociology, projective technique, intellect

- ★ Esin Roman (Krasnoyarsk, Russia)..... 210

#### Solution of the inverse problem of the theory of risk in the model CVaR

**Abstract.** In this paper, we present a method for solving the inverse problem with respect to CVaR, also considered characteristic attitude of investors to risk, and as a key concept — risk aversion. A procedure to determine the parameter of risk aversion on the basis of a detailed questionnaire.

**Keywords.** Risk theory, coherent risk measures, risk aversion, CVaR

- ★ Glik Ludmila, Semenov Mikhail (Tomsk, Russia) . 165

#### Statistical Analysis of Intraday Fluctuations of Spread in Russian Stock Market

**Abstract.** This article is concerned with statistical analysis of intraday fluctuations of spread in Russian stock market. It contains review of existing models which are used to predict the future dynamics of securities quotations. The key characteristics possessed by the time series of increase in the spread are identified. In conclusion the most appropriate model to predict a price movement is selected, which is based on the obtained results.

**Keywords.** Stock market, spread, dynamics of prices, stationarity, autocorrelation.

- ★ Gnezdilova Margaret, Sadovsky Michael (Krasnoyarsk) 169

#### A case study of the efficiency of political ads

**Abstract.** We studied, studied, studied, and finally over-studied. . .

**Keywords.** Frequency, money, leisure, more money, more and more and more money.

- ★ Gordeeva Ksenia (Krasnoyarsk) ..... 175

#### Consumer demand research by the methods of associative rules

**Abstract.** The basic conceptions, objects and fields of the Data Mining are considered in the article. The two association rules method is described: the Apriori method and FP-method. Then it is showed the opportunities of foregoing methods for practical application.

**Keywords.** Data Mining, association rules method, consumer demand.

- ★ Gorokhov Alexandr and Raspopov Vitali (Krasnoyarsk) ..... 180

#### Numerical identification of the right part a special type system of two parabolic equations

**Abstract.** In this paper we numerically solved the inverse problem of coefficient for a system of differential equations of parabolic type with the source functions of a special type. The algorithm of numerical solution of the inverse problem, created a software product performed numerical experiments.

**Keywords.** The inverse problem, the boundary conditions, the conditions override the compatibility conditions, direct differential problem.

- ★ Grigoriev Yuri (Saint-Petersburg) ..... 183

#### Reflexions concerning fate of our high school science

**Abstract.** Reasons concerning the processes what are going on a high school science are considered. The opinion on the transfer inadmissibility of western scientometric standards on the Russian ground is expressed.

**Keywords.** oneself organized system, scientometrics, high school science, citation index, rating.

- ★ Grigoriev Yuri (Saint-Petersburg) ..... 195

#### Hankel determinants and polynomial regression

**Abstract.** On an example of the polynomial regression a use of Hankel determinants in a problem of a correlation ratio recurrent



evaluation and optimum experimental designs construction is showed.

**Keywords.** Hankel determinants, correlation ratio, polynomial regression, canonical moments.

★ Fedorov Eugeny (Krasnoyarsk) ..... 348

**Restoration of model of business processes of distance learning on the basis of data of information system. Problem definition**

**Abstract.** Systems of distance learning in the course of the work keep all data on passing of courses by trainees. Thanks to it there is an opportunity to restore model of processes of distance learning and to carry out its analysis. The Prom system allows to carry out the analysis of the processes presented in the form of a log, it is necessary to lead only data of system of distance learning to a demanded look. обучения к требуемому виду.

**Keywords.** System of distance learning, moodle, manifesto of the analysis of processes, process mining, database, log.

★ Imomov Azam (Karshi city, Uzbekistan), Shahlo Gulyamovna (Karshi city, Uzbekistan) ..... 224

**On convergence to the exponential law and a differential analogue of the basic lemma of the theory of critical Galtona-Watson processes**

**Abstract.** In the present note one criterion of justice of the theorem of convergence to the exponential law and locally-differential analogue of the basic lemma of the theory of critical Galton-Watson processes are investigated. It is discussed their some applications.

**Keywords.** Exponential law, locally-differential analogue; Galton-Watson process, Q-processes.

★ Kireev Vitaliy (Krasnoyarsk, Russia) ..... 227

**On the numerical solution of an isothermal model of an incompressible viscous fluid filtration in porous media**

**Abstract.** Significant scale of oil entry into the environment during the development of oil and gas resources lead to the fact that this type of pollution is central to many areas of oil production. In this case, the developing of the methods for studying problems of ecological safety in the oil pollution is needed. In this paper, a mathematical model for the monitoring of the dynamics of hydrocarbon contamination is considered. These results of numerical experiments are consistent with the literature data.

**Keywords.** Porous media, oil pollution, Darcy's law, the two-phase fluid flow model, implicit iterative methods.

★ Kirilyuk Anna (Krasnoyarsk) ..... 231

**Development and application of electronic educational resource "Algorithms of fast computing"**

**Abstract.** In modern society, high technology and universal use of the computer the ability to quickly and accurately produce in the mind of computing has not lost its relevance. To increase the computing culture need to have good computer skills. Improving the skills of quick calculations depends not only on how to organize a lesson, forms, controls, but also a lot on how the children are interested in this form of work. In developed teaching materials for training students in high-level comprehensive school of fast calculations using multimedia and electronic resources, also developed guidelines for the conduct classes. The use of computer technology in education makes it possible to use in teaching practice psycho-educational development, helps intensify the learning process, to realize the idea of developing training. Modern web-based information and communication technologies allow for the control of their own education, to build individual learning paths, that is, to independently choose the tempo, time and place of learning. Electronic-learning resource consists of two parts: information for teachers, and information for

students. Each unit contains modules for quick calculations of the fifth and sixth grade. The unit for the teacher presents summaries of lessons and presentations for lessons. In the block for students minimal structural unit is the thematic element. For each thematic element has three types of e-learning modules: information, practical training module, the control module. The module is based on practical knowledge of interactivity.

**Keywords.** Algorithm, fast computing, electronic educational resource, thematic element, multimedia, interactivity.

★ Klokova Xenija (Krasnoyarsk), Sadovsky Michael G. (Krasnoyarsk) ..... 237

**On the distribution of the nearest neighbours in Markov chains**

**Abstract.** Some strict results are presented on the study of a distribution of a distance between two neighbours in Markov chains. Lemma is proven describing the distribution rule, and some simulations are carried out. Markov chains of order 2, 3, 4 and 6 were used for the simulation, in four-letter alphabet.

**Keywords.** Frequency, ergodicity, relaxation, genetics.

★ Komorovski Vitold (Krasnoyarsk, Russia), Dorrer Georgy (Krasnoyarsk, Russia) ..... 244

**Cognitive modeling in wildfires risk management**

**Abstract.** This work focuses on the application of cognitive models in the management of forest fire risk. Problems of cognitive map construction, systems analyze with it, and to develop scenarios for the forest fire situation.

**Keywords.** Cognitive models, cognitive maps, forest fires, risks, controllability, stability

★ Kucherov Matvey (Krasnoyarsk, Russia) ..... 249

**Application of the methods of factor analysis for research socio-economic status of population**

**Abstract.** This paper considers concept of factor analysis and its appointment. Learn the basic techniques of classical factor analysis: principal component analysis, centroid and minimal residual method. Also this paper contains description of socio-economic statistic and solution of the practical task of research multifactor statistics using methods above.

**Keywords.** factor analysis, centroid, principal component analysis, correlation, factor

★ Kustitskaya Tatyana (Krasnoyarsk, Russia) ..... 255

**Definition of elliptic acceptance cones for some Holder norms)**

**Abstract.** In this paper a method of defining elliptic acceptance cones is enunciated for the class of generalized coherent risk measures. The method is based on estimating risk aversion in zero for the uniform distribution. The formulas for the axial function calculating were found for the most popular Holder norms.

**Keywords.** Generalized coherent risk measures, risk aversion, elliptic acceptance set, axial function

★ Lantsev Eugeny (Krasnoyarsk, Russia) ..... 261

**Developing e-learning agent-based simulation models using process mining**

**Abstract.** This paper describes agent-based simulation model of discrete mathematics courses based on the process mining results for the LMS Moodle event log. The results of simulation experiments for agent-based model are presented.

**Keywords.** Event analysis, agent-based simulation, AnyLogic, process mining.

★ Lukyanova Natalia (Krasnoyarsk, Russia) ..... 264

**Entropy of a set of events and its subset**



**Abstract.** In paper the author considers idea of equiprobable refinement of probabilistic distribution of a subset of events up to the dimension of the whole set. It needs to expand the concept of relative entropy. Determination of relative quasientropy is introduced. Relative entropy of full and incomplete probability distributions of a set of events and its subset is considered and illustrated on a simple example of any triplet of events.

**Keywords.** Entropy of a set of events, entropy of a subset of events, relative entropy, relative quasientropy.

★ Maksimova Kristina and Irina Baranova (Krasnoyarsk, Russia) ..... 272

#### Finding association rules in products selling statistics

**Abstract.** In article the notion and applications of the Data Mining are considered. The purpose of Data Mining is identify the hidden rules and patterns in data sets. In work are described two association rules methods: Apriori and AprioriSome. It is solved the practical problem of finding the association rules in products selling statistics by aids of both algorithms.

**Keywords.** Data Mining, tasks, association rules method, products selling statistics, Apriori, AprioriSome

★ Масаев Сергей Н. (Красноярск, Россия) ..... 29

#### Оценка управления в компании методом интегральных показателей

**Аннотация.** Метод интегральных показателей компании, основанный на вычислении простой корреляции между исторической серией расходов компании, используется для выработки управленческого решения.

**Ключевые слова.** Корреляция, процесс, системный анализ, интегральные показатели, управление.

★ Masaev Sergey (Krasnoyarsk, Russia) ..... 29

#### Company management system estimation on the basis of integrated indicators

**Abstract.** The method of the structure and indicators analysis of company business processes based on the calculation of simple correlation between historic series of expenses is offered.

**Keywords.** Correlation, adaptation, process, system analysis, management.

★ Mednitskiy Aleksey (Samara, Russia), Trusova Alla (Samara, Russia) ..... 277

#### The actuarial calculations on voluntary medical insurance

**Abstract.** In the article were considered actuarial calculations for voluntary medical insurance. For the calculation of the main part of the net rate and the risk premium was applied formulation of Shuett-Nessbitt. There were shown three methods of correction factors calculating. Also were constituted computations of deductible calculations for several distribution laws. In the environment of programming Borland Delphi was written program, which automates the actuarial calculations. All the methods which were described in the theoretical part, were considered on real facts.

**Keywords.** Insurance, actuarial calculations, voluntary medical insurance, risk, probability, net rate, gross rate, rates, correction factors, unprofitableness, risk premium, franchise.

★ Mednitskiy Aleksey (Samara, Russia), Trusova Alla (Samara, Russia) ..... 280

#### The calculation of the deductible for the beta, gamma, exponential distribution

**Abstract.** In this article were considered the principles of calculation of the deductible for the beta, gamma, exponential distribution. There were introduced probability and mathematical expectation for each of the distributions and these same indices

were computed for conditional franchise. On a basis of real facts there were chosen distributions, which more approximating the real process of receipt losses.

**Keywords.** Insurance, actuarial calculations, franchise, conditional franchise, unconditional franchise, probability, expectation, amount of loss, the beta distribution, gamma distribution, exponential distribution, the approximation.

★ Nartov Jakov (Krasnoyarsk, Russia) ..... 284

#### Poisson distribution visualisation and its mean probable event

**Abstract.** Class of mean probable events for Bernoulli scheme and Poisson distribution visualisation is considered.

**Keywords.** A mean probable events, probability distance, Bernoulli scheme, Poisson distribution.

★ Nifontov Alexander (Krasnoyarsk, Russia) ..... 287

#### Monte-Carlo algorithms for generating distributions of sets of events and their application in the portfolio analysis of currency markets and the precious metals markets

**Abstract.** The direct eventological problem of portfolio analysis of events, the event-analogue of the Markowitz problem of classical portfolio analysis, is considered. The generalized formulation of the direct eventological Markowitz problem is given. It describes a method for selecting a uniform probability distribution and a method of uniform choice of probability distribution with given probabilities of events. It is solved generalized eventological Markowitz direct problem, based on statistical data from the currency markets and the precious metals markets.

**Keywords.** Eventology, event, probability, portfolio analysis, portfolio analysis of events, currency market, precious metals market.

★ Novosyolov Arcady (Krasnoyarsk, Russia) ..... 293

#### Risk sharing among factors and assets

**Abstract.** The paper is devoted to sharing a system risk among elements of the system and exterior factors, which influence the system behavior under risk. Two-way sharing between elements and factors is also considered. For the sake of clarity the system is represented by a financial portfolio governed by a factor risk-model of financial market. Similar results are also true for complex systems behavior in other branches of risk management.

**Keywords.** Risk, factor, asset, risk-model, covariance matrix, risk sharing, Euler theorem.

★ Novosyolov Arcady (Krasnoyarsk, Russia) ..... 301

#### Some problems of fire risk control

**Abstract.** The paper is devoted to fire risk control using restricted resources. We present methods of solution, compare different methods, describe numeric experiments, and make a few recommendations on using methods of control.

**Keywords.** Risk, fire, control, optimization, dynamic programming.

★ Oleynikov Boris and Andrey Shalabay (Krasnoyarsk, Russia) ..... 305

#### The algorithm of thematic science texts classification

**Abstract.** This article describes the original algorithm for automatic classification of scientific texts. Its main advantages are speed and the ability to store data in a SQL database. The analysis of the accuracy of the algorithm

**Keywords.** classification of scientific texts, UDC, algorithm, SQL

- ★ Pazenko Olga(Krasnoyarsk), Sadovsky Michael (Krasnoyarsk) ..... 308

«Beeline» Ads perception and attitude in small settlements of Krasnoyarsk Region

**Abstract.** We studied the peculiarities of the perception of commercial ads in small settlements of Krasnoyarsk region. Some advanced techniques have been implemented. Preliminary results are shown and discussed.

**Keywords.** Perception, advertising, image, cohesion

- ★ Poddubny Vasilii and Alexandr Pekhterev (Tomsk, Russia) ..... 311

Copulas of smoothed empirical distributions at the presence of samples ties (coincidences) and its using in the simulation modeling

**Abstract.** We consider one approach to simulation modeling of frequency features of objects for categorizations and recognitions. The approach is based on the using of copula-functions of smoothed empirical distributions of learning samples with ties (coincidences of values). On the base of the piecewise linear smoothing of marginal empirical distribution functions at a presence of ties, the algorithm of the simulation modeling of random vectors with statistical characteristics (including rank correlations) close to vectors of learning samples is offered. From the condition of equality both of the model empirical and gaussian copulas the multivariate distribution function of simulated vectors is found. As example the simulation modeling of vectors with characteristics of learning samples of the relative frequencies of the syntactic words in texts of Russian prose writers of 19th century is organized. By means of nonparametric goodness of fit testing of distribution laws both of the learning and model samples, their statistical adequacy is examined. It allows to use the simulation modeling for the estimation of the quality of algorithms of categorizations and recognitions in condition of the absence of real objects for the checking.

**Keywords.** Simulation modeling, learning sample, tie, empirical distribution, smoothing, rank correlation, copula, algorithm of modeling, checking of adequacy.

- ★ Popov Alexander (Krasnoyarsk, Russia) and Shabanov Vladislav Saidovich (Krasnoyarsk, Russia) ..... 321

Information retrieval systems for electronic libraries

**Abstract.** The article devoted to development information retrieval systems for electronic libraries. Authors review existed systems. The document model, methods of indexation and information retrieval are proposed. Representation of search results is discussed. The future ways of information retrieval system development are reviewed in conclusion.

**Keywords.** Information retrieval systems, electronic libraries, indexation, information retrieval, relevance, ranking.

- ★ Popova Elena (Krasnoyarsk, Russia) ..... 324

The history of the law of large numbers

**Abstract.** In 1713, the essay was published Bernoulli "Ars Conjectandi", containing proof that established the law of large numbers. SD Poisson later proved a general offer, calling it the law of large numbers. An important further generalization of the law was given by PL Chebyshev.

**Keywords.** The law of large numbers, the history of the law of large numbers, Bernoulli theorem, the 300-th anniversary of the law of large numbers.

- ★ Popova Olga (Krasnoyarsk, Russia) ..... 328

On the approaches to the formulation of optimization problems uncertain data

**Abstract.** This paper studies the the need to consider uncertainties in the practical optimization problems and decision-

making. We are considering the classification uncertainties and are describing the general approach to the formulation and solution of optimization problems in stochastic and interval uncertainty.

**Keywords.** Uncertain programming, stochastic optimization, expected model value, Numerical Probabilistic Analysis.

- ★ Sevastyanova Nelli (Krasnoyarsk, Russia) ..... 333

The generalization of Nikolskiy' results for an error complicated quadrature formulas in the case of a class of functions with fractional derivative.

**Abstract.** The purpose of this work is to research an error of approximate integration with complicated quadrature formulas for functions, which have a fractional derivative order  $\alpha$ , summing by the interval  $[a, b]$  to the power  $p$ .

**Keywords.** Quadrature formula, error, functional of errors, fractional derivative.

- ★ Semenova Daria and Yulia Kochanova (Krasnoyarsk, Russia) ..... 335

Eventological distribution of the random set of events based on associative functions

**Abstract.** In work new approach of the description of eventological distributions of a random set of events by means of the associative functions is offered. Feature of this approach is that for definition of eventological distribution of a random set of events is enough known  $N$  probabilities of events and the type of associative function.

**Keywords.** Random set of events, eventological distribution, associative functions.

- ★ Semenova Daria and Yulia Kochanova (Krasnoyarsk, Russia) ..... 341

Examples of families of associative random sets

**Abstract.** In work the families of associative eventological distributions of random sets of events are entered on the basis of known parametrical families of triangular norms. Examples for random sets of events of small capacity are reviewed.

**Keywords.** Random set of events, eventological distribution, associative functions.

- ★ Shiryayeva Ludmila (Samara, Russia) ..... 351

The distribution of the maximum likelihood-ratio statistic used for testing a normal sample for one outlier

**Abstract.** We find the exact distribution of the test statistic which uses in two-side discordancy test for an extreme outlier. We obtain recursive relationships for the distribution function of the statistic, which uses in two-side discordancy test. We receive the region of critical values of the statistic, where the significance level of criteria equals to the double significance level of the Grubbs criteria.

**Keywords.** Outlier, two-side discordancy test for an extreme outlier, integral distribution function, normal sample.

- ★ Shikhinskaya Reyhan (Baku, Azerbaijan) ..... 358

An example of processing fuzzy inference rules in expert systems

**Abstract.** In paper the rules of inference with application of a defuzzification method WABL is offered.

**Keywords.** Fuzzy modeling, rules of inference.

- ★ Veretnova Xeniya (Krasnoyarsk, Russia) ..... 92

The problem of constructing a set of the most independent events that give rise to a given partition of terraced universal elementary outcomes

**Abstract.** Event-driven idea of principal component, which aims to reduce the dimension of the factor events affecting the collection of hazardous events. Considered the first step of the algorithm of eventological principal component analysis.

**Keywords.** The principal components analysis of dimension reduction, terraced distribution, independent events.

★ Vorobyev Oleg (Krasnoyarsk, Russia) ..... 97

**In search of a primary source: remaking the paper (1975) where at the first time a definition of the lattice (Vorob'ev) expectation of a random set was given**

**Abstract.** Remaking the primary source of a one old good idea of the "lattice expectation", published in 1975 at the first time and subsequently, especially in the western literature on stochastic geometry and the theory of random sets, named from a light hand of Dietrich Stoyan (1994) "Vorob'ev expectation"<sup>1</sup>; had an only historical and methodic value for eventology and probability theory once more reminding how and "...from what rubbish verses grow, not knowing shame".

**Keywords.** Eventology, probability theory, random set, mean measure set, lattice expectation, Vorob'ev expectation.

★ Воробьев Олег Юрьевич (Красноярск) ..... 33

**В поисках первоисточника: римейк статьи (1975), где впервые дается определение решетчатого (воробьевского) ожидания случайного множества**

**Аннотация.** Римейк первоисточника одной хотя и старой, но оказавшейся вполне здоровой идеи "решетчатого ожидания случайного множества", впервые опубликованной мною в 1975 [1, 2] и впоследствии, особенно в западной литературе по стохастической геометрии и теории случайных множеств, названной с легкой руки Дитриха Штойяна [3, 1994] "Vorob'ev expectation" — "воробьевским ожиданием", или "ожиданием по Воробьеву"<sup>2</sup>; имеет лишь историческую и методическую ценность для эвентологии и теории вероятностей, еще раз напоминая, как и "...из какого сора растут стихи, не ведая стыда".

**Ключевые слова.** Эвентология, теория вероятностей, случайное множество, среднее множество, решетчатое ожидание, воробьевское ожидание, ожидание по Воробьеву.

★ Vorobyev Oleg (Krasnoyarsk, Russia) ..... 104

**Applicable eventology of safety: inconclusive totals**

**Abstract.** Totals of the eventological [1, 3] safety system modeling [2, 10, 3, 6, 4, 7] is considered for examples and illustrations, which are intended to demonstrate the main features of the algorithm for calculating the risk of a dangerous event at the company under established the event-related circumstances based on the portfolio of identification indicators

<sup>1</sup>I always tried to name this notion so that the term hints at its sense a little, in the beginning (1975) as the "lattice expectation" of a random finite set to emphasize the finiteness of situation (which has allowed for the first time to open it), and then (1984) as the "mean measure set" of a random (measurable, closed, compact) set. But probably "against the nature you will not trample", and now I'd remember English variant "Vorob'ev expectation" though a special desire to become an eponym behind itself I did not notice.

<sup>2</sup>Всегда старался называть это понятие так, чтобы имя немного намекало на его смысл, вначале (1975) "решетчатым ожиданием" случайного конечного множества, чтобы подчеркнуть конечность ситуации (кстати, позволившей впервые обнаружить это понятие), а затем (1984) "среднемерным множеством" случайного (измеримого, замкнутого, компактного) множества. Но видимо "против природы не попрешь", и теперь придется помнить и об англ. варианте "Vorob'ev expectation", хоть особого желания стать эпонимом за собой не замечал.

of company safety; inter alia the examples and illustrations show the role and functions (in calculating the risk) of the three main event-based figurants in the safety eventological system: the total subject, the total object and the total barrier; and most importantly they reveal the key of eventological approach applicability for the field of safety in the methods for selecting the optimal portfolio of identification indicators of safety providing specified accuracy of estimating risk of the dangerous event for this company by minimal expert costs.

**Keywords.** Eventology, applicable eventology, probability theory, event, probability, set of events, algebra of events, mean probable event, value of an event, Gibbsean event-based model, event identification, total subject, total object, total barrier, portfolio of identification indicators, accuracy of estimating risk, minimum cost of expert.

★ Воробьев Олег Юрьевич (Красноярск) ..... 40

**Применимая эвентология безопасности: неокончательные итоги**

**Аннотация.** Итоги эвентологического моделирования системы безопасности рассматриваются в примерах и иллюстрациях, которые предназначены для демонстрации основных особенностей оценивания риска опасного события на предприятии при сложившихся событийных обстоятельствах на основе портфеля идентификационных признаков безопасности; кроме того, эти примеры и иллюстрации показывают роль и функции в оценивании риска трех основных событийных фигурантов эвентологической системы безопасности: совокупного субъекта, совокупного объекта и совокупного барьера; и, самое главное, обнаруживают ключ применимости эвентологического подхода в области безопасности в методах выбора оптимального портфеля идентификационных признаков безопасности, обеспечивающего заданную точность оценки риска опасного события на данном предприятии минимальными экспертными затратами.

**Ключевые слова.** Эвентология, применимая эвентология, теория вероятностей, событие, вероятность, множество событий, алгебра событий, среднее вероятное событие, ценность события, гиббсовская событийная модель, событийная идентификация, совокупный субъект, совокупный объект, совокупный барьер, портфель идентификационных признаков, точность оценки риска, минимальные экспертные затраты.

★ Vorobyev Oleg (Krasnoyarsk, Russia) ..... 115

**Pigeon and Mind: an eventological transduction of the Herrnstein pigeon experiment and a probability choice problem**

**Abstract.** Notes for lectures on eventology. Eventological transduction of the known Herrnstein pigeon experiment and its connection to the making-decision tasks under uncertainty and to a probability choice problem are considered. There are several options of event-accurate staging pigeon experiment in the Skinner box, similar to the one that was first held by Herrnstein [1961], to finally prove or disprove the Lefebvre hypothesis on the decisive role of a probabilistic choice in the emergence of mind in living organisms. The strict event-staging of the pigeon experiment, the results of which it is possible to draw definitive conclusions about the fate of the Lefebvre hypothesis, is formulated. The new model of the matching reactions of living organisms to reinforcements of the environment, which is called the Gibbs-matching and that sums up what in cognitive psychology called the matching law. Reference sections on the eventological H-theorem, relative entropy, isomorphic sets of events, Möbius inverting probability distributions of a set of events, and cognitive psychology are offered too.

**Keywords.** Eventology, probability theory, elementary outcome, event, probability, set of events, algebra of events, eventological transduction, Herrnstein pigeon experiment, making-decision,



uncertainty, probability choice, matching law, Gibbsean matching, eventological  $H$ -theorem, relative entropy, isomorphic sets of events, Möbius inversing, cognitive psychology, Lefebvre hypothesis.

★ Vorobyev Oleg (Krasnoyarsk, Russia) ..... 133

#### Eventological features of choice of a point in the Frechet-polyhedron of distributions simplex

**Abstract.** Some eventological features of the algorithm of choice of a probability distribution of the set of events in the Frechet-polyhedron of distribution simplex are considered.

**Keywords.** Eventology, probability theory, elementary outcome, event, probability, set of events, algebra of events, uniform choice, distribution simplex, Frechet-polyhedron.

★ Vorobyev Oleg (Krasnoyarsk, Russia) ..... 142

#### Measuring multivariate event-based dependencies in eventological portfolio analysis of financial and actuarial, humanitarian, and socio-economic markets

**Abstract.** New eventological methods for measuring multivariate event-based dependence in the portfolio analysis of financial and actuarial, humanitarian and socio-economic markets explores. We introduce the concept of covariance bullet that is new for portfolio analysis and describes the structure of multivariate dependencies of market events portfolio.

**Keywords.** Eventology, probability theory, elementary outcome, event, probability, set of events, algebra of events, covariance measure of multivariate dependence, set-scattering, eventological portfolio analysis, portfolio event, structure of multivariate dependencies of event portfolio.

★ Vorobyev Oleg (Krasnoyarsk, Russia) ..... 150

#### Gibbsean characterization of sets of events

**Abstract.** Gibbsean characterization and other tools of Gibbsean analysis of probability and value distributions of sets of events are considered.

**Keywords.** Eventology, probability theory, elementary outcome, event, probability, algebra of events, set of events, eventological distribution, Gibbsean characterization of eventological distribution, lay portret of a set of events.

★ Vorobyev Oleg (Krasnoyarsk, Russia) ..... 158

#### Approximation of the set of events in the mean-probable

**Abstract.** The approximation of a set of events in the mean-probable is considered that matches the set of events by the set of mean-probable events.

**Keywords.** Eventology, probability theory, elementary outcome, state, event, probability, set of events, set of states, algebra of states, subalgebra of states, algebra of events, mean-probable event, approximation in the mean-probable.

★ Vorobyev Oleg (Krasnoyarsk, Russia), Nifontov Alexander (Krasnoyarsk, Russia) ..... 162

#### Portfolio analysis of events: An expanded version

**Abstract.** Two basic problems of the portfolio analysis of events (direct and inverse) are considered. The first of them is an analogue of the event based problem from the classical Markowitz portfolio analysis, and the second one solves the problem in inverse Markowitz event-formulation, when under fixed share of capital invested in a portfolio of events the probability distribution of the set of portfolio events under the same restrictions as in direct problem has to be found. The formulation of the inverse problem of an eventological portfolio is an eventology innovation in portfolio analysis: up to now its statement, apparently, not even considered in the framework of the classical portfolio analysis, though, just like the forward, the inverse portfolio problem has obvious practical value.

**Keywords.** Eventology, probability theory, event, probability, portfolio analysis, portfolio analysis of events

★ Tsitsiashvili Gurami (Vladivostok, Russia) ..... 349

#### Image recognition by multivariate intervals

**Abstract.** The paper is devoted to the image recognition by a system of multivariate intervals. Such recognition is more complicated in calculations but more accuracy.

**Keywords.** Algorithm, multivariate interval, image recognition.

★ Yumagulov Dim (Ufa) ..... 359

#### Functional moment management strategy of an investment portfolio

**Abstract.** This paper presents one way to keep the money on Russian stock market. They are functional strategies moment at which rewrites the portfolio at regular intervals according to formal rules.

**Keywords.** Securities market, investment portfolio.

★ Zhatkina Kristina, Baranova Irina (Krasnoyarsk, Russia) ..... 220

#### Solution of the scoring based on the statistics of different types of data

**Abstract.** In article the methods of eventological scoring and bipartite sets of events are considered. It is solved the practical problem of determining the characteristics of borrowers favoring return loan.

**Keywords.** Scoring, loan, event, probability, bipartite set of events

★ Zyryanov A. A. (Krasnoyarsk, Russia) ..... 215

#### Calculation of numerical characteristics of homogeneous GERT-network on the basis of equivalent transformations

**Abstract.** The methods of calculating the numerical characteristics of homogeneous GERT-networks such as the mean and variance. The technique involves conversion algorithm GERT-network to an equivalent arc with the conversion of numerical characteristics of random variables of the arcs on the proposed formulas.

**Keywords.** GERT, stochastic networks, business process, equivalent transformation.

## **П р и г л а ш е н и е**

*Российская академия наук  
Министерство образования и науки РФ  
Институт математики и фундаментальной информатики  
Торгово-экономический институт  
Сибирский федеральный университет  
Научно-исследовательский институт проблем пожарной безопасности  
Институт вычислительного моделирования СО РАН*

# **XIII международная конференция ФИНАНСОВО~АКТУАРНАЯ МАТЕМАТИКА и ЭВЕНТОЛОГИЯ БЕЗОПАСНОСТИ ФАМЭБ'2014**

**18 ~ 19 апреля 2014 года, Красноярск**

### **Т е м ы**

- Финансово-актуарная математика
- Конвергенция нано-био-инфо-когно наук и технологий для совершенствования человеческой деятельности
- Эвентологическая математика в гуманитарных, социо-экономических и естественнонаучных областях
- Эвентология безопасности и риска
- Теория вероятностей и математическая статистика
- Эвентология и математическая эвентология
- Философская эвентология и философия вероятности
- Практическая эвентология
- Эвентологическая экономика
- Эвентологическая психология
- Эвентологические проблемы искусственного интеллекта
- Принятие решений в условиях риска и неопределённости
- Измерение и модели риска
- Теория нечётких событий, обобщенная теория неопределённости
- Математические начала хаоса в экономике
- Системный анализ и управление случайными событиями
- Теория механизмов создания распределений событий

### **Важнейшие даты:**

- 1 марта 2014 — завершение приема заявок, докладов, секций
- 10 марта 2014 — подтверждение и рассылка приглашений

### **Инструкции:**

- <http://fam.conf.sfu-kras.ru/submission.php>



# *I n v i t a t i o n*

*R u s s i a n   A c a d e m y   o f   S c i e n c e s*  
*M i n i s t r y   o f   E d u c a t i o n   a n d   S c i e n c e   o f   R F*  
*I n s t i t u t e   o f   M a t h e m a t i c s   a n d   C o m p u t e r   S c i e n c e*  
*T r a d e   E c o n o m y   I n s t i t u t e*  
*S i b e r i a n   F e d e r a l   U n i v e r s i t y*  
*R e s e a r c h   I n s t i t u t e   f o r   F i r e   S a f e t y   P r o b l e m s*  
*I n s t i t u t e   o f   C o m p u t a t i o n a l   M o d e l i n g   o f   S B   R A S*

## **XIII**

### **International Conference**

### **FINANCIAL and ACTUARIAL MATHEMATICS and EVENTOLOGY of SAFETY**

## **FAMES'2014**

**18 ~ 19 April 2014, Krasnoyarsk**

#### Themes

- Financial and actuarial mathematics
- Converging nano-bio-info-cogno sciences and technologies for improving human performance
- Eventological mathematics in the humanities, socioeconomic and natural sciences
- Eventology of safety and risk
- Probability theory and statistics
- Eventology and mathematical eventology
- Philosophical eventology and philosophy of probability
- Practical eventology
- Eventological economics
- Eventological psychology
- Eventological problems of artificial intelligence
- Decision-making under risk and uncertainty
- Risk measurement and risk models
- Theory of fuzzy events and generalized theory of uncertainty
- Mathematical onset to chaos in economy
- System analysis and events management
- Events distributions mechanism design theory

#### Important dates:

- 1 March 2014 — deadline for papers, and sessions
- 10 March 2014 — notification about acceptance/rejection; distribution of invitations

#### Instructions:

- <http://fam.conf.sfu-kras.ru/submission-e.php>

*Научное издание*

Т р у д ы

XII международной ФАМЭБ'2013 конференции

по финансово-актуарной математике

и эвентологии безопасности

*Под редакцией Олега Воробьёва*

Сибирский федеральный университет  
Институт математики и фундаментальной информатики  
Свободный пр., 79, Красноярск, 660041  
Электронное издание • <http://fam.conf.sfu-kras.ru>

ISBN 978-5-9903358-3-7







## **ЭМ и ФАМ семинары**

Сибирский федеральный университет  
Институт математики и фундаментальной информатики  
Торгово-экономический институт  
НИИ проблем пожарной безопасности  
Институт вычислительного моделирования СО РАН

<http://fam.conf.sfu-kras.ru>  
<http://em.conf.sfu.kras.ru>  
[famemconf@gmail.com](mailto:famemconf@gmail.com)

ISBN 978-5-9903358-3-7



9 785990 335837