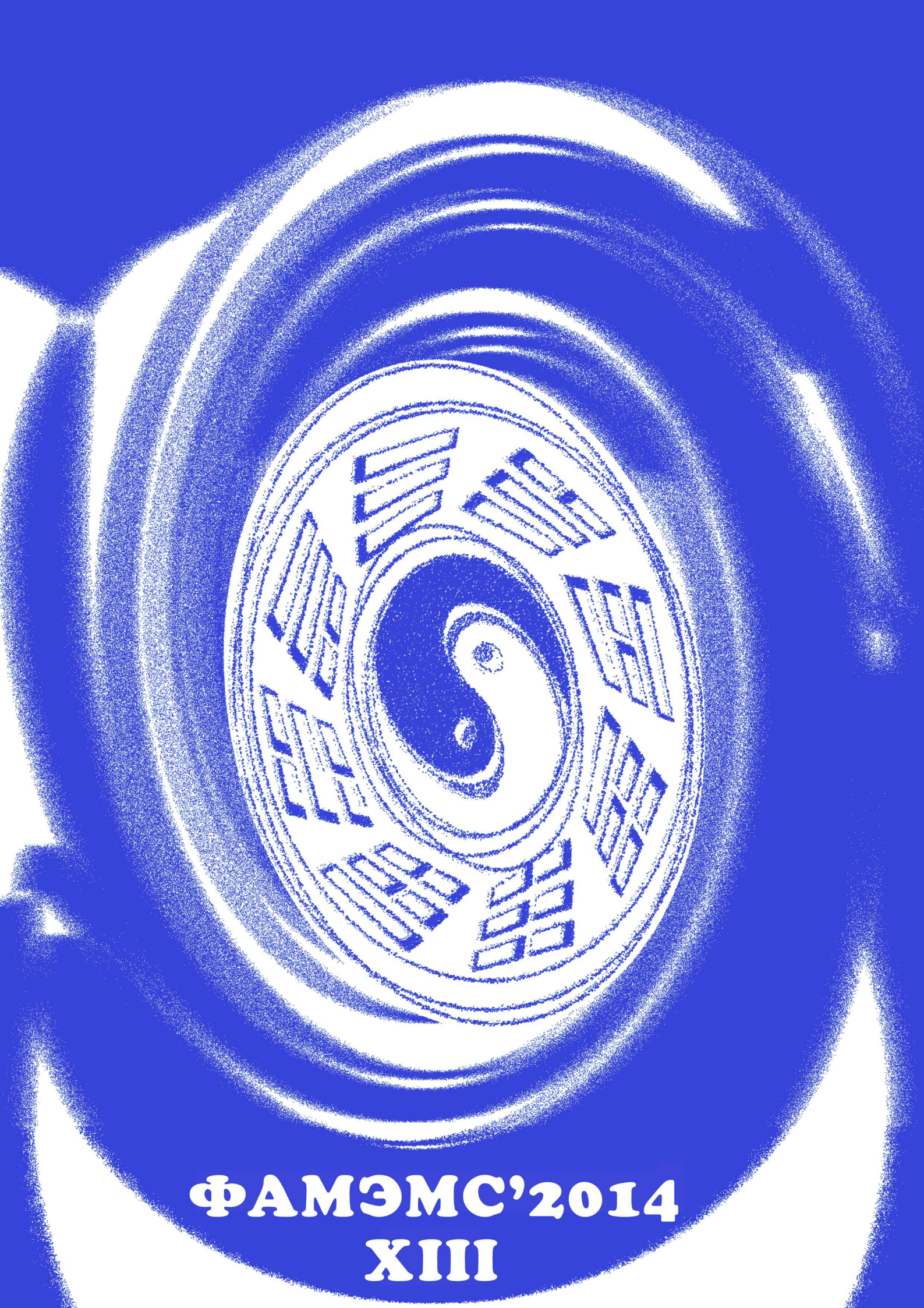


ФАМЭМС'2014
XIII

Российская академия наук
Министерство образования и науки РФ
Сибирский федеральный университет
Институт математики и фундаментальной информатики
Торгово-экономический институт
Институт вычислительного моделирования СО РАН
Научно-исследовательский институт проблем пожарной безопасности

ТРУДЫ

ТРИНАДЦАТОЙ КОНФЕРЕНЦИИ ПО ФИНАНСОВО–АКТУАРНОЙ МАТЕМАТИКЕ И ЭВЕНТОЛОГИИ МНОГОМЕРНОЙ СТАТИСТИКИ

Красноярск
2014

УДК 519.248: [004.8+33+301+159.9]

Т 78

Труды XIII ФАМЭМС'2014 конференции. Под ред. Олега Воробьёва. — Красноярск: СФУ, 2014. — 251 с.

ISBN 978–5–9903358–4–4

Конференция поддержана
Институтом математики и фундаментальной информатики
Торгово-экономическим институтом

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)

Голденок 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)

© Сибирский федеральный университет, 2014
© Институт математики и фундаментальной информатики, 2014
© Торгово–экономический институт, 2014
© Институт вычислительного моделирования СО РАН, 2014
© НИИ проблем пожарной безопасности, 2014

ISBN 978–5–9903358–4–4

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

PROCEEDINGS

OF THE XIII CONFERENCE ON FINANCIAL AND ACTUARIAL MATHEMATICS AND EVENTOLOGY OF MULTIVARIATE STATISTICS

Krasnoyarsk
2014

УДК 519.248: [004.8+33+301+159.9]

Т78

Proceedings of the XIII FAMEMS'2014 Conference. Oleg Vorobyev, ed. — Krasnoyarsk: SFU, 2014. — 251 p.

ISBN 978–5–9903358–4–4

The conference is supported by
Institute of Mathematics and Computer Science
Institute for Trade and Economics

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)

© Siberian Federal University, 2014
© Institute of Mathematics and Computer Science, 2014
© Institute for Trade and Economy, 2014
© Institute of Computational Modeling of RAS, 2014
© Research Institute for Fire Safety Problems, 2014

ISBN 978–5–9903358–4–4

ТРИНАДЦАТАЯ КОНФЕРЕНЦИЯ ПО ФИНАНСОВО–АКТУАРНОЙ МАТЕМАТИКЕ И ЭВЕНТОЛОГИИ МНОГОМЕРНОЙ СТАТИСТИКИ

18 ∼ 19 апреля 2014 года, Красноярск

Темы

- Финансово-актуарная математика
- Математика в гуманитарных, социо-экономических и естественнонаучных областях.
- Эвентология многомерной статистики
- Эвентология безопасности
- Эвентология риска и принятия решений в условиях риска и неопределенности
- Эвентология и математическая эвентология
- Философская эвентология и философия вероятности
- Практическая эвентология
- Эвентологическая экономика и психология
- Эвентологические проблемы искусственного интеллекта
- Эвентоконвергенция наук и технологий
- Теория вероятностей и математическая статистика
- Многомерный статистический анализ
- Принятие решений в условиях риска и неопределенности
- Измерение и модели риска
- Теория нечетких событий, обобщенная теория неопределенности
- Математические начала хаоса в экономике
- Системный анализ и управление случайными событиями

Председатель конференции

Воробьёв О.Ю. (Красноярск)

Программный комитет

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)

Организационный комитет

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

Контакты

- Оргкомитет XIII ФАМЭМС'2014 конференции: Торгово-экономический институт СФУ, ул. Лиды Прушинской 2, к.7-40, Красноярск, 660075, тел. +73912219842, Голденок 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.
- Председатель конференции: Воробьёв Олег Юрьевич,
  т. +73912494795; e-mail: oleg.yu.vorobyev@gmail.com, url: sfu-kras.academia.edu/OlegVorobyev
- Секретарь оргкомитета: Семёнова Дарья Владиславовна,
  т. +79029468205; e-mail: famemconf@gmail.com

THE XIII CONFERENCE ON FINANCIAL AND ACTUARIAL MATHEMATICS AND EVENTOLOGY OF MULTIVARIATE STATISTICS

18 ∼ 19 April, 2014, Krasnoyarsk

Topics

- Financial and actuarial mathematics
- Mathematics in the humanities, socio-economic and natural sciences
- Eventology of multivariate statistics
- Eventology of safety
- Eventology of risk and decision-making under risk and uncertainty
- Eventology and mathematical eventology
- Philosophical eventology and philosophy of probability
- Eventology and the new humanity
- Practical eventology
- Eventological economics and psychology
- Eventological problems of artificial intelligence
- Converging sciences and technologies
- Probability theory and statistics
- Multivariate statistical analysis
- 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

Conference Chair

Vorobyev Oleg (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. (Tomsk)
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)
Szmidt 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 XIII FAMEMS'2014 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 Chair: Vorobyev Oleg,
  tel. +73912494795; e-mail: oleg.yu.vorobyev@gmail.com, url: sfu-kras.academia.edu/OlegVorobyev
- Organizing Secretary: Semenova Daria,
  tel. +79029468205; e-mail: famemconf@gmail.com

Содержание

Abdushukurov A. A., Kakadjanova L. R. *(Tashkent, Uzbekistan)* 11
On special empirical processes of independence with application

Abdushukurov A. A., Chuyanov Kh. *(Tashkent, Uzbekistan)* 13
Estimating of survival function in cox model under competing risks

Demotchkine Val *(London, United Kingdom)* 15
Hedge Fund Strategies, Financial Instability, Fat Tails and Leverage Cycles

Imomov A.A., Almanova D. S. *(Karshi, Uzbekistan)* 19
On estimation of main parameter in the long-surviving Markov branching process

Imomov A. A., Jumakulov K. Q., Tukhtaev E. E. *(Karshi, Uzbekistan)* 21
On estimation of main parameter in Q-process

Абдушукуров А. А. и Р. С. Мурадов *(Ташкент, Узбекистан)* 23
Асимптотическая гауссовость копула оценок функции выживания при зависимом случайном цензурировании

Баранова И. В. *(Красноярск)* 27
Ковариационный анализ разнотипных данных с помощью метода двудольных множеств событий

Баранова И. В. и Д. М. Богодухов *(Красноярск)* 35
Применение генетических алгоритмов для решения задачи кластеризации данных

Баранова И. В. и Д. В. Князь *(Красноярск)* 40
Методы кластеризации многомерных статистических данных

Воробьев О.Ю. *(Красноярск)* 44
Вне привычных измерений статистических взаимосвязей: эвентологически пересмотренный подход к измерению общности и ассоциации (статистической зависимости) множеств событий

Воробьев О.Ю. *(Красноярск)* 95
Сет-дистограммные измерители разобщенности событий

Воробьев О.Ю. *(Красноярск)* 112
К обоснованию эвентологического метода

Воробьев О.Ю. *(Красноярск)* 122
Эвентологическое обобщение границ Фреше для множества со́-бытий

Голденок Е. Е. *(Красноярск)* 131
Оценка общности покупательских сегментов по чековой статистике

Григорьев Ю. Д. *(Санкт-Петербург)* 136
Полярное уравнение и ϕ_p-оптимальные планы эксперимента для полиномиальной регрессии

Джамирзаев А. А. *(Ташкент, Узбекистан)* 142
О предельном распределении последовательностей случайных величин со случайным индексом

Додонова М. М. *(Красноярск)* 143
Изучение различных постановок задачи о рюкзаке и методов их решения

Есин Р. В. *(Красноярск)* 147
Решение обратной задачи теории риска для меры возмущенной вероятности

Зупаров Т. М. и О. Сафаров *(Ташкент, Узбекистан)* 153
Центральная предельная теорема для случайных сумм линейного процесса, порожденного последовательностью m-зависимых случайных

Клоков А. В. и В. Е. Распопов *(Красноярск)* 155
Численная идентификация старших коэффициентов в системе параболических уравнений

Крупкина Т. В., Мажаров В. Ф., Горный Б. Э. и С. В. Бабенышев *(Красноярск, Железногорск)* *158*
О различных подходах к построению интегральных показателей

Кустицкая Т .А. *(Красноярск)* *163*
Когерентные меры риска. Примеры, модификации, обобщения

Максимова К. И. и И. В. Баранова *(Красноярск)* *170*
Разработка вероятностного алгоритма поиска ассоциативных правил в статистике покупок

Махонин И. В., Демиденко В. А. *(Красноярск)* *178*
Проблемы сбалансированности знаковых графов

Нартов Я. В. *(Красноярск)* *180*
Регрессия как условное средневероятное событие

Нифонтов А. С. *(Красноярск)* *186*
Средне-феноменный портфельный анализ

Новоселов А. А *(Красноярск)* *193*
Стресс-тестирование риск-моделей: некоторые подводные камни

Новоселов А. А *(Красноярск)* *196*
Обобщение понятия границ Фреше

Новоселов А. А *(Красноярск)* *200*
Измеримость случайного множества событий

Поддубный В. В. и А. А. Поликарпов *(Томск, Москва)* *213*
Непрерывная стохастическая динамическая модель развития полисемии и смыслового объема ансамбля знаков естественного языка

Рахимова Г. Г., Сагидуллаев К. С. и Г. Т. Турсунов *(Ташкент, Узбекистан)* *202*
Скорость сходимости в последовательном оценивании интервалами фиксированной ширины асимптотической дисперсии ранговых оценок параметра сдвига

Садовский М. Г. и Я. В. Гребнев *(Красноярск)* *204*
Правила Чаргаффа и супер-симметрия в геномах

Садовский М. Г. и К. А. Никитина *(Красноярск)* *208*
Поиск и изучение структур в символьных последовательностях, заданных случайными процессами

Саркисов В. Г. *(Самара)* *219*
Формирование объединенных портфелей инвесторов при различных критериях риска

Семенова Д. В. и А. И. Иванова *(Красноярск)* *222*
Нечеткая и сет-регрессионная модели распределения потребительских предпочтений между фирмами

Семенова Д. В. и Г. Кушбак кызы *(Красноярск)* *228*
Корреляционное отношение между случайными множествами событий

Семенова Д. В., Лукьянова Н. А., и Л. Ю. Шангареева *(Красноярск)* *234*
Семейство случайных множеств событий Франка

Семенова Д. В., Лукьянова Н. А., и Л. Ю. Шангареева *(Красноярск)* *239*
Оценка параметра ассоциативной функции методом наименьших квадратов для семейств Франка и Али-Михаэля-Хака случайных множеств событий

Список авторов *244*

List of authors *244*

Abstracts *245*

Contents

Abdushukurov Abdurakhim and Leila Kakadjanova *(Tashkent, Uzbekistan)* 11
On special empirical processes of independence with application

Abdushukurov Abdurakhim and Khurshid Chuyanov *(Tashkent, Uzbekistan)* 13
Estimating of survival function in cox model under competing risks

Demotchkine Val *(London, United Kingdom)* 15
Hedge Fund Strategies, Financial Instability, Fat Tails and Leverage Cycles

Imomov Azam and Dilnoza Almanova *(Karshi, Uzbekistan)* 19
On estimation of main parameter in the long-surviving Markov branching process

Imomov Azam, Khurshid Jumakulov and Erkin Tukhtaev *(Karshi, Uzbekistan)* 21
On estimation of main parameter in Q-process

Abdushukurov Abdurahim and Rustamjon Muradov *(Tashkent, Uzbekistan)* 23
The asymptotic Gaussian properties of copula estimating the survival function for dependent random censoring (in Russian)

Baranova Irina *(Krasnoyarsk, Russia)* 27
Covariance analysis of polytypic data using the bipartite sets of events method (in Russian)

Baranova Irina and Dmitriy Bogodukhov *(Krasnoyarsk, Russia)* 35
Application of genetic algorithms to solve the problem of data clustering (in Russian)

Baranova Irina and Dimitriya Knyaz' *(Krasnoyarsk, Russia)* 40
Clustering methods of multivariate statistical data (in Russian)

Vorobyev Oleg *(Krasnoyarsk, Russia)* 44
Beyond usual measuring statistical interconnections: An eventologically revised approach to measurement for agreement and association (statistical dependence) of a set of events (in Russian)

Vorobyev Oleg *(Krasnoyarsk, Russia)* 95
Set-distogram measurers of disagreement of events (in Russian)

Vorobyev Oleg *(Krasnoyarsk, Russia)* 112
On the foundations of the eventological method (in Russian)

Vorobyev Oleg *(Krasnoyarsk, Russia)* 122
Eventological generalization of Fréchet bounds for a set of có-events (in Russian)

Goldenok Ellen *(Krasnoyarsk, Russia)* 131
Assessment of an association of consumer segments for check market statistics (in Russian)

Grigoriev Yuri *(Санкт-Петербург)* 136
Polar equation and experimental ϕ_p-designs for polynomial regression (in Russian)

Djamirzayev Abdusunnat *(Ташкент, Узбекистан)* 142
About the limiting distribution sequences of random variables with random indices (in Russian)

Dodonova Maria *(Krasnoyarsk, Russia)* 143
Research of various formulations of the knapsack problem and methods of problem solution (in Russian)

Esin Roman *(Krasnoyarsk, Russia)* 147
Solve the inverse problem for the theory of risk measures distorted probability (in Russian)

Zuparov Talat and Oybek Safarov *(Ташкент, Узбекистан)* 153
The central limit theorem for random sums of a linear process generated sequence of m-dependent random variables (in Russian)

Klokov Andrey and Vitaly Raspopov *(Krasnoyarsk, Russia)* 155
Numerical identification of the leading coefficients of the system of parabolic equations (in Russian)

Krupkina Tatiana, Mazharov Vladimir, Gornyi Boris, and Babenyshev Sergey (*Krasnoyarsk, Zheleznogorsk, Russia*) ... *158*
About various approaches to the construction of integral indicators (in Russian)

Kustitskaya Tatyana (*Krasnoyarsk, Russia*) ... *163*
Coherent risk measures. Examples, variations, generalizations (in Russian)

Maksimova Kristina and Irina Baranova (*Krasnoyarsk, Russia*) ... *170*
The problem of finding the association rules method in Data Mining (in Russian)

Makhonin Igor and Vitaliy Demidenko (*Krasnoyarsk, Russia*) ... *178*
Problems of balance of sign graphs (in Russian)

Nartov Yakov (*Krasnoyarsk, Russia*) ... *180*
Regression as a Conditional Mean Probability Event (in Russian)

Nifontov Aleksander (*Krasnoyarsk, Russia*) ... *186*
Mean-phenomenon portfolio analysis (in Russian)

Novosyolov Arcady (*Krasnoyarsk, Russia*) ... *193*
Stress-testing of risk models: some pitfalls (in Russian)

Novosyolov Arcady (*Krasnoyarsk, Russia*) ... *196*
A generalization of Frechet bounds concept (in Russian)

Novosyolov Arcady (*Krasnoyarsk, Russia*) ... *200*
Measurability of random set of events (in Russian)

Poddubny Vasily and Anatoly Polikarpov (*Томск, Москва*) ... *213*
Continuous Stochastic Dynamic Model for the Evolution of Polysemy and Sense Volume of Signs Ensembles of Natural Language (in Russian)

Rakhimova Gulnoza, Sagidullaev Kalmurza and Gafurjon Tursunov (*Ташкент, Узбекистан*) ... *202*
The rate of convergence in the sequential estimation of fixed-width intervals of asymptotic variance of rank estimators of shift (in Russian)

Sadovsky Michael and Yaroslav Grebnev (*Krasnoyarsk, Russia*) ... *204*
Chargaff's rules and super-symmetry in genomes (in Russian)

Sadovsky Michael and Ksenya Nikitina (*Krasnoyarsk, Russia*) ... *208*
Search and study of structures in the character sequences defined by random processes (in Russian)

Sarkisov Vigen (*Самара*) ... *219*
Creating Investors' Pooled Portfolios at Various Risk Criteria (in Russian)

Semenova Daria and Anastasiya Ivanova (*Krasnoyarsk, Russia*) ... *222*
Fuzzy and set-regression model of distribution of consumer preferences between firms (in Russian)

Semenova Daria and Guliza Kushbak kyzy (*Krasnoyarsk, Russia*) ... *228*
Correlation ratio between random sets of events (in Russian)

Semenova Daria, Lukyanova Natalia and Ludmila Shangareeva (*Krasnoyarsk, Russia*) ... *234*
Family of random sets of events of Frank (in Russian)

Semenova Daria, Lukyanova Natalia and Ludmila Shangareeva (*Krasnoyarsk, Russia*) ... *239*
Estimate of the parameter of associative function by the least squares method for associative random sets of events of Frank and Ali-Mikhail-Haq (in Russian)

List of authors (in Rissian) ... *244*

List of authors (in English) ... *244*

Abstracts ... *245*

On special empirical processes of independence with application

Abdurahim Axmedovich Abdushukurov
National University of Uzbekistan
Uzbekistan
a_abdushukurov@rambler.ru

Leyla Reshitovna Kakadjanova
National University of Uzbekistan
Uzbekistan
leyla_tvms@rambler.ru

Abstract. *In this paper we investigate the asymptotic properties of one class of empirical processes for certain classes of integrable functions.*

Keywords. *Empirical processes; metric entropy; Glivenko-Cantelli's and Donsker's theorems.*

1 Introduction

In this paper we investigate the limit properties of a class of empirical processes of independence indexed on a set of measurable functions. Consider a following model of experiments in which observed pairs are consists of $\{(X_k, A_k), k \geq 1\}$, where X_k are random elements defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with values in a measurable space $(\mathfrak{X}, \mathfrak{B})$. Events A_k have a common probability $p \in (0,1)$. Let $\delta_k = I(A_k)$ is indicator of the event A_k. At the $n-$ th step of experiments is observed a sample $\mathbb{S}^{(n)} = \{(X_k, \delta_k),\ 1 \leq k \leq n\}$. Each pair in the sample $\mathbb{S}^{(n)}$ induced a statistical model with sample space $\mathfrak{X} \otimes \{0,1\}$, σ-algebra of sets of the form $B \times D$ and induced distribution

$$\mathbb{Q}^*(B \times D) = \mathbb{P}(X_k \in B, \delta_k \in D),$$

where $B \in \mathfrak{B}$, $D \subset \{0,1\}$. Define submeasures

$$\mathbb{Q}_1(B) = \mathbb{Q}^*(B \times \{1\}), \qquad \mathbb{Q}_0(B) = \mathbb{Q}^*(B \times \{0\})$$

and

$$\mathbb{Q}(B) = \mathbb{Q}^*(B \times \{0,1\}) = \mathbb{Q}_0(B) + \mathbb{Q}_1(B),$$

$B \in \in \mathfrak{B}$ and its estimates

$$\mathbb{Q}_{1n}(B) = \frac{1}{n}\sum_{k=1}^{n} \delta_k I(X_k \in B),$$

$$\mathbb{Q}_{0n}(B) = \frac{1}{n}\sum_{k=1}^{n} (1-\delta_k) I(X_k \in B),$$

$$\mathbb{Q}_n(B) = \frac{1}{n}\sum_{k=1}^{n} I(X_k \in B) = \mathbb{Q}_{0n}(B) + \mathbb{Q}_{1n}(B)$$

for all $B \in \mathfrak{B}$. Consider the hypothesis $\mathcal{H}$ of independence of X_k and A_k for each $k \geq 1$.

We consider general classes of specially normalized empirical processes of independence indexed by the class $\mathcal{F}$ of measurable functions $f : \mathfrak{X} \to \mathbb{R}$. For a signed measure $\mathbb{G}$ and function $f \in \mathcal{F}$ denote the integral

$$\mathbb{G}f = \int_{\mathfrak{X}} f d\mathbb{G}.$$

Define $\mathcal{F}$ - indexed empirical process $\mathbb{G}_n : \mathcal{F} \in \mathbb{R}$ as:

$$f \mapsto \mathbb{G}_n f = \sqrt{n}(\mathbb{Q}_n - \mathbb{Q})f =$$

$$= n^{-1/2}\sum_{k=1}^{n} (f(X_k) - \mathbb{Q}f),\ f \in \mathcal{F}.$$

Note that $\mathbb{G}_n f = \mathbb{G}_{0n} f + \mathbb{G}_{1n} f$, where $\mathbb{G}_{jn} f = \sqrt{n}(\mathbb{Q}_{jn} - \mathbb{Q}_j) f$, $j = 0,1$. Donsker-type theorems provide a general conditions on $\mathcal{F}$, under which

$$\mathbb{G}_n f \Rightarrow \mathbb{G}f \ \text{ in } \ l^\infty(\mathcal{F}), \tag{1}$$

where $l^\infty(\mathcal{F})$ - the space of all bounded functions $f : \mathfrak{X} \to \mathbb{R}$ equipped with the supremum — norm $\|f\|_{\mathcal{F}}$ and $\Rightarrow$ means the weak convergence. Limiting field $\{\mathbb{G}f,\ f \in \mathcal{F}\}$ called $\mathbb{Q}$ — Brownian bridge. In connection with the problem of testing the hypothesis $\mathcal{H}$, we introduce $\mathcal{F}$ - processes $\Lambda f = \mathbb{Q}_1 f - p\mathbb{Q}f$, $\Lambda_n f = = \mathbb{Q}_{1n} f - p_n \mathbb{Q}_n f$, $f \in \mathcal{F}$ and

$$\Delta_n f = \int_{\mathfrak{X}} f d\Delta_n = \left(\frac{n}{p_n(1-p_n)}\right)^{1/2} (\Lambda_n - \Lambda) f, \tag{2}$$

$f \in \mathcal{F}$. One of important properties of the process (2) is its convergence to the same $\mathbb{Q}$-Brownian bridge $\{\mathbb{G}f,\ f \in \mathcal{F}\}$ under validity of $\mathcal{H}$.

To present the basic theorems we define the complexity or entropy of class $\mathcal{F}$. Bracketing (or covering) number $N_{[]}(\varepsilon, \mathcal{F}, \mathcal{L}_q(\mathbb{Q}))$ is the minimum number of ε - brackets in $\mathcal{L}_q(\mathbb{Q})$ needed to cover $\mathcal{F}$ (see Shorack and Wellner [4], Van der Vaart and Wellner [5]):

$$N_{[]}(\varepsilon, \mathcal{F}, \mathcal{L}_q(\mathbb{Q})) = \min\left\{ k : \text{ for some } f_1, ..., f_k \in \mathcal{L}_q(\mathbb{Q}), \mathcal{F} \subset \bigcup_{i,j} [f_i, f_j] : \|f_j - f_i\|_{\mathbb{Q},q} \leq \varepsilon. \right.$$

For weak convergence of $\mathcal{F}$ - indexed empirical processes (2) we need the integral of the metric entropy with bracketing to be

$$J_{j[]}^{(q)}(\delta) = J_{j[]}(\delta; \mathcal{F}; \mathcal{L}_q(\mathbb{Q}_j)) =$$

$$= \int_0^\delta (H_{jq}(\varepsilon))^{1/2} d\varepsilon, j = 0,1, \text{ for } 0 < \delta < 1,$$

© 2014 A.A. Abdushukurov and L.R. Kakadjanova

Oleg Vorobyev (ed.), Proc. XIII FAMEMS'2014, Krasnoyarsk: SFU

where $H_{jq}(\varepsilon) = \log N_{[]}(\varepsilon, \mathcal{F}, \mathcal{L}_q(\mathbb{Q}_j))$ is metric entropy of class $\mathcal{F}$ in $\mathcal{L}_q(\mathbb{Q}_j)$, $j = 0, 1$. We introduce the following conditions:

(i) Let the class $\mathcal{F}$ such that

$$\mathcal{F} \subset \mathcal{L}_2(\mathbb{Q}_j) \text{ and } J^{(2)}_{j[]}(1) < \infty, \ j = 0, 1. \tag{3}$$

Теорема 1. *Under the conditions (3) for* $n \to \infty$

$$\Delta_n f \Rightarrow \Delta f \text{ in } l^\infty(\mathcal{F}), \tag{4}$$

where $\{\Delta f, \ f \in \mathcal{F}\}$ *is a Gaussian field with zero mean and under validity of the hypothesis* $\mathcal{H}$*, it coincides in distribution with* $\mathbb{Q}$ *- Brownian bridge.*

Now concider case of random sample size. Let the sequence N_n of Poisson r.v.-s with mean n. Suppose that sequences $\{N_n, \ n \geq 1\}$ and $\{(X_k, \delta_k), \ k \geq 1\}$ are independent. Let $\left\{\Delta^*_{N_n} f, \ f \in \mathcal{F}\right\}$ be sequence of normalized empirical processes of independence obtained from (2) by replacing upper index n of all summation to a random sequence N_n .

Теорема 2. *Under the conditions (3) at* $n \to \infty$

$$\Delta^*_{N_n} f \Rightarrow \Delta^* f \text{ in } l^\infty(\mathcal{F}), \tag{5}$$

where by hypothesis $\mathcal{H}$*,* $\Delta^* f \stackrel{d}{=} \mathbb{W}(f)$*,* $f \in \mathcal{F}$*. Here* $\{\mathbb{W}(f), f \in \mathcal{F}\}$ *is Brownian sheet.*

2 Application to random censoring

Let's consider a right random censoring model, where $X_i = \min\{T_i, C_i\}$ and $A_i = \{T_i \leq C_i\}$. Here r.v.-s T_i and C_i denotes life times and censoring times, mutually independent with common continuous distribution functions F and G respectively ($F(0) = G(0) = 0$). Then in observing data $\mathbb{S}^{(n)} = \{(X_i, \delta_i), 1 \leq i \leq n\}$ with $\delta_i = I(A_i)$, r.v.-s of interest T_i are observed when A_i occurs, i.e. $\delta_i = 1$. Observe that the X_i have common distribution function $H = 1 - (1 - F)(1 - G)$ and subdistributions defined as

$$\mathbb{Q}_0(B) = \mathbb{P}(X_k \in B, \delta_k = 0) =$$

$$= \mathbb{P}(C_k \in B \cap [0, T_k)) \int_B (1 - F(t))G(dt),$$

$$\mathbb{Q}_1(B) = \mathbb{P}(X_k \in B, \delta_k = 1) =$$

$$= \mathbb{P}(T_k \in B \cap [0, C_k]) \int_B (1 - G(t))F(dt). \tag{6}$$

Now consider simple proportional hazards model (PHM) or Koziol-Green model which has been found in statistical literature to be very useful and interesting in practice. In PHM assume the parametric relation

$$1 - G = (1 - F)^\beta \text{ for some } \beta > 0. \tag{7}$$

Under (7) it is easy seen that $1 - F = (1 - H)^p$, where $p = \frac{1}{1+\beta} = \mathbb{P}(A_k)$. One of basic property of PHM that (7) holds when r.v.-s X_k and δ_k are independent and only this case. Such characterization of PHM plays a basic role in constructing and investigating of estimators of many functionals of distribution F. Consider sufficient maximum likelihood ACL-estimator of F:

$$F_n(t) = 1 - (1 - H_n(t))^{p_n}, \tag{8}$$

where $H_n(t) = \frac{1}{n}\sum_{k=1}^n I(X_k \leq t)$ and $p_n = \frac{1}{n}\sum_{k=1}^n \delta_k$ are independent empirical estimators of $H(t)$ and p respectively.

There are a rich literature devoted to statistical analysis of F_n. These papers are concerned with the superiority of methods for estimation and the testing in PHM are based on F_n rather than the product-limit estimator of Kaplan-Meier. Most of the outstanding references can be found in [2]. Hence we have a natural question: when can the advantages of the PHM be used? In other words, there is now a need for testing of validity of PHM, i.e. for the composite hypothesis described by relation (7). But this relation is equivalent to hypothesis $\mathcal{H}$ on independence of r.v.-s $(X_1, ..., X_n)$ and $(\delta_1, ..., \delta_n)$. Consider following special empirical process of (2):

$$\Delta_n(t) = \left(\frac{n}{p_n(1 - p_n)}\right)^{1/2} (H_{1n}(t) - p_n H_n(t)),$$

$$-\infty < t < \infty, \tag{9}$$

where $H_{1n}(t) = \frac{1}{n}\sum_{k=1}^n I(X_k \leq t, \delta_k = 1)$. Then we have next consequence from Theorem 1: If $\mathcal{H}$ holds then as $n \to \infty$

$$\Delta_n(\cdot) \Rightarrow \mathbb{B}(H(\cdot)), \tag{10}$$

where $\{\mathbb{B}(y), 0 \leq y \leq 1\}$ is a Brownian bridge.

References

[1] Абдушукуров А.А., Какаджанова Л.Р. Об одном классе эмпирических процессов. *Труды XVII Международной конференции по эвентологической математике и смежным вопросам. Красноярск.*, pages 30–31, 2012.

[2] Csörgő S., Faraway J.J. The paradoxical nature of the proportional hazards model of random censorship. *Statistics*, pages 67–78, 1998.

[3] Gaensler P., Stute W. Empirical processes: a survey of results for independent and identically distributed random variables. *Ann. Probab*, 7, №2:193–243, 1979.

[4] Shorack G.R., Wellner J.A. *Empirical processes with applications to statistics.* JOHN WILEY and SONS, 1986.

[5] Van der Vaart A. W., Wellner J.A. *Weak convergence and empirical processes.* Springer, 1996.

Estimating of survival function in cox model under competing risks

Abdushukurov Abdurahim Akhmedovich
National University of Uzbekistan
Uzbekistan, Tashkent
a_abdushukurov@rambler.ru

Chuyanov Khurshid
National University of Uzbekistan
Uzbekistan, Tashkent

Abstract. *We propose three types of and corresponding semiparametric estimators for survival functions in Cox model in presence of competing risks.*

Keywords. *Survival function, Cox model, Competing risks.*

1 Introduction

In this article we consider Cox model with failure time Z and k competing risks

$$\Lambda^{(i)}(x/v) = \Lambda_0^{(i)}(x)\exp\left(\left(\beta^T, v\right)\right), \quad x \geq 0,\ i = \overline{1,k}, \tag{1}$$

which postulates that i-th conditional cumulative hazard function

$$\Lambda^{(i)}(x/v) = \lim_{h\downarrow 0} P\left(Z \leq x+h,\ A^{(i)}/Z \geq x, \mathbb{V} = v\right),$$

can be written as a product of an exponential factor depending on the p-vectors of covariate $\mathbb{V} = v$, regression parameterβ and a i-th baseline cumulative hazard function $\Lambda_0^{(i)}(x) = \Lambda^{(i)}(x/0)$. Here $A^{(1)}, ..., A^{(k)}$ are pairwise disjoint events. Let $1 - H^{(i)}(x/v)$ the conditional and $1 - H_0^{(i)}(x)$ the baseline survival functions of i-th risk, where $H_0^{(i)}(x) = H^{(i)}(x/0)$ and

$$1 - H^{(i)}(x/v) = \left[1 - H_0^{(i)}(x)\right]^{\exp\left(\left(\beta^T, v\right)\right)}, \tag{2}$$

$x \geq 0, i = \overline{1,k}$.

2 Main results

We consider following three functionals of exponential, product and relative - risk power types of cumulative hazard functions:

$$1 - H_1^{(i)}(x/v) = \exp\left\{-\Lambda_{0c}^{(i)}(x)\exp\left(\left(\beta^T, v\right)\right)\right\} \cdot \prod_{u\leq x}\exp\left\{-\Delta\Lambda_0^{(i)}(u)\exp\left(\left(\beta^T, v\right)\right)\right\},$$

$$1 - H_2^{(i)}(x/v) = \exp\left\{-\Lambda_{0c}^{(i)}(x)\exp\left(\left(\beta^T, v\right)\right)\right\} \cdot \prod_{u\leq x}\left[1 - \Delta\Lambda_0^{(i)}(u)\exp\left(\left(\beta^T, v\right)\right)\right] \tag{3}$$

$$1 - H_3^{(i)}(x/v) = \left[1 - H(x - /v)\right]^{R^{(i)}(x)},$$

where

$$1 - H(x/v) = \prod_{i=1}^{k}\left(1 - H^{(i)}(x/v)\right),$$

$$\Delta\Lambda_0^{(i)}(u) = \Lambda_0^{(i)}(u) - \Lambda_0^{(i)}(u-),$$

$$\Lambda_{0c}^{(i)}(x) = \Lambda_0^{(i)}(x) - \sum_{u\leq x}\Delta\Lambda_0^{(i)}(u),$$

$$R^{(i)}(x) = \frac{\Lambda_0^{(i)}(x)}{\sum\limits_{i=1}^{k}\Lambda_0^{(i)}(x)}, i = \overline{1,k}.$$

Let's introduce functionals:

$$\omega_1^{(i)}(x,v) = \frac{1}{2}\exp\left\{-\Lambda_{0c}^{(i)}(x)\exp\left(\left(\beta^T, v\right)\right)\right\} \cdot \sum_{u\leq x}\Delta\Lambda_0^{(i)}(u)\exp\left(\left(\beta^T, v\right)\right),$$

$$\omega_2^{(i)}(x,v) = \sum_{u\leq x}\frac{\Delta\Lambda_0^{(i)}(u)\exp\left(\left(\beta^T, v\right)\right)}{\left[1 - \Delta\Lambda_0^{(i)}(u)\exp\left(\left(\beta^T, v\right)\right)\right]},$$

$$\omega_3(x,v) = \sum_{u\leq x}\frac{\left[\Delta\Lambda_H(u/v)\right]^2}{\left[1 - \Delta\Lambda_H(u/v)\right]} - \Delta\Lambda_H(x/v),$$

$$\Lambda_H(x/v) = \int\limits_{[0,x]}\frac{dH(u/v)}{1 - H(u - /v)},$$

and

$$\omega_4^{(i)}(x,v) = \omega_2^{(i)}(x,v) - \omega_3(x,v).$$

Note that functionals (3) are very useful in estimating of survival functions (2) under various types of censoring (see, for example [1, 3]). Therefore next we formulate theorem which gives some exact inequalities between functionals (3). Let $T_H = \inf\{x : H(x/v) = 1\}$.

Theorem 1. *For all* $(x,v) \in [0, T_H] \times \mathbb{R}^p$ *and* $i = 1, ..., k$:

(I) $0 \leq H_2^{(i)}(x/v) - H_1^{(i)}(x/v) \leq \omega_1^{(i)}(x,v)$;

© 2014 A.A. Abdushukurov, Kh. Chuyanov

Oleg Vorobyev (ed.), Proc. XIII FAMEMS'2014, Krasnoyarsk: SFU

(II) $0 < -\log\left(1-H_2^{(i)}(x/v)\right) + \log\left(1-H_1^{(i)}(x/v)\right) < \omega_2^{(i)}(x,v);$

(III) $0 < \log\left(1-H_3^{(i)}(x/v)\right) + \log\left(1-H_1^{(i)}(x/v)\right) < \omega_3(x,v);$

(IV) $\left|-\log\left(1-H_2^{(i)}(x/v)\right) + \log\left(1-H_3^{(i)}(x/v)\right)\right| < \omega_4^{(i)}(x,v);$

(V) $0 \le H_3^{(i)}(x/v) - H_1^{(i)}(x/v) < \omega_3(x,v);$

(VI) $\left|H_2^{(i)}(x/v) - H_3^{(i)}(x/v)\right| < \omega_4^{(i)}(x,v).$

Inequalities (I) - (VI) can be used in investigating of three types of estimators of survival functions (2). In order to construct these estimators we use approaches from [1, 3]. Let the baseline cumulative hazard functions $\left\{\Lambda_0^{(i)}(x),\ i=1,...,k\right\}$ are continuous, $\mathbb{V}=(V_1,...,V_p)$ is the vector of covariates and $\beta=(\beta_1,...,\beta_p)$ is corresponding vector of regression parameters. Our final goal is to estimation of baseline survival functions $\left\{1-H_0^{(i)}(x)=\exp\left\{-\Lambda_0^{(i)}(x)\right\},\ i=1,...,k\right\}$ based on independent observations of $(p+k+1)$ - variate vector

$$\xi=\left(Z,\mathbb{V},\delta^{(1)},...,\delta^{(k)}\right),$$

where $\delta^{(i)}=I\left(A^{(i)}\right)$ is indicator of the event $A^{(i)}$. Let $S^{(n)}=(\xi_1,...,\xi_n)$ is observed sample of independent replicas of vector ξ, where $\xi_i=\left(Z_i,\mathbb{V}_i,\delta_i^{(1)},...,\delta_i^{(k)}\right)$, $\delta_i^{(j)}=I\left(A_j^{(i)}\right)$. In order to construct estimators of baseline survival functions we use three types of functionals (3) and representations (2). Let $\pi(\nu)=\mathbb{P}(\mathbb{V}\le\nu)=\mathbb{P}(V_1\le\nu_1,...,V_p\le\nu_p)$, $\nu=(\nu_1,...,\nu_p)\in\mathbb{R}^p$, is joint distribution function of vector $\mathbb{V}$ and

$$\pi_n(t)=\frac{1}{n}\sum_{i=1}^{n} I(V_{1i}\le\nu_1,...,V_{pi}\le\nu_p),\quad \nu\in\mathbb{R}^p,$$

is its empirical estimator. For $1-H_0^{(i)}(x)$ we propose the following three types of intermediate estimators:

$$1-H_{mon}^{(i)}(x)=\Psi_{\mathrm{m}\,n}^{(i)}(x;\beta_n), \tag{4}$$

$m=1,2,3;\ i=1,...,k,$

$$\Psi_{1n}^{(i)}(x;\beta)=\exp\left\{-\Lambda_{0n}^{(i)}(x;\beta)\right\},$$

$$\Psi_{1n}^{(i)}(x;\beta)=\prod_{u\le x}\left(1-\left(\Lambda_{0n}^{(i)}(u;\beta)-\Lambda_{0n}^{(i)}(u-;\beta)\right)\right),$$

$$\Psi_{1n}^{(i)}(x;\beta)=\left[\omega_n^{(i)}(x;\beta)\right]^{\mathbb{R}_n^{(i)}(x;\beta)},$$

$$R_n^{(i)}(x;\beta)=\Lambda_{0n}^{(i)}(x;\beta)\left(-\int_{[0;x]}\frac{d\omega_n^{(i)}(u;\beta)}{\omega_n^{(i)}(u;\beta)}\right)^{-1},$$

$$\omega_n^{(i)}(x;\beta)=\frac{1}{n}\sum_{k=1}^{n}\exp\left(\beta^T;\nu\right)\delta_k^{(i)}I(Z_k\ge x),$$

$$\Lambda_{0n}^{(i)}(x;\beta)=\frac{1}{n}\sum_{k=1}^{n}\delta_k^{(i)}\frac{I(Z_k\le x)}{\omega_n^{(i)}(Z_k;\beta)}.$$

In considered model are assumed satisfied following conditions: (C1) Joint distribution of vector $\left(Z,\mathbb{V},\delta^{(1)},...,\delta^{(k)}\right)$ is such that for a numbers $T^{(i)}$, $\Lambda_0^{(i)}\left(T^{(i)}\right)<\infty$, we have

$$\inf_{x\le T^{(i)}}\int_{\mathbb{R}^p}\mathbb{P}\left(Z\ge x,\ A^{(i)}/\mathbb{V}=\nu\right)d\pi(\nu)>0,$$

$i=1,\dots,k;$

(C2) Distribution function $\pi(\nu)$ is absolutely continuous and for any $a\in\mathbb{R}^p$:

$$\sup_{t\in\mathbb{B}(\varepsilon;\beta)}\mathbb{E}\left[a\mathbb{V}^T\mathbb{V}a^T\cdot exp\left(\left(t^T;\mathbb{V}\right)\right)\right]^2\le\mathbb{K}<\infty,$$

where $\mathbb{B}(\varepsilon,\beta)$ - is the ball with diameter 2ε and the center at a point β (the true value). Let $Z_{(n)}=\max(Z_1,\dots,Z_n)$.

Theorem 2. *Under conditions (C1), (C2) for all* $x\in\left[0,T^{(i)}\right]$, $i=1,...,k$ *and each* $m=1,2,3$ *:*

$$(I)\ \sup_{\substack{x\le T^{(i)}\\ x<Z_{(n)}}}\left|H_{mon}^{(i)}(x)-H_0^{(i)}(x)\right|\underset{n\to\infty}{\overset{p}{\to}}0;$$

$$(II)\ n^{1/2}\left(H_{mon}^{(i)}(x)-H_0^{(i)}(x)\right)\underset{n\to\infty}{\overset{\mathbb{D}}{\to}}\chi^{(i)}(x),\ in\ \mathbb{D}[0,T^{(i)}].$$

where $\chi^{(i)}(x)$ *— is centered Gaussion process.*

References

[1] Abdushukurov A.A. *Statistics of incomplete observations.* Tashkent. Univ. Press. (In Russian), 2009.

[2] Abdushukurov A.A., Chuyanov Kh. Cox model under random censoring. *Materials of scientific-applied conference «Statistics and its applications». Tashkent. National University of Uzbekistan*, pages 53–54, 2013.

[3] Abdushukurov A.A., Muradov R.S. Estimating of survival function in cox regression model under random censoring from both sides. *Commun. Statist.: Theory&Methods*, 2014.

Hedge Fund Strategies, Financial Instability, Fat Tails and Leverage Cycles

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

Abstract. *The article discusses hedge funds in connection to the latest financial crisis. Hedge funds will continue to use leverage to enhance their alpha generation capabilities, and while we do not believe that leverage is a requirement, we do think that those funds that can obtain and use it appropriately may have a competitive advantage going forward. Indeed, recent market turbulence highlights the importance of focusing on high quality hedge fund firms that have developed well-structured, robust investment processes to address both investment risk and operational risks that can easily translate into investment risk, all while achieving their investment objective through a variety of market environments.*

Keywords. *Hedge Funds, Performance fees, Fat tails, leverage, speculators, risk, volatility, financial markets, Currency speculations, Bank Run, Global Financial Crisis, eventology, Financial Instability, Nash Equilibrium, Government intervention.*

1 Introduction

In [3], 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 [8] 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], [6], we discussed a connection between the hedge funds and the latest financial crisis and 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.

Recent events in financial markets have underscored the dangerous consequences of the use of excessive credit. At the most basic level the problem is obvious: If a firm buys assets with borrowed money, then under extreme market conditions it may owe more money than it has and default. If this happens on a sufficiently wide scale then it can severely stress creditors and cause them to fail as well [7].

3 Hedge Funds and Leverage

Leverage should be an important consideration in the evaluation of hedge fund managers. While leverage generally amplifies both return and risk, higher leverage does not always indicate higher risk, as it must be understood in the context of other features of an investment strategy.

It is not prudent to use leverage to magnify the returns of a low-return position or strategy, or to increase the size of the hedge fund balance sheet without having a great deal of conviction in the underlying investment ideas.

There are many ways for funds to obtain leverage, each of which offer unique attributes with respect to term, stability, cost, and availability. The amount of leverage used by hedge funds can vary substantially, dictated by the strategy being utilized and several other variables.

© 2014 V. Demotchkine

Oleg Vorobyev (ed.), Proc. XIII FAMEMS'2014, Krasnoyarsk: SFU

Analyzing hedge fund balance sheets can help an investor understand how difficult market circumstances can affect funds differently, with some funds able to weather such environments while others suffer greatly or even go out of business.

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.

In a financial system where balance sheets are continuously marked to market, asset price changes show up immediately as changes in net worth, and elicit responses from financial intermediaries who adjust the size of their balance sheets. There is evidence that marked-to-market leverage is strongly procyclical [1]. Such behavior has aggregate consequences. Changes in dealer repos — the primary margin of adjustment for the aggregate balance sheets of intermediaries — forecast changes in financial market risk as measured by the innovations in the VIX index. Aggregate liquidity can be seen as the rate of change of the aggregate balance sheet of the financial intermediaries.

Procyclical leverage can be seen as a consequence of the active management of balance sheets by financial intermediaries who respond to changes in prices and measured risk. For financial intermediaries, their models of risk and economic capital dictate active management of their overall Value-at-Risk (VaR) through adjustments of their balance sheets.

The use of leverage in the economy is not just an esoteric matter relating to funds: It is unavoidable. It is the mechanism through which most people are able to own homes and corporations do business. Credit (and thus leverage) is built into the fabric of society. The current financial crisis perfectly illustrates the dangers of too much leverage followed by too little leverage [7].

4 Mortgage Backed Securities and Leverage

By taking out a collateralized loan a buyer of stocks or mortgage backed securities can put together a portfolio that is worth a multiple of the cash he has available for their purchase. In 2006 this multiple or «leverage» reached 60 to 1 for AAA rated mortgage securities, and 16 to 1 for what are now called the toxic mortgage securities. The outstanding volume of these leveraged asset purchases reached many trillions of dollars.

Leverage has fluctuated up and down in long cycles over the last 30 years. Conventional credit is for a fixed amount and a fixed maturity, extending over the period the borrower needs the money. In a collateralized loan with margin calls, the debt is guaranteed not by the reputation (or punishment) of the borrower, but by an asset which is confiscated if the loan is not repaid. Typically the loan maturity is very short, say a day, much shorter than the length of time the borrower anticipates needing the money. The contract usually specifies that after the daily interest is paid, as long as the loan to asset value ratio remains below a specified threshold, the debt is rolled over another day (up to some final maturity, when the threshold ratio might be changed). If, however, the collateral asset value falls, the lender makes a margin call and the borrower is expected to repay part of the debt and so roll over a smaller loan to maintain the old loan to value threshold. Quite often the borrower will obtain the cash for this extra downpayment by selling some of the collateral. The nature of the collateralized loan contract thus sometimes turns buyers of the collateral into sellers, even when they might think it is the best time to buy.

While the hedge fund community's access to and use of leverage is evolving, we believe that it continues to represent a valid tool in the hedge fund manager's toolbox and that managers with better access to it may benefit. Furthermore, while the use of leverage has in many cases augmented the ability of certain managers and strategies to generate attractive returns, we do not believe that in most cases access to leverage is a prerequisite to alpha generation within the hedge fund space.

However, high leverage with low basis risk may be less risky than low leverage with high basis risk. Importantly, strategies involving less basis risk will tend to call for more leverage in order to generate returns.

5 Different Hedge Funds Strategies and The Leverage

Long/Short вГУ In this strategy, hedge fund managers can either purchase stocks that they feel are undervalued or sell short stocks they deem to be overvalued. In most cases, the fund will have positive exposure to the

equity markets вҐY for example, having 70% of the funds invested long in stocks and 30% invested in the shorting of stocks. In this example, the net exposure to the equity markets is 40% (70%-30%) and the fund would not be using any leverage (Their gross exposure would be 100%). If the manager, however, increases the long positions in the fund to, say, 80% while still maintaining a 30% short position, the fund would have gross exposure of 110% (80%+30% = 110%), which indicates leverage of 10%.

Market Neutral — In this strategy, a hedge fund manager applies the same basic concepts mentioned in the previous paragraph, but seeks to minimize the exposure to the broad market. This can be done in two ways. If there are equal amounts of investment in both long and short positions, the net exposure of the fund would be zero. For example, if 50% of funds were invested long and 50% were invested short, the net exposure would be 0% and the gross exposure would be 100%.

Market-neutral funds are often compared to 130/30 funds. In these funds, the portfolio manager holds a long position equal to 130% of assets and a short position equal to 30%. These funds would be expected to have a market beta, but potentially more alpha than a typical 100% long-only fund, as they provide more opportunities for a manager to exercise stock selection abilities.

There are many ways to construct and manage portfolios, but in the world of high finance, one active management strategy has caught the eye of many top players, and for good reason. This particular strategy is called «enhanced active equity», but you'll find most people just call it a 130/30 (or sometimes a 120/20). Read on to learn how enhanced active equity can make your portfolio more efficient and profitable.

Enhanced Indexing: A Background to Enhanced Active Equity

Enhanced active equity is similar to a concept you may already be familiar with called enhanced indexing. Enhanced indexing has two goals in mind. The first is to expose your portfolio to an index (or some other given benchmark) to realize market returns. The second goal is to provide exposure to returns derived from active management. This combination of goals can be achieved by exposing the portfolio to an index using a combination of equity and derivative securities. This leaves a portion of the original capital available for managers to use to generate long-only (in other words, buy and hold) active returns.

The combination of equity and derivatives provides full exposure to market returns, while managers can use the remaining capital to generate actively managed returns.

For example, if an investor has a $1,000 portfolio and wants to invest it like an enhanced index fund, the process might go something like this: The investor should ensure that the $1,000 is exposed to market returns. To do this, he or she should buy a combination of the equity that makes up a given index and derivatives that represent the index. The capital used to do this should not equal 100% of the original capital because of the leverage provided by the derivatives. It is important to note that the position still exposes the investor to returns equal to 100% of the original capital. For this example, let's suppose that the investor used $800. The remaining $200 would then be invested in securities that the investor feels will outperform the index.

Enhanced indexing is still subject to a few constraints, though, such as the long-only constraint (which limits investors to long positions) and a restricted amount of residual risk to which the overall portfolio may be exposed. The amount of capital exposed to residual risk is constrained by the amount of capital not used to track the index. These constraints suggest that enhanced indexing may be a suboptimal approach.

6 Leverage and Futures Contracts

When you open a futures contract, the futures exchange will state a minimum amount of money that you must deposit into your account. This original deposit of money is called the initial margin. When your contract is liquidated, you will be refunded the initial margin plus or minus any gains or losses that occur over the span of the futures contract. In other words, the amount in your margin account changes daily as the market fluctuates in relation to your futures contract. The minimum-level margin is determined by the futures exchange and is usually 5% to 10% of the futures contract. These predetermined initial margin amounts are continuously under review: at times of high market volatility, initial margin requirements can be raised.

7 Leverage: The Double-Edged Sword

In the futures market, leverage refers to having control over large cash amounts of commodities with comparatively small levels of capital. In other words, with a relatively small amount of cash, you can enter into a futures contract that is worth much more than you initially have to pay (deposit into your margin account). It is said that in the futures market, more than any other form of investment, price changes are highly leveraged, meaning a small change in a futures price can translate into a huge gain or loss.

Futures positions are highly leveraged because the initial margins that are set by the exchanges are relatively small compared to the cash value of the contracts in question (which is part of the reason why the futures market is useful but also very risky). The smaller the margin in relation to the cash value of the futures

contract, the higher the leverage. So for an initial margin of $5,000, you may be able to enter into a long position in a futures contract for 30,000 pounds of coffee valued at $50,000, which would be considered highly leveraged investments.

Hedge fund leverage has soared over the past 12 months, according to new data from the Financial Conduct Authority.

According to an anonymous poll of London-based hedge funds, leverage has risen from 54 times in March 2013 to 64 times in March 2014 the size of net fund asset values.

The study also showed that the UK's 20 largest hedge funds controlled 82 per cent of the total assets under management. Global hedge fund assets rose 17 percent in 2013 to a record $2.63 trillion.

8 Conclusion

For many of fund managers, leverage will remain an important tool for achieving their return goals. Now more than ever, investors must take a prudent and thoughtful approach to assessing not just leverage, but a variety of risks imbedded in hedge funds. We still think, however, that leverage is useful tool in skilful hands.

Acknowledgements

You can acknowledge help and support you received in an unnumbered section at the end of the main text, just before the references.

This section can also be reserved for additional information on the author; for example, that the paper was written while the first author worked at night at the University of Great Knowledge.

References

[1] Adrian, Tobias, and Hyun S. Shin. *Liquidity and Leverage*. Tech. Rep. 328 Federal Reserve Bank of New York., 2008.

[2] Val Demotchkine. A comparison of event driven hedge fund strategies. In *XIII International EM Conference on Eventological Mathematics and Related Fields, Krasnoyarsk, Siberia, Russia*, 2009.

[3] Val Demotchkine. A comparison of hedge fund strategies. In *8th International FAM Conference on Financial and Actuarial Mathematics and Related Fields, Krasnoyarsk, Russia*, 2009.

[4] Val Demotchkine. A comparison of hedge fund strategies. part 3. In *XV International EM Conference on Eventological Mathematics and Related Fields, Krasnoyarsk, Siberia, Russia*, 2011.

[5] Val Demotchkine. Hedge fund strategies, financial instability, and global financial crisis. In *11th International Conference on FAM and Eventology of Safety, Krasnoyarsk, Siberia, Russia*, 2012.

[6] Val Demotchkine. Hedge fund strategies, financial instability, and stock prices stabilization. In *12th International Conference on FAM and Eventology of Safety, Krasnoyarsk, Siberia, Russia*, 2013.

[7] Stefan Thurner, J. Doyne Farmer and John Geanakoplos. *Leverage Causes Fat Tails And Clustered Volatility*. 2010.

[8] Vorobyev, Oleg Yu., Goldblatt, Joe Jeff and Finkel, Rebecca. Eventological theory of decision making. *Journal of Iberian Federal University. Mathematics & Physics*, 1:3–16, 2009.

On estimation of main parameter in the long-surviving Markov branching process

Azam A. Imomov
Karshi State University
17, Kuchabag street, 180100 Karshi city, Uzbekistan
imomov_azam@mail.ru

Dilnoza S. Almanova
Karshi State University
17, Kuchabag street, 180100 Karshi city, Uzbekistan
imomov_azam@mail.ru

Abstract. *In this work a problem of estimation of key parameter of Markov Q-process is considered. One estimator function is introduced. Asymptotic properties of its dispersion is investigated.*

Keywords. *Markov Q-processes, Markov Branching processes, main parameter estimation.*

1 Introduction

We consider a Markov branching process, its trajectory will run on in remote future. Let $Z(t),\ t \geq 0$ represents the population size to the moment t in homogeneous continuous time Markov Branching process (MBP) with transition probabilities

$$P_{ij}(t) := \mathbf{P}\left\{Z(t+\tau) = j \,|Z(\tau) = i\right\}, \quad \tau \geq 0,$$

where $i, j \in \mathbf{N}_0 = \{0\} \cup \mathbf{N}$. These probabilities are equal to i-fold convolution of distribution $P_{1j}(t)$, that is

$$P_{ij}(t) = \sum_{j_1+\cdots+j_i=j} P_{1j_1}(t) \cdot \ldots \cdot P_{1j_i}(t).$$

Probabilities $P_{1j}(t)$, in turn, are calculated using of "local densities" $\{a_j,\ j \in \mathbf{N}_0\}$ via the relation

$$P_{1j}(\varepsilon) = \delta_{1j} + a_j\varepsilon + o(\varepsilon), \quad \varepsilon \to 0, \tag{1}$$

where δ_{1j} is the Kronecker delta, the densities $a_j \geq 0$ for $j \in \mathbf{N}_0 \backslash \{1\}$ and $0 < a_0 < -a_1$, moreover $\sum_{j\in\mathbf{N}_0} a_j = 0$; see Sevastyanov [3, pp. 11-12, 26].

Imomov [1] investigated some asymptotic properties of so-called Markov Q-process (MQP) $\{W(t),\ t \geq 0\}$, defined as "long surviving" MBP, whose transition probabilities are

$$\begin{aligned} Q_{ij}(t) &:= \mathbf{P}\left\{W(t+\tau) = j \mid W(\tau) = i\right\} \\ &= \frac{jq^{j-i}}{i\beta^t} P_{ij}(t). \end{aligned} \tag{2}$$

In (2) $\beta := \exp\left\{\sum_{j\in\mathbf{N}} ja_jq^{j-1}\right\}$, and q is extinction probability of MBP. Using (1) and (2) probabilities $Q_{1j}(\varepsilon)$ may be represented as

$$Q_{1j}(\varepsilon) = \delta_{1j} + p_j\varepsilon + o(\varepsilon), \quad \varepsilon \to 0,$$

with transition probability densities $p_1 = a_1 - \ln\beta < 0$ and $p_j = jq^{j-1}a_j \geq 0$ для $j \in \mathbf{N_0}\backslash\{1\}$. Therefore, infinitesimal generating function (GF)

$$g(x) := \sum_{j\in\mathbf{N}} p_j x^j = x\left[f'(qx) - f'(q)\right]$$

completely defines the considering MQP, where $f(x)$ is infinitesimal GF that define MBP $Z(t)$, that is

$$f(x) = \sum_{j\in\mathbf{N}_0} a_j x^j.$$

In this designation $\beta := \exp\{f'(q)\}$ and $f(q) = 0$; see Imomov [1].

It is known that in the theory of MBP the quantity $f'(1)$ plays a role of regulating parameter of its evolution. Depending on its value three types of MBP is allocated: the MBP is called subcritical, critical and supercritical if $f'(1) < 0$, $f'(1) = 0$ and $f'(1) > 0$ respectively. We note that evolution of MQП is regulated essentially by positive parameter β and it is known that $\beta = 1$ as $f'(1) = 0$ and $\beta < 1$ if $f'(1) \neq 0$; see Imomov [1] and literature therein. Thus, we divide two types of MQП in dependence of values of parameter β. Told visually affirms in the following local limit theorem.

So, let $\{W(t),\ t \geq 0\}$ is the MQП generated by GF $g(x)$ with the known first moment $b := g'(1)$.

Theorem A [1]. *Let $b < \infty$.*

- *If $\beta = 1$, then*

$$t^2Q_{11}(t) \to \frac{2}{bf(0)}, \quad t \to \infty. \tag{3}$$

- *If $\beta < 1$, then*

$$Q_{11}(t) \to \frac{|\ln\beta|\, A}{f(0)}, \quad t \to \infty, \tag{4}$$

where constant

$$A = q\exp\left\{\int_0^q \left[\frac{1}{s-q} - \frac{f'(q)}{f(s)}\right] ds\right\}.$$

It is easy to be convinced that at $q = 1$ the positive number A represents the Sevastyanov's constant from the theory of subcritical MBP; see Sevastyanov [3, c.54].

© 2014 A.A.Imomov, and D.S.Almanova

Oleg Vorobyev (ed.), Proc. XIII FAMEMS'2014, Krasnoyarsk: SFU

Now put into consideration the probability GF of distribution of states of MQP

$$\begin{aligned} G_i(t;x) &= \mathbf{E}_i x^{W(t)} := \mathbf{E}\left[x^{W(t)}\middle| W(0)=i\right] \\ &= \sum_{j\in\mathbf{N}} Q_{ij}(t)x^j. \end{aligned}$$

As was shown by Imomov [1],

$$G_i(t;x) = \left[\frac{\Phi(t;qx)}{q}\right]^{i-1} G(t;x), \tag{5}$$

where

$$G(t;x) = x\exp\left\{\int_0^t b\left(\frac{\Phi(\tau;qx)}{q}\right)d\tau\right\},$$

$$\Phi(t;x) = \sum_{j\in\mathbf{N}_0} P_{1j}(t)x^j.$$

and

$$b(x) = \frac{g(x)}{x} = f'(qx) - f'(q).$$

By differentiation in a point of $x=1$ from (5) will be received

$$\mathbf{E}_i W(t) = (i-1)\,\beta^t + \mathbf{E}_1 W(t), \tag{6}$$

where

$$\mathbf{E}_1 W(t) = \begin{cases} bt+1 & , \ \beta = 1, \\ 1+\gamma(1-\beta^t)\,, & \beta < 1. \end{cases}$$

Repeatedly differentiation yields of the dispersion $\mathbf{D}_i W(t) = \mathbf{D}\left[W(t)\,|\,W(0)=i\right]$ in form of

$$\mathbf{D}_i W(t) = \begin{cases} bti & , \ \beta = 1, \\ \left[\gamma+\delta(i)\beta^t\right](1-\beta^t) & , \ \beta < 1, \end{cases} \tag{7}$$

where $\delta(i) = (i-1)(1+\gamma)$ и $\gamma = b/|\ln\beta|$.

2 Main results

In view of all above-stated, we purpose to estimate parameter β by observing values of $W(t)$. From (5) it is obvious that

$$\mathbf{E}\left[x^{W(t+1)}\middle| W(t)\right] = \left[\frac{\Phi(1;qx)}{q}\right]^{W(t)-1} G(1;x),$$

where $G(t;x)$ is define in (5). From here we will receive

$$\mathbf{E}\left[W(t+1)\,|W(t)\right] = [W(t)-1]\cdot\beta + \mathbf{E}_1 W(1).$$

This equality we will use to represent $W(t)$ by stochastic equation

$$W(t+1) = [W(t)-1]\cdot\beta + W(1) + \varepsilon(t),$$

with the error term $\varepsilon(t)$ has zero mean: $\mathbf{E}\varepsilon(t) = 0$. Considering this equation, we put forward the following estimator for β, when $\mathbf{E}_1 W(1)$ is known:

$$\widehat{\beta}(t) = \frac{W(t+1) - \mathbf{E}_1 W(1)}{W(t)-1}, \quad t > 1.$$

According to the formula of full probability and considering homogeneity property of MQП with account of (6) we have

$$\begin{aligned} \mathbf{E}\widehat{\beta}(t) &= \sum_j \left[Q_{1j}(t)\cdot\left(\frac{\mathbf{E}_j W(1) - \mathbf{E}_1 W(1)}{j-1}\right)\right] \\ &= \beta\cdot\sum_j Q_{1j}(t) = \beta. \end{aligned}$$

Following theorems characterize properties of dispersions of estimator $\widehat{\beta}(t)$.

Theorem 1. *Let $b<\infty$. If $\beta = 1$, then*

$$\frac{t}{2}\cdot\mathbf{D}\widehat{\beta}(t) = 1 + O\left(\frac{\ln^2 t}{t}\right), \quad t\to\infty.$$

Theorem 2. *Let $b<\infty$. If $\beta < 1$, then*

$$\mathbf{D}\widehat{\beta}(t) = O(1), \quad t\to\infty.$$

3 Ideas of proofs of Theorems

Using (7) and homogeneity property of MQП we receive

$$\mathbf{D}\widehat{\beta}(t) = T_1\cdot\sum_{k\in\mathbf{N}}\frac{Q_{1k+1}(t)}{k} + T_2\cdot\sum_{k\in\mathbf{N}}\frac{Q_{1k+1}(t)}{k^2},$$

where constants T_1 and T_2 depend on $\mathbf{D}_k W(1)$. On the other hand we obtain

$$\sum_{k\in\mathbf{N}}\frac{Q_{1k+1}(t)}{k} = \int_0^1 \frac{G(t;x) - Q_{11}(t)x}{x^2}dx \tag{8}$$

and

$$\sum_{k\in\mathbf{N}}\frac{Q_{1k+1}(t)}{k^2} = \int_0^1 \frac{v(t;x)}{x}dx, \tag{9}$$

where

$$v(t;x) = \int_0^x \frac{G(t;s) - Q_{11}(t)s}{s^2}ds.$$

So we should estimate integrals in (8) and (9). In further discussion we will essentially use results of Imomov [2] in combination with assertions (3) and (4).

References

[1] Imomov A.A. On markov continuous time analogue of q-processes. *Journal Theory of Probability and Mathematical Statistics*, 84:57–64, 2012.

[2] Imomov A.A. A differential analog of the main lemma of the theory of markov branching processes and its applications. *Ukrainian Mathematical Journal*, 2005:307–315, 57(2).

[3] Sevastyanov B.A. *Branching processes.* Moscow, Nauka, 1971.

On estimation of main parameter in Q-process

Azam A. Imomov
Karshi State University
17, Kuchabag street, 180100 Karshi city, Uzbekistan
imomov_azam@mail.ru

Khurshid Q. Jumakulov
Kokand Pedagogical Institute
28, Turon street, 150000 Kokand city, Uzbekistan
xurshid81@gmail.com

Erkin E. Tukhtaev
Karshi State University
17, Kuchabag street, 180100 Karshi city, Uzbekistan
tuxtayev@umail.uz

Abstract. *We consider a problem of estimation of key parameter in Q-process. One estimator function is introduced. Asymptotic properties of its dispersion is investigated.*

Keywords. *Q-processes, Branching processes, main parameter estimation.*

1 Preliminaries and results

Let $Z_0 = 1$ and the sequence of random variables $\{Z_n,\ n \in \mathbf{N}_0\}$ represents a population sizes in Galton-Watson branching Process (GWP), where $\mathbf{N}_0 = \{0\} \cup \{\mathbf{N} = 1, 2, \ldots\}$. Setting a reproduction law of the single particle $p_k := \mathbf{P}\{Z_1 = k \in \mathbf{N}_0\}$ we completely define a distribution of Z_n for all $n \in \mathbf{N}$. In fact probability generating function (GF) $F_n(x) := \mathbf{E}x^{Z_n}$ is n-th functional iteration of GF $F(x) = \sum_{k\in\mathbf{N}_0} p_k x^k$; see [3]. Throughout the paper we will assume that $p_0 \neq 0$ и $p_0 + p_1 \neq 1$.

Consider the sequence $\{W_n,\ n \in \mathbf{N}_0\}$ described by family of GFs

$$\left\{ Y_n(x) := \mathbf{E}x^{W_n} = x\frac{F_n'(qx)}{\beta^n},\quad n \in \mathbf{N}_0 \right\}, \qquad (1)$$

where $\beta = F'(q)$ and quantity $q \in (0; 1]$ is the extinction probability of the GWP, which is smallest non-negative root of the equation $x = F(x)$. The above-defined sequence means states of the Q-process and forms a homogeneous Markov chain with initial state $W_0 \stackrel{d}{=} Z_0$ on set of non-negative integers. Classification of these states depends on the value of parameter $A := F'(1)$ - the per capita mean number of offspring in GWP. This chain is transient if $A = 1$ and, recurrent if $A \neq 1$. And its transition probabilities $\mathbf{P}\{W_{n+k} = j | W_k = i\} = Q_{ij}^{(n)}$ are form of

$$Q_{ij}^{(n)} = \frac{jq^{j-i}}{i\beta^n}\mathbf{P}\{Z_{n+k} = j | Z_k = i\},\ \ i, j \in \mathbf{N}, \qquad (2)$$

for any $k \in \mathbf{N}_0$; see [3, pp.58-59]. Considering (1) and (2) it is easily to see that GF

$$\begin{aligned} Y_n^{(i)}(x) &= \mathbf{E}_i x^{W_n} := \mathbf{E}\left[x^{W_n} \middle| W_0 = i\right] \\ &= \sum_{j\in\mathbf{N}} Q_{ij}^{(n)} x^j \end{aligned}$$

be represented as

$$Y_n^{(i)}(x) = \left[\frac{F_n(qx)}{q}\right]^{i-1} Y_n(x), \qquad (3)$$

where GF $Y_n(x)$ is introduced in (1). As GF $F_n(x)$ obey the iteration law, for definition of $Y_n^{(i)}(x)$ it is enough to introduce

$$Y(x) := Y_1(x) = x\frac{F'(qx)}{\beta}.$$

We know that the number A presents as regulating parameter in evolution of GWP and divides three types of GWP: it is called as subcritical, critical and supercritical, if $A < 1$, $A = 1$ and $A > 1$, respectively. Investigation shows that evolution of Q-process in essence regulates by positive parameter β. It is known $\beta < 1$ at $A \neq 1$ and $\beta = 1$ if $A = 1$; see [2]. Thus we will subdivide two types of Q-processes. The told visually affirms by the subsequent remarks. Further we suppose $Y'(1) < \infty$. Then by differentiation from (3) will easily be found

$$\mathbf{E}_i W_n = (i-1)\beta^n + \mathbf{E}_1 W_n, \qquad (4)$$

where

$$\mathbf{E}_1 W_n = \begin{cases} F''(1)n + 1 & ,\ \ \beta = 1, \\ 1 + \gamma(1 - \beta^n), & \beta < 1. \end{cases}$$

and $\gamma = qF''(q)/\beta(1-\beta)$.

In view of above facts we purpose to estimate the parameter β based on the observations W_n. From (3) we see

$$\mathbf{E}\left[x^{W_{n+1}} | W_n\right] = \left[\frac{F(qx)}{q}\right]^{W_n - 1} Y(x),$$

and therefore

$$\mathbf{E}\left[W_{n+1} | W_n\right] = [W_n - 1]\cdot\beta + \mathbf{E}W_1.$$

The last relation allows us to write a following stochastic equation:

$$W_{n+1} = [W_n - 1]\cdot\beta + \mathbf{E}W_1 + \varepsilon_n,$$

with an error term ε_n, has zero mean: $\mathbf{E}\varepsilon_n = 0$. We use this equation to establish a following estimator function for β when $\mathbf{E}W_1$ is known:

$$\widehat{\beta}_n = \frac{W_{n+1} - \mathbf{E}W_1}{W_n - 1},\quad n > 1.$$

© 2014 A.A.Imomov, Kh.Q.Jumakulov and E.E.Tukhtaev
Oleg Vorobyev (ed.), Proc. XIII FAMEMS'2014, Krasnoyarsk: SFU

Considering the homogeneity property of Q-process and (4) we see

$$\begin{aligned}\mathbf{E}\widehat{\beta}_n &= \sum_k \left[Q_{1k}^{(n)} \cdot \left(\frac{\mathbf{E}_k W_1 - \mathbf{E}_1 W_1}{k-1} \right) \right] \\ &= \beta \cdot \sum_k Q_{1k}^{(n)} = \beta.\end{aligned}$$

Following theorems hold.

Theorem 1. *Let $Y'(1) < \infty$. If $\beta < 1$, then*

$$\mathbf{D}\widehat{\beta}_n = O(1), \quad n \to \infty.$$

Theorem 2. *Let $Y'(1) < \infty$. If $\beta = 1$, then*

$$\frac{n}{2} \cdot \mathbf{D}\widehat{\beta}_n = 1 + O\left(\frac{\ln^2 n}{n} \right), \quad n \to \infty.$$

2 Ideas of proofs of Theorems

By formula of full probability and homogeneity property of Q-process we have

$$\mathbf{D}\widehat{\beta}_n = \sum_{k \in \mathbf{N}} \frac{Q_{1k+1}^{(n)}}{k^2} \mathbf{D}_{k+1} W_1,$$

where $\mathbf{D}_i W_n = \mathbf{D}[W_n \,|\, W_0 = i]$. Almost obvious that under condition of $\mathbf{P}\{W_0 = k\} = 1$ the following relation holds:

$$W_1 = 1 + \zeta_1 + \zeta_2 + \cdots + \zeta_{k-1} + \eta,$$

where ζ_k are independent variables with common GF $\mathbf{E}x^{\zeta_1} = F(qx)/q$ and $\mathbf{E}x^{\eta} = F'(qx)/\beta$. Using last equality we easy will calculate $\mathbf{D}_k W_n$ and after simple transformations we will write down

$$\mathbf{D}\widehat{\beta}_n = T_1 \cdot \sum_{k \in \mathbf{N}} \frac{Q_{1k+1}^{(n)}}{k} + T_2 \cdot \sum_{k \in \mathbf{N}} \frac{Q_{1k+1}^{(n)}}{k^2}, \qquad (5)$$

where constants T_1 and T_2 depend on $\mathbf{D}_k W_1$. On the other hand we obtain

$$\sum_{k \in \mathbf{N}} \frac{Q_{1k+1}^{(n)}}{k} = \int_0^1 \frac{Y_n(x) - Q_{11}^{(n)} x}{x^2} dx \qquad (6)$$

and

$$\sum_{k \in \mathbf{N}} \frac{Q_{1k+1}^{(n)}}{k^2} = \int_0^1 \frac{v_n(x)}{x} dx, \qquad (7)$$

where

$$v_n(x) = \int_0^x \frac{Y_n(s) - Q_{11}^{(n)} s}{s^2} ds.$$

So integrals in (6) and (7) are subject to estimate.

Using (1) and (2) we receive

$$h_n(x) := \frac{1}{\beta^n} \frac{f_n'(x) - P_{11}^{(n)}}{x},$$

where $f_n(x) = \sum_{k \in \mathbf{N_0}} P_{1k}^{(n)} x^k$ is n-fold functional iteration of GF

$$f(x) = \frac{F(qx)}{q}.$$

Now we will find an asymptotic representation for $h_n(x)$ considering the discrete time analogue results for the asymptotic relation of the function $f_n'(x)$; see [5] and [6]. We use it in (6) and (7). In the further reasoning we follow the methods from [4] and [2]. The received estimations we put together in (5).

References

[1] Бадалбаев И.С. и Мухитдинов А.А. *Статистические задачи многотипных ветвящихся процессов.* Ташкент, Фан, 1990.

[2] Форманов Ш.К. и Имомов А.А. Об асимптотических свойствах q-процессов. *Узбекский математический журнал*, 3:175–183, 2011.

[3] Athreya K.B. and Ney P.E. *Branching processes.* Springer, New York, 1972.

[4] Нагаев А.В. Об оценке среднего числа непосредственных потомков частицы в ветвящемся случайном процессе. *Теория вероятностей и ее применения*, 67, 2:363–369, 1967.

[5] Imomov A.A. On markov continuous time analogue of q-processes. *Journal Theory of Probability and Mathematical Statistics*, 84:57–64, 2012.

[6] Imomov A.A. A differential analog of the main lemma of the theory of markov branching processes and its applications. *Ukrainian Mathematical Journal*, 2005:307–315, 57(2).

Асимптотическая гауссовость копула оценок функции выживания при зависимом случайном цензурировании

Абдушукуров Абдурахим Ахмедович
Национальный Университет Узбекистана имени Мирзо Улугбека
г. Ташкент, Узбекистан
a_abdushukurov@rambler.ru

Мурадов Рустамжон Собитхонович
Национальный Университет Узбекистана имени Мирзо Улугбека
г. Ташкент, Узбекистан
r_muradov1985@rambler.ru,

Аннотация. *В данной работе рассматривается задача оценивания функции выживания при случайном зависимом цензурировании наблюдений справа. Для оценок доказаны свойства асимптотической гауссовости.*

Ключевые слова. *Случайное цензурирование, функция выживания, Архимедовы копулы.*

1 Введение

В медико-биологических, социологических исследованиях, в инженерии, в страховом деле, в финансовой математике и других областях исследований практического характера нас часто интересуют совместные свойства двух и более возможно зависимых случайных величин (с.в.). Так в медицинских исследованиях это могут быть совместные времена жизни близких родственников, в инженерии - совместные времена безотказной работы нескольких технических объектов, в страховом деле размеры групповых страховых выплат страховой компанией клиентам в связи с осуществлением какого-либо страхового случая и т.д.. В этих случаях интерес представляют совместные распределения нескольких неотрицательных зависимых с.в.. Для того, чтобы делать какие-то статистические выводы о совместном распределении этих с.в. необходимо проводить наблюдения над ними. Однако, на практике такие зависимые наблюдения к тому же могут быть и неполными. Неполнота данных может быть результатом цензурирования одной совокупности с.в. другой возможно зависимой от первоначальной совокупностью с.в.. Например, на самом простом случае, когда одна с.в. подвергается случайному цензурированию, скажем справа другой зависимой от нее с.в., получаем повторную двумерную выборку, по которой требуются оценить одномерное распределение или его функционалы. При построении статистических оценок для функционалов от распределений по полным и неполным многомерным зависимым наблюдениям более популярными и с практической точки зрения удобными становятся методы, основанные на так называемые копула функции, которых также называют и функциями зависимости. В основе копула методологии находится идея отделения маргинальных распределений из структуры зависимости, связывая их с совместным распределением через копула функцию.

© 2014 А.А. Абдушукуров, Р.С.Мурадов

Олег Воробьев (ред.), Труды XIII ФАМЭМС'2014, Красноярск: СФУ

2 Оценивание с использованием копул

На вероятностном пространстве $(\Omega, \mathcal{A}, P)$ рассмотрим $\{(X_k, Y_k),\ k \geq 1\}$ — последовательность независимых и одинаково распределенных (н.о.р.) пар неотрицательных с.в. с общей совместной функцией распределения (ф.р.) $H(x,y) = P(X_1 \leq x,\, Y \leq y),\ (x,y) \in \overline{R}^{+\,2}$. Предположим, что маргинальные ф.р. $F(x) = P(X_1 \leq x) = H(x, +\infty)$ и $G(y) = P(Y_1 \leq y) = H(+\infty, y),\ x, y \in \overline{R}^{+}$, являются непрерывными и $F(0) = G(0) = 0$. Пусть последовательность $\{X_k,\ k \geq 1\}$ цензурируется справа последовательностью $\{Y_k,\ k \geq 1\}$ и в n-м шаге эксперимента наблюдению доступна повторная выборка $\mathbb{V}^{(n)} = \{(Z_k, \delta_k),\ 1 \leq k \leq n\}$, где $Z_k = \min(X_k, Y_k)$ и $\delta_k = I(Z_k = X_k)$, т.е. интересующие нас с.в. X_k наблюдаемы лишь в случае $\delta_k = 1$. Отметим, что здесь не требуется независимость последовательностей $\{X_k\}$ и $\{Y_k\}$. Задача состоит в оценивании функционалов от ф.р. F по выборке $\mathbb{V}^{(n)}$ при мещающей ф.р. G. Одним из таких основных функционалов является функция выживания $S^X(x) = P(X_1 > x) = 1 - F(x),\quad x \in \overline{R}^{+}$, означающая, что отказ (смерть) испытываемого на выживаемость объекта произойдет в промежутке времени $(x, +\infty)$.

Рассмотрим оценку F_n из работ [3, 1, 2, 4] для ф.р. F. Пусть $C(u,v)$ — копула функция(см. [6]), соответствующая паре (X_1, Y_1) и $C^*(u,v) = u + v - 1 + C(1-u,\ 1-v),\ (u,v) \in [0,1]^2$, отвечающая ей копула выживания является архимедовой: $C^*(u,v) = \varphi^{-1}[\varphi(u) + \varphi(v)],\ (u,v) \in [0,1]^2$. Здесь φ - функция-генератор копулы $\varphi : [0,1] \to \overline{R}^{+}$ — непрерывная, строго убывающая функция такая, что $\varphi(1) = 0,\ \varphi(0) = \infty$ и φ^{-1} - обратная для φ функция. Определим оценку F_n формулой $F_n(x) = 1 - S_n^X(x)$, где

$$S_n^X(x) = \varphi^{-1}\ \left[\varphi\left(S_n^Z(x)\right)\mu_n(x)\right],\ x \in \bar{R}^{+}, \qquad (1)$$

при этом в формулу (1) входят статистики:

$$\mu_n(x) = \frac{\varphi\left(\tilde{S}_n^X(x)\right)}{\varphi\left(\tilde{S}_n^Z(x)\right)},\ x \in \bar{R}^+,$$

$$S_n^Z(x) = \frac{1}{n}\sum_{k=1}^{n} I(Z_k > x) = 1 - H_n(x),$$

$$\varphi(\tilde{S}_n^X(x)) = -\int_0^x I(S_n^Z(t-) > 0) S_n^Z(t-)\varphi'(S_n^Z(t))d\Lambda_n(t),$$

$$\varphi(\tilde{S}_n^Z(x)) = -\int_0^x I(S_n^Z(t-) > 0) S_n^Z(t-)\varphi'(S_n^Z(t))d\Lambda_n^Z(t),$$

$$\Lambda_n^Z(x) = \int_0^x \frac{dH_n(t)}{S_n^Z(t-)}, \Lambda_n(x) = \int_0^x \frac{dH_{1n}(t)}{S_n^Z(t-)},$$

$$H_{1n}(t) = \frac{1}{n}\sum_{k=1}^{n} I(Z_k \le t,\ \delta_k = 1).$$

Отметим, что предложенные ранее оценки для S^X Зенга-Клайна [8] и ее приближение Ривеста-Уэллса [7] используют часть наблюдений из выборки $\mathbb{V}^{(n)}$, налагая при этом и условие на схему цензуривания.

3 Основной результат

Для формулировки результатов нам необходимы некоторые обозначения и условия. Пусть

$$\Lambda(x) = \int_0^x \lim_{\Delta\downarrow 0}\frac{1}{\Delta}P(u < X_1 \le u+\Delta / X_1 > u, Y_1 > u)du,$$

$$\Lambda^Z(x) = \int_0^x \lim_{\Delta\downarrow 0}\frac{1}{\Delta}P(u < Z_1 \le u+\Delta / X_1 > u, Y_1 > u)du,$$

— интегральные функции интенсивности и $Z^{(n)} = \sup\{x \ge 0 : \mathbb{J}_n(x) > 0\}$, $T_X = \sup\{x \ge 0 :\ S^X(x) > 0\}$, $T_Y = \sup\{x \ge 0 : S^Y(x) > 0\}$, $T_Z = \sup\{x \ge 0 : S^Z(x) > 0\}$, $\Psi(x) = -x\varphi'(x)$, где $\mathbb{J}_n(x) = nS_n^Z(x-)$, $S^Y(x) = P(Y_1 > x)$, $S^Z(x) = P(Z_1 > x)$. Введем следующие условия регулярности относительно S^X, S^Z и генератора копулы φ где под Λ^* подразумеваются одновременно и Λ и Λ^Z:

(C1) Функция $\varphi(\cdot)$ строго убывает на (0,1] и является достаточно гладкой в следующем смысле: первые две производные функций $\varphi(x)$ и $\Psi(x)$ ограничены для $x \in [\varepsilon,\ 1]$, где $\varepsilon > 0$ произвольное число. Более того, первое производное φ' ограничено и отделено от нуля на $[0,1]$;

(C2) $0 < \int_0^{T_Z}\left[\Psi\left(S^Z(x)\right)\right]^2 d\Lambda^*(x) < \infty$;

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

(C4) $\limsup_{x\to T_Z}\int_x^{T_Z}\frac{\Psi(S^Z(t))}{S^Z(t)}d\Lambda^*(t) = 0$;

(C5) $S^X(\cdot)$ – непреревна на $[0,\ T_Z]$ если $T_Z < \infty$, в противном случае $S^X(\infty) = \lim_{x\to\infty} S^X(x)$.

Отметим, что условия (C1)-(C5) справедливы, например, для копула-генераторов Клейтона-Френка.

В работах [1, 2, 4] доказано следующее утверждение о состоятельности оценок (1) на отрезке $[0,\ T]$, где $T = T_Z$, если $T_Z < \infty$ и $T = Z^{(n)}$, если $T_Z = \infty$. На самом деле, эти результаты верны и на всей полупрямой $[0,\infty)$ и при $T_Z = \infty$, т.к. $S_n^X(x) = 0$ при $x > Z^{(n)}$, $Z^{(n)} \xrightarrow{p} \infty$ и $S^X\left(Z^{(n)}\right) \xrightarrow{p} S^X(\infty) = 0$ при $n \to \infty$.

Теорема 1. *Пусть выполнены условия (C1)-(C3). Тогда при* $n \to \infty$

$$\sup_{0\le x\le T}\left|S_n^X(x) - S^X(x)\right| \xrightarrow{P} 0.$$

В данной работе докажем свойство слабой сходимости оценок (1) к гауссовскому процессу. Рассмотрим "остановленный"процесс

$$\mathbb{D}_n(x) = n^{1/2}(\varphi(S_n^X(x\wedge Z^{(n)})) - \varphi(S^X(x\wedge Z^{(n)}))),$$

где $a \wedge b = \min(a,b)$. Рассмотрим ковариационную функцию при $x_1,\ x_2 \in [0,\ T_Z)$ предельного гауссовского процесса $\mathbb{D}(x)$ для $\mathbb{D}_n(x)$:

$$Z(x_1,x_2) = Cov(\mathbb{D}(x_1), \mathbb{D}(x_2)) =$$

$$= \int_0^{x_1\wedge x_2} S^Z(t)[\varphi'(S^Z(t))]^2 d\Lambda(t) +$$

$$+2\int_0^{x_1\wedge x_2}\int_0^t S^Z(t)(1-S^Z(s))\cdot$$

$$\cdot\Psi'(S^Z(t))\Psi'(S^Z(s))d\Lambda(s)d\Lambda(t) +$$

$$+2\int_0^{x_1\wedge x_2}\int_0^t \varphi'(S^Z(s))S^Z(t)\Psi'(S^Z(t))d\Lambda(s)d\Lambda(t) +$$

$$+\int_{x_1\wedge x_2}^{x_1\vee x_2} S^Z(t)\Psi'(S^Z(t))d\Lambda(t)\cdot$$

$$\cdot\int_0^{x_1\wedge x_2}\left[\left(1-S^Z(s)\right)\Psi'\left(S^Z(s)\right) + \varphi'(S^Z(s))\right]d\Lambda(s), \tag{2}$$

где $x_1 \vee x_2 = \max(x_1,x_2)$ и функцию дисперсии $Z(x,x) = \sigma^2(x)$. Пусть

$$\sigma_0^2 = \lim_{x\to T_Z}\sigma^2(x) < \infty, \tag{3}$$

и для каждого $x \in [0,\ T_Z)$

$$C(x) = \lim_{t \to T_Z} Z(t,x) < \infty\ . \tag{4}$$

Определим с.в. $\xi_0 \stackrel{D}{=} N\left(0,\ \sigma_0^2\right)$ и рассмотрим также процессы

$$\mathbb{D}(x) = I[0,\ T_Z)\, e(x) + I\{T_Z\} \cdot \xi_0,$$

$$q_n(x) = n^{1/2}\left(S_n^X\left(x \wedge Z^{(n)}\right) - S^X\left(x \wedge Z^{(n)}\right)\right),$$

$$q(x) = e(x)\left[\varphi'\left(S^X(x)\right)\right]^{-1} + \xi_0\left[\varphi'\left(S^X(T_Z)\right)\right]^{-1},$$

где $e(x)$ –гауссовский процесс с нулевым средним и ковариацией (2) и $\operatorname{cov}(e(x),\ \xi_0) = C(x)$.

Сформулируем основное утверждение.

Теорема 2. *Пусть выполнены условия (C1)-(C5), (3) и (4). Тогда при* $n \to \infty$

$$q_n(x) \stackrel{D}{\Rightarrow} q(x) \quad в\ \ D[0,T_Z], \tag{5}$$

где $D[0,T_Z]$ *- пространство Скорохода функций на множестве* $[0,T_Z]$.

Доказательство теоремы 2. Сперва исследуем процесс

$$\mathbb{D}_n(x) = n^{1/2}(\varphi(S_n^X(x \wedge Z^{(n)})) - \varphi(S^X(x \wedge Z^{(n)}))) \tag{6}$$

и покажем, что при $n \to \infty$

$$\mathbb{D}_n(x) \stackrel{D}{\Rightarrow} \mathbb{D}(x) \quad в\ \ D[0,T_Z].$$

Согласно представлению (19) из [1]

$$\mathbb{D}_n(x) = \sum_{m=1}^{3} n^{1/2}\mathbb{A}_{mn}^*\left(x \wedge Z^{(n)}\right). \tag{7}$$

Поскольку $Z^{(n)} \stackrel{\text{п.н.}}{\to} T_Z$ при $n \to \infty$, то согласно (33) из [1]

$$\sup_{0 \le x \le T_Z} n^{1/2}\left|\mathbb{A}_{mn}^*(x \wedge Z^{(n)})\right| = \mathrm{O}_p(n^{-1/2}), \quad m = 2,3. \tag{8}$$

Поэтому для установления (6), с учетом (7) и (8), достаточно доказать, что

$$n^{1/2}\mathbb{A}_{1n}^*\left(x \wedge Z^{(n)}\right) \stackrel{D}{\Rightarrow} \mathbb{D}(x) \quad в\ \ D[0,T_Z]. \tag{9}$$

Используя формулы (2.1.6), (2.1.14) и (2.1.16) из [2], имеем

$$n^{1/2}\mathbb{A}_{1n}^*\left(x \wedge Z^{(n)}\right) = n^{1/2}(-\frac{1}{n}\int_0^{x \wedge Z^{(n)}} I\left(S_n^Z(t) > 0\right) \cdot$$

$$\cdot S_n^Z(t-)\,\varphi'\left(S_n^Z(t)\right) d\Lambda_n(t) + \int_0^{x \wedge Z^{(n)}} I\left(S_n^Z(t) > 0\right) \cdot$$

$$\cdot S^Z(t)\,\varphi'\left(S^Z(t)\right) d\Lambda(t)) =$$

$$= n^{1/2}(-\frac{1}{n}\int_0^{x \wedge Z^{(n)}} I(S_n^Z(t) > 0) \cdot \varphi'\left(S_n^Z(t)\right)) d\mathbb{M}_n(t) +$$

$$+ \int_0^{x \wedge Z^{(n)}} I(S_n^Z(t) > 0)[\Psi(S_n^Z(t)) - \Psi(S^Z(t))] d\Lambda(t))) +$$

$$+ \mathrm{O}(\frac{1}{n}), \tag{10}$$

где также использовали формулу (16) из [1] и равенство $S_n^Z(t-) = S_n^Z(t) + \frac{1}{n}$. Здесь

$$\mathbb{M}_n(t) = \sum_{k=1}^{n} I\left(Z_k \le t, \delta_k = 1\right) - \int_0^t \mathbb{J}_n(u)\, d\Lambda(u),$$

— квадратично-интегрируемый мартингал с квадратической характеристической

$$< \mathbb{M}_n, \mathbb{M}_n > (t) = \int_0^t \mathbb{J}_n(u)\, d\Lambda(u).$$

Согласно равномерной оценке

$$\sup_{t \in R}\left|S_n^Z(t) - S(t)\right| \stackrel{\text{п.н.}}{=} O\left(\left(n^{-1}\ln n\right)^{1/2}\right)$$

для эмпирической функции S_n^Z с учетом условия (C1), получаем

$$\varphi'(S_n^Z(t)) = \varphi'(S^Z(t)) + \mathrm{O}_p((n^{-1}\ln n)^{1/2}),$$

$$\Psi(S_n^Z(t)) - \Psi(S^Z(t)) =$$

$$= \Psi'(S^Z(t))(S_n^Z(t) - S^Z(t)) + \mathrm{O}_p(n^{-1}\ln n). \tag{11}$$

Из соотношений (10) , (11) имеем

$$n^{1/2}\mathbb{A}_{1n}^*\left(x \wedge Z^{(n)}\right) = B_{1n}(x) + B_{2n}(x) + o_p(1), \tag{12}$$

где

$$B_{1n}(x) = -n^{1/2}\int_0^{x \wedge Z^{(n)}} \varphi'\left(S^Z(t)\right) d\mathbb{M}_n(t),$$

$$B_{2n}(x) = \int_0^{x \wedge Z^{(n)}} \Psi'\left(S^Z(t)\right) \varepsilon_n(t)\, d\Lambda(t).$$

$$\varepsilon_n(t) = n^{1/2}\left(S_n^Z(t) - S^Z(t)\right).$$

Согласно (12) сходимость (9) следует из

$$B_n(x) \stackrel{D}{\Rightarrow} \mathbb{D}(x) \ \text{ в } D[0,T_Z], \tag{13}$$

где $B_n(x) = B_{1n}(x) + B_{2n}(x)$. В работе [5] установлено, что для любого x_0 такого, что $S^Z(x_0) > 0$, $B_n(x)$

слабо сходится к $\mathbb{D}(x)$ в $D[0,x_0]$. Тогда для доказательства (13) согласно критерию слабой сходимости достаточно установить плотность $B_n(x)$, т.е. следующее соотношение при $k=1,2$ для любого $\varepsilon>0$:

$$\lim_{y\to T_Z}\limsup_{n\to\infty} P\left(\sup_{x\in(y,T_Z)}|B_{kn}(x)-B_{kn}(y)|>\varepsilon\right)=0. \tag{14}$$

При $k=1$

$$B_{1n}(x)-B_{1n}(y)=-n^{-1/2}\int\limits_{y\wedge Z^{(n)}}^{x\wedge Z^{(n)}}\varphi'\left(S^Z(t)\right)d\mathbb{M}_n(t). \tag{15}$$

Заметим что (15) является мартингалом интегрального вида с моментом остановки $Z^{(n)}$ и тогда согласно неравенству Ленгляра $\forall\varepsilon,\eta>0$ имеем

$$P\left(\sup_{x\in[y,T_Z]}|B_{1n}(x)-B_{1n}(y)|>\varepsilon\right)\le\frac{\eta}{\varepsilon^2}+$$

$$+P\left(\frac{1}{n}\int\limits_{y\wedge Z^{(n)}}^{T_Z}\left(\varphi'\left(S^Z(t)\right)\right)^2\mathbb{J}_n(t)\,d\Lambda(t)>\eta\right)\le\frac{\eta}{\varepsilon^2}+$$

$$+P\left(\int\limits_{y}^{T_Z}\left(\varphi'\left(S^Z(t)\right)^2S_n^Z(t)\right)d\Lambda(t)>\eta\right). \tag{16}$$

При этом согласно теореме Гливенко-Кантелли при $n\to\infty$

$$\int\limits_{y}^{T_Z}(\varphi'(S^Z(t)))^2S_n^Z(t)d\Lambda(t)\xrightarrow{P}\int\limits_{y}^{T_Z}\left(\frac{\Psi(S^Z(t))}{S^Z(t)}\right)^2d\Lambda(t).$$

Следовательно, второе выражение в (15)при $y\to T_Z$ ввиду условия (C4) и произвольности $\eta>0$ сходится к нулю, т.е. (14) при $k=1$ справедливо. Поскольку по теореме Дуба-Донскера эмпирический процесс $\varepsilon_n(t)$ слабо сходится в $D[0,T_Z]$ к броуновскому мосту $B\left(1-S^Z(t)\right)$, то с учетом условия (C3) и представления

$$B_{2n}(x)-B_{2n}(y)=\int\limits_{y\wedge Z^{(n)}}^{x\wedge Z^{(n)}}\Psi'\left(S^Z(t)\right)\varepsilon_n(t)\,d\Lambda(t),$$

убедимся в справедливости (14) и в случае $k=2$. Теперь плотность $B_n(x)$ следует из (14) по неравенству треугольника. Таким образом, (6) справедливо. Для доказательства (5) достаточно заметить, что согласно условию (C1)

$$q_n(x)=\mathbb{D}_n(x)\left[\varphi'\left(S^X\left(x\wedge Z^{(n)}\right)\right)\right]^{-1}+o_p(1)=$$

$$=\mathbb{D}_n(x)\left\{\left[\varphi'\left(S^X(x)\right)\right]^{-1}I\left[0,Z^{(n)}\right)+\right.$$

$$\left.+\left[\varphi'\left(S^X\left(Z^{(n)}\right)\right)\right]^{-1}I\left\{Z^{(n)}\right\}\right\}+o_p(1). \tag{17}$$

Теперь (5) следует из (17) при $n\to\infty$. Теорема 2 доказана.

Замечание 1. В ходе доказательстве теоремы 2 использовена слабая сходимость

$$B_n(x)\Rightarrow\mathbb{D}\text{ в }D[0,x_0],$$

где гауссовский процесс $\mathbb{D}(x)$ имеет ковариацию (2), вычисление которой имеется в [5] поэтому оно здесь не приводится.

Замечание 2. Следует отметить, что результаты данной работы авторами обобщены и для многомерного аналога оценки (1) с использованием многомерной архимедова копула функции.

Список литературы

[1] Абдушукуров А.А., Душатов Н.Т., Мурадов Р.С. Оценивание функционалов многомерного распределения по цензурированным наблюдениям при помощи копула функций. *Сб. Статистические методы оценивания и проверки гипотез. Пермь. Пермский Госуниверситет.*, вып. 23:132–143, 2011.

[2] Абдушукуров А.А., Мурадов Р.С. Модель одномерного зависимого случайного цензурирования справа и оценки функции выживания. *УзМЖ*, 3:3–12, 2011.

[3] Абдушукуров А.А. *Статистика неполных наблюдений. Асимптотическая теория оценивания для неклассических моделей.* Ташкент. "Университет 2009.

[4] Abdushukurov A.A. and Muradov R.S. Dependent random censoring from the right and the survival function copula estimates. *Proceedings of XII International Conf. on Financial and Actuarial Mathematics and Eventology of Safety. Krasnoyarsk. Russia*, pages 15–19, 2013.

[5] Li Y. , Tiwari R.C., Guha S. Mixture cure survival models with dependent censoring. *J.Royal Statist. Soc. B.*, 69. Part 3.:285–306, 2007.

[6] Nelsen R.B. *An introduction to copulas.* Springer, New York., 1999 (Second edition in 2006).

[7] Rivest L.-P. and Wells M.T. A martingale approach to the copula-graphic estimator for the survival function under dependent censoring. *J. Multivar. Anal.*, 79:138–155, 2001.

[8] Zeng M. and Klein J.P. A self-consistent estimator of marginal survival functions based on dependent competing risk data an assumed copula. *Commun. Statist.: Theory and Meth.*, 23:2299–2311, 1994.

Ковариационный анализ разнотипных данных с помощью метода двудольных множеств событий

Ирина Владимировна Баранова
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 понятия классической и эвентологической ковариации. Изучаются зависимости между двудольными множествами событий с помощью ковариации.*

Ключевые слова. *Событие, вероятность, ковариация, двудольное множество событий, множество двудольных множеств*

1 Введение

Измерение вероятностной зависимости между событиями — фундаментальная проблема теории вероятности и математической статистики, которая решается с помощью изучения зависимостей случайных величин или зависимостей более общих случайных элементов линейных пространств. Традиционными мерами зависимости случайных линейных элементов считается ковариация и корреляция.

Целью данной работы является изучение зависимости между разнотипными данными, описывающими поведение сложной системы, с помощью нахождения ковариационной связи между ними.

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 был предложен метод двудольных множеств случайных событий [1], [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.

1.1 Ковариация в теории вероятности

Определение 1. *Вероятностным пространством называется тройка* $(\Omega, \mathcal{F}, \mathbf{P})$, *где* Ω *— пространство элементарных событий,* $\mathcal{F}$ *— алгебра событий и* $\mathbf{P}$ *— вероятность, определенная на элементах алгебры* $\mathcal{F}$ *— случайных событиях* $x, y, \ldots \in \mathcal{F}$.

Определение 2. *Конечное множество избранных событий* $\mathfrak{X} \subseteq \mathcal{F}$, *выбранных из алгебры вероятностного пространства* $(\Omega, \mathcal{F}, \mathbf{P})$ *и состоящее из* $N = |\mathfrak{X}|$ *событий, называется множеством случайных событий.*

Определение 3. *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: $x \subseteq y$, $y \subseteq x$.

Определение 4. *Если каждые два события из* $\mathfrak{X}$ *вложенные, то* $\mathfrak{X}$ *называется множеством вложенных событий или множеством случайных событий с вложенной эвентологической структурой.*

Определение 5. *Случайная величина определяется как измеримое отображение*

$$\xi : (\Omega, \mathcal{F}, \mathbf{P}) \to (\mathbb{R}, \mathcal{B}),$$

где $\mathbb{R}$ *— вещественная прямая,* $\mathcal{B}$ *— борелевская алгебра подмножеств* $\mathbb{R}$.

Рассмотрим понятие ковариации в теории вероятности. Пусть $\mathbf{X}$ и $\mathbf{Y}$ — две случайные величины, определенные на одном и том же пространстве элементарных событий. Тогда $\mathbf{X}+\mathbf{Y}$ и $\mathbf{XY}$ суть новые случайные величины, распределение которых можно найти простой группировкой членов совместного распределения $\mathbf{X}$ и $\mathbf{Y}$. Наша цель состоит теперь в вычислении $D(\mathbf{X}+\mathbf{Y})$. Для этого введем новое понятие ковариации. Если $\{p(x_j, y_k)\}$ есть совместное распределение величин $\mathbf{X}$ и $\mathbf{Y}$, то математическое ожидание $\mathbf{XY}$ определяется формулой

$$\mathbf{E(XY)} = \sum \mathbf{x_j y_k p(x_j, y_k)},$$

в предположении, конечно, что ряд абсолютно сходится. Далее, $|x_j y_k| \leq (x_j^2 + y_k^2)/2$, и поэтому $\mathbf{E(XY)}$ обязательно существует, если существует $\mathbf{E(X^2)}$ и

$\mathbf{E}(\mathbf{Y^2})$. В этом случае существует также и математическое ожидание

$$\mu_x = \mathbf{E}(\mathbf{X}), \mu_\mathbf{y} = \mathbf{E}(\mathbf{Y})$$

а случайные величины $\mathbf{X}-\mu_x$ и $\mathbf{Y}-\mu_y$ имеют нулевые математические ожидания. Для их произведения по теореме сложения получаем

$$\mathbf{E}((\mathbf{X}-\mu_\mathbf{x})(\mathbf{Y}-\mu_\mathbf{y})) =$$

$$= \mathbf{E}(\mathbf{XY})-\mu_\mathbf{x}\mathbf{E}(\mathbf{Y})-\mu_\mathbf{y}\mathbf{E}(\mathbf{X})+\mu_\mathbf{y}\mu_\mathbf{x} = \mathbf{E}(\mathbf{XY})-\mu_\mathbf{y}\mu_\mathbf{x}.$$

Определение 6. *Ковариацией случайных величин* $\mathbf{X}$ *и* $\mathbf{Y}$ *называется*

$$Cov(\mathbf{X},\mathbf{Y}) = \mathbf{E}((\mathbf{X}-\mu_\mathbf{x})(\mathbf{Y}-\mu_\mathbf{y})) = \mathbf{E}(\mathbf{XY}) - \mu_\mathbf{y}\mu_\mathbf{x}.$$

2 Эвентологический подход к ковариации

2.0.1 Плетные ковариации событий

Плетные ковариации — это обычные ковариации, используемые в теории вероятностей. Будем называть их просто ковариациям.

Обозначим

$$1_x = 1_x(\omega) = \begin{cases} 1, & \omega \in x \\ 0, & \text{иначе,} \end{cases}$$

— индикаторы событий $x \in \mathfrak{X}$—булевы случайные величины на вероятностном пространстве $(\Omega, \mathcal{F}, \mathbf{P})$.

Определение 7. *Ковариацией* $\mathfrak{X}$*, плетной ковариацией* $\mathfrak{X}$*, N-плетной ковариацией событий из* $\mathfrak{X}$*,* $\mathfrak{X}$*-плетной ковариацией называется величина*

$$Cov_\mathfrak{X} = \begin{cases} \mathbf{E}\left(\prod_{\mathbf{x}\in\mathfrak{X}}\left(\mathbf{1_x}-\mathbf{p_x^1}\right)\right), & \mathfrak{X}\neq\emptyset, \\ 1, & \textit{иначе,} \end{cases}$$

— центральный смешанный момент порядка $N = |\mathfrak{X}|$ *индикаторов событий из* $\mathfrak{X}$*.*

2.0.2 Арные ковариации событий

В [5] вводится новое понятие — это арная ковариация событий.

Определение 8. *Арной ковариацией* $\mathfrak{X}$*, N-арной ковариацией событий из* $\mathfrak{X}$*,* $\mathfrak{X}$*-арной ковариацией называется величина*

$$Kov_\mathfrak{X} = \begin{cases} \mathbf{E}\left(\prod_{\mathbf{x}\in\mathfrak{X}}\mathbf{1_x}\right) - \prod_{\mathbf{x}\in\mathfrak{X}}\mathbf{p_x^1}, & \mathfrak{X}\neq\emptyset, \\ 0, & \textit{иначе.} \end{cases}$$

Из определения следует, что

$$Kov_\mathfrak{X} = \begin{cases} \mathbf{P}\left(\bigcap_{\mathbf{x}\in\mathfrak{X}}\mathbf{x}\right) - \prod_{\mathbf{x}\in\mathfrak{X}}\mathbf{P}(\mathbf{x}), & \mathfrak{X}\neq\emptyset, \\ 0, & \text{иначе.} \end{cases} =$$

$$= \begin{cases} p_\mathfrak{X}(\mathfrak{X}) - ind_\mathfrak{X}(\mathfrak{X}) & \mathfrak{X}\neq\emptyset, \\ 0, & \text{иначе.} \end{cases}$$

—это разность между вероятностями пересечения и независимого пересечения событий мультиплета $\mathfrak{X}$.

3 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 $x, y, \ldots \in \mathcal{F}$ — элементы алгебры $\mathcal{F}$, на которой определена вероятность $\mathbf{P}$. Множество избранных случайных событий $\mathfrak{X} \in \mathcal{F}$.

Определение 9. *Случайное множество событий под* $\mathfrak{X}$ *определяется как измеримое отображение*

$$K : (\Omega, \mathcal{F}, \mathbf{P}) \to \left(2^\mathfrak{X}, 2^{2^\mathfrak{X}}\right),$$

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 10. *Двудольным множеством случайных элементов называется множество* $\{\boldsymbol{\xi}, \mathbf{K}\}$*, представимое в следующем виде:*

$$\{\boldsymbol{\xi}, \mathbf{K}\} = \boldsymbol{\xi} \cup \mathbf{K} = \{\xi_a,\, a \in A,\, K_\beta,\, \beta \in B\}, \qquad (1)$$

здесь первая доля — это случайные величины $\boldsymbol{\xi} = \{\xi_a,\, a \in A\}$*, 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 11. *Двудольное множество случайных событий представляет собой объединение двух множеств — множества событий, которое определяется случайными величинами, и множества событий, которое определяется случайными множествами событий:*

$$\{\mathcal{Y}, \mathfrak{X}\} = \{\mathcal{Y}_a,\, \mathfrak{X}_\beta,\, a \in A,\, \beta \in B\}. \qquad (2)$$

Полной характеристикой двудольного множества случайных событий служит его эвентологическое распределение. В работе [5] были найдены все формы эвентологического распределения двудольного множества случайных событий. Приведем ниже вид для одного из них.

Распределение вероятностей событий двудольного множества $\{\mathcal{Y}, \mathfrak{X}\}$:

$$p(\mathbf{r}, \mathbf{X}) = \mathbf{P}\Big(\bigcap_{a\in A} \Big\{ \bigcap_{y\neq r} \{\xi = y\}^c \Big\} \bigcap_{x_\beta \in X} x_\beta \bigcap_{x_\beta \in X^c} x_\beta^c,\ a \in A,\ \beta \in B \Big), \tag{3}$$

$$x_\beta \in \mathfrak{X},\ \beta \in B,$$
$$\mathbf{r} = \{r_a, r_a \in \mathbb{R}\},$$
$$\mathbf{X} = \{X_\beta, X_\beta \subseteq \mathfrak{X}_\beta\}.$$

На практике работать с такими распределения достаточно сложно. Поэтому в [3] было предложено использовать для их оценивания сет-операции по Минковскому.

4 Сет–операции по Минковскому над двудольными множествами случайных событий

В работе [3] было введено понятие произвольной операции по Минковскому над двумя двудольными множествами событий и ее вероятность.

Пусть имеются два двудольных множества $s^1 = \Big\{\mathcal{Y}_a^1,\ \mathfrak{X}_\beta^1,\ a \in A,\ \beta \in B\Big\}$ и $s^2 = \Big\{\mathcal{Y}_a^2,\ \mathfrak{X}_\beta^2,\ a \in A,\ \beta \in B\Big\}$.

Определение 12. *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:*

$$s1(\mathcal{O})s2 = \Big\{\mathcal{Y}_a^1(\mathcal{O})\mathcal{Y}_a^2,\ \mathfrak{X}_\beta^1(\mathcal{O})\mathfrak{X}_\beta^2,\ a \in A,\ \beta \in B\Big\} = \Big\{\mathcal{Y}_a^1(r_a)\mathcal{O}\mathcal{Y}_a^2(r_a),\ X_\beta^1 \mathcal{O} X_\beta^2,\ X_\beta \subseteq \mathfrak{X}_\beta,\ r_a \in \mathcal{R}_a,\ a \in A,\ \beta \in B\Big\}. \tag{4}$$

Здесь $\mathcal{Y}_a^1, \mathcal{Y}_a^2 \subseteq \mathcal{Y}_a$, $\mathfrak{X}_\beta^1, \mathfrak{X}_\beta^2 \subseteq \mathfrak{X}_\beta$. Вероятность произвольной операции определяется как

$$\mathbf{P}\left(s^1(\mathcal{O})s^2\right) = \frac{1}{|A|}\sum_{a\in A}\frac{1}{|\mathcal{Y}_a|}\sum_{r_a\in\mathcal{R}_a}\mathbf{P}\Big(\mathcal{Y}_a^1(r_a)\mathcal{O}\mathcal{Y}_a^2(r_a)\Big) + \frac{1}{|B|}\sum_{\beta\in B}\frac{1}{|\mathfrak{X}_\beta|}\sum_{X_\beta\subseteq\mathfrak{X}_\beta}\mathbf{P}\left(X_\beta^1 \mathcal{O} X_\beta^2\right). \tag{5}$$

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:

$$s1(\Delta)s2 = \Big\{\mathcal{Y}_a^1(r_a)\Delta\mathcal{Y}_a^2(r_a),\ X_\beta^1 \Delta X_\beta^2,\ X_\beta \subseteq \mathfrak{X}_\beta,\ r_a \in \mathcal{R}_a,\ a \in A,\ \beta \in B\Big\}. \tag{6}$$

Вероятность симметрической разности равна

$$\mathbf{P}\left(s^1(\Delta)s^2\right) = \frac{1}{|A|}\sum_{a\in A}\frac{1}{|\mathcal{Y}_a|}\sum_{r_a\in\mathbb{R}}\mathbf{P}\Big(\mathcal{Y}_a^1(r_a)\Delta\mathcal{Y}_a^2(r_a)\Big) + \frac{1}{|B|}\sum_{\beta\in B}\frac{1}{|\mathfrak{X}_\beta|}\sum_{X_\beta\subseteq\mathfrak{X}_\beta}\mathbf{P}\left(X_\beta^1 \Delta X_\beta^2\right). \tag{7}$$

В [5] 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.

5 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 13. *Событие-терраска двудольного множества случайных событий s представляет собой набор непересекающиеся событий, где каждое событие является подмножеством соответствующего множества событий $\mathcal{Y}_a$ или $\mathfrak{X}_\beta$:*

$$\begin{aligned}\mathrm{ter}(s) &= \mathrm{ter}\{\mathcal{Y}_{s_A}, \mathfrak{X}_{s_B}\} = \\ &= \bigcap_{a\in s_A}\mathrm{ter}(\mathcal{Y}_a)\bigcap_{\beta\in s_B}\mathrm{ter}(\mathfrak{X}_\beta) = \\ &= \bigcap_{a\in s_a}\mathcal{Y}_a(r_a)\bigcap_{\beta\in s_B}\Big(\bigcap_{x_\beta\in X_\beta}x_\beta\bigcap_{x_\beta\in X_\beta^c}x_\beta^c\Big),\end{aligned} \tag{8}$$

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 ξ_a:

$$\mathcal{R}_a = \left\{r_{a_1}, \ldots, r_{a_{N_a}}\right\} \subset \mathbb{R},\ a \in A.$$

6 Множество двудольных множеств событий

В работе [6] было предложено понятие множества двудольных множеств событий. Введем более строгое определение для данного понятия.

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:*

$$S = \left\{ s^1, \ldots, s^n \right\}. \tag{9}$$

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):

$$S = \Big\{ s^1 = \{\mathcal{Y}_a^1, \mathfrak{X}_\beta^1, a \in A, \beta \in B\}, \ldots,$$
$$s^n = \{\mathcal{Y}_a^n, \mathfrak{X}_\beta^n, a \in A, \beta \in B\} \Big\}.$$

   Этот случай является простейшим. Именно эта ситуация была рассмотрена в работах [5] и [6].

2. Случай, когда двудольные множества событий s^i являются подмножествами одного и того же двудольного множества событий Z:

$$S = \left\{ s^1 = \left\{ \mathcal{Y}_{s_A^1}, \mathfrak{X}_{s_B^1}, s_A^1 \subseteq A, s_B^1 \subseteq B \right\}, \ldots, \right.$$
$$\left. s^n = \{\mathcal{Y}_{s_A^n}, \mathfrak{X}_{s_B^n}, s_A^n \subseteq A, s_B^n \subseteq B\} \right\}.$$

3. Случай, когда двудольные множества событий s^i являются подмножествами различных двудольных множеств событий $Z^i = \{\mathcal{Y}^i, \mathfrak{X}^i\} = \{\mathcal{Y}_a^i, \mathfrak{X}_\beta^i, a \in A^i, \beta \in B^i\}$:

$$S = \left\{ s^1 = \left\{ \mathcal{Y}_{k_A^1}, \mathfrak{X}_{k_B^1}, k_A^1 \subseteq A^1, k_B^1 \subseteq B^1 \right\}, \ldots, \right.$$
$$\left. s^n = \{\mathcal{Y}_{k_A^n}, \mathfrak{X}_{k_B^n}, k_A^n \subseteq A^n, k_B^n \subseteq B^n\} \right\}.$$

   Данный случай является самым общим и самым сложным по своей структуре.

Приведем для каждого случая практические примеры и набор возможных задач, приводящих к описываемым случаям.

7 Применение множеств двудольных множеств событий

7.1 Множество однотипных двудольных множеств событий

Как уже было сказано выше, наиболее простым (и чаще всего встречающимся) является случай 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 $\{\mathcal{Y}, \mathfrak{X}\}$. Далее анализ поведения элементов системы сводится к анализу эвентологических распределений соответствующих им двудольных множеств событий. Таким образом, сложной системе ставится в соответствие множество

$$S = \Big\{ s^1 = \{\mathcal{Y}_a^1, \mathfrak{X}_\beta^1, a \in A, \beta \in B\}, \ldots,$$
$$s^n = \{\mathcal{Y}_a^n, \mathfrak{X}_\beta^n, a \in A, \beta \in B\} \Big\}.$$

Примерами таких систем являются банковская система (состояние каждого банка оценивается числовыми и множественными показателями), система здравоохранения территориальных единиц региона, система кредитных продуктов и т.д.

Приведем практический пример множества однотипных двудольных множеств событий.

Представление банковской системы

Банковская система представляет собой множество банков и состоит из n элементов. Финансовое положение и репутация каждого банка характеризуется множеством показателей, одни из которых является числовыми, а другие — множественными. Например, к числовым показателям могут быть отнесены: резервы в Центробанке, уставный капитал, эмиссионный доход, прибыльность по различным операциям, объем кредитов, предоставленных коммерческим предприятиям и организациям, частным лицам и т.д., ко множественным — обменная валюта банка, ассортимент банковских услуг, множество акционеров (участников), владеющих акциями банка. Очевидно, что банковская система является примером сложной системы, а банки — ее элементы. В работе [5] рассматривалось 39 показателей финансового положения банка.

Финансовое состояние каждого банка можно представить с помощью двудольного множества событий: первая доля которого определяется случайными величинами (числовыми показателями), вторая

— случайными множествами событий (множественными показателями):

$$s^i = \{\mathcal{Y}_a^i, \mathfrak{X}_\beta^i, a \in A, \beta \in B\},\ i = 1, \ldots, n.$$

Всей банковской системе можно поставить в соответствие множество двудольных множеств событий

$$S = \Big\{ s^1 = \{\mathcal{Y}_a^1,\, \mathfrak{X}_\beta^1,\, a \in A,\, \beta \in B\}, \ldots,$$
$$s^n = \{\mathcal{Y}_a^n, \mathfrak{X}_\beta^n,\, a \in A,\, \beta \in B\} \Big\}.$$

Таким образом, множество всех показателей имеет вид

$$Z = \{\mathcal{Y}_a, \mathfrak{X}_\beta\} = \Big\{\mathcal{Y}_1, \mathcal{Y}_2, \mathcal{Y}_3, \ldots, \mathcal{Y}_{36}, \mathfrak{X}_1, \mathfrak{X}_2, \mathfrak{X}_3\Big\}, \tag{10}$$

где $\mathcal{Y}_1$ — величина денежных резервов в Центробанке (в тыс. рублей), $\mathcal{Y}_2$ — размер уставного капитала (в тыс. рублей), $\mathcal{Y}_3$ — эмиссионный доход, и т.д. А $\mathfrak{X}_1$ — обменная валюта банка, $\mathfrak{X}_2$ — ассортимент банковских услуг, $\mathfrak{X}_3$ — множество акционеров банка.

Поскольку все двудольные множества s^i порождаются одним и тем же двудольным множеством случайных элементов Z, значит:

$$|\mathcal{Y}_a| = |\mathcal{Y}_a^1| = |\mathcal{Y}_a^2| = \ldots = |\mathcal{Y}_a^n|,$$

$$|\mathfrak{X}_\beta| = |\mathfrak{X}_\beta^1| = |\mathfrak{X}_\beta^2| = \ldots = |\mathfrak{X}_\beta^n|.$$

Вернемся к расмотрению множества однотипных двудольных множеств событий.

Полной характеристикой множества двудольных множеств случайных событий служит его эвентологическое распределение. Для этого необходимо задать распределение вероятностей событий множества двудольных множеств.

Пусть задано распределение вероятностей событий двудольного множества s:

$$p(s) = p(\mathcal{Y}_a,\, \mathfrak{X}_\beta) = p(\mathbf{r}, \mathbf{X}) = \mathbf{P}\Big(\bigcap_{a \in A} \{\mathcal{Y}_a(r_a)\} \bigcap_{x_\beta \in X} x_\beta \bigcap_{x_\beta \in X^c} x_\beta^c,\ a \in A,\, \beta \in B\Big), \tag{11}$$

$$x_\beta \in \mathfrak{X},\, \beta \in B,$$

$$\mathbf{r} = \{r_a, r_a \in \mathbb{R}\},$$

$$\mathbf{X} = \{X_\beta, X_\beta \subseteq \mathfrak{X}_\beta\}.$$

Тогда распределение вероятностей событий множества двудольных множеств событий $S = \Big\{s^1, \ldots, s^n\Big\}\, p(S^t), S^t \subseteq S$ имеет вид:

$$p(S^t) = \mathbf{P}\left(\bigcap_{s^i \in S^t} s \right) = \mathbf{P}\Big(\bigcap_{s^i \in S^t} \Big\{ \bigcap_{a \in A} \mathcal{Y}_a^i(r_a) \bigcap_{x_\beta \in X} x_\beta \bigcap_{x_\beta \in X^c} x_\beta^c \Big\}\Big),\ S^t \subseteq S. \tag{12}$$

В работах [3, 5] были решены несколько задач системного анализа — определение экстремального элемента системы (наилучшего или наихудшего), классификация и ранжирование элементов системы. Все эти задачи предполагали сравнение только пар двудольных множеств, что и выполнялось с помощью представленного в формуле (7) расстояния между эвентологическими распределениями двудольных множеств событий (как вероятности сет-операции симметрической разности по Минковскому).

Кроме того, к аналогичному виду множества двудольных множеств событий приводят задачи кластерного анализа, визуализации (построение карт визуализации), банковского (страхового) скоринга и некоторые другие.

В работе [6] было предложено применить метод двудольных множеств событий для проведения факторного анализа сложных систем. Здесь требовалось определить степень влияния факторов на различие между элементами и найти внутрисистемное различие (т.е. здесь необходимо сравнивать одновременно все элементы системы).

Очевидно, что для практических задач достаточно сложно оценить вероятности распределения $p(S^t), S^t \subseteq S$. Поэтому в [6] были введены понятия произвольной сет-операции по Минковскому над множеством двудольных множеств событий и ее вероятности.

Определение 14. *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:*

$$\underset{s \in S}{\mathcal{O}} s = \Big\{ \underset{s \in S}{\mathcal{O}} \Big\{\mathcal{Y}_a\Big\}, \underset{s \in S}{\mathcal{O}} \Big\{\mathfrak{X}_\beta\Big\},\ a \in A,\ \beta \in B\Big\} = \Big\{ \underset{s \in S}{\mathcal{O}} \Big\{\mathcal{Y}_a(r_a)\Big\}, \underset{s \in S}{\mathcal{O}} \Big\{X_\beta\Big\},\ r_a \in \mathcal{R}_a,\ a \in A,\ X_\beta \subseteq \mathfrak{X}_\beta,\ \beta \in B\Big\}. \tag{13}$$

Вероятность произвольной сет–операции по Минковскому над множеством двудольных множеств собы-

тий определим следующим образом:

$$\mathbf{P}\left(\left(\bigcirc\limits_{s\in S}\right)s\right) = \\ = \frac{1}{|A|}\sum_{a\in A}\frac{1}{|\mathcal{Y}_a|}\sum_{r_a\in\mathcal{R}_a}\mathbf{P}\left(\bigcirc\limits_{s\in S}\left\{\mathcal{Y}_a(r_a)\right\}\right) + \\ \frac{1}{|B|}\sum_{\beta\in B}\frac{1}{|\mathfrak{X}_\beta|}\sum_{X_\beta\subseteq\mathfrak{X}_\beta}\mathbf{P}\left(\bigcirc\limits_{s\in S}\left\{X_\beta\right\}\right). \quad (14)$$

Приведем определение и вероятность для сет-операции симметрической разности по Минковскому над множеством двудольных множеств.

Определение 15. *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:*

$$\left(\bigtriangleup\limits_{s\in S}\right)s = \left\{\left(\bigtriangleup\limits_{s\in S}\right)\left\{\mathcal{Y}_a\right\}, \left(\bigtriangleup\limits_{s\in S}\right)\left\{\mathfrak{X}_\beta\right\}\right\} = \\ \left\{\bigtriangleup\limits_{s\in S}\left\{\mathcal{Y}_a(r_a)\right\}, \bigtriangleup\limits_{s\in S}\left\{X_\beta\right\}, r_a\in\mathcal{R}_a,\, a\in A,\right. \\ \left. X_\beta\subseteq\mathfrak{X}_\beta,\, \beta\in B\right\}. \quad (15)$$

Определение 16. *Под оператором общей симметрической разности* $\bigtriangleup\limits_{s\in S}$ *над множеством событий* S *понимается среднее событие по множеству, вероятность которого определяется следующим образом:*

$$\mathbf{P}\left(\bigtriangleup\limits_{s\in S}s\right) = \frac{1}{|S|}\sum_{s,t\in S}P(s\Delta t). \quad (16)$$

Вероятность общей симметрической разности по Минковскому над множеством двудольных множеств событий выглядит следующим образом:

$$\mathbf{P}\left(\left(\bigtriangleup\limits_{s\in S}\right)s\right) = \\ \frac{1}{|A|}\sum_{a\in A}\frac{1}{|\mathcal{Y}_a|}\sum_{r_a\in\mathcal{R}_a}\mathbf{P}\left(\bigtriangleup\limits_{r_a\in\mathcal{R}_a}\left\{\mathcal{Y}_a(r_a)\right\}\right) + \\ +\frac{1}{|B|}\sum_{\beta\in B}\frac{1}{|\mathfrak{X}_\beta|}\sum_{X_\beta\subseteq\mathfrak{X}_\beta}\mathbf{P}\left(\bigtriangleup\limits_{s\in S}\left\{X_\beta\right\}\right). \quad (17)$$

7.2 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:

$$S = \left\{s^1 = \left\{\mathcal{Y}_{s^1_A}, \mathfrak{X}_{s^1_B},\, s^1_A\subseteq A, s^1_B\subseteq B\right\}, \ldots, \right. \\ \left. s^n = \left\{\mathcal{Y}_{s^n_A}, \mathfrak{X}_{s^n_B},\, s^n_A\subseteq A, s^n_B\subseteq B\right\}\right\}.$$

К изучению множеств такого вида могут привести задачи регрессионного или ковариационного анализа. Например, когда требуется найти ковариацию или корреляцию между двудольными множествами событий. В качестве практического примера можно привести задачу изучения взаимосвязи между состоянием здоровья населения и антропогенным загрязнением воздуха и почвы.

Задача исследования зависимости между антропогенным загрязнением воздуха-почвы и заболеваемостью населения

В работе [2] решалась задача, которая заключалась в выявлении зависимости между антропогенным загрязнением атмосферы и почвы, коэффициентом опасности выбрасываемых в атмосферу веществ, жесткостью погоды и заболеваемостью населения Красноярского края.

Здесь двудольное множество Z состоит из трех числовых и трех множественных показателей

$$Z = \{\mathcal{Y}, \mathfrak{X}\} = \{\mathcal{Y}_a, \mathfrak{X}_\beta, a\in A, \beta\in B\} = \\ \{\mathcal{Y}_1, \mathcal{Y}_2, \mathcal{Y}_3, \mathfrak{X}_1, \mathfrak{X}_2, \mathfrak{X}_3\},$$

где

$\mathcal{Y}_1$ — коэффициент опасности (числовой показатель);

$\mathcal{Y}_2$ — коэффициент жесткости погоды по Осокину за летний период (числовой показатель);

$\mathcal{Y}_3$ — коэффициент жесткости погоды по Осокину за зимний период (числовой показатель);

$\mathfrak{X}_1$ — множество выбросов вредных веществ в атмосферу (множественный показатель);

$\mathfrak{X}_2$ — химический состав почвы (множественный показатель);

$\mathfrak{X}_3$ — множество заболеваний (множественный показатель).

Рассматривались двудольные множества

$$s = \{\mathcal{Y}_1, \mathcal{Y}_2, \mathcal{Y}_3, \mathfrak{X}_1, \mathfrak{X}_2\}, \quad (18)$$

$$t = \{\mathfrak{X}_3\}. \quad (19)$$

Они являются подмножествами двудольного множества Z $(s, t\subseteq Z)$.

Однако для исследования интересны лишь часть терасок показателей, поэтому в качестве исследуемого множества в работе использовалось не все множество, а множество подмножеств двудольных множеств событий — наиболее «плохие» терраски (с наибольшими значениями выбросов и наибольшей опасностью) и наиболее «хорошие» (с наименьшими значениями).

К аналогичному виду множества двудольных множеств событий приводят задачи поиска ассоциатив-

ных правил (отыскания закономерностей между множествами событий), прогнозирования поведения системы и задачи принятия решений.

Распределение вероятностей событий множества двудольных множеств событий $S = \left\{s^1, \ldots, s^n\right\}$ $p\left(S^t\right), S^t \subseteq S$ для данного случая имеет следующий вид:

$$\begin{aligned} p\left(S^t\right) &= \mathbf{P}\left(\bigcap_{s^i \in S^t} s^i\right) = \\ &= \mathbf{P}\Big(\bigcap_{s^i \in S^t}\Big\{\bigcap_{a \in s_A^i} \mathrm{ter}(\mathcal{Y}_a) \bigcap_{\beta \in s_B^i} \mathrm{ter}(\mathfrak{X}_\beta)\Big\} = \\ &= \mathbf{P}\Big(\bigcap_{a \in s_a^i} \mathcal{Y}_a(r_a) \bigcap_{\beta \in s_B^i}\Big(\bigcap_{x_\beta \in X_\beta} x_\beta \bigcap_{x_\beta \in X_\beta^c} x_\beta^c\Big)\Big), \\ s_A^i \subseteq A, s_B^i \subseteq B,\ S^t \subseteq S. \end{aligned} \tag{20}$$

В работе [4] 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*

$$\begin{aligned} Kov_{\mathrm{ter}(s),\mathrm{ter}(t)} &= \mathbf{P}\left(\mathrm{ter}(s) \cap \mathrm{ter}(t)\right) - \\ &\mathbf{P}\left(\mathrm{ter}(s)\right)\mathbf{P}\left(\mathrm{ter}(t)\right). \end{aligned}$$

Можно распространить данное понятие и на множество событий-террасок двудольных множеств.

Определение 18. *Ковариацией множества $S^t \subseteq S$, где*

$$\begin{aligned} S = \Big\{ & s^1 = \Big\{\mathcal{Y}_{s_A^1}, \mathfrak{X}_{s_B^1},\ s_A^1 \subseteq A, s_B^1 \subseteq B\Big\}, \ldots, \\ & s^n = \left\{\mathcal{Y}_{s_A^n}, \mathfrak{X}_{s_B^n},\ s_A^n \subseteq A, s_B^n \subseteq B\right\}\Big\}. \end{aligned}$$

называется

$$Kov_{S^t} = p\left(S^t\right) - \prod_{s^i \in S^t} p\left(s^i\right), \tag{21}$$

где $p\left(s^i\right)$ — распределение вероятностей событий двудольного множества s^i, а распределение множества двудольных множеств $p\left(S^t\right)$ находится по формуле (20).

Распределения $p\left(s^i\right)$ имеют вид, представленный выше:

$$\begin{aligned} & p\left(s^i\right) = p\left(\mathcal{Y}_a^i,\ \mathfrak{X}_\beta^i\right) = p\left(\mathbf{r}, \mathbf{X}\right) = \\ & \mathbf{P}\Big(\bigcap_{a \in A}\left\{\mathcal{Y}_a(r_a)\right\} \\ & \bigcap_{x_\beta \in X} x_\beta \bigcap_{x_\beta \in X^c} x_\beta^c,\ a \in A,\ \beta \in B\Big), \end{aligned} \tag{22}$$

$$x_\beta \in \mathfrak{X},\ \beta \in B,$$

$$\mathbf{r} = \left\{r_a, r_a \in \mathbb{R}\right\},$$

$$\mathbf{X} = \left\{X_\beta, X_\beta \subseteq \mathfrak{X}_\beta\right\}.$$

7.3 Множество подмножеств различных двудольных множеств событий

Теперь рассмотрим третий случай, когда двудольные множества событий s^i являются подмножествами различных двудольных множеств событий $Z^i = \{\mathcal{Y}^i, \mathfrak{X}^i\} = \{\mathcal{Y}_a^i, \mathfrak{X}_\beta^i, a \in A^i, \beta \in B^i\}$:

$$\begin{aligned} S = \Big\{ & s^1 = \Big\{\mathcal{Y}_{k_A^1}, \mathfrak{X}_{k_B^1},\ k_A^1 \subseteq A^1, k_B^1 \subseteq B^1\Big\}, \ldots, \\ & s^n = \left\{\mathcal{Y}_{k_A^n}, \mathfrak{X}_{k_B^n},\ k_A^n \subseteq A^n, k_B^n \subseteq B^n\right\}\Big\}, \end{aligned}$$

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 $\mathcal{Y}_{k_A^1} \subseteq \mathcal{Y}^i$, а $\mathfrak{X}_{k_B^1} \subseteq \mathfrak{X}_\beta^i$.

К изучению множеств такого вида могут привести задачи регрессионного, ковариационного анализа, поиска ассоциативных правил (отыскания закономерностей между множествами событий), прогнозирования поведения системы и принятия решений. Как мы видим, этот перечень задач аналогичен ситуации с вторым видом распределения множества двудольных множеств событий. Однако отличием является то, что во втором случае все множества являются подмножествами одного и того же множества Z, в то время как для данного случая происходит работа не с одним, а с целым набором разных областей Z^i. Большинство задач регрессионного и ковариационного анализа относятся именно к данному виду множества двудольных множеств.

Распределение вероятностей событий множества двудольных множеств событий $S = \left\{s^1, \ldots, s^n\right\}$ $p\left(S^t\right), S^t \subseteq S$ для данного случая имеет следующий

вид:

$$
\begin{aligned}
&p\Big(S^t\Big) = \mathbf{P}\left(\bigcap_{s^i\in S^t} s^i\right) = \\
&= \mathbf{P}\Big(\mathrm{ter}\left(\mathcal{Y}_{k_A^1}\right)\bigcap\mathrm{ter}\left(\mathcal{Y}_{k_A^2}\right)\ldots\bigcap\mathrm{ter}\left(\mathcal{Y}_{k_A^t}\right) \\
&\mathrm{ter}\left(\mathfrak{X}_{k_B^1}\right)\bigcap\mathrm{ter}\left(\mathfrak{X}_{k_B^2}\right)\ldots\bigcap\mathrm{ter}\left(\mathfrak{X}_{k_B^t}\right)\Big) = \\
&= \mathbf{P}\Big(\bigcap_{k_A^1\in A^1}\mathcal{Y}_{k_A^1}\ldots\bigcap_{k_A^t\in A^t}\mathcal{Y}_{k_A^t} \\
&\bigcap_{k_B^1\in B^1}\Big(\bigcap_{x_\beta\in X_{k_B^1}} x_\beta\bigcap_{x_\beta\in X^c_{k_B^1}} x^c_\beta\Big)\ldots \\
&\bigcap_{k_B^t\in B^t}\Big(\bigcap_{x_\beta\in X_{k_B^t}} x_\beta\bigcap_{x_\beta\in X^c_{k_B^t}} x^c_\beta\Big), \\
&\mathcal{Y}_{k_A^1}\subseteq\mathcal{Y}^i,\ \mathfrak{X}_{k_B^1}\subseteq\mathfrak{X}^i_\beta,\ S^t\subseteq S.
\end{aligned}
\tag{23}
$$

В данной формуле для простоты записи для подмножества S^t указано подмножество $\{s^1,\ldots,s^t\}$ со множествами индексов для числовой доли $A^1,\ldots,A^t$ и множествами индексов для множественной доли $B^1,\ldots,B^t$, а также с соответствующими им индексами $k_A^1,\ldots,k_A^t$ и $k_B^1,\ldots,k_B^t$. 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 помножестве $S^t\subseteq S$, с мощностью равной t (т.е. $|S^t|=t$).

Приведем еще раз список прикладных задач, приводящих к работе со множеством двудольных множеств событий.

7.4 Основные задачи применения множества двудольных множеств событий

К числу задач, при решении которых используется рассмотренное в работе понятие множества двудольных множеств событий относятся задачи:

- определение наилучшего или наихудшего элемента системы,
- классификация элементов или групп,
- ранжирование элементов системы,
- факторный анализ,
- кластерный анализ,
- построение карт визаулизации данных,
- поиск ассоциативных правил (отыскание закономерностей между множествами событий),
- ковариационный анализ разнотипных данных (выявление связи между разнотипными данными),
- регрессионный анализ разнотипных данных (построение функции зависимости между данными),
- прогнозирование поведения системы,
- принятие решений при нечетких данных,
- банковский (страховой) скоринг,
- определение информативности признаков (с помощью энтропии системы).

Заключение

В работе рассматривалось понятие множества двудольных множеств событий, каждое из которых представляет собой объединение множеств событий, первая доля которых определяется случайными величинами, а вторая — случайными множествами. Приведены понятия классической и эвентологической ковариации. Рассмотрены виды распределений множества двудольных множеств. Изучены зависимости между двудольными множествами событий с помощью ковариации.

Список литературы

[1] О. Ю. Воробьев и И. В. Баранова. *Метод двудольных множеств событий в эвентологическом анализе сложных систем.* Ин-т естеств. и гуманит. наук, Красноярск, 2007.

[2] И.В. Баранова и Ю.А. Коновалова. Ковариация двудольных множеств случайных событий. In *Труды IX международной конференции по финансово-актуарной математике и эвентоконвергенции*, pages 163–167. СФУ, Красноярск, 2010.

[3] И. В. Баранова. Метод двудольных множеств событий в эвентологическом анализе социально-экономических систем. *Вестник Красноярского государственного университета: физико–математические науки*, 1:142–152, 2006.

[4] И. В. Баранова. Ковариация и регрессия двудольных множеств событий. In *Труды VIII Международной конференции по финансово-актуарной математике и смежным вопросам*, pages 62–67. СФУ, Красноярск, 2009.

[5] О. Ю. Воробьев. *Эвентология.* СФУ, Красноярск, 2007.

[6] 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 v. 1, pages 259–262. Azerbaijan National Academy of Sciences, Baku, Azerbaijan, 2008.

Применение генетических алгоритмов для решения задачи кластеризации данных

Ирина Владимировна Баранова
Сибирский Федеральный Университет
Институт математики и фундаментальной информатики
Красноярск
irinabar@yandex.ru

Дмитрий Михайлович Богодухов
Сибирский Федеральный Университет
Институт математики и фундаментальной информатики
Красноярск
d2991a@mail.ru

Аннотация. *Целью работы является разработка генетических алгоритмов, позволяющих находить решение задачи кластеризации данных с заданным числом кластеров. Приводятся основные понятия и постановка задачи кластеризации данных. Рассматриваются основные методы кластеризации данных. В работе дается описание одного из наиболее популярных методов кластеризации данных: метода k-средних. Излагаются принципы работы генетических алгоритмов. Приводится подробное описание разрабатываемого генетического алгоритма, решающего описанную задачу кластеризации. Проведено сравнение генетического алгоритма и метода k-средних.*

Ключевые слова. *Кластер, данные, кластеризация, метрика, функционал, генетический алгоритм*

1 Введение

Кластеризация — это задача разбиения множества объектов на сравнительно однородные группы, называемые кластерами [1], [2]. Однородность кластеров означает, что объекты, отнесенные к одному кластеру, должны быть близки относительно выбранной метрики. Объекты из разных кластеров должны существенно отличаться.

Когда набор данных кластеризован, каждая точка включается в определенный кластер и может быть охарактеризована одной описательной точкой, как правило, средней точкой в кластере.

Основная цель кластеризации данных — уменьшение размера и сложности данных. Сокращение данных осуществляется заменой координаты каждой точки в кластере координатами описательной точки кластера. Кластеризованные данные требуют значительно меньше ресурсов памяти и ими можно манипулировать гораздо быстрее, чем исходными данными. Значение того или иного метода кластеризации будет зависеть от того, насколько точно точки представляют собой данные, а также, насколько быстро работает программа.

Спектр применений кластерного анализа очень широк: его используют в медицине, психологии, химии, биологии, социологии, экономике и других дисциплинах. В кластерном анализе существуют разные виды задач, мы будем рассматривать задачу кластерного анализа типа «объект – признак» с заданным числом кластеров. Приведем формальную постановку задачи кластеризации данных.

2 Постановка задачи

Пусть $X = \{x_1, x_2, \ldots, x_m\}$ — множество объектов, заданных значениями в пространстве признаков $P = \{P_1, P_2, \ldots, P_n\}$ (т.е. каждый объект $x_i = (x_i^1, x_i^2, \ldots, x_i^n)$, $i = 1, \ldots, m$) и задана функция расстояния (метрика) между объектами $\rho(x_i, x_j)$, x_i, $x_j \in X$.

Определение 1. *Функцией кластеризации называется функция $f : X \to Y$ которая любому объекту $x \in X$ ставит в однозначное соответствие номер $y \in Y = \{1, \ldots, k\}$, $k \leq m$.*

Определение 2. *Множество кластеров $C = \{C_1, C_2, \ldots, C_k\}$, $k \leq m$ представляет собой разбиение множества объектов X такое, что кластер $C_i = \{x \in X : f(x) = i,\}$ $C_i \cap C_j = \emptyset$. Причем для C справедливо следующее: если $x_i, x_j \in C_i$, то $\rho(x_i, x_j) \to min$. Если $x_i \in C_i$, $x_j \in C_j$, то $\rho(x_i, x_j) \to max$.*

Тогда постановку задачу кластеризации данных можно сформулировать следующим образом.
Требуется найти такую функцию кластеризации f^*, чтобы

$$Q(f^*, C, \rho) = min_f Q(f, C, \rho),$$

где $Q(f, C, \rho)$ — выбранный критерий качества кластеризации. В данной работе в качестве критерия кластеризации рассматривается следующий функционал:

$$Q = \sum_i \frac{1}{|C_i|} \sum_{x \in C_i} \rho(x, \mu_i)$$

где $\mu_i = \frac{1}{|C_i|} \sum\limits_{x \in C_i} x$— центр кластера C_i.

© 2014 И.В.Баранова и Д.М.Богодухов

Олег Воробьев (ред.), Труды XIII ФАМЭМС'2014, Красноярск: СФУ

В качестве метрики между объектами будем использовать евклидову метрику

$$\rho(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}.$$

Как уже было сказано выше, каждый объект описывается набором своих характеристик, называемых признаками. Признаки могут быть следующих типов:

- бинарный признак: $P_i = \{0,1\}$;
- номинальный (качественный) признак: P_i — конечное множество;
- порядковый признак: P_i — конечное упорядоченное множество;
- количественный признак: $P_i = \mathcal{R}$ — множество действительных чисел.

Самой распространенной ситуацией является кластеризация объектов, у которых все признаки являются количественными. В работе мы будем рассматривать именно такую ситуацию.

3 Методы кластеризации

Общепринятой классификации методов кластеризации не существует, но можно выделить ряд общепринятых групп подходов [2]:

1. Вероятностный подход: алгоритм K-средних (K-means), K-медиан (K-medians), EM-алгоритм, алгоритмы семейства FOREL, дискриминантный анализ;
2. Подходы на основе систем искусственного интеллекта: метод нечеткой кластеризации C-средних (C-means), самоорганизующиеся карты Кохонена (нейронные сети), генетический алгоритм;
3. Логический подход: деревья решений;
4. Теоретико-графовый подход: графовые алгоритмы кластеризации;
5. Иерархический подход: метод ближайшего соседа, метод Варда, алгоритма Ланса-Уильямса, методы взвешенного и невзвешенного попарного среднего и т.п.;
6. Другие методы:
   - Статистические алгоритмы кластеризации,
   - Ансамбль кластеризаторов,
   - Алгоритмы семейства KRAB,
   - Алгоритм, основанный на методе просеивания.

Выбор метода кластеризации зависит от количества данных и от того, есть ли необходимость работать одновременно с несколькими типами данных. Теоретико-графовый и иерархический подходы иногда объединяют под названием структурного или геометрического подхода, обладающего большей формализованностью понятия близости. Несмотря на начительные различия между перечисленными методами все они опираются на исходную «гипотезу компактности»: в пространстве объектов все близкие объекты должны относиться к одному кластеру, а все различные объекты соответственно должны находиться в различных кластерах.

3.1 Метод k–средних

Метод k-средних является самым популярным методом кластеризации данных. Целью метода является разделение m наблюдений (из пространства R^n) на k кластеров, при этом каждое наблюдение относится к тому кластеру, к центроиду которого оно ближе всего.

Определение 3. *Центроид — точка, представляющее собой центр масс точек кластера, т.е. покоординатные средние точек из кластеров:*

$$\mu_j = (\mu_{j1}, \mu_{j2}, \ldots, \mu_{jn}), \tag{1}$$

где

$$\mu_{jl} = \frac{\sum\limits_{x_j \in C_j} x_{jl}}{|C_j|},\ l = 1,\ldots,n,\ j = 1,\ldots,k.$$

Описание алгоритма:

Пусть имеется множество точек данных $X = \{x_1,\ldots,x_m\}$, где $x_i = (x_{i1}, x_{i2}, \ldots, x_{in}) \in \mathbb{R}^n$.

1) Задается количество кластеров k, и на первом шаге производится задание центроидов μ_j, $j = 1,\ldots,k$ — «центров масс» кластеров C_j, $j = 1,\ldots,k$. Каждому кластеру соответствует один центр. Выбор начальных центроидов может осуществляться следующим образом:

- выбор k-наблюдений для максимизации начального расстояния;
- случайный выбор k-наблюдений;
- выбор первых k-наблюдений.

Кластеры $C_j = \{\emptyset\}$, $j = 1,\ldots,k$.
2) Производится распределение объектов по кластерам. Точка x_i, $i = 1,\ldots,n$ относится к ближайшему

кластеру, т.е. $x_i \in C_{j*}$, где

$$\rho(x_i, \mu_{j*}) = \min_{j=1,\dots,k} \rho(x_i, \mu_j).$$

В результате каждый объект назначен определенному кластеру.

3) Вычисляются новые центры кластеров μ_j, $j = 1, \dots, k$ как центры масс новых кластеров C_j, $j = 1, \dots, k$, полученных на предыдущем этапе.

4)Продолжать итерационный процесс вычисления центров и перераспределения объектов до тех пор, пока не выполнится одно из условий:

- кластерные центры μ_j стабилизировались (перестали изменяться);
- число итераций равно максимальному числу итераций (ограничение на число итераций).

4 Введение в генетические алгоритмы

Генетические алгоритмы (ГА) — это адаптивные методы поиска, в которых используются аналоги, как механизма генетического наследования, так и аналоги естественного отбора [3], [4]. Основной идеей генетических алгоритмов является организация «борьбы за существование» и «естественного отбора» среди этих пробных решений. Введем основные понятия, применяемые в генетических алгоритмах.

Определение 4. *Вектор — упорядоченный набор чисел, называемых компонентами вектора. Так как вектор можно представить в виде строки его координат, то в дальнейшем понятия вектора и строки считаются идентичными.*

Определение 5. *Хромосома — вектор (или строка) из каких-либо чисел. Если этот вектор представлен бинарной строкой из нулей и единиц, например,* 1010011*, то он получен либо с использованием двоичного кодирования, либо кода Грея. Каждая позиция (бит) хромосомы называется геном.*

Определение 6. *Индивидуум (генетический код, особь) — набор хромосом (вариант решения задачи). Обычно особь состоит из одной хромосомы, поэтому в дальнейшем особь и хромосома идентичные понятия.*

Определение 7. *Популяция — совокупность индивидуумов.*

Определение 8. *Функция приспособленности — функция, позволяющая оценить степень приспособленности конкретных особей в популяции и выбрать из них наиболее приспособленные (т.е. имеющие наибольшие значения функции приспособленности) в соответствии с эволюционным принципом выживания ксильнейшихъ (лучше всего приспособившихся).*

Терминология ГА представляет собой синтез генетических и искусственных понятий [3].

Рассмотрим популярные генетические операторы:

1. Оператор селекции — отбирает популяцию для дальнейшего размножения. Выбор производится согласно принципу естественного отбора, по которому наибольшие шансы на участие в создании новых особей имеют хромосомы с наибольшими значениями функции приспособленности.
2. Оператор скрещивания — его целью служит, распространение хороших генов по популяции. При скрещивании отобранные индивиды (родители) по заданному правилу передают части своих хромосом. Потомок может унаследовать только те гены, которые есть у его родителей.
3. Оператор мутации изменяет значение гена в хромосоме на любое другое возможное значение. Мутация рассматривается как метод восстановления потерянного генетического материала, а не как поиск лучшего решения. Обычно мутация применяется к генам с очень низкой вероятностью $p_m \in [0.0001, 0.01]$.

Однако следует отметить, что оператор мутации играет явно второстепенную роль по сравнению с оператором скрещивания. Это означает, что скрещивание в классическом генетическом алгоритме производится практически всегда, тогда как мутация — достаточно редко. Вероятность скрещивания, как правило, достаточно велика (обычно $0.5 \leq p_c \leq 1.0$), тогда как вероятность мутации устанавливается весьма малой (чаще всего $p_m \in [0.0001, 0.01]$). Это следует из аналогии с миром живых организмов, где мутации происходят чрезвычайно редко [4].

5 Генетический алгоритм для решения задачи кластеризации

В работе предлагается генетический алгоритм для решения задачи кластеризации данных с детерминированным числом кластеров. Алгоритм основан на идее объединения в кластер объектов в областях их наибольшего сгущения.

Рассмотрим подробно основные этапы работы генетического алгоритма:

1. Инициализация.
   Генерируем начальную популяцию. Для каждой точки случайным образом устанавливаем номер кластера. В качестве исходного решения берется хромосома (вектор) $v^j = (v_1^j, v_2^j, \dots, v_m^j)$, m —

число точек, $v_i^j \in [1,k]$ — номер кластера C_j, в который попадает точка x_i. Популяцию заполняется n хромосомами.

2. Вычисление коэффициента пригодности.
Оценка приспособленности особей в популяции заключается в вычислении значения функции пригодности для каждого члена популяции. Чем выше это значение, тем больше особь отвечает требованиям решаемой задачи. В качестве коэффициент пригодности используется функционал суммы средних внутрикластерных расстояний:

$$Q = \sum_i \frac{1}{|C_i|} \sum_{x \in C_i} \rho(x, \mu_i)$$

где $\mu_i = \frac{1}{|C_i|} \sum\limits_{x \in C_i} x$— центр кластера C_i, а $\rho(v_j^i, \mu_i)$ — евклидово расстояние от точки v_j^i до центра, к которому она принадлежит.

3. Селекция.
Отобрать пару хромосом-родителей v^i и v^j , у которых значение выше среднего по популяции.

4. Скрещивание.
Проводится скрещивание двух родителей с вероятностью p_c, производя потомка, т.е. создается новое решение на основе рекомбинации из существующих. Целью рекомбинации является получение такой комбинации промежутков хромосом, при которой особь будет представлять собой наилучшее из решений, возможное при текущем генетическом материале. Подобный механизм решения задач кластеризации данных, в отличие от существующих методов, не осуществляет замену одного решения на другое, а получает новые возможные решения посредством обмена информацией между ними. Выбираются хромосомы-родители $v^i = (v_1^i, v_2^i, \ldots, v_m^i)$ и $v^j = (v_1^j, v_2^j, \ldots, v_m^j)$, выбирается случайная точка разрыва t, и хромосомы обмениваются своими частями, в результате получаем два потомка: $v^{i*} = (v_1^i, v_2^i, \ldots, v_t^i, \Updownarrow v_{t+1}^j, v_{t+2}^j, \ldots, v_m^j)$ и $v^{j*} = (v_1^j, v_2^j, \ldots, v_t^j, \Updownarrow v_{t+1}^i, v_{t+2}^i, \ldots, v_m^i)$.

5. Мутация.
Производится мутация потомков с вероятностью p_m. В рамках решения задачи кластеризации данных наибольшее значение имеют генные мутации, которые в большинстве случаев затрагивают один или несколько генов. В рассматриваемом генетическом алгоритме гены содержат в себе целые числа от 1 до k, где k — число кластеров. Коэффициент мутации p_m определяет долю генов, подвергнутых мутации на текущей итерации, в расчете на их общее количество. Если коэффициент мутации слишком мал, то получится ситуация, при которой множество полезных генов просто не будут существовать в популяции. В то же время использование большого значения коэффициента мутации приведет к множеству случайных возмущений и значительно увеличит время поиска. Потомки перестанут походить на родителей, алгоритм больше не будет иметь возможности обучаться на основе сохранения наследственных признаков. К выбранным генам применяют преобразование, которое вызывает изменение значения текущего гена на любое другое возможное.

6. Генерация нового поколения.
Важнейшим этапом разработанного генетического алгоритма является формирование новой популяции. Насколько сильно выросла численность популяции, определяется общим коэффициентом мутации p_m и коэффициентом скрещивания p_c. Этап формирования новой популяции предназначен для того, чтобы вернуть численность популяции к ее исходному значению. Значение функций пригодности хромосом–родителей, которые уже находились в популяции, известно. Алгоритм осуществляет оценку пригодности хромосом-потомков, полученных в результате операции скрещивания, а также оценку особей, которые были получены в результате работы оператора мутации. На основе полученных значений функции пригодности для каждой особи новой популяции производится удаление тех особей, у которых значения функции пригодности наименьшие. Алгоритм реализует данную задачу путем последовательного удаления особи с минимальным значением функции пригодности до тех пор, пока численность популяции не вернется к начальному значению.

7. Выбор наилучшей хромосомы.
На данном этапе выполняется поиск наилучшего решения — той особи, у которой максимальное значение функции пригодности. Эта операция выполняется после того, как численность популяции установилась до исходной величины.
После выбора наилучшей особи алгоритм передает управление процессом следующему этапу — проверке остановки алгоритма.

8. Пункты 3-7 повторять до тех пор, пока не будет достигнут критерий остановки алгоритма.

Определение критерия остановки генетического алгоритма напрямую зависит от специфики решаемой задачи и имеющихся сведений об объекте поиска [3]. В рамках решения задачи кластеризации данных, для которой не известно оптимальное значение функции пригодности, остановку алгоритма можно организовать следующим образом:

1. При отсутствии изменения функции пригодности наилучшей особи в течение определенного

числа итераций, которое задается в качестве параметра генетического алгоритма (т.е. полученная наилучшая особь совпадающей с наилучшей особью прошлого поколения).

2. Кроме этого, алгоритм предусматривает остановку своей работы по истечении определенного числа итераций, которое также задается в качестве параметра.

В том случае, если условие остановки выполнено, то алгоритм выдает в качестве оптимального решения то, которое представлено наилучшей особью, определенной на этапе формирования новой популяции. Если же условие не выполнено, то алгоритм передает управление этапу селекции.

6 Сравнение предложенного алгоритма с методом k-средних

В работе было проведено сравнение предложенного алгоритма с методом k-средних. Работа алгоритмов сравнивалась на ряде тестовых примеров. Было выявлено следующее:

Алгоритм k-средних крайне чувствителен к выбору начальных приближений центров, а также к выбросам, которые могут искажать среднее. Недостатком алгоритма также является необходимость задавать число кластеров, исследователь может при этом исходить из некоторых априорных знаний или соображений. Большинство популярных алгоритмов кластеризации требуют введения дополнительных параметров, которые затем изменяется в ту или иную сторону. Таким образом, получается хорошее разбиение, но не всегда — самое оптимальное. Главным достоинством генетических алгоритмов в данном применении является то, что они ищут глобальное оптимальное решение. Операторы скрещивания и мутации выбивают популяцию из локального экстремума и позволяют получить решения, не похожие на исходные.

Предложенный генетический алгоритм, эффективно функционирует при обработке массивов большой размерности, поскольку в нем оптимально сочетаются целенаправленный поиск и элементы случайности, направленные на выбивание целевой функции из локальных минимумов. Никаких предварительных условий для его использования не требуется.

7 Заключение

В данной работе приведено описание разработанного генетического алгоритма для решения задачи кластеризации данных. Алгоритм имеет ряд существенных достоинств, а именно:

1. Разработанный генетический алгоритм не требует задания дополнительных параметров, которые необходимы для решения задачи кластеризации с помощью большинства популярных алгоритмов кластеризации.

2. Главным достоинством генетических алгоритмов является то, что они ищут глобальное оптимальное решение.

3. Предложенный генетический алгоритм эффективно функционирует при обработке массивов большой размерности, поскольку в нем оптимально сочетаются целенаправленный поиск и элементы случайности, такие как оператор скрещивания и мутации, направленные на выбивание популяции из локальных экстремумов и позволяющие получить решения, непохожие на исходные.

Все эти преимущества свидетельствуют о необходимости решении задачи кластеризации данных предложенным методом.

Список литературы

[1] Н. Г. Загоруйко. *Прикладные методы анализа данных и знаний.* Новосибирск, ИМ СО РАН, 1999.

[2] Б. Г. Миркин. *Методы кластер-анализа для поддержки принятия решений: обзор.* Москва, Изд. дом НИУ Высшая школа экономики, 2011.

[3] Д.И. Батищев. *Генетические алгоритмы решения экстремальных задач.* Нижний Новгород, НГУ, 1995.

[4] Т.В. Панченко. *Генетические алгоритмы.* Астрахань, Астраханский университет, 2007.

Методы кластеризации многомерных статистических данных

Ирина Владимировна Баранова
Сибирский Федеральный Университет
Институт математики и фундаментальной информатики
Красноярск
irinabar@yandex.ru

Димитрия Владимировна Князь
Сибирский Федеральный Университет
Институт математики и фундаментальной информатики
Красноярск
dimitriya.knyaz@bk.ru

Аннотация. *В статье рассматриваются основные понятия, алгоритмы и методы кластеризации. Целью кластерного анализа является выбор необходимых средств и методов для обработки и анализа данных. Применительно к конкретному исследовательскому проекту выбираются те статистические процедуры, которые наиболее подходят для решения поставленных задач применительно к изучаемому вопросу.*

Ключевые слова. *Кластерный анализ, объект, кластеризация, типология, кластер, метод k-средних, кратчайший незамкнутый путь, алгоритм Ланса–Уильямса.*

1 Введение

Кластерный анализ (cluster analysis) [3] — многомерная статистическая процедура, выполняющая сбор данных, содержащих информацию о выборке объектов, и затем упорядочивающая объекты в сравнительно однородные группы. Задача кластеризации относится к статистической обработке, а также к широкому классу задач обучения без учителя [1]. Термин кластеризация впервые ввел Трион в 1939 году.

Под кластером обычно понимается часть данных (в типичном случае — подмножество объектов или переменных), которая выделяется из остальной части наличием некоторой однородности ее элементов. В простейшем случае речь идет о похожести элементов, в идеальном случае — о совпадающих значениях основных переменных или иного рода близости, выражаемой геометрической близостью соответствующих объектов.

2 Постановка задачи кластеризации

Пусть $X = \{x^1, \ldots, x^m\}$ — множество объектов, Y — множество номеров кластеров. Задана функция расстояния между объектами (метрика) $\rho(x, x')$.
Требуется разбить объекты на непересекающиеся подмножества, называемые кластерами, так, чтобы каждый кластер состоял из объектов, близких по метрике ρ, а объекты разных кластеров существенно отличались.

При этом каждому объекту $x_i \in X^m$ приписывается номер кластера Y_i.

Алгоритм кластеризации — это функция $a : X \to Y$, которая любому объекту $x \in X$ ставит в соответствие номер кластера $y \in Y$. Множество Y в некоторых случаях известно заранее, однако чаще ставится задача определить оптимальное число кластеров, с точки зрения того или иного критерия качества кластеризации.

Кластеризация (обучение без учителя) отличается от классификации (обучения с учителем) тем, что метки исходных объектов Y_i изначально не заданы, и даже может быть неизвестно само множество Y.

Кластерный анализ предъявляет следующие требования к данным:

1. показатели не должны коррелировать между собой;
2. показатели должны быть безразмерными;
3. распределение показателей должно быть близко к нормальному;
4. показатели должны отвечать требованию «устойчивости», под которой понимается отсутствие влияния на их значения случайных факторов;
5. выборка должна быть однородна.

Для вычисления расстояния между объектами используются различные меры сходства (меры подобия), называемые также метриками или функциями расстояний. Приведем наиболее популярные метрики:

- Наиболее популярной является *евклидова метрика*. Евклидова метрика между точками x и y — это длина отрезка $\overline{xy}$. В декартовых координатах, если $x = (x_1, x_2, \ldots, x_n)$ и $y = (y_1, y_2, \ldots, y_n)$ — две точки в евклидовом пространстве, евклидова метрика имеет вид:

$$\rho(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}. \qquad (1)$$

© 2014 И.В. Баранова и Д.В. Князь

Олег Воробьев (ред.), Труды XIII ФАМЭМС'2014, Красноярск: СФУ

Для придания большего значения более отдаленным друг от друга объектам, можно использовать *квадрат евклидова расстояния*:

$$\rho(x,y) = \sum_{i=1}^{n}(x_i - y_i)^2. \quad (2)$$

- *Хеммингово расстояние.* Оно также называется манхэттенским или расстоянием городских кварталов. Это расстояние является разностью по координатам. В большинстве случаев эта мера расстояния приводит к таким же результатам, как и для обычного расстояния Евклида. Однако стоит отметить, что для этой меры влияние отдельных больших разностей (выбросов) уменьшается (так как они не возводятся в квадрат). Хеммингово расстояние вычисляется по формуле:

$$\rho(x,y) = \sum_{i=1}^{n}(|x_i| - |y_i|). \quad (3)$$

- *Расстояние Чебышева.* Принимает значение наибольшего модуля разности между значениями соответствующих свойств (признаков) объектов:

$$\rho(x,y) = \max_{i=1,\ldots,n}|x_i - y_i|. \quad (4)$$

Это расстояние стоит использовать, когда необходимо определить два объекта как «различные», если они отличаются по какому-то одному измерению.

3 Методы кластеризации

Общепринятой классификации методов кластеризации не существует, но можно выделить ряд общепринятых групп подходов [2]:

1. Вероятностный подход: алгоритм K-средних (K-means), K-медиан (K-medians), EM-алгоритм, алгоритмы семейства FOREL, дискриминантный анализ;
2. Подходы на основе систем искусственного интеллекта: метод нечеткой кластеризации C-средних (C-means), самоорганизующиеся карты Кохонена (нейронные сети), генетический алгоритм;
3. Логический подход: деревья решений;
4. Теоретико-графовый подход: графовые алгоритмы кластеризации;
5. Иерархический подход: метод ближайшего соседа, метод Варда, алгоритма Ланса-Уильямса, методы взвешенного и невзвешенного попарного среднего и т.п.;
6. Другие методы:
   - Статистические алгоритмы кластеризации,
   - Ансамбль кластеризаторов,
   - Алгоритмы семейства KRAB,
   - Алгоритм, основанный на методе просеивания.

Выбор метода кластеризации зависит от количества данных и от того, есть ли необходимость работать одновременно с несколькими типами данных. Теоретико-графовый и иерархический подходы иногда объединяют под названием структурного или геометрического подхода, обладающего большей формализованностью понятия близости. Несмотря на начительные различия между перечисленными методами все они опираются на исходную «гипотезу компактности»: в пространстве объектов все близкие объекты должны относиться к одному кластеру, а все различные объекты соответственно должны находиться в различных кластерах.

В основе теоретико-графовых методов лежит принцип представления данных в виде вершин графа, а расстояние между ними имеет смысл ребер графа. Для последующей кластеризации используется сравнительно широкий спектр методов, разработанный для анализа графов. Типичными представителями графовых алгоритмов кластеризации являются: алгоритм кратчайшего незамкнутого пути, выделения связных компонент, ФОРЭЛ и другие.

Суть иерархической кластеризации состоит в последовательном объединении меньших кластеров в большие или разделение больших кластеров на меньшие. К иерархическим методам относятся метод ближайшего соседа (или одиночная связь), метод наиболее удаленных соседей (или полная связь), метод Варда, метод невзвешенного попарного среднего, метод взвешенного попарного среднего, невзвешенный центроидный метод,взвешенный центроидный метод и другие. Сложности иерархических методов кластеризации: ограничение объема набора данных; выбор меры близости; негибкость полученных классификаций. Преимущество этой группы методов — их наглядность и возможность получить детальное представление о структуре данных.

При большом количестве наблюдений иерархические методы кластерного анализа не пригодны. В таких случаях используют неиерархические методы, основанные на разделении, которые представляю собой итеративные методы дробления исходной совокупности. В процессе деления новые кластеры формируются до тех пор, пока не будет выполнено правило остановки. Неиерархические методы выявляют более высокую устойчивость по отношению к шумам и выбросам, некорректному выбору метрики, включению незначимых переменных в набор, участвующий в кластеризации.

В данной работе рассматриваются три наиболее популярных алгоритма кластеризации данных: метод К-средних, кратчайшего незамкнутого пути и Ланса-Уильямса.

3.1 Метод k-средних

Алгоритм k-средних является самым популярным методом, относящимся к неиерархическому подходу. Также этот метод называют быстрым кластерным анализом. В отличие от иерархических методов, которые не требуют предварительных предположений относительно числа кластеров, для возможности использования этого метода необходимо иметь гипотезу о наиболее вероятном количестве кластеров.

Число «k» в названии метода означает количество кластеров, на которое производится разбиение данных. Выбор числа K может базироваться на результатах предшествующих исследований, теоретических соображениях или интуиции.

Слово «средние» в названии метода относится к центроидам кластеров.

Определение 1. *Центроид — точка, представляющее собой центр масс точек кластера, т.е. покоординатные средние точек из кластеров:*

$$\mu_j = (\mu_{j1}, \mu_{j2}, \ldots, \mu_{jn}), \tag{5}$$

где

$$\mu_{jl} = \frac{\sum\limits_{x_j \in S_j} x_{jl}}{|S_j|},\ l = 1, \ldots, n,\ j = 1, \ldots, k.$$

Описание алгоритма:

Пусть имеется множество точек данных $X = \{x_1, \ldots, x_m\}$, где $x_i = (x_{i1}, x_{i2}, \ldots, x_{in}) \in \mathbb{R}^n$.

1) Задается количество кластеров k, и на первом шаге производится задание центроидов μ_j, $j = 1, \ldots, k$ — «центров масс» кластеров S_j, $j = 1, \ldots, k$. Каждому кластеру соответствует один центр. Выбор начальных центроидов может осуществляться следующим образом:

- выбор k-наблюдений для максимизации начального расстояния;
- случайный выбор k-наблюдений;
- выбор первых k-наблюдений.

Кластеры $S_j = \{\emptyset\}$, $j = 1, \ldots, k$.
2) Производится распределение объектов по кластерам. Точка x_i, $i = 1, \ldots, n$ относится к ближайшему кластеру, т.е. $x_i \in S_{j*}$, где

$$\rho(x_i, \mu_{j*}) = \min_{j=1,\ldots,k} \rho(x_i, \mu_j).$$

В качестве метрики используется одна из приведенных выше метрик, чаще всего евклидова. В результате каждый объект назначен определенному кластеру.

3) Вычисляются новые центры кластеров μ_j, $j = 1, \ldots, k$ как центры масс новых кластеров S_j, $j = 1, \ldots, k$, полученных на предыдущем этапе.

4)Продолжать итерационный процесс вычисления центров и перераспределения объектов до тех пор, пока не выполнится одно из условий:

- кластерные центры μ_j стабилизировались (перестали изменяться);
- число итераций равно максимальному числу итераций (ограничение на число итераций).

Выбор числа кластеров является сложным вопросом. Если нет предположений относительно этого числа, рекомендуют создать 2 кластера, затем 3, 4, 5 и т.д., сравнивая полученные результаты.

3.2 Алгоритм КНП — «Кратчайший Незамкнутый Путь»

Имеется множество точек данных $X = \{x_1, \ldots, x_m\}$. Задается количество кластеров k. Представляем данные в виде вершин графа, расстояние между двумя точками данных имеет смысл ребер графа.

Описание алгоритма:

1. Найти пару вершин x_i, x_j с наименьшим расстоянием $\rho(x_i, x_j)$ между вершинами, и соединить их ребром;
2. Пока в выборке остаются изолированные точки (не соединенные ребрами с другими вершинами) выполнять:
   - Найти изолированную точку, ближайшую к некоторой неизолированной (в смысле метрики ρ);
   - Соединить эти две точки ребром;
3. Удалить $k - 1$ самых длинных ребер.

3.3 Алгоритм Ланса-Уильямса

Имеется множество точек данных $X = \{x_1, \ldots, x_m\}$. В алгоритме обозначение C_t используется для обозначения множества кластеров, полученных на t-ой итерации алгоритма, под R_t понимается расстояние между кластерами на t-ой итерации.

Описание алгоритма:

1. Сначала все кластеры одноэлементные: $t = 1$, $C_t = \{\{x_1\}, \ldots, \{x_m\}\}$

2. Для всех t=2,...,l (t — номер итерации);
3. Найти в C_{t-1} два ближайших кластера $U, V \in C_{t-1}$:
$$(U, V) = \arg\min_{U \neq V} R(U, V).$$
4. Слить их в один кластер:
5. Для всех вычислить $R(W, S)$ по формуле Ланса–Уильямса.

Формула Ланса-Уильямса имеет следующий вид:

$$R(U \cup V, S) = \alpha_U(U, S) + \alpha_V R(V, S) + \\ + \beta R(U, V) + \gamma |R(U, S) - R(V, S)|, \quad (6)$$

где α_U, α_V, β и γ — числовые параметры.

Приведем наиболее употребляемые частные случаи формулы Ланса-Уильямса:

1. Расстояние ближнего соседа:
$$R(W, S) = \min_{w \in W,\, s \in S} \rho(w, s),$$
т.е. в формуле (6) $\alpha_U = \alpha_V = \frac{1}{2}$, $\beta = 0$ и $\gamma = -\frac{1}{2}$.
2. Расстояние дальнего соседа:
$$R(W, S) = \max_{w \in W,\, s \in S} \rho(w, s),$$
в формуле (6) $\alpha_U = \alpha_V = \frac{1}{2}$, $\beta = 0$ и $\gamma = \frac{1}{2}$.
3. Групповое среднее расстояние:
$$R(W, S) = \frac{1}{|W||S|} \sum_{w \in W} \sum_{s \in S} \rho(w, s),$$
в формуле (6) $\alpha_U = \frac{|U|}{|W|}$, $\alpha_V = \frac{|V|}{|W|}$, $\beta = \gamma = 0$.
4. Расстояние между центрами:
$$R(W, S) = \rho^2 \left(\sum_{w \in W} \frac{|w|}{|W|} \sum_{s \in S} \frac{|s|}{|S|} \right),$$
в формуле (6) $\alpha_U = \frac{|U|}{|W|}$, $\alpha_V = \frac{|V|}{|W|}$, $\beta = -\alpha_U \alpha_V$, $\gamma = 0$.

4 Практическая задача кластеризации

В работе решается практическая задача кластеризации ведущих российских банков по основным показателям их финансовой стабильности. Для решения задачи использовалась официальная статистика отчетности кредитных организаций РФ по показателям их деятельности, публикуемая на сайте Банка России. В статистике рассматриваются следующие основные показатели деятельности банков: активы нетто, чистая прибыль, кредитный портфель, просроченная задолженность в кредитном портфеле, вклады физических лиц, вложение в ценные бумаги, средства предприятий и организаций, привлеченные межбанковские кредиты, векселя и облигации и др.

Кластеризация данных проводилась с помощью трёх методов: метода K-средних (вероятностный подход); с использованием алгоритма КНП — «Кратчайший незамкнутый путь» (теоретико-графовый подход), и алгоритма Ланса-Уильямса (иерархический подход). Для алгоритма Ланса–Уильямса рассматривались частные случаи формулы Ланса-Уильямса, такие как: а) расстояние ближнего соседа; б) групповое среднее расстояние; в) расстояние между центрами.

5 Заключение

В работе на основе статистических данных решена практическая задача кластеризации ведущих российских банков по основным показателям финансовой стабильности. Задача решена при помощи кластерного анализа. Определена принадлежность каждого объекта к определенному кластеру. Проведено сравнение кластеров, полученных в результате работы каждого метода.

Список литературы

[1] Н. Г. Загоруйко. *Прикладные методы анализа данных и знаний.* Новосибирск, ИМ СО РАН, 1999.

[2] Б. Г. Миркин. *Методы кластер-анализа для поддержки принятия решений: обзор.* Москва, Изд. дом НИУ Высшая школа экономики, 2011.

[3] И.Д. Мандель. *Кластерный анализ.* Москва, Финансы и статистика, 1988.

Вне привычных измерений статистических взаимосвязей: эвентологически пересмотренный подход к измерению общности и ассоциации (статистической зависимости) множеств событий

Олег Юрьевич Воробьев
Институт математики и фундаментальной информатики
Сибирский федеральный университет
Красноярск
email: oleg.yu.vorobyev@gmail.com
url: sfu-kras.academia.edu/OlegVorobyev

Аннотация. *Автор развил общую теорию измерений статистической взаимосвязи между событиями, называемую "математическая эвентология" [3, 2007], [4, 2011]. Эта работа представляет свежие "эвентологические" идеи, которые кардинально пересматривают основы прежнего подхода к измерению статистических взаимосвязей событий и выходят за пределы традиционных измерений в теории вероятностей и математической статистике, добавляя к уже известным методам измерения ассоциации (association), или статистической зависимости (statistical dependency), еще одну новую группу методов измерения общности (agreement) множеств событий. Показано, что при помощи ряда полезных измерителей взаимосвязи множества событий, названных* ***общность***, ***ассоциация***, ***ковариация***, ***сет-расстояние***, ***сет-близость*** *и* ***сет-кучность***, *формулы в этом контексте могут быть упрощены и элегантно обобщены для многих приложений, в том числе, для корректного обобщения таких традиционных показателей статистической взаимосвязи событий, как корреляция Пирсона, ассоциация Юла, а также популярного показателя общности — каппа Коэна [9, 1960]. Предлагаемые новые методы демонстрируют растущий потенциал "эвентологических" технологий в исследовании структур взаимосвязи (общности, ассоциации, или статистической зависимости) событий, который распространяется на измерения как в номинальных (категориальных), так и в упорядоченных и числовых (разностей, отношений) шкалах.*

Ключевые слова. *Категориальный анализ, номинальная шкала, множество событий, ассоциация, общность, разобщенность, статистическая зависимость, статистическая независимость, многомерные сопряженные таблицы, многие эксперты, многие категории, ковариация, сет-расстояние, меры общности, сет-близость, сет-кучность.*

© 2014 О.Ю.Воробьев

Олег Воробьев (ред.), Труды XIII ФАМЭМС'2014, Красноярск: СФУ

Понятие *сет-расстояния множества событий* (м.с.) — показателя разобщенности его событий в пространстве элементарных исходов — было введено в эвентологическую теорию [1, 20, 2004] как развитие известного понятия *вероятностного расстояния* дуплета событий — вероятности симметрической разности двух событий — для произвольного м.с. Это продвижение сразу позволило, например, естественным образом распространить *принцип метрической интерпретации* парной ковариации [2, 10, 2002] на ковариацию произвольного м.с. [3, 17, 2007], и оказалось довольно удачным, чтобы сформулировать в [3, 17, 2007] общее определение *сет-расстояния м.с.*, дополненное доказательством для него неравенства сет-треугольника в [4, 18]. Дальнейшее развитие этого понятия в виде ***д***истограммной теории измерения разобщенности м.с. в пространстве элементарных исходов дано в [5, 19].

Разработка и использование в разнообразных приложениях нового понятия сет–расстояния м.с. рано или поздно должны были так или иначе соприкоснуться с обширной областью современной многомерной статистики, которая занимается измерением межэкспертной общности оценок в категориальных (номинальных) шкалах. Результаты этого драматического соприкосновения изложены ниже, где в качестве "мальчика для битья (whipping boy)" выбрана, пожалуй самая популярная группа таких статистик — каппа-показатели межэкспертной общности оценок в категориальных шкалах.

1 Каппа Коэна как измеритель межэкспертной общности в категориальной шкале

В более чем многочисленной литературе, посвященной каппа-показателям межэкспертной общности (interrater agreement), с легкой руки отца-основателя Джекоба Коэна [9, 1960] продолжает рекламироваться как основное достоинство каппа-показателей их так называемая шанс-откорректированность (chance-corrected), или шанс-отрегулированность (chance-

adjusted), которое традиционно интерпретируется как удаление из них “учета случайной общности экспертов”. Сам факт появления в середине прошлого века в работах по межэкспертной общности (interrater agreement) терминов “корректировка случая (chance-corrected)” и “регулировка случая (chance-adjusted)” заслуживает пристального внимания и подробного толкования. Разумеется, эта терминология может привести и приводит к недоразумениям и заблуждениям. Тем более интересно изучить причины ее возникновения и пути возможного оправдания в рамках эвентологического подхода.

В традиционных описаниях задачи оценки общности мнений экспертов в категориальной шкале и в интерпретации каппа-показателя общности есть два пункта[1], спорность которых не ускользает от внимания эвентолога:

1) в постановке задачи часто подчеркивается предположение, что в рамках эксперимента по оценке общности эксперты действуют *“независимо друг от друга”*;

2) в интерпретации каппа-показателя за редким исключением утверждается, что этот показатель *“освобожден от учета случайной общности”*.

Конечно, спорность для эвентолога этих двух пунктов во многом объясняется всего лишь неудачей в использовании терминологии, которая однако, мне кажется, скрывает фундаментальные расхождения в понимании вероятности в рамках колмогоровского и байесовского подходов.

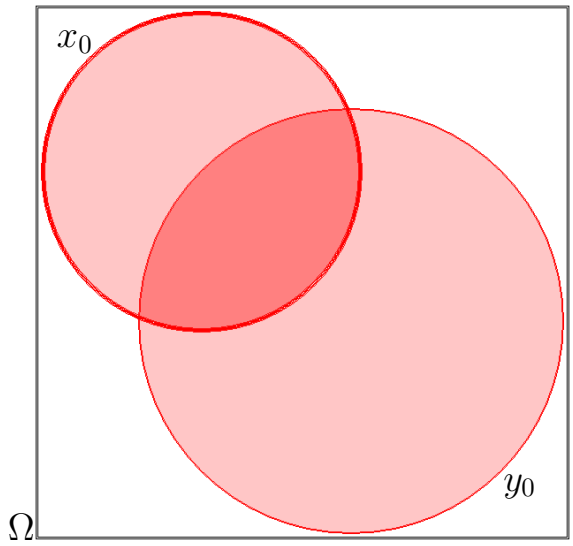


Рисунок 1: Диаграмма Венна дуплета *произвольно-зависимых* событий $\mathfrak{X}_\circ = \{x_0, y_0\}$ ($4 = 2^2$ террасных события). Событие x_0 обведено толстой линией. Пространство всеобщих элементарных событий Ω — бесцветный квадрат.

[1] Первые работы по измерению общности экспертов, оценивающих предъявляемые им события, начали появляться в середине прошлого века. Цитаты из них, иллюстрирующие эти два спорных пункта, см. в Приложении на стр. 91.

Явно неудачную терминологию: “корректировка случая”, по-видимому, породило желание сравнить поведения экспертов с *независимым* подбрасыванием смещенных (bias) монет. Для подробного изучения этого казуса уместно внимательно рассмотреть совместную диаграмму Венна трех m-множеств событий, характеризующих три возможных способа поведения m экспертов, в каждом из которых решающую роль играет *случай*:

- **первое (где источником случая служат са-**

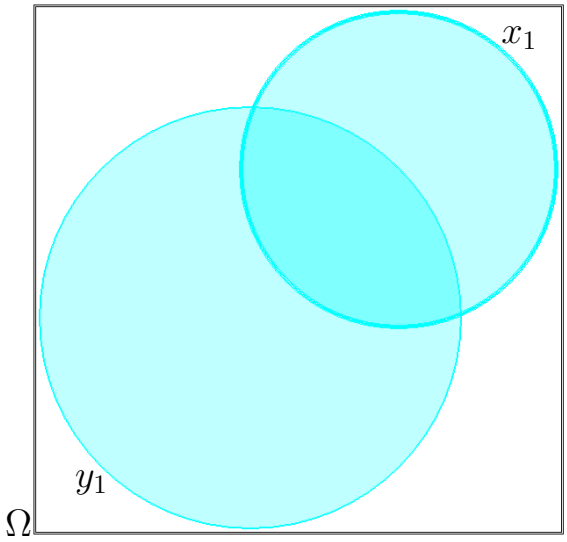


Рисунок 2: Диаграмма Венна дуплета *произвольно-зависимых* событий $\mathfrak{X}_1 = \{x_1, y_1\}$ ($4 = 2^2$ террасных события), имеющих то же вероятностное распределение, что и дуплет событий $\mathfrak{X}_0$; т.е. совпадают вероятности всех 4-х соответствующих террасных событий, порожденных дуплетами событий $\mathfrak{X}_0$ и $\mathfrak{X}_1$. Событие x_1 обведено толстой линией. Пространство всеобщих элементарных событий Ω — бесцветным квадрат.

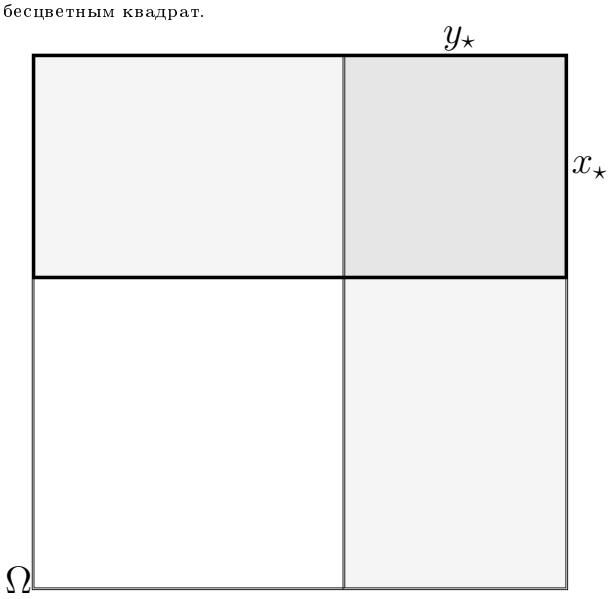


Рисунок 3: Диаграмма Венна дуплета *независимых* событий $\mathfrak{X}_\star = \{x_\star, y_\star\}$ ($4 = 2^2$ террасных события), происходящих с теми же вероятностями, что и события из $\mathfrak{X}_0$. Событие $x_\star$ — горизонтальный прямоугольник, обведенный толстой линией; событие $y_\star$ — вертикальный прямоугольник справа. Пространство всеобщих элементарных событий Ω — бесцветный квадрат.

ми *эксперты со своими мнениями*) характеризует экспертов, которые оценивают наступление предъявляемого им события "по совести" — на основе своих знаний, опыта и интуиции;

- **второе (где источником случая служит *независимо подбрасываемые смещенные монеты*)** характеризует экспертов, которые для оценки наступления предъявляемого им события прибегают к *независимому друг от друга* подбрасыванию смещенных (bias) монет, выпадающих наступлением предъявляемого события с вероятностью, равной вероятности их собственной оценки его наступления;
- и, наконец, **третье (где источником случая служит *зависимо подбрасываемые смещенные монеты*)** характеризует экспертов, которые для оценки наступления предъявляемого им события прибегают к *зависимому друг от друга* подбрасыванию смещенных (bias) монет, выпадающих наступлением предъявляемого события с вероятностью, равной вероятности их собственной оценки его наступления, причем структура зависимости третьего множества событий*совпадает со структурой зависимости первого множества событий, характеризующего оценки "по совести"*.

Диаграммы Венна этих m-м.с. при $m = 2$, показанные на рисунках 1, 2, 3, 4, позволяют дать следующие эвентологические комментарии к двум упомянутым спорным пунктам:

K1) чем выше квалификация экспертов, тем более их мнения об одном и том же событии *"**статистически зависимы** друг от друга"*; а якобы естественное предположение о ***независимом*** поведении экспертов — это застарелый казус классического заблуждения первых статистиков рубежа $XIX - XX$ в.в., которые пытались выводить причинно-следственную связь из статистической зависимости; забывая, что это всего лишь, например, наступление лета заставляет всех "сведущих в лете экспертов" почти разом, т.е. ***статистически зависимо***, перейти на летние одежды.

K2) *"освобождение от учета случайной общности"* — это иллюзия, за которой скрывается всего лишь измерение отклонения наблюдаемого поведения экспертов от ***статистически независимого*** поведения.

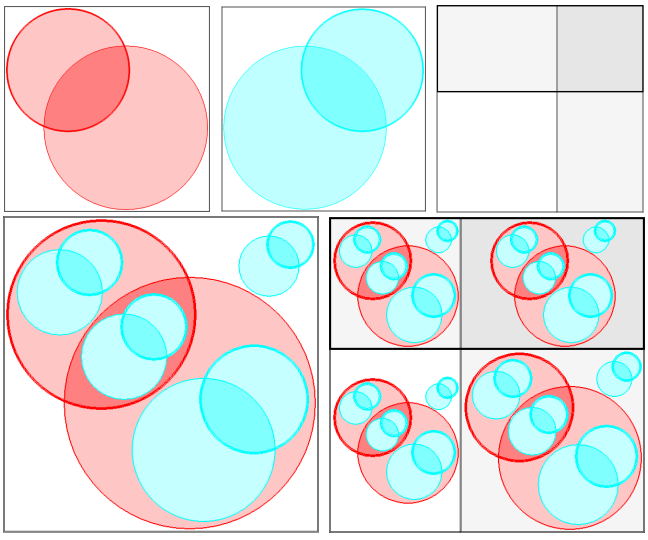


Рисунок 4: Пять диаграмм Венна (слева направо, сверху вниз) дуплетов событий $\mathfrak{X}_\circ = \{x_0, y_0\}$ (красный), $\mathfrak{X}_1 = \{x_1, y_1\}$ (аква) и $\mathfrak{X}_\star = \{x_\star, y_\star\}$ (серый) и несвязные диаграммы Венна квартета событий $\mathfrak{X}_\circ + \mathfrak{X}_1$ ($16 = 2^4$ террасных событий) и секстета событий $\mathfrak{X}_\circ + \mathfrak{X}_1 + \mathfrak{X}_\star$ ($64 = 2^6$ террасных события), составленных из этих дуплетов. События x_0, x_1 и $x_\star$ обведены на всех диаграммах толстыми линиями соответствующего цвета. Пространство всеобщих элементарных событий Ω — бесцветный квадрат.

2 Эвентологические показатели общности множества событий (м.с.)

2.1 Измерение вероятностной общности и зависимости событий

События взаимосвязаны. Одни сильнее, другие слабее. Одни события происходят вместе чаще, чем "можно было предвидеть"; другие — реже. Измерение взаимосвязей событий в рамках выбранного конечного м.с. — это наше повседневное вынужденное занятие. Чтобы измерять взаимосвязи событий требуется подходящая мера. Теория вероятностей, работая в рамках вероятностного пространства $(\Omega, \mathcal{A}, \mathbf{P})$, предлагает измерять *вероятностные* взаимосвязи между событиями из выбранного конечного м.с. $\mathfrak{X} \subseteq \mathcal{A}$, т.е. взаимосвязи, которые определяются вероятностной мерой $\mathbf{P}$.

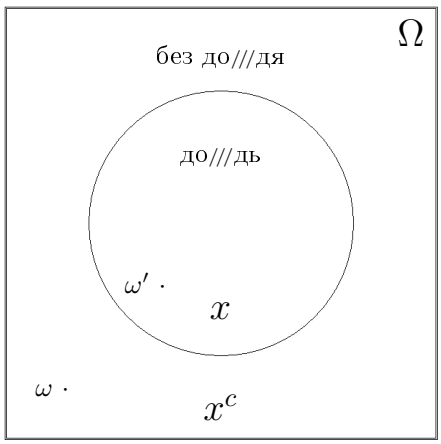


Рисунок 5: Диаграмма Венна одного события "до///дь": $x \subseteq \Omega$, которое происходит, когда $\omega' \in x$ (элементарный исход $\omega' \in \Omega$ "попадает" в x); и его дополнения "без до///дя": $x^c = \Omega - x \subseteq \Omega$, которое происходит, когда $\omega \in x^c$ (элементарный исход $\omega \in \Omega$ "попадает" в x^c).

2.1.1 Вероятностная зависимость и независимость событий и "независимая точка отсчета"

В качестве мер вероятностной взаимосвязи в теории вероятностей обычно используются отклонения вероятностных характеристик событий относительно **"независимой точки отсчета"**, которой служит значение аналогичных вероятностных характеристик *независимых* событий. Данный подход привел теорию вероятностей к измерению такой взаимосвязи событий, которую называют либо *вероятностной зависимостью*, либо *вероятностной независимостью*, и давал основания говорить, что единственным отличием теории вероятности от общей теории меры можно считать понятия *зависимости и независимости событий.*

> *При измерении зависимости м.с.* ***"независимая точка отсчета"*** *— значение показателя для тотально-независимого м.с., происходящих с теми же вероятностями.*

2.1.2 Вероятностная общность и разобщенность событий и "средневероятная точка отсчета"

В рамках нового направления в теории вероятностей, которое называется "математическая эвентология", были предложены другие меры взаимосвязей событий из данного м.с., которые измеряют отклонения вероятностных характеристик событий от другой **"точки отсчета" — средней вероятности событий** из этого м.с. Данный подход приводит "математическую эвентологию" к измерению такой взаимосвязи событий, которая называется либо *вероятностной общностью*, либо *вероятностной разобщенностью*, и дает основания говорить, что еще одним отличием теории вероятности, включающей математическую эвентологию, от общей теории меры можно считать понятие *общности и разобщенности событий.*

> *При измерении общности м.с.* ***"средневероятная точка отсчета"*** *— средняя вероятность событий в этом м.с.*

2.1.3 Две стороны и две шкалы измерения взаимосвязей событий

Вероятностные взаимосвязи событий — *зависимость / независимость, общность / разобщенность* — можно измерять с двух сторон, в рамках двух шкал относительно двух соответствующих точек отсчета.

Первая шкала позволяет анализировать ту сторону взаимосвязей событий, которая измеряется *вероятностями пересечения событий* либо относительно вероятностей пересечения независимых событий ("независимой точки отсчета") и называется *зависимостью / независимостью*, либо относительно средней вероятности ("средневероятной точки отсчета") и называется *общностью / разобщенностью.*

Вторая шкала — ту сторону, которая измеряется *вероятностями объединения событий* либо относительно вероятностей объединения независимых событий ("независимой точки отсчета") и называется *зависимостью / независимостью*, либо относительно средней вероятности ("средневероятной точки отсчета") и называется *общностью / разобщенностью.*

Шкалой вероятностей пересечений событий измеряются взаимосвязи в тех частях событий (в тех терасных событиях), где происходит много событий и которые на диаграмме Венна расположены в стороне, "тяготеющей" к пересечению всех событий из данного м.с.

Шкалой вероятностей объединений событий измеряются взаимосвязи в тех частях событий (в тех терасных событиях), где происходит мало событий и которые на диаграмме Венна расположены в иной стороне, "тяготеющей" к пересечению всех дополнений этих событий.

Таким образом каждая из двух сторон взаимосвязей событий измеряется своими односторонними частичными несимметричными измерителями в соответствующих шкалах. А для измерения взаимосвязи данного м.с. в целом, требуется симметричная конструкция этих измерителей, которая измеряет сразу обе стороны взаимосвязей событий, используя обе шкалы.

2.1.4 Ковариация — мера зависимости и независимости событий

П-ковариация. Распространенной мерой *вероятностной зависимости событий* является *П-ковариация* м.с. $X \subseteq \mathfrak{X}$, которая измеряет отклонение вероятности пересечения м.с. от "независимой точки отсчета" — произведения вероятностей событий, определяется формулой

$$\mathsf{Kov}^{\cap}_X = p_X - p^{\star}_X, \tag{2.1.1}$$

где

$$p_X = \mathbf{P}\left(\bigcap_{x\in X} x\right)$$

— вероятность пересечения событий из м.с. $X \subseteq \mathfrak{X}$,

$$p^{\star}_X = \prod_{x\in X} p_x$$

— вероятность пересечения тех же событий, будь они независимы, а

$$p_x = \mathbf{P}(x)$$

— вероятности событий $x \in X$.

Положительная П-ковариация м.с. $X \subseteq \mathfrak{X}$ означает, что события из этого м.с. *происходят вместе чаще*, чем будь они независимыми; отрицательная П-ковариация м.с. $X \subseteq \mathfrak{X}$ означает, что события из этого м.с. *происходят вместе реже*, чем будь они независимыми; нулевая П-ковариация м.с. $X \subseteq \mathfrak{X}$ означает, что события из этого м.с. *независимы*[2].

В частности, парная П-ковариация дуплета событий $X = \{x, y\} \subseteq \mathfrak{X}$ совпадает с обычной ковариацией индикаторов этих событий, как булевых с.в. $\mathbf{1}_x$ и $\mathbf{1}_y$:

$$\mathsf{Kov}^{\cap}_{\{x,y\}} = p_{xy} - p_x p_y =$$
$$= \mathbf{E}\left(\mathbf{1}_x \mathbf{1}_y\right) - \mathbf{E}\mathbf{1}_x \mathbf{E}\mathbf{1}_y = \mathsf{Cov}(\mathbf{1}_x, \mathbf{1}_x).$$

О-ковариация. Менее распространенной мерой *вероятностной зависимости событий* является *О-ковариация* м.с. $X \subseteq \mathfrak{X}$, которая измеряет отклонение вероятности объединения м.с. от "независимой точки отсчета" — вероятности объединения тех же событий, будь они независимы, определяется формулой

$$\mathsf{Kov}^{\cup}_X = u^{\star}_X - u_X, \tag{2.1.2}$$

где

$$u_X = \mathbf{P}\left(\bigcup_{x \in X} x\right)$$

— вероятность объединения событий из м.с. $X \subseteq \mathfrak{X}$,

$$u^{\star}_X = 1 - \prod_{x \in X} (1 - p_x)$$

— вероятность объединения тех же событий, будь они независимы, а

$$p_x = \mathbf{P}(x)$$

— вероятности событий $x \in X$.

Положительная О-ковариация м.с. $X \subseteq \mathfrak{X}$ означает, что события из этого м.с. *происходят вместе чаще*, чем будь они независимыми; отрицательная О-ковариация м.с. $X \subseteq \mathfrak{X}$ означает, что события из этого м.с. *происходят вместе реже*, чем будь они независимыми; нулевая О-ковариация м.с. $X \subseteq \mathfrak{X}$ означает, что события из этого м.с. *независимы*[3].

В частности, парная О-ковариация дуплета событий $X = \{x, y\} \subseteq \mathfrak{X}$ совпадает с обычной ковариацией булевых с.в. $1 - \mathbf{1}_x$ и $1 - \mathbf{1}_y$:

$$\mathsf{Kov}^{\cup}_{\{x,y\}} = 1 - (1 - p_x)(1 - p_y) - u_{xy} =$$
$$= \mathbf{E}\left((1 - \mathbf{1}_x)(1 - \mathbf{1}_y)\right) - \mathbf{E}(1 - \mathbf{1}_x)\mathbf{E}(1 - \mathbf{1}_y) =$$
$$= \mathsf{Cov}(1 - \mathbf{1}_x, 1 - \mathbf{1}_x).$$

[2]Заметим, что для *тотальной независимости событий* из м.с. $X \subseteq \mathfrak{X}$ этого, разумеется, недостаточно и требуется большее, чтобы П-ковариации всех его подмножеств обращались в нуль: $\mathsf{Kov}_Y = 0$ для $Y \subseteq X$.

[3]Заметим, что для *тотальной независимости событий* из м.с. $X \subseteq \mathfrak{X}$ этого, разумеется, недостаточно и требуется большее, чтобы П-ковариации всех его подмножеств обращались в нуль: $\mathsf{Kov}_Y = 0$ для $Y \subseteq X$.

2.1.5 Сет-расстояние и близость м.с. — меры разобщенности и общности событий

О-сет-расстояние м.с. Первой мерой *вероятностной разобщенности событий*, введенной в математической эвентологии, было *О-сет-расстояние* м.с. $X \subseteq \mathfrak{X}$ [4, 2011], которое определяется формулой

$$\Delta^{\cup}_X = u_X - \langle p \rangle_X, \tag{2.1.2}$$

где

$$u_X = \mathbf{P}\left(\bigcup_{x \in X} x\right)$$

— вероятность объединения событий из м.с. $X \subseteq \mathfrak{X}$,

$$\langle p \rangle_X = \frac{1}{|X|} \sum_{x \in X} p_x$$

— средняя вероятность событий из м.с. $X \subseteq \mathfrak{X}$, а

$$p_x = \mathbf{P}(x)$$

— вероятности событий $x \in X$.

Нулевое О-сет-расстояние м.с. $X \subseteq \mathfrak{X}$ означает, что события из этого м.с. совпадают, т.е. *минимально разобщены*; максимально возможное значение из $[0, 1]$ означает, что события из этого м.с. *максимально разобщены*, т.е. наименее пересекаются или же не пересекаются, если им "позволяют" их вероятности.

В частности, О-сет-расстояние дуплета событий $X = \{x, y\} \subseteq \mathfrak{X}$ совпадает с половиной *вероятностного расстояния* между ними:

$$\Delta^{\cup}_{\{x,y\}} = u_{xy} - \frac{1}{2}(p_x + p_y) = \frac{1}{2}(p_x + p_y) - p_{xy} =$$
$$= \frac{1}{2}(p_x + p_y - 2p_{xy}) = \frac{1}{2}\mathbf{P}(x \Delta y).$$

Сет-расстояние м.с. относительно пересечения. Если определение О-сет-расстояния (2.1.2) применить не к м.с. $X \subseteq \mathfrak{X}$, а к его М-дополнению

$$X^{(c)} = \{x^c,\ x \in X\},$$

то нетрудно получить следующую формулу

$$\Delta_{X^{(c)}} = \langle p \rangle_X - p_X, \tag{2.1.3}$$

так как работают законы де Моргана: $u_{X^{(c)}} = 1 - p_X$, а средние вероятности событий из X и их дополнений из $X^{(c)}$ связаны соотношением: $\langle 1 - p \rangle_{X^{(c)}} = 1 - \langle p \rangle_X$.

Ради краткости сет-расстояние м.с. $X^{(c)}$ (2.1.3) называется *П-сет-расстоянием* м.с. $X \subseteq \mathfrak{X}$ и имеет специальное обозначение:

$$\Delta^{\cap}_X = \langle p \rangle_X - p_X = \Delta_{X^{(c)}}. \tag{2.1.3'}$$

Нулевое П-сет-расстояние м.с. $X \subseteq \mathfrak{X}$ означает, что события из этого м.с. совпадают, т.е. *минимально разобщены*.

Максимально возможное значение из $[0,1]$ означает, что события из этого м.с. *максимально разобщены*, т.е. наименее пересекаются или же не пересекаются, если им "позволяют" их вероятности.

В частности, П-сет-расстояние дуплета событий $X = \{x,y\} \subseteq \mathfrak{X}$ совпадает с его О-сет-расстоянием:

$$\Delta^{\cap}_{\{x,y\}} = \frac{1}{2}(p_x + p_y) - p_{xy} = \frac{1}{2}(p_x + p_y - 2p_{xy}) =$$

$$= \frac{1}{2}\mathbf{P}(x\Delta y) = \Delta^{\cup}_{\{x,y\}}.$$

2.1.6 Сет-расстояние

Определение сет-расстояния относительно объединения. *Сет-расстояние м.с. $\mathfrak{X}$ относительно объединения (сет-О-расстояние)* определяется как величина

$$\Delta^{\cup}_{\mathfrak{X}} = \sum_{\emptyset \subset X \subseteq \mathfrak{X}} \left(1 - \frac{|X|}{|\mathfrak{X}|}\right) p(X /\!\!/ \mathfrak{X}) = u_{\mathfrak{X}} - \langle p \rangle_{\mathfrak{X}},$$

которая интерпретируется как "средний штраф за несовпадение" событий из $\mathfrak{X}$.

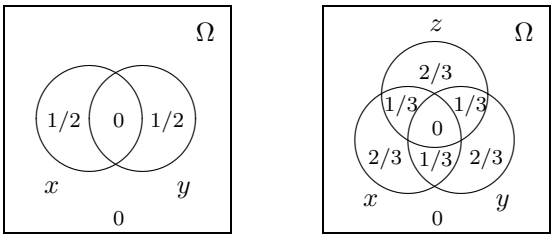


Рисунок 6: Террасные «штрафы за несовпадение» дуплета событий $\{x,y\}$ (слева) и триплета событий $\{x,y,z\}$ (справа) для сет-расстояний этих множеств: Δ_{xy} и Δ_{xyz} соответственно (сет-О-расстояний этих множеств: $\Delta^{\cup}_{xy}$ и $\Delta^{\cup}_{xyz}$).

Определение сет-расстояния относительно пересечения. *Сет-расстояние м.с. $\mathfrak{X}$ относительно пересечения (сет-П-расстояние)* определяется как величина

$$\Delta^{\cap}_{\mathfrak{X}} = \sum_{\emptyset \subseteq X \subset \mathfrak{X}} \frac{|X|}{|\mathfrak{X}|} p(X /\!\!/ \mathfrak{X}) = \langle p \rangle_{\mathfrak{X}} - p_{\mathfrak{X}}.$$

которая интерпретируется как "средний штраф за несовпадение" событий из $\mathfrak{X}$.

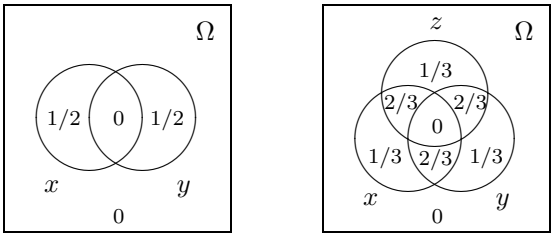


Рисунок 7: Террасные «штрафы за несовпадение» М-дополнения $\{x,y\}^{(c)}$ дуплета событий $\{x,y\}$ (слева) и М-дополнения $\{x,y,z\}^{(c)}$ триплета событий $\{x,y,z\}$ (справа) для сет-расстояний М-дополнений: $\Delta_{\{x,y\}^{(c)}}$ и $\Delta_{\{x,y,z\}^{(c)}}$ соответственно (сет-П-расстояний этих множеств: $\Delta^{\cap}_{xy}$ и $\Delta^{\cap}_{xyz}$).

2.1.7 Ковариация

Определение ковариации относительно объединения. *Ковариация м.с. $X \subseteq \mathfrak{X}$ относительно объединения (О-ковариация)* определяется как величина

$$\mathsf{Kov}^{\cup}_X = u^{\star}_X - u_X,$$

которая интерпретируется как отклонение вероятности объединения u_X м.с. X от "независимой точки отсчета" $u^{\star}_X$ — вероятности объединения тотально независимого м.с. $X^{\star}$, имеющих те же вероятности.

Определение ковариации относительно пересечения. *Ковариация м.с. $X \subseteq \mathfrak{X}$ относительно пересечения (П-ковариация)* определяется как величина

$$\mathsf{Kov}^{\cap}_X = p_X - p^{\star}_X,$$

которая интерпретируется как отклонение вероятности пересечения p_X м.с. X от "независимой точки отсчета" $p^{\star}_X$ — вероятности пересечения тотально независимого м.с. $X^{\star}$, имеющих те же вероятности.

2.1.8 Фреше-корреляция

Определение Фреше-корреляции относительно объединения. *Фреше-корреляция м.с. $X \subseteq \mathfrak{X}$ относительно объединения (О-ковариация)* определяется как величина

$$\mathsf{Kov}^{\cup}_X = u^{\star}_X - u_X,$$

которая интерпретируется как отклонение вероятности объединения u_X м.с. X от "независимой точки отсчета" $u^{\star}_X$ — вероятности объединения тотально независимого м.с. $X^{\star}$, имеющих те же вероятности.

Определение Фреше-корреляции относительно пересечения. *Фреше-корреляция м.с. $X \subseteq \mathfrak{X}$ относительно пересечения (П-ковариация)* определяется как величина

$$\mathsf{Kov}^{\cap}_X = p_X - p^{\star}_X,$$

которая интерпретируется как отклонение вероятности пересечения p_X м.с. X от "независимой точки отсчета" $p^{\star}_X$ — вероятности пересечения тотально независимого м.с. $X^{\star}$, имеющих те же вероятности.

3 Сет-расстояние м.с. — показатель разобщенности м.с.

3.1 Вероятностная метрика на алгебре событий

3.1.1 Метрика

Определение (метрика). *Метрика* на множестве $\mathfrak{X}$ (*расстояние между двумя элементами x, y множества $\mathfrak{X}$*) — это действительная числовая функция $\rho(x,y)$, определенная на $\mathfrak{X} \times \mathfrak{X}$ и удовлетворяющая для любых $x, y, z \in \mathfrak{X}$ двум аксиомам:

1) $\rho(x,y)=0 \iff x=y$ *(аксиома нуля)*,
2) $\rho(x,z)+\rho(y,z)\geq\rho(x,y)$ *(аксиома треугольника)*.

Замечание 1. Из определения вытекают два свойства расстояния:

- $\rho(x,y)\geq 0$ *(неотрицательность)*,
- $\rho(x,y)=\rho(y,x)$ *(симметричность)*.

Действительно, если в аксиоме треугольника, во-первых, положить $y=x$, то получим *неотрицательность*: $2\rho(x,y)\geq 0$, а во-вторых, положить $z=x$, то получим *симметричность*, так как из возникающего неравенства $\rho(y,x)\geq\rho(x,y)$ сразу же следует ему противоположное: $\rho(x,y)\geq\rho(y,x)$.

Замечание 2. Если для функции ρ выполнена только одна аксиома треугольника, то она называется *псевдометрикой (квазиметрикой)*, или *псевдорасстоянием (квазирасстоянием)*.

Замечание 3. Когда $\mathfrak{X}\subseteq\mathcal{F}$ — конечное м.с., то действительная числовая функция

$$\Delta_{xy}=\frac{1}{2}\mathbf{P}(x\Delta y),\quad x,y\in\mathfrak{X},\qquad(\star)$$

— *вероятность симметрической разности двух событий* x и y, будет метрикой на $\mathfrak{X}$ всякий раз, когда вероятность $\mathbf{P}$ такова, что

$$\mathbf{P}(x)=0\quad\Longleftrightarrow\quad x=\emptyset.\qquad(\star\star)$$

Иначе $\mathbf{P}(x\Delta y)$ — псевдометрика.

Далее вероятность будет предполагаться *регулярной*, т.е. для которой выполнено $(\star\star)$, а метрика $(\star)$ будет называться *вероятностной метрикой*, или *вероятностным расстоянием*.

3.1.2 Метрическая интерпретация парной ковариации событий

Парная ковариация событий имеет метрическую интерпретацию:

$$2\mathsf{Kov}_{xy}=\mathbf{P}^i(x\Delta y)-\mathbf{P}(x\Delta y),\qquad(3.1.2.2)$$

где

$$\mathbf{P}(x\Delta y)=\mathbf{P}(x\cup y)-\mathbf{P}(x\cap y)$$

— *вероятностное расстояние* между событиями x и y в общем случае, а

$$\mathbf{P}^i(x\Delta y)=\mathbf{P}(x\cup y)-\mathbf{P}(x)\mathbf{P}(y)$$

— *вероятностное расстояние* между независимыми событиями, имеющими те же вероятности, что и события x и y.

Соотношение (3.1.2.1) означает, что удвоенная парная ковариация событий — это отличие вероятностного расстояния между этими событиями от вероятностного расстояния между независимыми событиями с такими же вероятностями. Причем если вероятностное расстояние между событиями больше вероятностного расстояния между независимыми событиями с такими же вероятностями (отрицательная парная ковариация), то значит события «вероятностно отталкиваются», в противном случае (положительная парная ковариация) — «вероятностно притягиваются».

Определим *дополнение по Минковскому (M-дополнение) м.с.* $X\subseteq\mathfrak{X}$ как множество

$$X^{(c)}=\{x^c:x\in X\}\subseteq\mathcal{F},$$

составленное из дополнений событий из X. Рассмотрим

$$\{x,y\}^{(c)}=\{x^c,y^c\}$$

— дополнение по Минковскому дуплета событий $\{x,y\}\subseteq\mathfrak{X}$, и парную ковариацию этого м.с.:

$$\mathsf{Kov}_{\{x,y\}^{(c)}}=\mathbf{P}(x^c\cap y^c)-\mathbf{P}(x^c)\mathbf{P}(y^c),\qquad(3.1.2.3)$$

которая, как легко показать, совпадает с парной ковариацией дуплета событий $\{x,y\}$:

$$\mathsf{Kov}_{\{x,y\}^{(c)}}=\mathsf{Kov}_{\{x,y\}}=\mathsf{Kov}_{xy}.$$

Парная ковариация M-дополнения дуплета событий также имеют метрическую интерпретацию:

$$2\mathsf{Kov}_{\{x,y\}^{(c)}}=\mathbf{P}^i(x^c\Delta y^c)-\mathbf{P}(x^c\Delta y^c),\qquad(3.1.2.4)$$

где $\mathbf{P}(x^c\Delta y^c)$ — вероятностное расстояние между событиями x^c и y^c, а $\mathbf{P}^i(x^c\Delta y^c)$ — вероятностное расстояние между событиями x^c и y^c в независимом случае.

Итак, *парная ковариация* Kov_{xy}, а также *парная ковариация* $\mathsf{Kov}_{\{x,y\}^{(c)}}$ M-дополнения $\{x,y\}^{(c)}$ дуплета событий $\{x,y\}$ являются мерами зависимости событий из дуплета, а их метрические интерпретации имеют смысл разности вероятностных расстояний между событиями дуплета в общем и независимом случаях.

3.2 Сет-расстояние м.с.

Обычно расстояние определяется лишь для пар элементов, а не для произвольных множеств элементов.

Определим на $2^{\mathfrak{X}}$ такую *сет-функцию* $\Delta^{\cup}_X$ — функцию множества $X\subseteq\mathfrak{X}$, которая бы, во-первых, обобщала понятие *вероятностного расстояния* между двумя событиями $x,y\in\mathfrak{X}$, т.е. была пропорциональна вероятности их симметрической разности, когда $X=\{x,y\}$ — дуплет:

$$\Delta^{\cup}_{xy}=\frac{1}{2}\mathbf{P}(x\Delta y),$$

а, во-вторых, для произвольного $X \subseteq \mathfrak{X}$ обладала бы свойствами, напоминающими привычные свойства расстояния. Как выяснится ниже, этим условиям вполне удовлетворяет сет-функция Δ_X, определяемая для $X \subseteq \mathfrak{X}$ как

$$\Delta^{\cup}_X = \begin{cases} \mathbf{P}\left(\bigcup\limits_{x\in X} x\right) - \dfrac{1}{|X|}\sum\limits_{x\in X}\mathbf{P}(x), & X \neq \emptyset, \\ 0, & X = \emptyset, \end{cases} \tag{3.2.e0}$$

которая называется *сет-расстоянием м.с.* $X \subseteq \mathfrak{X}$.

Пример 1. Пусть м.с. $X = \{x, y\}$ — дуплет событий. Тогда

$$\Delta^{\cup}_{xy} = u_{xy} - \frac{1}{2}(p_x + p_y) = \frac{1}{2}(p_x + p_y - 2p_{xy}) \tag{3.2.e1}$$

— *сет-расстояние м.с.* $\{x, y\}$.

Пример 2. Пусть м.с. $\mathfrak{X}_i = \{x_i^1, \ldots, x_i^m\}$ — i-е категориальное m-м.с. для сопряженной (m, M)-таблицы. Тогда

$$\Delta^{\cup}_{\mathfrak{X}_i} = u_{\mathfrak{X}_i} - \frac{1}{m}\sum_{x\in\mathfrak{X}_i} p_x \tag{3.2.e2}$$

— *сет-расстояние м.с.* $\mathfrak{X}_i$.

Пример 3. Пусть м.с. $\mathfrak{X}_i^\star = \{x_i^1, \ldots, x_i^m\}$ — тотально-независимое i-е категориальное m-м.с. для сопряженной (m, M)-таблицы. Тогда

$$\Delta^{\Sigma\star}_{\mathfrak{X}_i} = u^\star_{\mathfrak{X}_i} - \frac{1}{m}\sum_{x\in\mathfrak{X}_i} p_x =$$

$$= 1 - \prod_{x\in\mathfrak{X}_i}(1 - p_x) - \frac{1}{m}\sum_{x\in\mathfrak{X}_i} p_x \tag{3.2.e3}$$

— *сет-расстояние тотально-независимого м.с.* $\mathfrak{X}_i^\star$.

3.2.1 Сет-метрика

Два события. Рассмотрим два события $x, y \in \mathcal{F}$, образующие дуплет событий $X = \{x, y\} \subset \mathcal{F}$. *Вероятностным расстоянием* между двумя событиями служит полу-вероятность их симметрической разности $\Delta_{xy} = \frac{1}{2}\mathbf{P}(x \Delta y)$, которую можно представить в виде следующей взвешенной суммы вероятностей террасных событий:

$$\Delta^{\cup}_{xy} = 1/2\cdot\mathbf{P}(x\cap y^c) + 1/2\cdot\mathbf{P}(x^c\cap y) +$$
$$+0\cdot\mathbf{P}(x\cap y) + 0\cdot\mathbf{P}(x^c\cap y^c).$$

Веса вероятностей террасных событий имеют смысл «штрафов за несовпадение» событий. Наименьший «штраф» 0 имеет террасное событие $x \cap y$, так как его наступление означает одновременное наступление двух событий (т.е. события x и y совпадают на этом террасном событии); террасные события $x \cap y^c$ и $x^c \cap y$ штрафуются $1/2$, так как наступает только одно событие из двух (рис. 8 слева). Таким образом, *вес террасного события — это доля ненаступающих событий* при его наступлении.

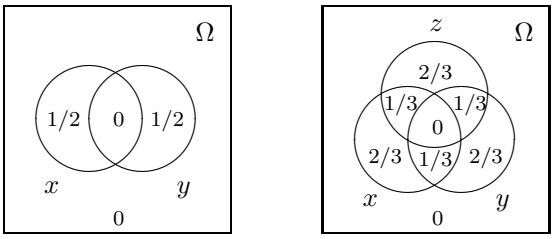


Рисунок 8: Террасные «штрафы за несовпадение» дуплета событий $\{x, y\}$ (слева) и триплета событий $\{x, y, z\}$ (справа) для сет-расстояний этих множеств: Δ_{xy} и Δ_{xyz} соответственно (сет-О-расстояний этих множеств: $\Delta^{\cup}_{xy}$ и $\Delta^{\cup}_{xyz}$).

Три события. Применим ту же систему «террасных штрафов» для террасных событий, образованных пересечением трех событий $x, y, z \in \mathcal{F}$ (рис. 8 справа). Тогда аналогичная взвешенная сумма, которую мы обозначим Δ_{xyz}, имеет вид

$$\Delta^{\cup}_{xyz} = 0\cdot\mathbf{P}(x\cap y\cap z) + 0\cdot\mathbf{P}(x^c\cap y^c\cap z^c) +$$
$$+1/3\cdot\mathbf{P}(x\cap y\cap z^c) + 1/3\cdot\mathbf{P}(x\cap y^c\cap z) + 1/3\cdot\mathbf{P}(x^c\cap y\cap z) +$$
$$+2/3\cdot\mathbf{P}(x\cap y^c\cap z^c) + 2/3\cdot\mathbf{P}(x^c\cap y\cap z^c) + 2/3\cdot\mathbf{P}(x^c\cap y^c\cap z).$$

Произвольное м.с.. Для произвольного м.с. $X \subset \mathcal{F}$ аналогичная взвешенная сумма равна

$$\Delta^{\cup}_X = \sum_{\emptyset\subset Y\subseteq X}\left(1 - \frac{|Y|}{|X|}\right)\mathbf{P}\left(\bigcap_{x\in Y} x \bigcap_{x\in X\setminus Y} x^c\right),$$

где суммирование распространяется на все непустые подмножества X. Сначала заметим, что

$$\mathbf{P}\left(\bigcap_{x\in Y} x \bigcap_{x\in X\setminus Y} x^c\right) = p(X, Y) = \mathbf{P}\left(K_X = Y\right)$$

— вероятность того, что среди м.с. X наступят только события из подмножества $Y \subseteq X$, т.е. вероятность того, что случайное множество наступивших событий K_X, определенное под X, примет значение Y.

Определение (сет-расстояние м.с., О-сет-расстояние). *Сет-расстояние м.с., О-сет-расстояние м.с.* — это сет-функция, определяемая для каждого м.с. $X \subset \mathfrak{X} \subseteq \mathcal{F}$ как величина

$$\Delta^{\cup}_X = \mathbf{P}\left(\bigcup_{x\in X} x\right) - \frac{1}{|X|}\sum_{x\in X}\mathbf{P}(x) = \Delta^{\cup}_X. \tag{3.2.2}$$

3.2.2 Сет-расстояние М-дополнения

Рассмотрим сет-расстояние $\Delta_{X^{(c)}}$ М-дополнения $X^{(c)}$ м.с. $X \subseteq \mathcal{F}$. Из (3.2.2) не трудно дать следующее

Определение (сет-расстояние М-дополнения м.с., П-сет-расстояние м.с.). *Сет-расстояние М-дополнения м.с.* X, *П-сет-расстояние м.с.* X — это

сет-функция, определяемая для каждого м.с. $X \subset \mathfrak{X} \subseteq \mathcal{F}$ как величина

$$\Delta_{X^{(c)}} = \frac{1}{|X|}\sum_{x\in X}\mathbf{P}(x) - \mathbf{P}\left(\bigcap_{x\in X} x\right) = \Delta^{\cap}_{X} \qquad (3.2.3)$$

Схема «террасных штрафов за несовпадение» М-дополнения показана на рис. 9.

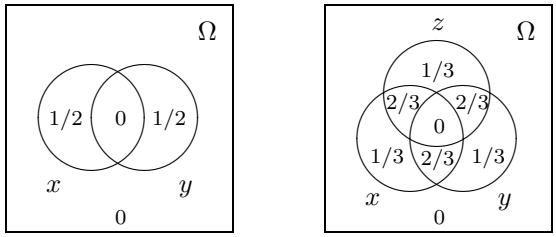


Рисунок 9: Террасные «штрафы за несовпадение» М-дополнения $\{x,y\}^{(c)}$ дуплета событий $\{x,y\}$ (слева) и М-дополнения $\{x,y,z\}^{(c)}$ триплета событий $\{x,y,z\}$ (справа) для сет-расстояний М-дополнений: $\Delta_{\{x,y\}^{(c)}}$ и $\Delta_{\{x,y,z\}^{(c)}}$ соответственно (сет-П-расстояний этих множеств: $\Delta^{\cap}_{xy}$ и $\Delta^{\cap}_{xyz}$).

Приведем еще формулу, справедливую для сет-расстояния М-дополнения м.с.:

$$\Delta_{X^{(c)}} = \sum_{\emptyset\subseteq Y\subset X}\frac{|Y|}{|X|}p(Y),$$

где $|Y|/|X|$ — "штраф за несовпадение" событий $x^c \in X^{(c)}$ при наступлении террасного события $\mathsf{ter}(Y), \emptyset \subseteq Y \subset X$.

Заметим, что когда $Y = X$, т.е. при наступлении террасного события $\mathsf{ter}(X)$, "штраф за несовпадение" полагается равным нулю. Подробнее:

$$\rho(Y/\!\!/X) = \begin{cases} 1-|Y|/|X|, & \emptyset \subset Y \subseteq X, \\ 0, & Y = \emptyset, \end{cases}$$

$$\rho^c(Y/\!\!/X) = \begin{cases} |Y|/|X|, & \emptyset \subseteq Y \subset X, \\ 0, & Y = X, \end{cases}$$

где обозначены:

- $\rho(Y/\!\!/X)$ — "штраф за несовпадение" событий $x \in X$ при наступлении террасного события $\mathsf{ter}(Y/\!\!/X)$;
- $\rho^c(Y/\!\!/X)$ — "штраф за несовпадение" событий $x^c \in X^{(c)}$ при наступлении террасного события $\mathsf{ter}(Y/\!\!/X)$.

Так что, во-первых, для $\emptyset \subseteq Y \subseteq X$

$$\rho(Y/\!\!/X) + \rho^c(Y/\!\!/X) = \begin{cases} 1, & \emptyset \subset Y \subset X, \\ 0, & \text{иначе}, \end{cases} \qquad (3.2.4)$$

и, во-вторых, сет-расстояния м.с. X и его М-дополнения $X^{(c)}$ выражаются через "штрафы за несовпадение" формулами:

$$\Delta^{\cup}_{X} = \sum_{Y\subseteq X}\rho(Y/\!\!/X)p(Y) = \mathbf{E}\rho(K/\!\!/X),$$

$$\Delta_{X^{(c)}} = \sum_{Y\subseteq X}\rho^c(Y/\!\!/X)p(Y) = \mathbf{E}\rho^c(K/\!\!/X)$$

и интерпретируются как средние значения "штрафов за несовпадение" событий из X и из $X^{(c)}$ соответственно. Из (3.2.4), в частности, следует, что

$$\Delta^{\cup}_{X} + \Delta_{X^{(c)}} = \sum_{\emptyset\subset Y\subset X} p(Y) =$$

$$= 1 - p(\emptyset) - p(X) = u_X - p_X. \qquad (3.2.5)$$

3.2.3 Неотрицательность сет-расстояния

Сет-расстояние — неотрицательная сет-функция:

$$\Delta^{\cup}_{X} \geq 0, \quad X \subseteq \mathfrak{X}.$$

Это следует из того, что

$$\Delta^{\cup}_{X} = \frac{1}{|X|}\sum_{x\in X}\left\{\mathbf{P}\left(\bigcup_{x\in X} x\right) - \mathbf{P}(x)\right\},$$

а для любого $X \subseteq \mathfrak{X}$ и каждого $x \in X$

$$\mathbf{P}\left(\bigcup_{x\in X} x\right) - \mathbf{P}(x) \geq 0.$$

3.3 Неравенство сет-треугольника

Лемма (неравенство сет-треугольника). *Сет-расстояние м.с. удовлетворяет неравенству сет-треугольника для произвольных подмножеств $X \subseteq \mathfrak{X}, Y \subseteq \mathfrak{X}$ любого конечного м.с. $\mathfrak{X}$ с различными именами:*

$$\Delta_X + \Delta_Y \geq \Delta_{X(\cup)Y}, \qquad (3.3.1)$$

где

$$X(\cup)Y = \{x \cup y : x \in X, y \in Y\}. \qquad (3.3.2)$$

— объединение множеств событий X и Y по Минковскому (М-объединение).

Доказательство см. в [18, 4].

3.4 Границы Фреше для сет-расстояния м.с.

Сет-расстояние м.с. $X \subseteq \mathfrak{X}$ определяется как величина

$$\Delta^{\cup}_{X} = u_{X/\!\!/X} - \frac{1}{|X|}\sum_{x\in X} p_x =$$

$$= 1 - p(\emptyset/\!\!/X) - \frac{1}{|X|}\sum_{x\in X} p_x,$$

где

$$u_{X/\!/X} = \mathbf{P}\left(\bigcup_{x\in X} x\right), \quad p(\emptyset/\!/X) = \mathbf{P}\left(\bigcap_{x\in X} x^c\right)$$

— вероятности террасных событий 4-го и 1-го родов соответственно.

Границы для сет-расстояния м.с. $X \subseteq \mathfrak{X}$ в неравенствах Фреше

$$\Delta_X^{\cup-} \le \Delta_X^{\cup} \le \Delta_X^{\cup+}$$

связаны с границами Фреше для объединения событий из X и определяются формулами

$$\Delta_X^{\cup-} = u_X^- - \frac{1}{|X|}\sum_{x\in X} p_x = \max_{x\in X} p_x - \frac{1}{|X|}\sum_{x\in X} p_x, \quad (3.4.1)$$

$$\Delta_X^{\cup+} = u_X^+ - \frac{1}{|X|}\sum_{x\in X} p_x =$$

$$= \begin{cases} \left(1-\dfrac{1}{|X|}\right)\displaystyle\sum_{x\in X} p_x, & \displaystyle\sum_{x\in X} p_x \le 1, \\ 1 - \dfrac{1}{|X|}\displaystyle\sum_{x\in X} p_x, & \displaystyle\sum_{x\in X} p_x > 1. \end{cases} \quad (3.4.2)$$

Если использовать удобную аббревиатуру

$$\langle p\rangle_X = \frac{1}{|X|}\sum_{x\in X} p_x, \quad (3.4.3)$$

для средней вероятности событий из м.с. X, то формулы (3.4.1) и (3.4.2) примут вид

$$\Delta_X^{\cup-} = \max_{x\in X} p_x - \langle p\rangle_X, \quad (3.4.4)$$

$$\Delta_X^{\cup+} = \begin{cases} \displaystyle\sum_{x\in X} p_x - \langle p\rangle_X, & \displaystyle\sum_{x\in X} p_x \le 1, \\ 1 - \langle p\rangle_X, & \displaystyle\sum_{x\in X} p_x > 1. \end{cases} \quad (3.4.5)$$

Более того, формулу (3.4.5) можно эквивалентным образом избавить и от суммы вероятностей:

$$\Delta_X^{\cup+} = \begin{cases} (|X|-1)\langle p\rangle_X, & \langle p\rangle_X \le \dfrac{1}{|X|}, \\ 1 - \langle p\rangle_X, & \langle p\rangle_X > \dfrac{1}{|X|}. \end{cases} \quad (3.4.5)$$

3.5 Фреше-нормализация функции м.с.

Определение (Фреше-нормализация функции м.с.). *Фреше-нормализацией* функции f_X м.с. X, для каждого значений которой f_X определены нижняя f_X^- и верхняя f_X^+ границы Фреше, называется функция

$$\lceil f_X \rceil^{\pm} = \begin{cases} f_X, & f_X^+ = f_X^-, \\ \dfrac{f_X - f_X^-}{f_X^+ - f_X^-}, & f_X^+ > f_X^-, \end{cases} \quad (3.5.1)$$

где функция $f_X^+ - f_X^-$ называется *размахом Фреше* функции f_X.

Свойство.

$$\lceil 1 - f_X \rceil^{\pm} = 1 - \lceil f_X \rceil^{\pm} =$$

$$= \begin{cases} 1 - f_X, & f_X^+ = f_X^-, \\ \dfrac{f_X^+ - f_X}{f_X^+ - f_X^-}, & f_X^+ > f_X^-. \end{cases} \quad (3.5.2)$$

Доказательство очевидно следует из определения.

4 Близость и кучность м.с. — показатели общности м.с.

4.1 Определение показателей общности и разобщенности м.с. на основе его сет-расстояния

На основе понятия сет-расстояния м.с. и Фреше-нормализации можно сконструировать несколько показателей общности и разобщенности м.с.: *Фреше-расстояние, О-близость, Фреше-О-близость, О-кучность и Фреше-О-кучность.*

4.1.1 Фреше-расстояние м.с.

Отклонение сет-расстояния от его нижней границы Фреше (минимально возможного значения при данных вероятностях событий) равно:

$$\Delta_X - \Delta_X^- = u_X - \max_{x\in X} p_x, \quad (4.1.1.1)$$

отклонение сет-расстояния от его верхней границы Фреше (максимально возможного значения при данных вероятностях событий) равно:

$$\Delta_X^{\cup+} - \Delta_X^{\cup} = \begin{cases} \displaystyle\sum_{x\in X} p_x - u_X, & \displaystyle\sum_{x\in X} p_x \le 1, \\ 1 - u_X, & \displaystyle\sum_{x\in X} p_x > 1, \end{cases} \quad (4.1.1.2)$$

размах Фреше[4] сет-расстояния м.с. X при данных вероятностях событий равен

$$\Delta_X^{\cup+} - \Delta_X^{\cup-} = \begin{cases} \displaystyle\sum_{x\in X} p_x - \max_{x\in X} p_x, & \displaystyle\sum_{x\in X} p_x \le 1, \\ 1 - \max_{x\in X} p_x, & \displaystyle\sum_{x\in X} p_x > 1, \end{cases} =$$

$$= \begin{cases} |X|\langle p\rangle_X - \max_{x\in X} p_x, & \langle p\rangle_X \le \dfrac{1}{|X|}, \\ 1 - \max_{x\in X} p_x, & \langle p\rangle_X > \dfrac{1}{|X|}. \end{cases} \quad (4.1.1.3)$$

[4] *Размах Фреше* сет-расстояния Δ_X м.с. X определяется как разность между его верней и нижней границей Фреше. (см. §3.5)

Фреше-нормализация сет-расстояния м.с. X равна по определению

$$\lceil \Delta_X \rceil^{\pm} = \begin{cases} \Delta_X, & \Delta_X^+ = \Delta_X^-, \\ \dfrac{\Delta_X - \Delta_X^-}{\Delta_X^+ - \Delta_X^-}, & \Delta_X^+ > \Delta_X^-. \end{cases} \tag{4.1.1.4}$$

Выполнение условия $\Delta_X^+ = \Delta_X^-$ в (4.1.1.4) зависит от средней вероятности событий из м.с. X и поэтому переписывается в виде

$$\begin{cases} \max\limits_{x \in X} p_x = |X| \langle p \rangle_X, & \langle p \rangle_X \le \dfrac{1}{|X|}, \\ \max\limits_{x \in X} p_x = 1, & \langle p \rangle_X > \dfrac{1}{|X|}. \end{cases} \tag{4.1.1.5}$$

Отсюда Фреше-нормализация сет-расстояния м.с. X равна

$$\lceil \Delta_X \rceil^{\pm} = \begin{cases} \Delta_X, & \max\limits_{x \in X} p_x = \sum\limits_{x \in X} p_x \le 1 \\ & \text{или} \\ & \max\limits_{x \in X} p_x = 1 < \sum\limits_{x \in X} p_x, \\ \dfrac{u_X - \max\limits_{x \in X} p_x}{\sum\limits_{x \in X} p_x - \max\limits_{x \in X} p_x}, & \max\limits_{x \in X} p_x < \sum\limits_{x \in X} p_x \le 1, \\ \dfrac{u_X - \max\limits_{x \in X} p_x}{1 - \max\limits_{x \in X} p_x}, & \max\limits_{x \in X} p_x < 1 < \sum\limits_{x \in X} p_x. \end{cases} \tag{4.1.1.6}$$

Из свойства Фреше-нормализации (3.5.2) следует, что Фреше-нормализация функции $1 - \Delta_X$ равна

$$\lceil 1 - \Delta_X \rceil^{\pm} = \begin{cases} 1 - \Delta_X, & \max\limits_{x \in X} p_x = \sum\limits_{x \in X} p_x \le 1 \\ & \text{или} \\ & \max\limits_{x \in X} p_x = 1 < \sum\limits_{x \in X} p_x, \\ \dfrac{\sum\limits_{x \in X} p_x - u_X}{\sum\limits_{x \in X} p_x - \max\limits_{x \in X} p_x}, & \max\limits_{x \in X} p_x < \sum\limits_{x \in X} p_x \le 1, \\ \dfrac{1 - u_X}{1 - \max\limits_{x \in X} p_x}, & \max\limits_{x \in X} p_x < 1 < \sum\limits_{x \in X} p_x. \end{cases} \tag{4.1.1.7}$$

Определение (Фреше-расстояние м.с.). *Фреше-расстоянием м.с.* X называется Фреше-нормализация его сет-расстояния

$$\mathbb{\Delta}_X^{\cup} = \lceil \Delta_X^{\cup} \rceil^{\pm}. \tag{4.1.1.7}$$

Пример 1. Пусть м.с. $X = \{x, y\}$ — дуплет событий. Тогда

$$\Delta_{xy}^{\cup} = u_{xy} - \frac{1}{2}(p_x + p_y) = \frac{1}{2}(p_x + p_y - 2p_{xy}) \tag{4.1.1.8}$$

— *сет-расстояние м.с.* $\{x, y\}$,

$$\mathbb{\Delta}_{xy}^{\cup} = \lceil \Delta_{xy}^{\cup} \rceil^{\pm} = \begin{cases} \Delta_{xy}, & \max\{p_x, p_y\} = p_x + p_y \le 1 \\ & \text{или} \\ & \max\{p_x, p_y\} = 1 < p_x + p_y, \\ \dfrac{\min\{p_x, p_y\} - p_{xy}}{p_x + p_y - \max\{p_x, p_y\}}, & \max\{p_x, p_y\} < p_x + p_y \le 1, \\ \dfrac{\min\{p_x, p_y\} - p_{xy}}{1 - \max\{p_x, p_y\}}, & \max\{p_x, p_y\} < 1 < p_x + p_y \end{cases} \tag{4.1.1.9}$$

— *Фреше-расстояние м.с.* $\{x, y\}$.

Пример 2. Пусть м.с. $\mathfrak{X}_i = \{x_i^1, \ldots, x_i^m\}$ — i-е категориальное m-м.с. для сопряженной (m, M)-таблицы. Тогда

$$\Delta_{\mathfrak{X}_i} = u_{\mathfrak{X}_i} - \frac{1}{m} \sum_{x \in \mathfrak{X}_i} p_x \tag{4.1.1.10}$$

— *сет-расстояние м.с.* $\mathfrak{X}_i$,

$$\mathbb{\Delta}_{\mathfrak{X}_i} = \lceil \Delta_{\mathfrak{X}_i} \rceil^{\pm} = \begin{cases} \Delta_{\mathfrak{X}_i}, & \max\limits_{x \in \mathfrak{X}_i} p_x = \sum\limits_{x \in \mathfrak{X}_i} p_x \le 1 \\ & \text{или} \\ & \max\limits_{x \in X} p_x = 1 < \sum\limits_{x \in X} p_x, \\ \dfrac{u_{\mathfrak{X}_i} - \max\limits_{x \in \mathfrak{X}_i} p_x}{\sum\limits_{x \in \mathfrak{X}_i} p_x - \max\limits_{x \in \mathfrak{X}_i} p_x}, & \max\limits_{x \in \mathfrak{X}_i} p_x < \sum\limits_{x \in \mathfrak{X}_i} p_x \le 1, \\ \dfrac{u_{\mathfrak{X}_i} - \max\limits_{x \in \mathfrak{X}_i} p_x}{1 - \max\limits_{x \in \mathfrak{X}_i} p_x}, & \max\limits_{x \in \mathfrak{X}_i} p_x < 1 < \sum\limits_{x \in \mathfrak{X}_i} p_x. \end{cases} \tag{4.1.1.11}$$

— *Фреше-расстояние м.с.* $\mathfrak{X}_i$.

Пример 3. Пусть м.с. $\mathfrak{X}_i^\star = \{x_i^1, \ldots, x_i^m\}$ — тотально-независимое i-е категориальное m-м.с. для сопряженной (m, M)-таблицы. Тогда

$$\Delta_{\mathfrak{X}_i}^\star = u_{\mathfrak{X}_i}^\star - \frac{1}{m} \sum_{x \in \mathfrak{X}_i} p_x =$$

$$= 1 - \prod_{x \in \mathfrak{X}_i} (1 - p_x) - \frac{1}{m} \sum_{x \in \mathfrak{X}_i} p_x \tag{4.1.1.12}$$

— *сет-расстояние тотально-независимого м.с.* $\mathfrak{X}_i^\star$,

$$\mathbb{\Delta}_{\mathfrak{X}_i}^\star = \lceil \Delta_{\mathfrak{X}_i}^\star \rceil^{\pm} =$$

$$= \begin{cases} \Delta^{\star}_{\mathfrak{X}_i}, & \max\limits_{x\in\mathfrak{X}_i} p_x = \sum\limits_{x\in\mathfrak{X}_i} p_x \le 1 \\ & \text{или} \\ & \max\limits_{x\in X} p_x = 1 < \sum\limits_{x\in X} p_x, \\ \dfrac{u^{\star}_{\mathfrak{X}_i} - \max\limits_{x\in\mathfrak{X}_i} p_x}{\sum\limits_{x\in\mathfrak{X}_i} p_x - \max\limits_{x\in\mathfrak{X}_i} p_x}, & \max\limits_{x\in\mathfrak{X}_i} p_x < \sum\limits_{x\in\mathfrak{X}_i} p_x \le 1, \\ \dfrac{u^{\star}_{\mathfrak{X}_i} - \max\limits_{x\in\mathfrak{X}_i} p_x}{1 - \max\limits_{x\in\mathfrak{X}_i} p_x}, & \max\limits_{x\in\mathfrak{X}_i} p_x < 1 < \sum\limits_{x\in\mathfrak{X}_i} p_x \end{cases} \tag{4.1.1.13}$$

— *Фреше-расстояние тотально-независимого м.с.* $\mathfrak{X}^{\star}_i$.

4.1.2 О-близость и Фреше-О-близость м.с.

Определение (О-близость м.с.). *О-Близостью м.с.* $X \subseteq \mathfrak{X}$ называется величина

$$1 - \Delta^{\cup}_X = \begin{cases} 1 - u_X + \langle p\rangle_X, & X \neq \emptyset, \\ 1, & X = \emptyset, \end{cases} \tag{4.1.2.1}$$

которая дополняет его сет-расстояние Δ_X до единицы, где

$$\langle p\rangle_X = \begin{cases} \dfrac{1}{|X|}\sum\limits_{x\in X} p_x, & X \neq \emptyset, \\ 0, & X = \emptyset \end{cases}$$

— удобная аббревиатура для средней вероятности событий из X.

Определение (Фреше-О-близость м.с.). *Фреше-О-близость м.с.* X определяется как *Фреше-нормализация его О-близости*:

$$1 - \mathbb{\Delta}^{\cup}_X = \lceil 1 - \Delta^{\cup}_X \rceil^{\pm} =$$

$$= \begin{cases} 1 - \Delta^{\cup}_X, & \max\limits_{x\in X} p_x = \sum\limits_{x\in X} p_x \le 1 \\ & \text{или} \\ & \max\limits_{x\in X} p_x = 1 < \sum\limits_{x\in X} p_x, \\ \dfrac{\sum\limits_{x\in X} p_x - u_X}{\sum\limits_{x\in X} p_x - \max\limits_{x\in X} p_x}, & \max\limits_{x\in X} p_x < \sum\limits_{x\in X} p_x \le 1, \\ \dfrac{1 - u_X}{1 - \max\limits_{x\in X} p_x}, & \max\limits_{x\in X} p_x < 1 < \sum\limits_{x\in X} p_x. \end{cases} \tag{4.1.2.2}$$

Пример 1. Пусть м.с. $X = \{x, y\}$ — дуплет событий. Тогда

$$1 - \Delta^{\cup}_{xy} = 1 - u_{xy} + \frac{1}{2}(p_x + p_y) =$$

$$= 1 - \frac{1}{2}(p_x + p_y - 2p_{xy}) \tag{4.1.2.7}$$

— *О-близость м.с.* $\{x, y\}$,

$$1 - \mathbb{\Delta}^{\cup}_{xy} = \lceil 1 - \Delta_{xy} \rceil^{\pm} =$$

$$= \begin{cases} 1 - \Delta_{xy}, & \max\{p_x, p_y\} = p_x + p_y \le 1 \\ & \text{или} \\ & \max\{p_x, p_y\} = 1 < p_x + p_y, \\ \dfrac{p_{xy}}{p_x + p_y - \max\{p_x, p_y\}}, & \max\{p_x, p_y\} < p_x + p_y \le 1, \\ \dfrac{1 - p_x - p_y + p_{xy}}{1 - \max\{p_x, p_y\}}, & \max\{p_x, p_y\} < 1 < p_x + p_y \end{cases} \tag{4.1.2.8}$$

— *Фреше-О-близость м.с.* $\{x, y\}$.

Пример 2. Пусть м.с. $\mathfrak{X}_i = \{x^1_i, \ldots, x^m_i\}$ — i-е категориальное m-м.с. для сопряженной (m, M)-таблицы. Тогда

$$1 - \Delta^{\cup}_{\mathfrak{X}_i} = 1 - u_{\mathfrak{X}_i} + \frac{1}{m}\sum_{x\in\mathfrak{X}_i} p_x \tag{4.1.2.9}$$

— *О-близость м.с.* $\mathfrak{X}_i$,

$$1 - \mathbb{\Delta}_{\mathfrak{X}_i} = \lceil 1 - \Delta_{\mathfrak{X}_i} \rceil^{\pm} =$$

$$= \begin{cases} 1 - \Delta^{\cup}_{\mathfrak{X}_i}, & \max\limits_{x\in\mathfrak{X}_i} p_x = \sum\limits_{x\in\mathfrak{X}_i} p_x \le 1 \\ & \text{или} \\ & \max\limits_{x\in X} p_x = 1 < \sum\limits_{x\in X} p_x, \\ \dfrac{\sum\limits_{x\in\mathfrak{X}_i} p_x - u_{\mathfrak{X}_i}}{\sum\limits_{x\in\mathfrak{X}_i} p_x - \max\limits_{x\in\mathfrak{X}_i} p_x}, & \max\limits_{x\in\mathfrak{X}_i} p_x < \sum\limits_{x\in\mathfrak{X}_i} p_x \le 1, \\ \dfrac{1 - u_{\mathfrak{X}_i}}{1 - \max\limits_{x\in\mathfrak{X}_i} p_x}, & \max\limits_{x\in\mathfrak{X}_i} p_x < 1 < \sum\limits_{x\in\mathfrak{X}_i} p_x. \end{cases} \tag{4.1.2.10}$$

— *Фреше-О-близость м.с.* $\mathfrak{X}_i$.

Пример 3. Пусть м.с. $\mathfrak{X}^{\star}_i = \{x^1_i, \ldots, x^m_i\}$ — тотально-независимое i-е категориальное m-м.с. для сопряженной (m, M)-таблицы. Тогда

$$1 - \Delta^{\cup\star}_{\mathfrak{X}_i} = 1 - u^{\star}_{\mathfrak{X}_i} + \frac{1}{m}\sum_{x\in\mathfrak{X}_i} p_x =$$

$$= \prod_{x\in\mathfrak{X}_i}(1 - p_x) + \frac{1}{m}\sum_{x\in\mathfrak{X}_i} p_x \tag{4.1.2.11}$$

— *О-близость тотально-независимого м.с.* $\mathfrak{X}^{\star}_i$,

$$1 - \mathbb{\Delta}^{\cup\star}_{\mathfrak{X}_i} = \lceil 1 - \Delta^{\cup\star}_{\mathfrak{X}_i} \rceil^{\pm} =$$

$$= \begin{cases} 1-\Delta^{\cup\star}_{\mathfrak{X}_i}, & \max\limits_{x\in\mathfrak{X}_i} p_x = \sum\limits_{x\in\mathfrak{X}_i} p_x \le 1 \\ & \text{или} \\ & \max\limits_{x\in X} p_x = 1 < \sum\limits_{x\in X} p_x, \\ \dfrac{\sum\limits_{x\in\mathfrak{X}_i} p_x - u^{\star}_{\mathfrak{X}_i}}{\sum\limits_{x\in\mathfrak{X}_i} p_x - \max\limits_{x\in\mathfrak{X}_i} p_x}, & \max\limits_{x\in\mathfrak{X}_i} p_x < \sum\limits_{x\in\mathfrak{X}_i} p_x \le 1, \\ \dfrac{1-u^{\star}_{\mathfrak{X}_i}}{1-\max\limits_{x\in\mathfrak{X}_i} p_x}, & \max\limits_{x\in\mathfrak{X}_i} p_x < 1 < \sum\limits_{x\in\mathfrak{X}_i} p_x. \end{cases} \quad (4.1.2.12)$$

— *Фреше-О-близость тотально-независимого м.с.* $\mathfrak{X}^{\star}_i$.

4.1.3 О-кучность и Фреше-О-кучность м.с.

Определение (О-кучность м.с.). *О-кучностью м.с.* X называется величина

$$1-\nabla^{\cup}_X = \begin{cases} 1-\dfrac{\Delta^{\cup}_X}{u_X} = \dfrac{\langle p\rangle_X}{u_X}, & X\neq\emptyset, \\ 1, & X=\emptyset \end{cases} \quad (4.1.3.1)$$

— отношение средней вероятности $\langle p\rangle_X$ событий из X к вероятности их объединения u_X, дополняющее до единицы О-разброс м.с. X (4.1.4.1).

Определение (границы Фреше кучности м.с.). *Границы Фреше кучности м.с.* X определяются как величины, образующие *неравенства Фреше для кучности* X при данных вероятностях событий из X:

$$(1-\nabla^{\cup}_X)^{-} \le 1-\nabla^{\cup}_X \le (1-\nabla_X)^{\cup+},$$

где

$$(1-\nabla_X)^{\cup-} = 1-\nabla^{\cup+}_X = \frac{\langle p\rangle_X}{u^{-}_X} = \frac{\langle p\rangle_X}{\max\limits_{x\in X} p_x},$$

$$(1-\nabla_X)^{\cup+} = 1-\nabla^{\cup-}_X = \frac{\langle p\rangle_X}{u^{+}_X} = \begin{cases} \dfrac{1}{|X|}, & \langle p\rangle_X \le \dfrac{1}{|X|}, \\ \langle p\rangle_X, & \langle p\rangle_X > \dfrac{1}{|X|}. \end{cases}$$

Определение (размах границ Фреше для кучности м.с.). *Размах границ Фреше для кучности м.с.* X определяется как величина

$$(1-\nabla^{\cup}_X)^{+} - (1-\nabla^{\cup}_X)^{-} = \nabla^{\cup+}_X - \nabla^{\cup-}_X = \langle p\rangle_X \frac{u^{+}_X - u^{-}_X}{u^{+}_X u^{-}_X},$$

которая совпадает с размахом границ Фреше для разброса объединения X (определение разброса объединения м.с. см. ниже).

Определение (Фреше-О-кучность м.с.). *Фреше-О-кучность м.с.* X определяется как Фреше-нормализация его кучности

$$1-\mathbb{V}^{\cup}_X = \lceil 1-\nabla^{\cup}_X \rceil^{\pm} =$$

$$= \begin{cases} 1-\nabla^{\cup}_X, & \dfrac{\max\limits_{x\in X} p_x}{|X|} = \langle p\rangle_X \le \dfrac{1}{|X|} \\ & \text{или} \\ & \dfrac{\max\limits_{x\in X} p_x}{|X|} = \dfrac{1}{|X|} < \langle p\rangle_X, \\ \dfrac{\max\limits_{x\in X} p_x}{u_X}\cdot\dfrac{\sum\limits_{x\in X} p_x - u_X}{\sum\limits_{x\in X} p_x - \max\limits_{x\in X} p_x}, & \dfrac{\max\limits_{x\in X} p_x}{|X|} < \langle p\rangle_X \le \dfrac{1}{|X|}, \\ \dfrac{\max\limits_{x\in X} p_x}{u_X}\cdot\dfrac{1-u_X}{1-\max\limits_{x\in X} p_x}, & \dfrac{\max\limits_{x\in X} p_x}{|X|} < \dfrac{1}{|X|} < \langle p\rangle_X. \end{cases} \quad (4.1.3.2)$$

Пример 1. Пусть м.с. $X=\{x,y\}$ — дуплет событий. Тогда

$$1-\nabla_{xy} = \frac{\langle p\rangle_{xy}}{u_{xy}} = \frac{1}{2}\cdot\frac{p_x+p_y}{p_x+p_y-p_{xy}} \quad (4.1.3.3)$$

— *О-кучность м.с.* $\{x,y\}$,

$$1-\mathbb{V}_{xy} =$$

$$= \begin{cases} 1-\nabla_{xy}, & \max\{p_x,p_y\} = p_x+p_y \le 1 \\ & \text{или} \\ & \max\{p_x,p_y\} = 1 < p_x+p_y, \\ \dfrac{\max\{p_x,p_y\}}{p_x+p_y-p_{xy}}\cdot\dfrac{p_{xy}}{\min\{p_x,p_y\}}, & \max\{p_x,p_y\} < p_x+p_y \le 1, \\ \dfrac{\max\{p_x,p_y\}}{p_x+p_y-p_{xy}}\cdot\dfrac{1-p_x-p_y+p_{xy}}{1-\max\{p_x,p_y\}}, & \max\{p_x,p_y\} < 1 < p_x+p_y \end{cases} \quad (4.1.3.4)$$

— *Фреше-О-кучность м.с.* $\{x,y\}$.

Пример 2. Пусть м.с. $\mathfrak{X}_i=\{x^1_i,\ldots,x^m_i\}$ — i-е категориальное m-м.с. для сопряженной (m,M)-таблицы. Тогда

$$1-\nabla_{\mathfrak{X}_i} = \frac{\langle p\rangle_{\mathfrak{X}_i}}{u_{\mathfrak{X}_i}} = \frac{1}{m\cdot u_{\mathfrak{X}_i}}\sum_{x\in\mathfrak{X}_i} p_x \quad (4.1.3.5)$$

— *О-кучность м.с.* $\mathfrak{X}_i$,

$$1-\mathbb{V}_{\mathfrak{X}_i} =$$

$$= \begin{cases} 1-\nabla_{\mathfrak{X}_i}, & \dfrac{\max\limits_{x\in\mathfrak{X}_i} p_x}{m} = \langle p\rangle_{\mathfrak{X}_i} \le \dfrac{1}{m} \\ & \text{или} \\ & \dfrac{\max\limits_{x\in\mathfrak{X}_i} p_x}{m} = \dfrac{1}{m} < \langle p\rangle_{\mathfrak{X}_i}, \\ \dfrac{\max\limits_{x\in\mathfrak{X}_i} p_x}{u_{\mathfrak{X}_i}} \cdot \dfrac{\sum\limits_{x\in\mathfrak{X}_i} p_x - u_{\mathfrak{X}_i}}{\sum\limits_{x\in\mathfrak{X}_i} p_x - \max\limits_{x\in\mathfrak{X}_i} p_x}, & \dfrac{\max\limits_{x\in\mathfrak{X}_i} p_x}{m} < \langle p\rangle_{\mathfrak{X}_i} \le \dfrac{1}{m}, \\ \dfrac{\max\limits_{x\in\mathfrak{X}_i} p_x}{u_{\mathfrak{X}_i}} \cdot \dfrac{1-u_{\mathfrak{X}_i}}{1-\max\limits_{x\in\mathfrak{X}_i} p_x}, & \dfrac{\max\limits_{x\in\mathfrak{X}_i} p_x}{m} < \dfrac{1}{m} < \langle p\rangle_{\mathfrak{X}_i}. \end{cases} \tag{4.1.3.6}$$

— *Фреше-О-кучность м.с.* $\mathfrak{X}_i$.

Пример 3. Пусть м.с. $\mathfrak{X}_i^\star = \{x_i^1, \dots, x_i^m\}$ — тотально-независимое i-е категориальное m-м.с. для сопряженной (m, M)-таблицы. Тогда

$$1-\nabla^\star_{\mathfrak{X}_i} = \frac{\langle p\rangle_{\mathfrak{X}_i^\star}}{u^\star_{\mathfrak{X}_i}} = \frac{\dfrac{1}{m}\sum\limits_{x\in\mathfrak{X}_i} p_x}{1-\prod\limits_{x\in\mathfrak{X}_i}(1-p_x)} \tag{4.1.3.7}$$

— *О-кучность тотально-независимого м.с.* $\mathfrak{X}_i^\star$,

$$1-\mathbb{V}^\star_{\mathfrak{X}_i} =$$

$$= \begin{cases} 1-\nabla^\star_{\mathfrak{X}_i}, & \dfrac{\max\limits_{x\in\mathfrak{X}_i} p_x}{m} = \langle p\rangle_{\mathfrak{X}_i} \le \dfrac{1}{m} \\ & \text{или} \\ & \dfrac{\max\limits_{x\in\mathfrak{X}_i} p_x}{m} = \dfrac{1}{m} < \langle p\rangle_{\mathfrak{X}_i}, \\ \dfrac{\max\limits_{x\in\mathfrak{X}_i} p_x}{u^\star_{\mathfrak{X}_i}} \cdot \dfrac{\sum\limits_{x\in\mathfrak{X}_i} p_x - u^\star_{\mathfrak{X}_i}}{\sum\limits_{x\in\mathfrak{X}_i} p_x - \max\limits_{x\in\mathfrak{X}_i} p_x}, & \dfrac{\max\limits_{x\in\mathfrak{X}_i} p_x}{m} < \langle p\rangle_{\mathfrak{X}_i} \le \dfrac{1}{m}, \\ \dfrac{\max\limits_{x\in\mathfrak{X}_i} p_x}{u^\star_{\mathfrak{X}_i}} \cdot \dfrac{1-u^\star_{\mathfrak{X}_i}}{1-\max\limits_{x\in\mathfrak{X}_i} p_x}, & \dfrac{\max\limits_{x\in\mathfrak{X}_i} p_x}{m} < \dfrac{1}{m} < \langle p\rangle_{\mathfrak{X}_i} \end{cases} \tag{4.1.3.8}$$

— *Фреше-О-кучность тотально-независимого м.с.* $\mathfrak{X}_i^\star$.

4.1.4 О-разброс и Фреше-О-разброс м.с.

Определение (О-разброс м.с.). *Разбросом объединения м.с.* X называется отношение его сет-расстояния к объединению

$$\nabla^\cup_X = \begin{cases} \dfrac{\Delta^\cup_X}{u_X} = 1 - \dfrac{\langle p\rangle_X}{u_X}, & X \neq \emptyset, \\ 0, & X = \emptyset, \end{cases} \tag{4.1.4.1}$$

которое дополняет его *О-кучность*[5] $1-\nabla^\cup_X = \dfrac{\langle p\rangle_X}{u_X}$ до единицы.

Определение (границы Фреше разброса объединения м.с.). *Границы Фреше разброса объединения м.с.* X определяют как величины, образующие *неравенства Фреше для разброса объединения* X при данных вероятностях событий из X:

$$\nabla^{\cup-}_X \le \nabla^\cup_X \le \nabla^{\cup+}_X,$$

где

$$\nabla^{\cup-}_X = 1 - \frac{\langle p\rangle_X}{u^-_X} = 1 - \frac{\langle p\rangle_X}{\max\limits_{x\in X} p_x},$$

$$\nabla^{\cup+}_X = 1 - \frac{\langle p\rangle_X}{u^+_X} = \begin{cases} 1 - \dfrac{1}{|X|}, & \sum\limits_{x\in X} p_x \le 1, \\ 1 - \langle p\rangle_X, & \sum\limits_{x\in X} p_x > 1, \end{cases} =$$

$$= \begin{cases} 1 - \dfrac{1}{|X|}, & \langle p\rangle_X \le \dfrac{1}{|X|}, \\ 1 - \langle p\rangle_X, & \langle p\rangle_X > \dfrac{1}{|X|} \end{cases}$$

— *нижняя и верхняя границы Фреше разброса объединения м.с.* X, определяемые, соответственно, через верхнюю и нижнюю границы для u_X — вероятности объединения событий из м.с. X.

Определение (размах границ Фреше для разброса объединения м.с.). *Размах границ Фреше для разброса объединения м.с.* X определяется как величина

$$(1-\nabla^\cup_X)^+ - (1-\nabla^\cup_X)^- = \nabla^{\cup+}_X - \nabla^{\cup-}_X = \langle p\rangle_X \frac{u^+_X - u^-_X}{u^+_X u^-_X},$$

которая совпадает с размахом границ Фреше для кучности X.

Определение (Фреше-О-разброс м.с.). *Фреше-разбросом объединения м.с.* X называется Фреше-нормализация его разброса объединения

$$\mathbb{V}^\cup_X = \lceil \nabla^\cup_X \rceil^\pm =$$

[5] См. определение О-кучности на стр. 56

$$= \begin{cases} \nabla^{\cup}_{X}, & \frac{\max\limits_{x\in X} p_x}{|X|} = \langle p\rangle_X \le \frac{1}{|X|} \\ & \text{или} \\ & \frac{\max\limits_{x\in X} p_x}{|X|} = \frac{1}{|X|} < \langle p\rangle_X, \\ \dfrac{\sum\limits_{x\in X} p_x}{u_X} \cdot \dfrac{u_X - \max\limits_{x\in X} p_x}{\sum\limits_{x\in X} p_x - \max\limits_{x\in X} p_x}, & \frac{\max\limits_{x\in X} p_x}{|X|} < \langle p\rangle_X \le \frac{1}{|X|}, \\ \dfrac{1}{u_X} \cdot \dfrac{u_X - \max\limits_{x\in X} p_x}{1 - \max\limits_{x\in X} p_x}, & \frac{\max\limits_{x\in X} p_x}{|X|} < \frac{1}{|X|} < \langle p\rangle_X. \end{cases} \quad (4.1.4.2)$$

Пример 1. Пусть м.с. $X = \{x, y\}$ — дуплет событий. Тогда

$$\nabla_{xy} = 1 - \frac{\Delta_{xy}}{u_{xy}} = 1 - \frac{\langle p\rangle_{xy}}{u_{xy}} =$$

$$= \frac{1}{2} \cdot \frac{p_x + p_y - 2p_{xy}}{p_x + p_y - p_{xy}} \quad (4.1.3.7)$$

— *О-разброс м.с.* $\{x, y\}$,

$$\mathbb{V}_{xy} = \begin{cases} \nabla_{xy}, & \max\{p_x, p_y\} = p_x + p_y \le 1 \\ & \text{или} \\ & \max\{p_x, p_y\} = 1 < p_x + p_y, \\ \dfrac{p_x + p_y}{p_x + p_y - p_{xy}} \cdot \dfrac{\min\{p_x, p_y\} - p_{xy}}{\min\{p_x, p_y\}}, & \max\{p_x, p_y\} < p_x + p_y \le 1, \\ \dfrac{1}{p_x + p_y - p_{xy}} \cdot \dfrac{\min\{p_x, p_y\} - p_{xy}}{1 - \max\{p_x, p_y\}}, & \max\{p_x, p_y\} < 1 < p_x + p_y \end{cases} \quad (4.1.4.3)$$

— *Фреше-О-разброс м.с.* $\{x, y\}$.

Пример 2. Пусть м.с. $\mathfrak{X}_i = \{x_i^1, \ldots, x_i^m\}$ — i-е категориальное m-м.с. для сопряженной (m, M)-таблицы. Тогда

$$\nabla_{\mathfrak{X}_i} = 1 - \frac{\langle p\rangle_{\mathfrak{X}_i}}{u_{\mathfrak{X}_i}} = 1 - \frac{1}{m \cdot u_{\mathfrak{X}_i}} \sum_{x\in\mathfrak{X}_i} p_x \quad (4.1.3.8)$$

— *О-разброс м.с.* $\mathfrak{X}_i$,

$$\mathbb{V}_{\mathfrak{X}_i} =$$

$$= \begin{cases} \nabla_{\mathfrak{X}_i}, & \frac{\max\limits_{x\in\mathfrak{X}_i} p_x}{m} = \langle p\rangle_{\mathfrak{X}_i} \le \frac{1}{m} \\ & \text{или} \\ & \frac{\max\limits_{x\in\mathfrak{X}_i} p_x}{m} = \frac{1}{m} < \langle p\rangle_{\mathfrak{X}_i}, \\ \dfrac{\sum\limits_{x\in\mathfrak{X}_i} p_x}{u_{\mathfrak{X}_i}} \cdot \dfrac{u_{\mathfrak{X}_i} - \max\limits_{x\in\mathfrak{X}_i} p_x}{\sum\limits_{x\in\mathfrak{X}_i} p_x - \max\limits_{x\in\mathfrak{X}_i} p_x}, & \frac{\max\limits_{x\in\mathfrak{X}_i} p_x}{m} < \langle p\rangle_{\mathfrak{X}_i} \le \frac{1}{m}, \\ \dfrac{1}{u_{\mathfrak{X}_i}} \cdot \dfrac{u_{\mathfrak{X}_i} - \max\limits_{x\in\mathfrak{X}_i} p_x}{1 - \max\limits_{x\in\mathfrak{X}_i} p_x}, & \frac{\max\limits_{x\in\mathfrak{X}_i} p_x}{m} < \frac{1}{m} < \langle p\rangle_{\mathfrak{X}_i}. \end{cases} \quad (4.1.4.4)$$

— *Фреше-О-разброс м.с.* $\mathfrak{X}_i$.

Пример 3. Пусть м.с. $\mathfrak{X}_i^\star = \{x_i^1, \ldots, x_i^m\}$ — тотально-независимое i-е категориальное m-м.с. для сопряженной (m, M)-таблицы. Тогда

$$\nabla^\star_{\mathfrak{X}_i} = 1 - \frac{\langle p\rangle_{\mathfrak{X}_i^\star}}{u^\star_{\mathfrak{X}_i}} = 1 - \frac{\frac{1}{m}\sum\limits_{x\in\mathfrak{X}_i} p_x}{1 - \prod\limits_{x\in\mathfrak{X}_i}(1 - p_x)} \quad (4.1.3.8)$$

— *О-разброс тотально-независимого м.с.* $\mathfrak{X}_i^\star$,

$$\mathbb{V}^\star_{\mathfrak{X}_i} =$$

$$= \begin{cases} \nabla^\star_{\mathfrak{X}_i}, & \frac{\max\limits_{x\in\mathfrak{X}_i} p_x}{m} = \langle p\rangle_{\mathfrak{X}_i} \le \frac{1}{m} \\ & \text{или} \\ & \frac{\max\limits_{x\in\mathfrak{X}_i} p_x}{m} = \frac{1}{m} < \langle p\rangle_{\mathfrak{X}_i}, \\ \dfrac{\sum\limits_{x\in\mathfrak{X}_i} p_x}{u^\star_{\mathfrak{X}_i}} \cdot \dfrac{u^\star_{\mathfrak{X}_i} - \max\limits_{x\in\mathfrak{X}_i} p_x}{\sum\limits_{x\in\mathfrak{X}_i} p_x - \max\limits_{x\in\mathfrak{X}_i} p_x}, & \frac{\max\limits_{x\in\mathfrak{X}_i} p_x}{m} < \langle p\rangle_{\mathfrak{X}_i} \le \frac{1}{m}, \\ \dfrac{1}{u^\star_{\mathfrak{X}_i}} \cdot \dfrac{u^\star_{\mathfrak{X}_i} - \max\limits_{x\in\mathfrak{X}_i} p_x}{1 - \max\limits_{x\in\mathfrak{X}_i} p_x}, & \frac{\max\limits_{x\in\mathfrak{X}_i} p_x}{m} < \frac{1}{m} < \langle p\rangle_{\mathfrak{X}_i}. \end{cases} \quad (4.1.4.4)$$

— *Фреше-О-разброс тотально-независимого м.с.* $\mathfrak{X}_i^\star$.

4.2 Нормализованные показатели общности и разобщенности м.с.

$$\kappa(\mathbb{X}) = \frac{\kappa_{\mathrm{diag}}(\mathbb{X}) - \kappa^\star_{\mathrm{diag}}(\mathbb{X})}{1 - \kappa^\star_{\mathrm{diag}}(\mathbb{X})} = \lceil \kappa_{\mathrm{diag}}(\mathbb{X})\rceil^\star$$

$$\mathcal{K}(\varkappa_\Delta),\ \mathcal{K}(\varkappa_{\blacktriangle}),\ \mathcal{K}(\varkappa_\nabla),\ \mathcal{K}(\varkappa_{\blacktriangledown}).$$

$$\lceil\varkappa_\Delta\rceil^\star,\ \lceil\lceil\varkappa_\Delta\rceil^\pm\rceil^\star,\ \lceil\varkappa_\nabla\rceil^\star,\ \lceil\lceil\varkappa_\nabla\rceil^\pm\rceil^\star.$$

$\lceil \varkappa_\Delta(\mathbb{X})\rceil^\star$, $\lceil\lceil \varkappa_\Delta(\mathbb{X})\rceil^\pm\rceil^\star$, $\lceil \varkappa_\nabla(\mathbb{X})\rceil^\star$, $\lceil\lceil \varkappa_\nabla(\mathbb{X})\rceil^\pm\rceil^\star$.

4.3 Нормализация показателей общности м.с.

4.3.1 Коэн-нормализация функции м.с.

Простейшим показателем общности (m, M)-сопряженного м.с. $\mathbb{X}$ служит его *диагональная каппа*, или *каппа-диагональ* $\kappa_{\mathrm{diag}}(\mathbb{X})$, которая вычисляется как сумма "диагональных" вероятностей соответствующей сопряженной таблицы, т.е. вероятностей пересечения всех m событий из i-го категориального м.с. $\mathfrak{X}_i = \{x_i^1, \ldots, x_i^m\}$, $i = 1, \ldots, M$:

$$\kappa_{\mathrm{diag}}(\mathbb{X}) = \sum_{i=1}^{M} \pi_{i\ldots i}^{1\ldots m}, \tag{4.2.1}$$

а для сопряженного м.с. $\mathbb{X} = \mathfrak{X}_1 + \ldots + \mathfrak{X}_M$ у которого все категориальные м.с. $\mathfrak{X}_i$ тотально независимы, равна

$$\kappa^\star_{\mathrm{diag}}(\mathbb{X}) = \sum_{i=1}^{M} \prod_{j=1}^{m} \pi_i^j, \tag{4.2.2}$$

где $p_i^j = \mathbf{P}(x_i^j)$ — вероятности событий сопряженного м.с.

$$\mathbb{X} = \sum_{i=1}^{M} \sum_{j=1}^{m} \{x_i^j\}.$$

Классическая каппа, или *каппа Коэна* сопряженного м.с. $\mathbb{X}$, определяется так [9], что ее можно рассматривать как результат преобразования $\lceil \kappa_{\mathrm{diag}}(\mathbb{X})\rceil^\star$ над каппа-диагональю $\kappa_{\mathrm{diag}}(\mathbb{X})$ м.с. $\mathbb{X}$:

$$\kappa(\mathbb{X}) = \lceil \kappa_{\mathrm{diag}}(\mathbb{X})\rceil^\star = \frac{\kappa_{\mathrm{diag}}(\mathbb{X}) - \kappa^\star_{\mathrm{diag}}(\mathbb{X})}{1 - \kappa^\star_{\mathrm{diag}}(\mathbb{X})}, \tag{4.2.3}$$

состоящего из "центрирования" каппа-диагонали $\kappa_{\mathrm{diag}}(\mathbb{X})$ тотально независимой каппа-диагональю $\kappa^\star_{\mathrm{diag}}(\mathbb{X})$ и "нормирования" "центрированной" величины $\kappa_{\mathrm{diag}}(\mathbb{X}) - \kappa^\star_{\mathrm{diag}}(\mathbb{X})$ на величину $1 - \kappa^\star_{\mathrm{diag}}(\mathbb{X})$, которая всегда больше "центрированной" величины.

Преобразование $\lceil\cdot\rceil^\star$, определяемое формулой (4.2.3), называется *Коэн-нормализацией*, а ее значение на каппа-диагонали κ_{diag}, равное классической каппа $\kappa(\mathbb{X}) = \lceil \kappa_{\mathrm{diag}}(\mathbb{X})\rceil^\star$, можно также назвать *Коэн-каппа-диагональю* сопряженного м.с. $\mathbb{X}$.

Определение (Коэн-нормализация функции м.с.). *Коэн-нормализацией* любого вероятностного показателя $f = f(X)$, как функции м.с. X со значениями из $[0, 1]$, для которой определено ее значение $f^\star$ для тотально независимой проекции $X^\star$ м.с. X, называется функция, определяемая формулой

$$\lceil f\rceil^\star = \begin{cases} f_X, & f_X^\star = 1, \\ \dfrac{f_X - f_X^\star}{1 - f_X^\star} & f_X^\star < 1; \end{cases} \tag{4.2.4}$$

4.3.2 Фреше-нормализация функции м.с.

Определение (Фреше-нормализация функции м.с.). *Фреше-нормализацией* любого вероятностного показателя $f = f(X)$, как функции м.с. X со значениями из $[0, 1]$, для которой определены верхняя и нижняя границы Фреше: f^+ и f^-, называется функция, определяемая формулой

$$\lceil f\rceil^\pm = \begin{cases} f_X, & f_X^+ = f_X^-, \\ \dfrac{f_X - f_X^-}{f_X^+ - f_X^-}, & f_X^+ > f_X^-; \end{cases} \tag{4.2.5}$$

4.3.3 Фреше-Коэн-нормализация функции м.с.

Внимательное рассмотрение формулы Коэн-нормализации (4.2.3) показывает, что величина $1 - \kappa^\star_{\mathrm{diag}}(\mathbb{X})$, на которую нормируется "центрированная" величина $\kappa_{\mathrm{diag}}(\mathbb{X}) - \kappa^\star_{\mathrm{diag}}(\mathbb{X})$, не является ее максимально возможным значением, и может быть равна меньшей величине:

$$\kappa^+_{\mathrm{diag}}(\mathbb{X}) - \kappa^\star_{\mathrm{diag}}(\mathbb{X}),$$

где

$$\kappa^+_{\mathrm{diag}}(\mathbb{X}) = \sum_{i=1}^{M} \kappa^+_{\mathrm{diag}}(\mathfrak{X}_i) = \sum_{i=1}^{M} \min_{x\in\mathfrak{X}_i} p_x \le 1.$$

— верхняя граница Фреше функции $\kappa_{\mathrm{diag}}(\mathbb{X})$ м.с. $\mathbb{X}$.

Отсюда возникает идея определения нормализации $\lceil\cdot\rceil$ функции м.с., более жесткой, чем Коэн-нормализация $\lceil\cdot\rceil^\star$, которая оказывается совпадает с суперпозицией Фреше- и Коэн- нормализаций $\lceil\lceil\cdot\rceil^\pm\rceil^\star$ (см. теорему ниже) и поэтому названа Фреше-Коэн-нормализацией функции м.с.

Определение (Фреше-Коэн-нормализация функции м.с.). *Верхней и нижней Фреше-Коэн-нормализациями* функции $f = f_X$ м.с. X, для каждого значений которой f_X определены нижняя f_X^- и верхняя f_X^+ границы Фреше и ее значение $f_X^\star$ на тотально независимой проекции $X^\star$ м.с. X, называются соответственно функции:

$$\lceil f\rceil = \begin{cases} f_X, & f_X^+ = f_X^\star, \\ \dfrac{f_X - f_X^\star}{f_X^+ - f_X^\star}, & f_X^+ > f_X^\star, \end{cases} \tag{4.2.6}$$

$$\lfloor f\rfloor = \begin{cases} f_X, & f_X^\star = f_X^-, \\ \dfrac{f_X^\star - f_X}{f_X^\star - f_X^-}, & f_X^\star > f_X^-, \end{cases} \tag{4.2.7}$$

где функция $f_X^+ - f_X^\star$ — *верхний Фреше-Коэн-размах*, а функция $f_X^\star - f_X^-$ — *нижний Фреше-Коэн-размах* функции $f = f_X$.

Теорема (о верхних нормализациях функции м.с.). *Пусть $f = f_X$ — функция м.с. X, тогда*

$$\lceil\lceil f\rceil^\pm\rceil^\star = \lceil f\rceil$$

— *верхняя Коэн-нормализация верхней Фреше-нормализации (3.5.1) функции f совпадает с ее верхней Фреше-Коэн-нормализацией (4.2.4).*

Доказательство следует из определений Фреше-нормализации (4.2.5) Коэн-нормализации (4.2.4) и верхней Фреше-Коэн-нормализации (4.2.6).

Замечание (нормализации показателей общности сопряженного м.с.). Можно говорить о Фреше-, Коэн- и Фреше-Коэн-нормализациях показателей общности и разобщенности сопряженного м.с. $\mathbb{X}$, которые определены выше. Например,

$$\lceil \kappa^{\cup}_{\mathrm{diag}}(\mathbb{X})\rceil^{\pm} = \frac{\kappa^{\cup}_{\mathrm{diag}}(\mathbb{X}) - \kappa^{\cup -}_{\mathrm{diag}}(\mathbb{X})}{\kappa^{\cup +}_{\mathrm{diag}}(\mathbb{X}) - \kappa^{\cup -}_{\mathrm{diag}}(\mathbb{X})}; \tag{4.2.8}$$

$$\lceil \kappa^{\cup}_{\mathrm{diag}}(\mathbb{X})\rceil^{\star} = \frac{\kappa^{\cup}_{\mathrm{diag}}(\mathbb{X}) - \kappa^{\cup \star}_{\mathrm{diag}}(\mathbb{X})}{1 - \kappa^{\cup \star}_{\mathrm{diag}}(\mathbb{X})}; \tag{4.2.8'}$$

$$\lceil \kappa^{\cup}_{\mathrm{diag}}(\mathbb{X})\rceil = \frac{\kappa^{\cup}_{\mathrm{diag}}(\mathbb{X}) - \kappa^{\cup \star}_{\mathrm{diag}}(\mathbb{X})}{\kappa^{\cup +}_{\mathrm{diag}}(\mathbb{X}) - \kappa^{\cup \star}_{\mathrm{diag}}(\mathbb{X})} \tag{4.2.8''}$$

— *Фреше-О-диагональ*, *Коэн-О-диагональ*, *Фреше-О-диагональ*,

$$\lceil \kappa_{\mathrm{diag}}(\mathbb{X})\rceil^{\pm} = \frac{\kappa_{\mathrm{diag}}(\mathbb{X}) - \kappa^{-}_{\mathrm{diag}}(\mathbb{X})}{\kappa^{+}_{\mathrm{diag}}(\mathbb{X}) - \kappa^{-}_{\mathrm{diag}}(\mathbb{X})}; \tag{4.2.9}$$

$$\lceil \kappa_{\mathrm{diag}}(\mathbb{X})\rceil^{\star} = \frac{\kappa_{\mathrm{diag}}(\mathbb{X}) - \kappa^{\star}_{\mathrm{diag}}(\mathbb{X})}{1 - \kappa^{\star}_{\mathrm{diag}}(\mathbb{X})}; \tag{4.2.9'}$$

$$\lceil \kappa_{\mathrm{diag}}(\mathbb{X})\rceil = \frac{\kappa_{\mathrm{diag}}(\mathbb{X}) - \kappa^{\star}_{\mathrm{diag}}(\mathbb{X})}{\kappa^{+}_{\mathrm{diag}}(\mathbb{X}) - \kappa^{\star}_{\mathrm{diag}}(\mathbb{X})} \tag{4.2.9''}$$

— *Фреше-П-диагональ*, *Коэн-П-диагональ*, *Фреше-П-диагональ*,

$$\lceil 1 - \Delta_{\mathbb{X}}\rceil^{\pm} = \frac{\Delta^{+}_{\mathbb{X}} - \Delta_{\mathbb{X}}}{\Delta^{+}_{\mathbb{X}} - \Delta^{-}_{\mathbb{X}}}; \tag{4.2.10}$$

$$\lceil 1 - \Delta_{\mathbb{X}}\rceil^{\star} = 1 - \frac{\Delta_{\mathbb{X}}}{\Delta^{\star}_{\mathbb{X}}}; \tag{4.2.10'}$$

$$\lceil 1 - \Delta_{\mathbb{X}}\rceil = \frac{\Delta^{\star}_{\mathbb{X}} - \Delta_{\mathbb{X}}}{\Delta^{\star}_{\mathbb{X}} - \Delta^{-}_{\mathbb{X}}} \tag{4.2.10''}$$

— *Фреше-О-близость*, *Коэн-О-близость*, *Фреше-Коэн-О-близость*,

$$\lceil 1 - \Delta^{\cap}_{\mathbb{X}}\rceil^{\pm} = \frac{\Delta^{\cap +}_{\mathbb{X}} - \Delta^{\cap}_{\mathbb{X}}}{\Delta^{\cap +}_{\mathbb{X}} - \Delta^{\cap -}_{\mathbb{X}}}; \tag{4.2.11}$$

$$\lceil 1 - \Delta^{\cap}_{\mathbb{X}}\rceil^{\star} = 1 - \frac{\Delta^{\cap}_{\mathbb{X}}}{\Delta^{\cap \star}_{\mathbb{X}}}; \tag{4.2.11'}$$

$$\lceil 1 - \Delta^{\cap}_{\mathbb{X}}\rceil = \frac{\Delta^{\cap \star}_{\mathbb{X}} - \Delta^{\cap}_{\mathbb{X}}}{\Delta^{\cap \star}_{\mathbb{X}} - \Delta^{\cap -}_{\mathbb{X}}} \tag{4.2.11''}$$

— *Фреше-П-близость*, *Коэн-П-близость*, *Фреше-Коэн-П-близость*,

$$\lceil 1 - \nabla_{\mathbb{X}}\rceil^{\pm} = \frac{\nabla^{+}_{\mathbb{X}} - \nabla_{\mathbb{X}}}{\nabla^{+}_{\mathbb{X}} - \nabla^{-}_{\mathbb{X}}}; \tag{4.2.12}$$

$$\lceil 1 - \nabla_{\mathbb{X}}\rceil^{\star} = 1 - \frac{\nabla_{\mathbb{X}}}{\nabla^{\star}_{\mathbb{X}}}; \tag{4.2.12'}$$

$$\lceil 1 - \nabla_{\mathbb{X}}\rceil = \frac{\nabla^{\star}_{\mathbb{X}} - \nabla_{\mathbb{X}}}{\nabla^{\star}_{\mathbb{X}} - \nabla^{-}_{\mathbb{X}}} \tag{4.2.12''}$$

— *Фреше-О-кучность*, *Коэн-О-кучность*, *Фреше-Коэн-О-кучность*.

$$\lceil 1 - \nabla^{\cap}_{\mathbb{X}}\rceil^{\pm} = \frac{\nabla^{\cap +}_{\mathbb{X}} - \nabla^{\cap}_{\mathbb{X}}}{\nabla^{\cap +}_{\mathbb{X}} - \nabla^{\cap -}_{\mathbb{X}}}; \tag{4.2.13}$$

$$\lceil 1 - \nabla^{\cap}_{\mathbb{X}}\rceil^{\star} = 1 - \frac{\nabla^{\cap}_{\mathbb{X}}}{\nabla^{\cap \star}_{\mathbb{X}}}; \tag{4.2.13'}$$

$$\lceil 1 - \nabla^{\cap}_{\mathbb{X}}\rceil = \frac{\nabla^{\cap \star}_{\mathbb{X}} - \nabla^{\cap}_{\mathbb{X}}}{\nabla^{\cap \star}_{\mathbb{X}} - \nabla^{\cap -}_{\mathbb{X}}} \tag{4.2.13''}$$

— *Фреше-П-кучность*, *Коэн-П-кучность*, *Фреше-Коэн-П-кучность*.

4.4 24 показателя общности категориального м.с.

Семейство из 24-х показателей общности категориального м.с. порождается всеми возможными комбинациями из

- *3-х типов показателей*:
  - "диагональ",
  - "близость",
  - "кучность";
- *2-х вариантов определения сет-расстояния*
  - относительно объединения ($\cup$),
  - относительно пересечения ($\cap$);
- *4-х вариантов нормализации*:
  - без нормализации,
  - Фреше-нормализация ($\lceil\cdot\rceil^{\pm}$),
  - Коэн-нормализация ($\lceil\cdot\rceil^{\star}$),
  - Фреше-Коэн-нормализация ($\lceil\cdot\rceil$).

4.5 Шесть показателей каппа м.с.

Рассмотрим сопряженную (m, M)-таблицу, т.е. m-мерную таблицу с M категориями по каждой размерности. В M^m клетках этой сопряженной таблицы взаимно однозначно расположены вероятности 2-го рода соответствующих $\{x^1_{i_1}, ..., x^m_{i_m}\}$-террасных событий 2-го рода

$$\mathsf{ter}_{\{x^1_{i_1},\ldots,x^m_{i_m}\}/\!/\mathbb{X}} = \bigcap_{j=1}^{m} x^j_{i_j},$$

порожденных сопряженным mM-м.с.

$$\mathbb{X} = \sum_{i=1}^{M}\sum_{j=1}^{m}\{x^j_i\} = \sum_{i=1}^{M}\mathfrak{X}_i = \sum_{j=1}^{m}\mathbf{X}^j,$$

где

$$\mathfrak{X}_i = \{x^1_i, \ldots, x^m_i\}$$

	$\bigcup$		$\bigcap$	
Диагональ	О-диагональ		П-диагональ	
	$\kappa^{\cup}_{\mathrm{diag}}$	$\lceil\kappa^{\cup}_{\mathrm{diag}}\rceil^{\pm}$	$\kappa^{\cap}_{\mathrm{diag}}$	$\lceil\kappa^{\cap}_{\mathrm{diag}}\rceil^{\pm}$
	$\lceil\kappa^{\cup}_{\mathrm{diag}}\rceil^{\star}$	$\lceil\kappa^{\cup}_{\mathrm{diag}}\rceil$	$\lceil\kappa^{\cap}_{\mathrm{diag}}\rceil^{\star}$	$\lceil\kappa^{\cap}_{\mathrm{diag}}\rceil$
Близость	О-близость		П-близость	
	$\varkappa^{\cup}_{\Delta}$	$\lceil\varkappa^{\cup}_{\Delta}\rceil^{\pm}$	$\varkappa^{\cap}_{\Delta}$	$\lceil\varkappa^{\cap}_{\Delta}\rceil^{\pm}$
	$\lceil\varkappa^{\cup}_{\Delta}\rceil^{\star}$	$\lceil\varkappa^{\cup}_{\Delta}\rceil$	$\lceil\varkappa^{\cap}_{\Delta}\rceil^{\star}$	$\lceil\varkappa^{\cap}_{\Delta}\rceil$
Кучность	О-кучность		П-кучность	
	$\varkappa^{\cup}_{\nabla}$	$\lceil\varkappa^{\cup}_{\nabla}\rceil^{\pm}$	$\varkappa^{\cap}_{\nabla}$	$\lceil\varkappa^{\cap}_{\nabla}\rceil^{\pm}$
	$\lceil\varkappa^{\cup}_{\nabla}\rceil^{\star}$	$\lceil\varkappa^{\cup}_{\nabla}\rceil$	$\lceil\varkappa^{\cap}_{\nabla}\rceil^{\star}$	$\lceil\varkappa^{\cap}_{\nabla}\rceil$

Таблица 1: 24 показателя общности категориального м.с.

— *категориальные м.с.* произвольного вида для $i = 1,\dots,M$, которые попарно не пересекаются *по диагонали*:

$$\mathfrak{X}_i[\cap]\mathfrak{X}_k = \{\underbrace{\emptyset,\dots,\emptyset}_{m}\}, \quad i \neq k,$$

а

$$\mathbf{X}^j = \{x^j_1,\dots,x^j_M\}$$

— *экспертные м.с.* специального вида, которые для каждого $j = 1,\dots,m$ образуют разбиение:

$$x^j_1 + \ldots + x^j_M = \Omega.$$

При этом

$$\sum_{i_1=1}^{M}\dots\sum_{i_m=1}^{M}\mathsf{ter}_{\{x^1_{i_1},\dots,x^m_{i_m}\}/\!/\mathbb{X}} = \Omega$$

— клеточные $\{x^1_{i_1},\dots,x^m_{i_m}\}$-террасные события образуют разбиение пространства исходов Ω на M^m фрагментов.

Таким образом, вероятностное распределение mM-м.с. $\mathbb{X}$ определяется M^m *клеточными вероятностями* $\{x^1_{i_1},\dots,x^m_{i_m}\}$-террасных событий 2-го рода

$$p_{\{x^1_{i_1},\dots,x^m_{i_m}\}/\!/\mathbb{X}} = \mathbf{P}\left(\mathsf{ter}_{\{x^1_{i_1},\dots,x^m_{i_m}\}/\!/\mathbb{X}}\right), \qquad (4.5.1)$$

расположенными в клетках сопряженной (m,M)-таблицы при

$$x^1_{i_1} \in \mathbf{X}^1,\dots,x^m_{i_m} \in \mathbf{X}^m,$$

где[6]

$$\sum_{i_1=1}^{M}\dots\sum_{i_m=1}^{M} p_{\{x^1_{i_1},\dots,x^m_{i_m}\}/\!/\mathbb{X}} = 1.$$

Вероятность каждого из mM событий $x^j_i \in \mathbb{X}$ определяется данным вероятностным распределением по формулам

$$\mathbf{P}(x^j_i) = \sum_{\substack{i_k=1\\k\neq j}}^{M} p_{\{x^1_{i_1},\dots,x^j_i,\dots,x^m_{i_m}\}/\!/\mathbb{X}} \qquad (4.5.2)$$

для $i = 1,\dots,M$, $j = 1,\dots,m$, где $(m-1)$-кратное суммирование производится по индексам $i_k \in \{i_1,\dots,i_m\} - \{i_j\}$, т.е. по всем всем индексам i_k, исключая j-тый, равный i.

Введем сокращенные обозначения для mM вероятностей

$$\pi^j_i = \mathbf{P}(x^j_i)$$

для $i = 1,\dots,M$, $j = 1,\dots,m$, которые образуют *двумерную*[7] $m \times M$ матрицу вероятностей событий x^j_i из сопряженного м.с. $\mathbb{X}$:

$$p(\mathbb{X}) = \begin{pmatrix} \pi^1_1 & \dots & \pi^1_i & \dots & \pi^1_M \\ \dots & \dots & \dots & \dots & \dots \\ \pi^j_1 & \dots & \pi^j_i & \dots & \pi^j_M \\ \dots & \dots & \dots & \dots & \dots \\ \pi^m_1 & \dots & \pi^m_i & \dots & \pi^m_M \end{pmatrix}. \qquad (4.5.3)$$

Определение (диагональная каппа сопряженного м.с.). *Диагональная каппа сопряженного м.с.* $\mathbb{X}$ для сопряженной (m,M)-таблицы определяется как величина

$$\kappa_{\mathrm{diag}}(\mathbb{X}) = \sum_{i=1}^{M} p_{\{x^1_i,\dots,x^m_i\}/\!/\mathbb{X}}, \qquad (4.5.4)$$

равная сумме *диагональных*[8] *вероятностей* 2-го рода из сопряженной (m,M)-таблицы.

Определение (диагональная каппа сопряженного м.с.). *Диагональная каппа сопряженного м.с.* $\mathbb{X}$ для сопряженной (m,M)-таблицы определяется как величина

$$\kappa_{\mathrm{diag}}(\mathbb{X}) = \sum_{i=1}^{M} \mathbf{P}(x^1_i \cap \ldots \cap x^m_i), \qquad (4.5.4)$$

[6] Тот факт, что сумма M^m вероятностей 2-го рода равна единице, объясняется специфической структурой ассоциаций и зависимостей событий в сопряженном mM-м.с. $\mathbb{X}$, однако для произвольных структур этот факт, разумеется, не имеет места.

[7] Не стоит путать эту *всегда (для любых сопряженных таблиц) двумерную* $m\times M$ матрицу объема mM с самой m-мерной сопряженной $\underbrace{M\times\ldots\times M}_{m}$ таблицей объема M^m (сокращенно: (m,M)-таблицей), в клетках которой расположены *клеточные вероятности* $\pi_{i_1\ldots i_m} = \pi^{1\ldots m}_{i_1\ldots i_m} = p_{\{x^1_{i_1},\dots,x^m_{i_m}\}/\!/\mathbb{X}}$ (4.5.1) и (6.2.17).

[8] расположенных в клетках главной диагонали сопряженной (m,M)-таблицы.

равная сумме *диагональных*[9] *вероятностей* 2-го рода из сопряженной (m, M)-таблицы.

Определение (классическая каппа (каппа Коэна) сопряженного м.с.). *Классическая каппа (каппа Коэна) сопряженного м.с.* $\mathbb{X}$ для сопряженной (m, M)-таблицы определяется как величина [9]

$$\kappa(\mathbb{X}) = \frac{\kappa_{\text{diag}}(\mathbb{X}) - \sum_{i=1}^{M}\prod_{j=1}^{m}\pi_i^j}{1 - \sum_{i=1}^{M}\prod_{j=1}^{m}\pi_i^j} = \\ = \frac{\kappa_{\text{diag}}(\mathbb{X}) - \kappa^{\star}_{\text{diag}}(\mathbb{X})}{1 - \kappa^{\star}_{\text{diag}}(\mathbb{X})}, \qquad (4.5.5)$$

равная *"специальным образом центрированной и нормированной"* диагональной каппа м.с. $\mathbb{X}$.

Определение (каппа-О-близость сопряженного м.с.). *Каппа-О-близость сопряженного м.с.*

$$\mathbb{X} = \mathfrak{X}_1 + \ldots + \mathfrak{X}_M$$

для сопряженной (m, M)-таблицы определяется как величина

$$\varkappa_{\Delta}(\mathbb{X}) = \frac{1}{M}\sum_{i=1}^{M}(1 - \Delta_{\mathfrak{X}_i}), \qquad (4.5.6)$$

равная среднему значению *близостей объединения* $1 - \Delta_{\mathfrak{X}_i}$ *(4.1.2.9) категориальных* m*-м.с.* $\mathfrak{X}_1, \ldots, \mathfrak{X}_M$, образующих $\mathbb{X}$.

Определение (каппа-Фреше-О-близость сопряженного м.с.). *Каппа-Фреше-О-близость сопряженного м.с.*

$$\mathbb{X} = \mathfrak{X}_1 + \ldots + \mathfrak{X}_M$$

для сопряженной (m, M)-таблицы определяется как величина

$$\varkappa_{\mathbb{\Delta}}(\mathbb{X}) = \frac{1}{M}\sum_{i=1}^{M}(1 - \mathbb{\Delta}_{\mathfrak{X}_i}), \qquad (4.5.7)$$

равная среднему значению *Фреше-близостей объединения* $1 - \mathbb{\Delta}_{\mathfrak{X}_i}$ *(4.1.2.10) категориальных* m*-м.с.* $\mathfrak{X}_1, \ldots, \mathfrak{X}_M$, образующих $\mathbb{X}$.

Определение (каппа-О-кучность сопряженного м.с.). *Каппа-О-кучность сопряженного м.с.*

$$\mathbb{X} = \mathfrak{X}_1 + \ldots + \mathfrak{X}_M$$

для сопряженной (m, M)-таблицы определяется как величина

$$\varkappa_{\nabla}(\mathbb{X}) = \frac{1}{M}\sum_{i=1}^{M}(1 - \nabla_{\mathfrak{X}_i}), \qquad (4.5.8)$$

равная среднему значению *кучностей* $1 - \nabla_{\mathfrak{X}_i}$ *(4.1.3.5) категориальных* m*-м.с.* $\mathfrak{X}_1, \ldots, \mathfrak{X}_M$, образующих $\mathbb{X}$.

Определение (каппа-Фреше-О-кучность сопряженного м.с.). *Каппа-Фреше-О-кучность сопряженного м.с.*

$$\mathbb{X} = \mathfrak{X}_1 + \ldots + \mathfrak{X}_M$$

для сопряженной (m, M)-таблицы определяется как величина

$$\varkappa_{\mathbb{\nabla}}(\mathbb{X}) = \frac{1}{M}\sum_{i=1}^{M}(1 - \mathbb{\nabla}_{\mathfrak{X}_i}), \qquad (4.5.9)$$

равная среднему значению *Фреше-кучностей* $1 - \mathbb{\nabla}_{\mathfrak{X}_i}$ *(4.1.3.6) категориальных* m*-м.с.* $\mathfrak{X}_1, \ldots, \mathfrak{X}_M$, образующих $\mathbb{X}$.

5 Постановка задачи общности мнений многих экспертов по многим категориям

5.1 Задача общности мнений двух экспертов по трем категориям: кошка, собака, заяц

Задача общности мнений двух экспертов по трем категориям: кошка, собака, заяц

	$x^2_{\text{кошка}}$	$x^2_{\text{собака}}$	$x^2_{\text{заяц}}$	Ω
$x^1_{\text{кошка}}$	$n^{12}_{\text{КК}} = 5$	$n^{12}_{\text{КС}} = 3$	$n^{12}_{\text{КЗ}} = 0$	$n^1_{\text{К}} = 8$
$x^1_{\text{собака}}$	$n^{12}_{\text{СК}} = 2$	$n^{12}_{\text{СС}} = 3$	$n^{12}_{\text{СЗ}} = 1$	$n^1_{\text{С}} = 6$
$x^1_{\text{заяц}}$	$n^{12}_{\text{ЗК}} = 0$	$n^{12}_{\text{ЗС}} = 2$	$n^{12}_{\text{ЗЗ}} = 11$	$n^1_{\text{З}} = 13$
Ω	$n^2_{\text{К}} = 7$	$n^2_{\text{С}} = 8$	$n^2_{\text{З}} = 12$	$n = 27$

Таблица 2: Двусторонняя сопряженная таблица для задачи общности мнений двух экспертов по трем категориям: кошка, собака, заяц (1-я статистика).

	$x^2_{\text{кошка}}$	$x^2_{\text{собака}}$	$x^2_{\text{заяц}}$	Ω
$x^1_{\text{кошка}}$	$n^{12}_{\text{КК}} = 6$	$n^{12}_{\text{КС}} = 3$	$n^{12}_{\text{КЗ}} = 0$	$n^1_{\text{К}} = 9$
$x^1_{\text{собака}}$	$n^{12}_{\text{СК}} = 0$	$n^{12}_{\text{СС}} = 3$	$n^{12}_{\text{СЗ}} = 0$	$n^1_{\text{С}} = 3$
$x^1_{\text{заяц}}$	$n^{12}_{\text{ЗК}} = 0$	$n^{12}_{\text{ЗС}} = 3$	$n^{12}_{\text{ЗЗ}} = 12$	$n^1_{\text{З}} = 15$
Ω	$n^2_{\text{К}} = 6$	$n^2_{\text{С}} = 9$	$n^2_{\text{З}} = 12$	$n = 27$

Таблица 3: Двусторонняя сопряженная таблица для задачи общности мнений двух экспертов по трем категориям: кошка, собака, заяц (2-я статистика).

[9] расположенных в клетках главной диагонали сопряженной (m, M)-таблицы.

6 Сопряженные таблицы и каппа-показатели общности

6.1 Сопряженные таблицы

В статистике и теории вероятностей *сопряженная таблица (перекрест-таблица, кросс-таблица, кросс-таб)*, является одним из видов таблиц в формате матрицы, создаваемой статистическим/аналитическим процессом *кросс-табуляции (перекрест-табуляции)*[10]

- результатов статистических наблюдений,
- структуры *сводных таблиц (pivot tables)*[11],
- структуры *электронных крупноформатных таблиц (spreadsheets)*[12],
- структуры многомерных вероятностных распределений

для удобства записи/визуализации/отображения (многомерного) частотного/вероятностного распределения многих переменных; термин "сопряженная таблица (contingency table)" был впервые использован Карлом Пирсоном [15, 1904].

Отыскание *табличной структуры зависимостей событий*, которая управляет многими переменными, образующими многомерную *сопряженную таблицу* — важнейшая проблема многомерной статистики, поскольку

- табличное исчисление вероятностей и статистик — это эффективно работающий инструмент для решения значительной части экспоненциально сложных задач многомерного анализа событий;
- обнаружение структур условных независимостей между событиями позволяет организовать более разумным способом хранение статистических данных.

[10] *Кросс-табуляция (перекрест-табуляция, кросс-таб)* представляет собой процедуру преобразования категориальных данных с целью создания *сопряженной таблицы*, которая отображает совместное распределение двух и более переменных; активно используется в бизнес-анализе, инженерных и научных исследованиях; обеспечивает визуализацию общей картины взаимосвязей между переменными и помогает обнаружить структуру взаимодействий между ними.

[11] *Сводная таблица* — это инструмент преобразования данных в программах визаулизации данных, таких как *электронные таблицы* (например, Mircosoft Excel) или в программах бизнес-аналитики (например, Spotfire); помимо суммирования и прочих функций, может автоматически сортировать данные, подсчитывать количество данных (записей), рассчитывать итоги и промежуточные итоги на основании данных, которые хранятся в одной таблице (электронной таблице, базе данных, в ином источнике), чтобы затем вывести результаты в другой таблице (которая собственно и называется "сводной"); особенно полезен для представления данных в виде *кросс-таблиц*; предназначен для создания и изменения структуры *сводной таблицы* посредством захвата и перетаскивания графических изображений ее клеток; визуальная *ротация* табличной структуры стала причиной английского названия *сводной таблицы — pivot table* (глагол *pivot* в переводе означает *вертеть, вращать, крутить*); понятие *сводной таблицы* стало общим брендом у многих производителей программного обеспечения, например, у корпорации Microsoft есть торговая марка *PivotTable*.

[12] *Электронные крупноформатные таблицы (spreadsheets)* — это интерактивные компьютерные программы (например, Microsoft Excel) для организации и анализа данных в табличной форме.

6.2 Статистика многомерных сопряженных таблиц

В ситуации "много экспертов — много категорий" исходная статистика наблюдений, определяющая *многомерную сопряженную таблицу*, состоит из набора *экспертных таблиц. Экспертная таблица* состоит из наблюдений (строк таблицы) за общим предметом одним из экспертов по всем категориям (столбцы таблицы). Каждое наблюдение (строка экспертной таблицы) проводится экспертом в соответствующих пространственно-временных условиях. При этом совокупность всех пространственно-временных условий (набор всех строк экспертной таблицы), в которых накапливается исходная статистика наблюдений, должна быть общей для всех экспертов, чтобы служить *единым плацдармом усреднения* их совместных оценок.

Итак, для ситуации "m экспертов, M категорий в n наблюдениях" *m-мерная сопряженная* $\underbrace{M\times\ldots\times M}_{m}$ *таблица* объема M^m определяется набором из m *экспертных* $n\times M$ *таблиц.*

Пример 1 (статистическая (2,2)-задача, событийный вариант). Для ситуации "2 экспертов, 2 категорий в n наблюдениях" 2*-мерная сопряженная* 2×2 *таблица* объема 2^2

	y	y^c	Ω
x	$n(xy)$	$n(x)$	n_x
x^c	$n(y)$	$n(\emptyset)$	n_{x^c}
Ω	n_y	n_{y^c}	n

(6.2.1)

определяется набором из 2-х *экспертных* $n\times 2$ *таблиц*:

	x	x^c	Ω
ω_1	$\mathbf{1}_x(\omega_1)$	$\mathbf{1}_{x^c}(\omega_1)$	1
$\ldots$	$\ldots$	$\ldots$	1
ω_i	$\mathbf{1}_x(\omega_i)$	$\mathbf{1}_{x^c}(\omega_i)$	1
$\ldots$	$\ldots$	$\ldots$	1
ω_n	$\mathbf{1}_x(\omega_n)$	$\mathbf{1}_{x^c}(\omega_n)$	1
Ω	n_x	n_{x^c}	n

,

	y	y^c	Ω
ω_1	$\mathbf{1}_y(\omega_1)$	$\mathbf{1}_{y^c}(\omega_1)$	1
$\ldots$	$\ldots$	$\ldots$	1
ω_i	$\mathbf{1}_y(\omega_i)$	$\mathbf{1}_{y^c}(\omega_i)$	1
$\ldots$	$\ldots$	$\ldots$	1
ω_n	$\mathbf{1}_y(\omega_n)$	$\mathbf{1}_{y^c}(\omega_n)$	1
Ω	n_y	n_{y^c}	n

(6.2.2)

где

$$n(xy)=\sum_{i=1}^{n}\mathbf{1}_x(\omega_i)\mathbf{1}_y(\omega_i)=\sum_{i=1}^{n}\mathbf{1}_{x\cap y}(\omega_i)=$$

$$=\sum_{i=1}^{n}\mathbf{1}_{\mathsf{ter}(xy)}(\omega_i),$$

$$n(x)=\sum_{i=1}^{n}\mathbf{1}_{x}(\omega_i)\mathbf{1}_{y^c}(\omega_i)=\sum_{i=1}^{n}\mathbf{1}_{x\cap y^c}(\omega_i)=$$

$$=\sum_{i=1}^{n}\mathbf{1}_{\mathsf{ter}(x)}(\omega_i),$$

$$n(y)=\sum_{i=1}^{n}\mathbf{1}_{x^c}(\omega_i)\mathbf{1}_{y}(\omega_i)=\sum_{i=1}^{n}\mathbf{1}_{x^c\cap y}(\omega_i)=$$

$$=\sum_{i=1}^{n}\mathbf{1}_{\mathsf{ter}(y)}(\omega_i),$$

$$n(\emptyset)=\sum_{i=1}^{n}\mathbf{1}_{x^c}(\omega_i)\mathbf{1}_{y^c}(\omega_i)=\sum_{i=1}^{n}\mathbf{1}_{x^c\cap y^c}(\omega_i)=$$

$$=\sum_{i=1}^{n}\mathbf{1}_{\mathsf{ter}(\emptyset)}(\omega_i).$$

$\star\star\star$

Пример 1a (статистическая (2,2)-задача, индексный вариант). Предыдущую ситуацию "2 экспертов, 2 категорий в n наблюдениях" можно также записать эквивалентным образом в виде следующей 2-*мерной сопряженной* 2×2 *таблицей* объема 2^2

	x_1^2	x_2^2	Ω
x_1^1	n_{11}^{12}	n_{12}^{12}	n_1^1
x_2^1	n_{21}^{12}	n_{22}^{12}	n_2^1
Ω	n_1^2	n_2^2	n

(6.2.3)

которая определяется набором из 2-х *экспертных* $n\times 2$ *таблиц*, также переписанных в эквивалентном виде:

	x_1^1	x_2^1	Ω
ω_1	$\mathbf{1}_{x_1^1}(\omega_1)$	$\mathbf{1}_{x_2^1}(\omega_1)$	1
...	...	...	1
ω_i	$\mathbf{1}_{x_1^1}(\omega_i)$	$\mathbf{1}_{x_2^1}(\omega_i)$	1
...	...	...	1
ω_n	$\mathbf{1}_{x_1^1}(\omega_n)$	$\mathbf{1}_{x_2^1}(\omega_n)$	1
Ω	n_1^1	n_2^1	n

(6.2.4)

	x_1^2	x_2^2	Ω
ω_1	$\mathbf{1}_{x_1^2}(\omega_1)$	$\mathbf{1}_{x_2^2}(\omega_1)$	1
...	...	...	1
ω_i	$\mathbf{1}_{x_1^2}(\omega_i)$	$\mathbf{1}_{x_2^2}(\omega_i)$	1
...	...	...	1
ω_n	$\mathbf{1}_{x_1^2}(\omega_n)$	$\mathbf{1}_{x_2^2}(\omega_n)$	1
Ω	n_1^2	n_2^2	n

(6.2.4)

где

$$n_{kl}^{12}=\sum_{i=1}^{n}\mathbf{1}_{x_k^1}(\omega_i)\mathbf{1}_{x_l^2}(\omega_i)=\sum_{i=1}^{n}\mathbf{1}_{x_k^1\cap x_l^2}(\omega_i)=$$

$$=\sum_{i=1}^{n}\mathbf{1}_{\mathsf{ter}(\{x_k^1,x_l^2\}/\!/\{x_k^1,x_l^2\})}(\omega_i)$$

для $k=1,2;l=1,2$.

$\star\star\star$

Пример (статистическая $(m,2)$-задача, каноническая сопряженная таблица). В силу того, что в ситуации "m экспертов и 2 категории" m-*мерная сопряженная* $\underbrace{2\times\ldots\times 2}_{m}$ *таблица* объема 2^m представляет собой табличный способ характеризации вероятностного распределения 1-го рода категориального м.с., такая таблица называется *канонической сопряженной таблицей* и рассматривается в отдельном параграфе 8.1.6.

$\star\star\star$

Пример 2 (статистическая $(2,M)$-задача). Для "2 экспертов, M категорий в n наблюдениях" 2-*мерная сопряженная* $M\times M$ *таблица* объема M^2

	x_1^1	...	x_k^1	...	x_M^1	Ω
x_1^2	n_{11}^{12}	...	n_{1k}^{12}	...	n_{1M}^{12}	n_1^2
...	...	...	...	...	...	...
x_k^2	n_{k1}^{12}	...	n_{kk}^{12}	...	n_{kM}^{12}	n_k^2
...	...	...	...	...	...	...
x_M^2	n_{M1}^{12}	...	n_{Mk}^{12}	...	n_{MM}^{12}	n_M^2
Ω	n_1^1	...	n_k^1	...	n_M^1	n

(6.2.5)

определяется набором из 2-х *экспертных* $n\times M$ *таблиц*:

	x_1^1	...	x_k^1	...	x_M^1	Ω
ω_1	$\mathbf{1}_{x_1^1}(\omega_1)$	...	$\mathbf{1}_{x_k^1}(\omega_1)$	...	$\mathbf{1}_{x_M^1}(\omega_1)$	1
...	...	...	...	...	...	...
ω_j	$\mathbf{1}_{x_1^1}(\omega_j)$	...	$\mathbf{1}_{x_k^1}(\omega_j)$	...	$\mathbf{1}_{x_M^1}(\omega_j)$	1
...	...	...	...	...	...	...
ω_n	$\mathbf{1}_{x_1^1}(\omega_n)$	...	$\mathbf{1}_{x_k^1}(\omega_n)$	...	$\mathbf{1}_{x_M^1}(\omega_n)$	1
Ω	n_1^1	...	n_k^1	...	n_M^1	n

(6.2.6)

	x_1^2	...	x_k^2	...	x_M^2	Ω
ω_1	$\mathbf{1}_{x_1^2}(\omega_1)$	...	$\mathbf{1}_{x_k^2}(\omega_1)$	...	$\mathbf{1}_{x_M^2}(\omega_1)$	1
...	...	...	...	...	...	...
ω_j	$\mathbf{1}_{x_1^2}(\omega_j)$	...	$\mathbf{1}_{x_k^2}(\omega_j)$	...	$\mathbf{1}_{x_M^2}(\omega_j)$	1
...	...	...	...	...	...	...
ω_n	$\mathbf{1}_{x_1^2}(\omega_n)$	...	$\mathbf{1}_{x_k^2}(\omega_n)$	...	$\mathbf{1}_{x_M^2}(\omega_n)$	1
Ω	n_1^2	...	n_k^2	...	n_M^2	n

(6.2.7)

где

$$n_{kl}^{12}=\sum_{j=1}^{n}\mathbf{1}_{x_k^1}(\omega_j)\mathbf{1}_{x_l^2}(\omega_j)=\sum_{j=1}^{n}\mathbf{1}_{x_k^1\cap x_l^2}(\omega_j)\qquad(6.2.8)$$

для $k=1,\ldots,M,l=1,\ldots,M$.

$\star\star\star$

Пример 3 (статистическая (m,M)-задача). Для общей ситуации "m экспертов, M категорий в n наблюдениях" элементы m-*мерной сопряженной*

$\underbrace{M\times\ldots\times M}_{m}$ *таблица* объема M^m выражаются через элементы m *экспертных* $n\times M$ *таблиц* вида (6.2.4) и (6.2.5) по аналогичной формуле:

$$n^{1\ldots m}_{k\ldots l}=\sum_{j=1}^{n}\underbrace{\mathbf{1}_{x^1_k}(\omega_j)\cdots\mathbf{1}_{x^m_l}(\omega_j)}_{m}=$$

$$=\sum_{j=1}^{n}\mathbf{1}_{\underbrace{x^1_k\cap\ldots\cap x^m_l}_{m}}(\omega_j). \qquad (6.2.9)$$

$\star\star\star$

Определение (мультиэкспертная / мультисубъектная M-категориальная таблица). Для краткости и удобства m *экспертных таблиц*[13] размера $n\times M$, которые определяют m-*мерную сопряженную* $\underbrace{M\times\ldots\times M}_{m}$ *таблицу*, обычно записывают в виде одной *мультиэкспертной/мультисубъектной* M-*категориальной* $n\times m$ *таблицы*[14]

	1	…	i	…	m
ω_1	$\mu^1(\omega_1)$	…	$\mu^i(\omega_1)$	…	$\mu^m(\omega_1)$
…	…	…	…	…	…
ω_j	$\mu^1(\omega_j)$	…	$\mu^i(\omega_j)$	…	$\mu^m(\omega_j)$
…	…	…	…	…	…
ω_n	$\mu^1(\omega_n)$	…	$\mu^i(\omega_n)$	…	$\mu^m(\omega_n)$

, (6.2.10)

которую можно записать покороче

	1	…	i	…	m
1	$\mu^1(1)$	…	$\mu^i(1)$	…	$\mu^m(1)$
…	…	…	…	…	…
j	$\mu^1(j)$	…	$\mu^i(j)$	…	$\mu^m(j)$
…	…	…	…	…	…
n	$\mu^1(n)$	…	$\mu^i(n)$	…	$\mu^m(n)$

, (6.2.10′)

где $j=1,\ldots,n$ — номер наблюдения,

$$\mu^i(j)=\mu^i(\omega_j),$$

и для $i=1,\ldots,m$

$$\mu^i(\omega_j)=\begin{cases}1, & \mathbf{1}_{x^i_1}(\omega_j)=1,\\ \ldots, & \ldots,\\ k, & \mathbf{1}_{x^i_k}(\omega_j)=1,\\ \ldots, & \ldots,\\ M, & \mathbf{1}_{x^i_M}(\omega_j)=1\end{cases} \qquad (6.2.11)$$

— *экспертная категориальная функция*

$$\mu^i:\Omega\to\{1,\ldots,M\}, \qquad (6.2.12)$$

$$\mu^i:\{1,\ldots,n\}\to\{1,\ldots,M\}, \qquad (6.2.12')$$

[13]Или две экспертных матрицы (6.2.6) и (6.2.7) при $m=2$.

[14](пример см. в Табл. 7)

определяемая индикаторами событий[15] $x^i_1,\ldots,x^i_M$ и возвращающая номер категории $\mu^i(\omega_j)=\mu^i(j)$, которой i-й эксперт оценивает j-й исход $\omega_j\in\Omega$ эксперимента, проводимого им для составления рейтинга.

Замечание. *Мультиэкспертная* M-*категориальная* $n\times m$ *таблица* (6.2.10) определяет соответствующую m-*мерную* $\underbrace{M\times\ldots\times M}_{m}$ *сопряженную таблицу*, элемент которой равны:

$$n_{\mu_1\ldots\mu_m}=n^{1\ldots m}_{\mu_1\ldots\mu_m}=$$
$$=|\{j:\{\mu_1(j),\ldots,\mu_m(j)\}=\{\mu_1,\ldots,\mu_m\};j=1,\ldots,n\}| \qquad (6.2.13)$$

— количествам наблюдений (строк) в мультиэкспертной M-категориальной $n\times m$ таблице, которые совпадают с наблюдением

$$\{\mu_1,\ldots,\mu_m\}, \qquad (6.2.14)$$

где

$$\mu_i\in\{1,\ldots,M\} \qquad (6.2.15)$$

— номер категории, выбранной i-м экспертом, $i=1,\ldots,m$.

При этом статистика $n_{\mu_1\ldots\mu_m}/n$ служит оценкой вероятностей

$$\pi_{\mu_1\ldots\mu_m}\approx n_{\mu_1\ldots\mu_m}/n, \qquad (6.2.16)$$

определяющих вероятностное распределение оцениваемого предмета с точки зрения m экспертов в рамках M категорий и составляющих соответствующую m-*мерную* $\underbrace{M\times\ldots\times M}_{m}$ *сопряженную таблицу вероятностей*. Для $m=2$ такая *сопряженная таблица вероятностей* соответствует *сопряженной таблице статистик* (6.2.5) и имеет вид

	x^1_1	…	x^1_k	…	x^1_M	Ω
x^2_1	π_{11}	…	π_{1k}	…	π_{1M}	π^2_1
…	…	…	…	…	…	…
x^2_k	π_{k1}	…	π_{kk}	…	π_{kM}	π^2_k
…	…	…	…	…	…	…
x^2_M	π_{M1}	…	π_{Mk}	…	π_{MM}	π^2_M
Ω	π^1_1	…	π^1_k	…	π^1_M	1

, (6.2.17)

где $\pi_{kl}=\pi^{12}_{kl}$ — вероятность того, что 1-й эксперт отнесет предмет к k-й, а 2-й — к l-й категории; π^1_k — вероятность того, 1-й эксперт отнесет предмет к k-й категории; π^2_l — вероятность того, 2-й эксперт отнесет предмет к l-й категории.

$\star\star\star$

[15]Заметим, что для каждого i-го эксперта ($i=1,\ldots,m$)

$$x^i_1+\ldots+x^i_M=\Omega,\quad\sum_{k=1}^{M}\mathbf{1}_{x^i_k}(\omega)=1.$$

Пример 4 (статистическая (6,20)-задача в 198 наблюдениях: кредитные рейтинги стран мира). Имеется m рейтинговых агенств[16], которые в каждой из n различных ситуаций оценивают один и тот же "предмет" в диапазоне одного и того же M-множества категорий (например, один и тот же финансовый инструмент по одной и той же шкале кредитных рейтингов). Итоговая статистика рейтингов всех m агентств имеет вид мультиэкспертной (мультисубъектной) M-категориальной $n \times m$ таблицы (6.2.10), где n — число оценок данного "предмета".

Выберем $m = 6$ рейтинговых агентств[17]:

1. Wikirating (Polling method)
2. Wikirating (SWI method)
3. Standard & Poor's
4. Moody's
5. Fitch
6. Dagong

Таблица 4: Шесть методов расчета кредитного рейтинга стран мира.

Эти агентства оценивают кредитные рейтинги $n = 198$ стран мира в рамках $M = 20$ кредитных категорий:

20. AAA	**10. BB+**
19. AA+	**9. BB**
18. AA	**8. BB−**
17. AA−	**7. B+**
16. A+	**6. B**
15. A	**5. B−**
14. A−	**4. CCC**
13. BBB+	**3. CC**
12. BBB	**2. C**
11. BBB−	**1. D**
	— not rated (n.r.)

Таблица 5: 20 кредитных категорий и их условные обозначения.

	$j\backslash i$	1	2	3	4	5	6
Afg.	1	B-	n.r.	n.r.	n.r.	n.r.	n.r.
...	...	...	...	...	...	...	...
Can.	31	BBB+	BB	AAA	AAA	AAA	AA+
...	...	...	...	...	...	...	...
Chin.	36	BBB+	BBB+	AA−	AA−	A+	AAA
...	...	...	...	...	...	...	...
Fra.	60	BBB	BB−	AA	AA+	AA+	A+
...	...	...	...	...	...	...	...
Ger.	64	BBB	BB+	AAA	AAA	AAA	AA+
...	...	...	...	...	...	...	...
Ita.	83	BB+	B	BBB	BBB	BBB+	BBB
...	...	...	...	...	...	...	...
Jap.	85	BBB+	BB	AA−	AA−	A+	AA−
...	...	...	...	...	...	...	...
Rus.	145	BBB	A−	BBB	BBB+	BBB	A
...	...	...	...	...	...	...	...
UK	188	BBB−	BB−	AAA	AA+	AA+	A+
USA	189	BB	BB−	AA+	AAA	AAA	A−
...	...	...	...	...	...	...	...
Zimb.	198	C	CC	n.r.	n.r.	n.r.	n.r.

Таблица 6: $m = 6$-экспертная 21-категориальная $n \times m = 198 \times 6$ таблица кредитных рейтингов $n = 198$ стран мира (в категориальных обозначениях на июль 2013).

	$j\backslash i$	1	2	3	4	5	6
Afg.	1	5	—	—	—	—	—
...	...	...	...	...	...	...	...
Can.	31	13	9	20	20	20	19
...	...	...	...	...	...	...	...
Chin.	36	13	13	17	17	16	20
...	...	...	...	...	...	...	...
Fra.	60	12	8	18	19	19	16
...	...	...	...	...	...	...	...
Ger.	64	12	10	20	20	20	19
...	...	...	...	...	...	...	...
Ita.	83	10	6	12	12	13	12
...	...	...	...	...	...	...	...
Jap.	85	13	9	17	17	16	17
...	...	...	...	...	...	...	...
Rus.	145	12	14	12	13	12	15
...	...	...	...	...	...	...	...
UK	188	11	8	20	19	19	16
USA	189	9	8	19	20	20	14
...	...	...	...	...	...	...	...
Zimb.	198	2	3	—	—	—	—

Таблица 7: $m = 6$-экспертная 21-категориальная $n \times m = 198 \times 6$ таблица кредитных рейтингов $n = 198$ стран мира (в цифровых обозначениях на июль 2013).

[16] *Рейтинговое агентство* — коммерческая организация, занимающаяся оценкой платежеспособности эмитентов, долговых обязательств, качества корпоративного управления, качества управления активами и т. п. Наиболее известный продукт рейтинговых агентств — это оценка платежеспособности — кредитный рейтинг. Он отражает риск невыплаты по долговому обязательству и влияет на величину процентной ставки, на стоимость и доходность долговых обязательств. При этом более высокий рейтинг соответствует меньшему риску невыплаты.

[17] Точнее, 6 методов расчета рейтинга.

7 Пример задачи общности мнений 6 рейтинговых агентств по 21 категории

7.1 Примеры 2-мерных сопряженных таблиц для кредитных рейтингов стран мира на июль 2013

11	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	13
2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2
1	1	3	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	6
-	-	-	5	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	7
-	-	-	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2
-	-	-	-	2	-	-	1	-	-	-	-	-	-	-	-	-	-	-	3
-	-	-	-	-	2	1	-	-	-	1	-	-	-	-	-	-	-	-	4
-	-	-	-	1	-	-	1	1	1	1	-	-	-	-	-	-	-	-	5
-	-	-	-	-	-	1	4	6	1	-	-	-	-	-	-	-	-	-	12
-	-	-	-	-	-	-	-	-	5	2	-	-	-	-	-	-	-	-	7
-	-	-	-	-	-	-	-	-	5	-	-	-	-	-	-	-	-	-	5
-	-	-	-	-	-	-	-	-	1	2	1	1	-	-	-	-	-	-	5
-	-	-	-	-	-	-	-	-	1	-	1	5	4	-	-	-	-	-	11
-	-	-	-	-	-	-	-	-	-	-	-	-	5	1	-	-	-	-	6
-	-	-	-	-	-	-	-	-	-	-	1	-	2	-	1	1	-	-	5
-	-	-	-	-	-	-	-	-	-	-	-	-	2	3	2	-	-	1	8
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	1
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	1
14	3	3	7	6	2	2	6	7	14	6	3	6	11	3	5	3	1	0	1	103

Таблица 8: S&P's—Moody's

11	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	13
2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2
-	1	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5
-	-	-	1	6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	7
-	-	-	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	2
-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
-	-	-	-	-	-	2	1	-	-	-	-	-	-	-	-	-	-	-	3
-	-	-	-	1	-	1	1	2	-	-	-	-	-	-	-	-	-	-	5
-	-	-	-	-	-	-	2	7	2	-	-	-	-	-	-	-	-	-	11
-	-	-	-	-	-	-	-	1	6	-	-	-	-	-	-	-	-	-	7
-	-	-	-	-	-	-	-	1	3	-	-	-	-	-	-	-	-	-	4
-	-	-	-	-	-	-	-	-	-	5	-	-	-	-	-	-	-	-	5
-	-	-	-	-	-	-	-	-	1	-	1	5	4	-	-	-	-	-	11
-	-	-	-	-	-	-	-	-	-	-	-	1	3	2	-	-	-	-	6
-	-	-	-	-	-	-	-	-	-	-	-	-	6	1	-	-	-	-	7
-	-	-	-	-	-	-	-	-	-	-	-	-	1	3	-	1	-	-	5
-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	1
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	1
13	3	4	2	8	1	3	4	11	12	5	1	7	8	8	5	0	1	0	0	96

Таблица 9: S&P's—Fitch

6	6	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	13
-	1	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	2
-	-	2	1	2	-	-	1	-	-	-	-	-	-	-	-	-	-	-	6
1	-	2	2	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	7
-	-	-	1	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	2
-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	1
-	-	-	-	1	1	1	-	-	-	-	-	-	-	-	-	-	-	-	3
-	-	-	-	-	-	1	-	2	1	-	1	-	-	-	-	-	-	-	5
-	-	-	-	-	2	3	2	2	1	-	-	-	-	-	-	-	-	-	10
-	-	-	-	-	1	-	1	1	-	1	-	2	-	-	-	-	-	-	6
-	-	-	-	-	-	-	-	-	2	-	1	1	-	-	-	-	-	-	4
-	-	-	-	-	-	-	-	-	-	1	1	-	-	1	-	-	-	-	3
-	-	-	-	-	-	-	-	-	1	3	-	5	2	-	-	-	-	-	11
-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	-	-	-	-	2
-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	-	-	-	-	3
-	-	-	-	-	-	-	-	-	-	1	-	1	-	2	2	1	-	-	7
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
7	7	4	4	5	5	7	4	5	6	6	2	9	3	4	4	2	1	0	0	85

Таблица 10: S&P's—Dagong

13	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	14
-	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3
-	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2
-	-	1	2	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	7
-	-	-	-	4	1	-	-	-	-	-	-	-	-	-	-	-	-	-	5
-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	1
-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	1
-	-	-	-	-	-	1	-	4	-	-	-	-	-	-	-	-	-	-	5
-	-	-	-	-	-	-	2	4	1	-	-	-	-	-	-	-	-	-	7
-	-	-	-	-	-	-	-	3	9	1	-	-	-	-	-	-	-	-	13
-	-	-	-	-	-	2	-	2	2	-	-	-	-	-	-	-	-	-	6
-	-	-	-	-	-	-	-	-	-	-	1	1	-	-	-	-	-	-	2
-	-	-	-	-	-	-	-	-	-	1	-	1	2	-	-	-	-	-	4
-	-	-	-	-	-	-	-	-	-	-	-	2	2	2	-	-	-	-	6
-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	1	-	-	-	2
-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	1	-	-	2
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	-	-	-	2
-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	1
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	1
13	3	4	2	8	1	3	4	11	12	4	1	5	5	2	5	0	1	0	0	84

Таблица 11: Moddy's—Fitch

5	7	1	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	14
1	-	-	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3
-	-	1	1	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	3
1	-	2	3	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	7
-	-	-	-	1	1	2	-	-	-	-	-	-	1	-	-	-	-	-	5
-	-	-	-	-	1	1	-	-	-	-	-	-	-	-	-	-	-	-	2
-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
-	-	-	-	-	2	1	-	1	1	-	-	-	-	-	-	-	-	-	5
-	-	-	-	-	-	2	1	3	1	-	-	-	-	-	-	-	-	-	7
-	-	-	-	-	-	-	2	-	2	2	2	2	-	-	-	-	-	-	11
-	-	-	-	-	-	-	-	2	1	-	-	1	-	-	-	-	-	-	4
-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	-	-	-	-	2
-	-	-	-	-	-	-	-	-	1	1	-	3	-	-	-	-	-	-	5
-	-	-	-	-	-	-	-	-	-	1	-	1	3	-	-	-	-	-	5
-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	1	-	-	-	2
-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	1	-	-	2
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1	-	-	3
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	1
7	7	4	4	5	5	7	4	6	6	5	2	8	3	2	4	2	1	0	0	82

Таблица 12: Moody's—Dagong

5	7	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	13
1	-	-	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3
-	-	2	-	1	-	-	1	-	-	-	-	-	-	-	-	-	-	-	4
-	-	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2
1	-	1	2	1	1	1	-	-	-	-	-	-	-	-	-	-	-	-	7
-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1
-	-	-	-	1	-	1	-	1	-	-	-	-	-	-	-	-	-	-	3
-	-	-	-	-	-	-	3	-	-	-	-	-	-	-	-	-	-	-	3
-	-	-	-	3	3	1	-	3	-	1	-	-	-	-	-	-	-	-	11
-	-	-	-	-	-	-	2	1	1	2	1	3	-	-	-	-	-	-	10
-	-	-	-	-	-	-	-	1	1	-	-	-	-	-	-	-	-	-	2
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	2	-	2	1	-	-	-	-	-	-	5
-	-	-	-	-	-	-	-	1	1	-	2	1	1	-	-	-	-	-	6
-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	1
-	-	-	-	-	-	-	-	-	-	-	-	-	-	3	-	1	-	-	4
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	1
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
7	7	4	3	5	4	7	4	5	6	6	2	7	2	2	3	1	1	0	0	76

Таблица 13: Fitch—Dagong

-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	1
-	-	1	1	-	-	-	1	-	-	-	-	1	-	-	-	-	-	-	4
2	2	3	2	1	1	3	3	4	1	2	-	1	-	-	-	-	-	-	26
-	-	-	-	-	1	3	2	3	2	1	3	-	-	-	-	-	-	-	15
-	-	-	-	-	1	-	-	2	-	1	-	2	-	-	-	-	-	-	6
-	-	-	-	-	1	1	-	1	1	3	3	1	-	2	-	-	-	-	13
-	-	-	-	-	-	-	-	2	1	1	1	2	1	-	-	1	-	-	9
-	-	-	-	-	-	2	-	-	2	-	3	-	2	2	-	1	-	-	12
-	-	-	-	-	-	-	-	-	1	3	1	2	3	-	-	-	-	-	10
-	-	-	-	-	-	-	-	-	2	1	-	1	-	-	-	-	-	-	4
-	-	-	-	-	-	-	-	1	-	-	-	-	1	-	-	-	1	-	3
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	1
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
2	2	5	3	1	3	5	6	12	11	11	14	11	9	8	0	2	2	1	0	108

Таблица 14: WR-Poll—WR-SWI

-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	1
2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	1
1	-	1	1	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	4
9	1	2	5	1	-	2	-	6	2	-	-	-	-	-	1	-	-	-	29
1	-	3	1	-	2	-	3	1	1	1	1	2	-	1	-	-	-	-	17
1	-	-	-	-	-	2	1	-	1	1	-	-	-	-	-	-	-	-	6
-	-	-	-	1	-	-	-	4	1	1	2	2	-	-	2	-	-	-	13
-	1	-	-	-	-	-	-	-	-	2	1	4	2	2	-	-	-	-	12
-	-	-	-	-	-	-	1	-	1	-	2	1	1	2	-	1	-	-	9
-	-	-	-	-	-	-	-	-	-	-	-	3	3	2	3	-	-	1	12
-	-	-	-	-	-	-	-	-	-	-	-	1	-	1	-	-	-	-	2
-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	2	-	-	-	3
-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	1
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
14	2	6	7	2	2	5	5	12	7	5	6	13	7	9	8	1	0	1	112

Таблица 15: WR-Poll—S&P's

-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	1
1	-	1	1	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	4
10	1	1	5	1	1	1	2	3	3	-	-	-	-	1	-	-	-	-	29
1	1	1	1	3	-	-	2	-	3	2	-	1	1	-	-	1	-	-	17
-	1	-	-	1	1	-	-	-	2	1	-	-	-	-	-	-	-	-	6
-	-	-	-	1	-	1	1	2	2	1	-	1	2	2	1	-	-	-	14
1	-	-	-	-	-	-	-	-	2	1	-	2	2	-	-	1	-	-	9
-	-	-	-	-	-	-	-	1	-	1	2	-	1	1	2	-	-	-	8
-	-	-	-	-	-	-	-	-	1	-	1	1	2	-	2	2	1	-	10
-	-	-	-	-	-	-	-	-	-	-	1	1	-	-	-	-	-	-	2
-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	1	2
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
14	3	3	7	6	2	2	5	7	14	6	4	6	8	3	7	4	1	0	103

Таблица 16: WR-Poll—Moody's

-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	1
1	-	-	-	1	-	-	-	-	1	-	-	-	-	-	-	-	-	-	3
9	1	2	2	4	1	2	1	5	2	-	-	-	-	-	-	-	1	-	30
1	1	2	-	2	-	1	1	2	2	1	-	-	2	-	1	-	-	-	16
-	1	-	-	1	-	-	1	-	2	-	-	-	-	-	-	-	-	-	5
-	-	-	-	-	1	-	1	2	2	2	-	-	2	-	1	-	-	-	11
1	-	-	-	-	-	-	-	-	1	1	-	2	2	3	-	-	-	-	10
-	-	-	-	-	-	-	-	1	2	1	1	-	-	1	1	-	-	-	7
-	-	-	-	-	-	-	-	-	1	-	-	4	1	-	1	-	-	-	7
-	-	-	-	-	-	-	-	-	-	-	-	1	-	1	-	-	-	-	2
-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	1	-	-	-	2
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
13	3	4	2	8	2	3	4	11	13	5	1	7	8	5	5	0	1	0	95

Таблица 17: WR-Poll—Fitch

-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	1
1	-	-	-	1	-	-	-	-	1	-	-	-	-	-	-	-	-	-	3
9	1	2	2	4	1	2	1	5	2	-	-	-	-	-	-	-	1	-	30
1	1	2	-	2	-	1	1	2	2	1	-	-	2	-	1	-	-	-	16
-	1	-	-	1	-	-	1	-	2	-	-	-	-	-	-	-	-	-	5
-	-	-	-	-	1	-	1	2	2	2	-	-	2	-	1	-	-	-	11
1	-	-	-	-	-	-	-	-	1	1	-	2	2	3	-	-	-	-	10
-	-	-	-	-	-	-	-	1	2	1	1	-	-	1	1	-	-	-	7
-	-	-	-	-	-	-	-	-	1	-	-	4	1	-	1	-	-	-	7
-	-	-	-	-	-	-	-	-	-	-	-	1	-	1	-	-	-	-	2
-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	1	-	-	-	2
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
13	3	4	2	8	2	3	4	11	13	5	1	7	8	5	5	0	1	0	95

Таблица 18: WR-Poll—Dagong

2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2
1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2
3	-	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5
-	-	2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3
-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
1	-	-	-	-	1	-	-	-	1	-	-	-	-	-	-	-	-	-	3
1	-	-	-	-	-	1	2	-	-	-	-	-	-	-	-	-	-	-	4
-	-	-	1	-	-	2	1	2	-	-	-	-	-	1	-	-	-	-	7
2	-	-	1	-	-	-	1	2	-	3	-	-	1	-	-	-	-	-	12
-	-	-	-	-	1	1	3	-	1	2	-	-	-	-	-	-	-	-	9
1	1	-	-	1	-	-	1	-	1	1	-	1	2	1	-	-	-	-	11
1	-	-	1	-	-	-	1	-	-	-	5	2	2	1	-	-	-	-	13
1	1	2	-	-	-	-	1	1	1	-	1	-	1	1	-	-	-	-	10
-	-	-	-	-	-	-	-	3	-	-	1	1	1	-	-	-	-	1	7
-	-	-	-	-	-	-	1	-	-	2	-	2	-	2	-	-	-	-	7
-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	1
-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	1	-	-	2
-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	1
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
13	2	5	6	2	2	4	5	10	7	4	6	10	8	8	6	1	0	0	100

Таблица 19: WR-SWI—S&P's

2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2
1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2
2	1	-	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5
1	-	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3
-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
1	-	-	-	1	-	-	-	1	-	-	-	-	-	-	-	-	-	-	3
1	-	-	-	-	-	1	2	-	-	-	-	-	-	-	-	-	-	-	4
-	-	-	1	-	-	2	1	2	-	-	-	-	-	1	-	-	-	-	7
2	-	-	1	-	-	-	1	2	-	3	-	-	1	-	-	-	-	-	12
-	-	-	-	1	1	1	3	-	1	2	-	-	-	-	-	-	-	-	10
1	1	-	-	1	-	-	1	-	1	1	2	1	-	-	-	-	-	-	11
1	-	-	1	-	-	-	1	-	-	-	5	2	2	1	-	-	-	-	13
1	1	2	-	-	-	-	1	1	1	-	1	-	1	1	-	-	-	-	10
-	-	-	-	-	-	-	-	3	-	-	1	1	1	-	-	-	-	1	7
-	-	-	-	-	-	-	1	-	-	2	-	2	-	2	-	-	-	-	7
-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	1
-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	1	-	-	2
-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	1
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
14	3	2	6	5	2	1	5	8	13	6	4	5	8	3	5	3	1	0	95

Таблица 20: WR-SWI—Moody's

2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2
1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2
2	1	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5
-	-	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2
-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
1	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	2
1	-	-	-	-	-	-	2	1	-	-	-	-	-	-	-	-	-	-	4
-	-	-	-	1	-	-	-	3	-	1	-	-	-	-	1	-	-	-	6
2	-	-	-	1	-	1	1	-	3	-	-	-	2	-	-	-	-	-	10
-	-	-	-	1	-	2	1	1	3	2	-	-	-	-	-	-	-	-	10
2	-	-	-	1	1	-	-	1	2	-	-	1	-	3	1	-	-	-	12
1	-	-	-	1	-	-	-	1	-	-	1	2	-	-	-	-	-	-	7
1	2	1	-	-	-	-	-	1	1	1	1	-	-	-	-	-	-	-	8
-	-	-	-	-	-	-	-	-	1	2	-	-	4	-	1	-	-	-	8
-	-	-	-	-	-	-	-	1	-	-	1	-	-	1	-	1	-	-	4
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	1
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	-	2
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	1
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
13	3	3	2	7	1	3	4	10	13	5	1	6	5	6	5	0	0	0	87

Таблица 21: WR-SWI—Fitch

2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2
-	1	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	2
2	1	1	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	5
-	-	2	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3
-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
-	1	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	2
1	-	-	-	-	1	-	-	2	-	-	-	-	-	-	-	-	-	-	4
1	-	-	-	-	-	-	1	-	1	-	1	-	-	-	1	-	-	-	5
1	1	-	-	2	-	-	-	-	1	1	3	1	-	-	-	-	-	-	11
-	-	-	-	-	-	2	1	4	1	-	-	-	-	-	-	-	-	-	8
-	2	-	-	-	-	2	-	-	1	1	-	1	-	-	1	-	-	-	8
-	1	-	1	-	1	-	-	-	2	-	1	1	1	-	-	-	-	-	8
-	-	-	-	3	-	2	-	1	1	-	-	1	-	-	-	-	-	-	8
-	-	-	-	-	1	-	1	1	-	-	1	1	-	-	-	-	-	-	5
-	-	-	-	-	-	-	1	-	1	-	-	-	1	2	2	-	-	-	7
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0
7	7	3	3	5	5	6	4	7	7	5	2	7	3	2	4	2	1	1	81

Таблица 22: WR-SWI—Dagong

8 Сопряженное м.с., определяющее сопряженные таблицы, и его структура (категориальные и экспертные м.с.)

8.1 Сопряженное м.с.: определения и обозначения

Определение (сопряженное, категориальные и экспертные м.с.). Задача общности мнений m экспертов в M категориях, поставленная для m-мерной сопряженной $\underbrace{M\times\ldots\times M}_{m}$ таблицы, характеризуется mM-м.с.

$$\mathbb{X}=\sum_{j=1}^{m}\sum_{i=1}^{M}\{x_i^j\}=\begin{Bmatrix} x_1^1, & \ldots, & x_M^1 \\ \ldots & \ldots & \ldots \\ x_1^m, & \ldots, & x_M^m \end{Bmatrix}, \qquad (8.1.1)$$

которое мы будем как и таблицу называть *сопряженным mM-м.с.*

Сопряженное mM-м.с. $\mathbb{X}$ можно разложить в сумму подмножеств двумя способами ("по столбцам" и "по строкам"):

$$\mathbb{X}=\begin{cases}\sum\limits_{i=1}^{M}\mathfrak{X}_i=\sum\limits_{i=1}^{M}\{x_i^1,\ldots,x_i^m\}, & (8.1.1a)\\ \sum\limits_{j=1}^{m}\mathbf{X}^j=\sum\limits_{j=1}^{m}\{x_1^j,\ldots,x_M^j\}, & (8.1.1b)\end{cases}$$

- либо в виде суммы из M *категориальных m-множеств событий*:

$$\mathfrak{X}_1,\ldots,\mathfrak{X}_M, \qquad (8.1.2)$$

i-тое из которых

$$\mathfrak{X}_i=\{x_i^1,\ldots,x_i^m\} \qquad (8.1.3)$$

— это m-м.с., характеризующее мнение всех m экспертов об одной i-той категории;

- либо в виде суммы из m *экспертных M-множеств событий*:

$$\mathbf{X}^1,\ldots,\mathbf{X}^m, \qquad (8.1.4)$$

где

$$\mathbf{X}^j=\{x_1^j,\ldots,x_M^j\} \qquad (8.1.5)$$

— это M-множество непересекающихся событий, образующих M-разбиение

$$\Omega=x_1^j+\ldots+x_M^j \qquad (8.1.6)$$

для каждого $j=1,\ldots,m$, которое характеризует мнение одного j-го эксперта обо всех M категориях.

Определение (клеточные м.с.). m-подмножества сопряженного м.с. $\mathbb{X}$ вида:

$$X_{i_1\ldots i_m}=X_{i_1\ldots i_m}^{1\ldots m}=\{x_{i_1}^1,\ldots,x_{i_m}^m\}\subset\mathbb{X},$$

где $i_k=1,\ldots,M, k=1,\ldots,m$, называются *клеточными подмножествами* м.с. $\mathbb{X}$

Понятие *клеточного подмножества* обобщает понятие *категориального*:

$$\mathfrak{X}_i=\{x_i^1,\ldots,x_i^m\}=X_{i\ldots i}^{1\ldots m}$$

для $i=1,\ldots,M$. Поэтому категориальные м.с. $\mathfrak{X}_1,\ldots,\mathfrak{X}_M$ можно также назвать *диагональными клеточными подмножествами* сопряженного м.с. $\mathbb{X}$.

8.1.1 Свойства клеточных м.с.

Свойство 1. Для любого клеточного m-м.с. $X_{i_1\ldots i_m}$ и любого экспертного м.с. $\mathbf{X}^j, j=1,\ldots,m$

$$\left|X_{i_1\ldots i_m}\cap\mathbf{X}^j\right|=1,$$

а поэтому

$$\sum_{j=1}^{m}\left|X_{i_1\ldots i_m}\cap\mathbf{X}^j\right|=m.$$

Доказательство следует из определения клеточного м.с.

Свойство 2. Террасные события 1-го и 2-го рода, порожденные м.с. $\mathbb{X}$ и занумерованные одним и тем же клеточным м.с. $X_{i_1\ldots i_m}$ совпадают

$$\mathsf{ter}(X_{i_1\ldots i_m}/\!\!/\mathbb{X})=\mathsf{ter}_{X_{i_1\ldots i_m}/\!\!/\mathbb{X}}. \qquad (8.1.7)$$

Доказательство. По определению террасных событий 1-го рода

$$\mathsf{ter}(X_{i_1\ldots i_m}/\!\!/\mathbb{X})=\bigcap_{x\in X_{i_1\ldots i_m}}x\bigcap_{x\in\mathbb{X}-X_{i_1\ldots i_m}}x^c.$$

Распишем первое пересечение:

$$\bigcap_{x\in X_{i_1\ldots i_m}}x=x_{i_1}^1\cap\ldots\cap x_{i_m}^m,$$

и, так как $\mathbb{X}=\mathbf{X}^1+\ldots+\mathbf{X}^j+\ldots+\mathbf{X}^m$, а $x_{i_j}^j\in\mathbf{X}^j$ для $j=1,\ldots,m$, — второе пересечение:

$$\bigcap_{x\in\mathbb{X}-X_{i_1\ldots i_m}}x^c=\bigcap_{j=1}^{m}\left(\bigcap_{x\in\mathbf{X}^j-\left\{x_{i_j}^j\right\}}x^c\right).$$

Подставим полученное в определение террасного события 1-го рода, получим

$$\mathsf{ter}(X_{i_1\ldots i_m}/\!\!/\mathbb{X})=\bigcap_{j=1}^{m}\left(x_{i_j}^j\cap\left(\bigcap_{x\in\mathbf{X}^j-\left\{x_{i_j}^j\right\}}x^c\right)\right).$$

Заметим, что для $j = 1, \ldots, m$

$$x^j_{i_j} = \bigcap_{x \in \mathbf{X}^j - \left\{x^j_{i_j}\right\}} x^c, \qquad (8.1.8)$$

в силу того, что экспертные м.с. $\mathbf{X}^j = \{x^j_1, \ldots, x^j_M\}$ образованы непересекающимися событиями, покрывающими Ω

$$\sum_{i=1}^{M} x^j_i = \Omega,$$

и поэтому пересечение дополнений всех событий из $\mathbf{X}^j$, кроме одного $x^j_{i_j}$, равно этому событию. В итоге,

$$\mathsf{ter}(X_{i_1 \ldots i_m} /\!\!/ \mathbb{X}) = \bigcap_{j=1}^{m} x^j_{i_j} = \bigcap_{x \in X_{i_1 \ldots i_m}} x = \mathsf{ter}_{X_{i_1 \ldots i_m} /\!\!/ \mathbb{X}}$$

— свойство 2 доказано.

8.1.2 Клеточные террасные события и клеточные вероятности

Определение (клеточные террасные события и клеточные вероятности). Террасные события 1-го и 2-го рода, порожденные сопряженным м.с. $\mathbb{X}$ и расположенные в клетках сопряженной таблицы, называются *клеточными террасными событиями*, а их вероятности — *клеточными вероятностями*:

$$\pi_{i_1 \ldots i_m} = \pi^{1 \ldots m}_{i_1 \ldots i_m} =$$

$$= \mathbf{P}\left(\mathsf{ter}(X_{i_1 \ldots i_m} /\!\!/ \mathbb{X})\right) = \mathbf{P}(\mathsf{ter}_{X_{i_1 \ldots i_m} /\!\!/ \mathbb{X}}).$$

Обозначим также

$$\pi^j_i = \mathbf{P}(x^j_i)$$

— вероятности событий $x^j_i \in \mathbb{X}, i = 1, \ldots, M,\ j = 1, \ldots, m$.

Из формул обращения Мёбиуса следует, что

$$\pi^j_i = \underbrace{\sum \ldots \sum}_{m-1} \pi_{i_1 \ldots i_m}, \qquad (8.1.9)$$

где $(m-1)$-кратное суммирование клеточных вероятностей $\pi_{i_1 \ldots i_m}$ распространяется на все индексы $i_1, \ldots, i_m$, за исключением одного индекса i_j, значение которого зафиксировано: $i_j = i$.

8.1.3 Пример "кошка–собака–заяц" в событийных обозначениях

В примере "кошка–собака–заяц" сопряженные (2,3)-таблицы клеточных террасных событий и клеточных вероятностей имеют вид, показанный на рис. 23.

	x^2_1	x^2_2	x^2_3	Ω
x^1_1	$x^1_1 \cap x^2_1$	$x^1_1 \cap x^2_2$	$x^1_1 \cap x^2_3$	x^1_1
x^1_2	$x^1_2 \cap x^2_1$	$x^1_2 \cap x^2_2$	$x^1_2 \cap x^2_3$	x^1_2
x^1_3	$x^1_3 \cap x^2_1$	$x^1_3 \cap x^2_2$	$x^1_3 \cap x^2_3$	x^1_3
Ω	x^2_1	x^2_2	x^2_3	1

	x^2_1	x^2_2	x^2_3	Ω
x^1_1	π^{12}_{11}	π^{12}_{12}	π^{12}_{13}	π^1_1
x^1_2	π^{12}_{21}	π^{12}_{22}	π^{12}_{23}	π^1_2
x^1_3	π^{12}_{31}	π^{12}_{32}	π^{12}_{33}	π^1_3
Ω	π^2_1	π^2_2	π^2_3	1

Таблица 23: (2,3)-таблица террасного (слева) и вероятностного (справа) распределений *сопряженного м.с.* $\mathbb{X} = \{x^1_1, x^2_1\} + \{x^1_2, x^2_2\} + \{x^1_3, x^2_3\} = \mathfrak{X}_1 + \mathfrak{X}_2 + \mathfrak{X}_3 = \mathbf{X^1} + \mathbf{X^2} = \{x^1_1, x^1_2, x^1_3\} + \{x^2_1, x^2_2, x^2_3\}$, составленного как из 3-х *категориальных м.с.* $\mathfrak{X}_1, \mathfrak{X}_2, \mathfrak{X}_3$, так и из 2-х *экспертных м.с.* $\mathbf{X^1}, \mathbf{X^2}$, в задаче общности мнений двух экспертов по трем категориям для примера "*кошка (1), собака (2), заяц (3)*"; где $\pi^{12}_{ij} = \mathbf{P}(x^1_i \cap x^2_j), \pi^1_i = \mathbf{P}(x^1_i), \pi^2_j = \mathbf{P}(x^2_j),\ i = 1, 2, 3,\ j = 1, 2$.

	x^2_1	x^2_2	x^2_2	Ω
x^1_1	$\square$ π^{12}_{11} $\square$	$\boxminus$ π^{12}_{12} $\square$	$\boxplus$ π^{12}_{13} $\square$	π^1_1
x^1_2	$\square$ π^{12}_{21} $\boxminus$	$\boxminus$ π^{12}_{22} $\boxminus$	$\boxplus$ π^{12}_{23} $\boxminus$	π^1_2
x^1_3	$\square$ π^{12}_{31} $\boxplus$	$\boxminus$ π^{12}_{32} $\boxplus$	$\boxplus$ π^{12}_{33} $\boxplus$	π^1_3
Ω	π^2_1	π^2_2	π^2_3	1

Таблица 24: Событийная структура сопряженной $(2,3)$-таблицы, которая характеризуется сопряженным 6-м.с. $\mathbb{X} = \mathfrak{X}_1 + \mathfrak{X}_2 + \mathfrak{X}_3$. События из трех категориальных дуплетов событий $\mathfrak{X}_1 = \{x^1_1, x^2_1\}, \mathfrak{X}_2 = \{x^1_2, x^2_2\}$ и $\mathfrak{X}_3 = \{x^1_3, x^2_3\}$ помечены символами $\square, \boxminus$ и $\boxplus$ соответственно: первый эксперт внизу клетки, второй эксперт вверху клетки.

8.1.4 Событийная структура сопряженного м.с.

8.1.5 Сопряженные м.с. с фиксированной "диагональной" вероятностью

Рассмотрим класс

$$\mathcal{C}_\alpha(m, M) = \left\{ \mathbb{X}(m, M) : \sum_{i=1}^{M} \pi^{1 \ldots m}_{i \ldots i} = \alpha \right\}$$

сопряженных м.с. $\mathbb{X} = \mathbb{X}(m, M)$, характеризуемых сопряженной (m, M)-таблицей, которые имеют "диагональную" вероятность, равную

$$\sum_{i=1}^{M} \pi^{1 \ldots m}_{i \ldots i} = \alpha,$$

где $\alpha \in [0, 1]$.

Ясно, что

$$\mathcal{C}_\alpha(m,M)=\sum_{\alpha_1+...+\alpha_M=\alpha}\mathcal{C}_{\alpha_1...\alpha_M}(m,M),$$

где

$$\mathcal{C}_{\alpha_1...\alpha_M}(m,M)=\left\{\mathbb{X}(m,M):\pi^{1...m}_{i...i}=\alpha_i,i=1,...,M\right\}$$

подкласс сопряженных м.с. $\mathbb{X}=\mathbb{X}(m,M)$, характеризуемых сопряженной (m,M)-таблицей, "диагональные" вероятности которых равны

$$\pi^{1...m}_{1...1}=\alpha_1,\ \ldots,\ \pi^{1...m}_{M...M}=\alpha_M.$$

8.1.6 Каноническое сопряженное м.с. для канонической многомерной сопряженной таблицы

Определение (каноническая сопряженная таблица, $(m,2)$-задача общности). В ситуации "m экспертов и 2 категории" *m-мерная сопряженная* $\underbrace{2\times\ldots\times2}_{m}$ *таблица* объема 2^m, т.е. сопряженная $(m,2)$-таблица, представляет собой табличную характеризацию вероятностного распределения 1-го рода категориального м.с. $\mathfrak{X}_1$ и называется *m-мерной канонической сопряженной таблицей*, а соответствующее ей сопряженное м.с. $\mathbb{X}=\mathfrak{X}_1+\mathfrak{X}_2$ называется *каноническим сопряженным $2m$-м.с.*

Лемма (дополнительность по Минковскому категориальных м.с. для m-мерной канонической сопряженной таблицы). *Оба категориальных м.с. $\mathfrak{X}_1=\{x^1_1,x^2_1\}$ и $\mathfrak{X}_2=\{x^1_2,x^2_2\}$ для m-мерной канонической сопряженной таблицы взаимно дополнительны по Минковскому:*

$$\mathfrak{X}_2=\mathfrak{X}_1^{(c)},$$

где $\mathfrak{X}_1^{(c)}=\{(x^1_1)^c,(x^2_1)^c\}$ — дополнение м.с. $\mathfrak{X}_1$ по Минковскому.

Доказательство. По определению (8.1.6) экспертные м.с. $\mathbf{X}^1=\{x^1_1,x^1_2\}$ и $\mathbf{X}^2=\{x^2_1,x^2_2\}$ для m-мерной сопряженной таблицы состоят из непересекающихся событий:

$$x^1_2=\Omega-x^1_1=(x^1_1)^c,\quad x^2_2=\Omega-x^2_1=(x^2_1)^c.$$

Отсюда

$$\mathfrak{X}_2=\{x^1_2,x^2_2\}=\{(x^1_1)^c,(x^2_1)^c\}=\mathfrak{X}_1^{(c)},$$

что и требуется.

Сет-расстояние категориальных м.с. Для m-мерной канонической сопряженной таблицы значения показателей общности категориальных множеств, которые основаны на понятии сет-расстояния, определяются вероятностным распределением только одного из них в силу утверждения леммы. Например, оба сет-расстояния определяются, либо по формулам:

$$\Delta_{\mathfrak{X}_1}=u_{\mathfrak{X}_1}-\langle p\rangle_{\mathfrak{X}_1}$$

$$\Delta_{\mathfrak{X}_2}=\langle p\rangle_{\mathfrak{X}_1}-p_{\mathfrak{X}_1},$$

либо по формулам

$$\Delta_{\mathfrak{X}_2}=u_{\mathfrak{X}_2}-\langle p\rangle_{\mathfrak{X}_2},$$

$$\Delta_{\mathfrak{X}_1}=\langle p\rangle_{\mathfrak{X}_2}-p_{\mathfrak{X}_2}.$$

Действительно, по определению сет-расстояния

$$\Delta_{\mathfrak{X}_2}=\Delta_{(\mathfrak{X}_1)^{(c)}}=u_{(\mathfrak{X}_1)^{(c)}}-\langle p\rangle_{(\mathfrak{X}_1)^{(c)}}=$$

$$=1-p_{\mathfrak{X}_1}-1+\langle p\rangle_{\mathfrak{X}_1}=\langle p\rangle_{\mathfrak{X}_1}-p_{\mathfrak{X}_1},$$

$$\Delta_{\mathfrak{X}_1}=\Delta_{(\mathfrak{X}_2)^{(c)}}=u_{(\mathfrak{X}_2)^{(c)}}-\langle p\rangle_{(\mathfrak{X}_2)^{(c)}}=$$

$$=1-p_{\mathfrak{X}_2}-1+\langle p\rangle_{\mathfrak{X}_2}=\langle p\rangle_{\mathfrak{X}_2}-p_{\mathfrak{X}_2}.$$

8.2 Случайный эксперимент, порождающий сопряженное м.с.

Рассмотрим случайный эксперимент, который порождает (m,M)-задачу общности мнений m экспертов по M категориям, определяемую m-мерной сопряженной $\underbrace{M\times...\times M}_{m}$ таблицей клеточных вероятностей.

Такой случайный эксперимент, порождающий соответствующее сопряженное M^m-м.с. $\mathbb{X}$, вероятностное распределение которого составлено из M^m *вероятностей клеточных террасных событий (клеточных вероятностей)*, расположенных в клетках сопряженной таблицы, будем называть *сопряженным случайным (m,M)-экспериментом.*

К участникам сопряженного случайного эксперимента относятся m экспертов и один экспериментатор. Результат сопряженного случайного эксперимента определяется совокупным поведением экспертов и экспериментатора в рамках эксперимента, которое выражается в оценке экспертами категорий предметов, выставляемых экспериментатором.

Полезно заметить, что сопряженный случайный эксперимент, где экспериментатору известны категории предметов, которые он выставляет экспертам для оценки, следует отличать от сопряженного случайного эксперимента, где экспериментатору эти категории неизвестны.

В первой ситуации экспериментатора можно считать либо экспертом, который безошибочно оценивает категории выставляемых предметов, либо учителем, который в состоянии сравнить результаты экспертных оценок с эталоном, либо просто генератором выставляемых предметов с известным вероятностным распределением их по категориям. В этой ситуации возникает задача оценки *правильности* мнений экспертов о категориях выставляемых предметов, которая в данной работе не рассматривается.

Во второй ситуации экспериментатор выступает в роли случайного генератора выставляемых предметов,

вероятностное распределение которых по категориям неизвестно. В этой ситуации и возникает задача оценки *общности* мнений экспертов о категориях выставляемых предметов, которой посвящена данная работа.

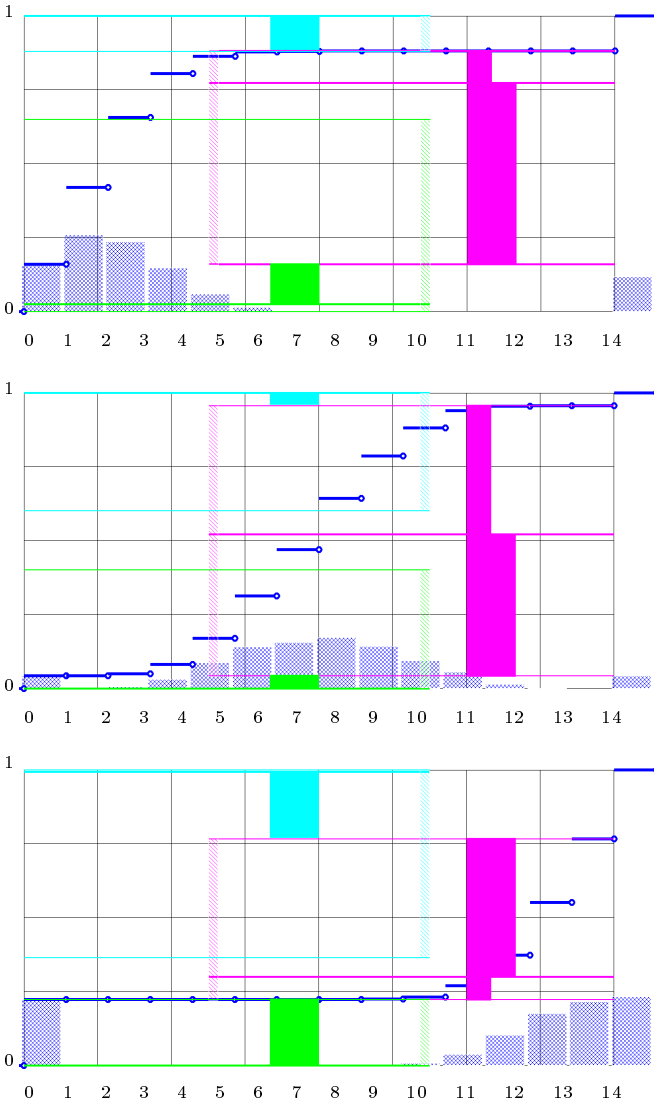


Рисунок 10: Шкалы зависимости и общности для 14-м.с. (фуксия — общность, аква — зависимость пересечения, лайм — зависимость объединения) и их сравнение с вероятностями и ф.р. его мощностного индикатора (синий). Сверху вниз — три варианта вероятностного распределения м.с. с малыми, средними и большими вероятностями событий.

8.2.1 Несимметричные и симметричные определения показателей вероятностной зависимости и общности событий

Несимметричные определения ковариации, Фреше-корреляции и сет-расстояния. Ковариации относительно пересечения и объединения:

$$\mathsf{Kov}^{\cap}_X = p_X - p^{\star}_X,$$

$$\mathsf{Kov}^{\cup}_X = u^{\star}_X - u_X.$$

Сет-расстояния относительно пересечения и объединения:

$$\Delta^{\cap}_X = \langle p\rangle_X - p_X,$$

$$\Delta^{\cup}_X = u_X - \langle p\rangle_X,$$

$$\Delta^{\cap\star}_X = \langle p\rangle_X - p^{\star}_X,$$

$$\Delta^{\cup\star}_X = u^{\star}_X - \langle p\rangle_X.$$

Метрическая интерпретация несимметричных ковариаций:

$$\mathsf{Kov}^{\cap}_X = \Delta^{\cap\star}_X - \Delta^{\cap}_X,$$

$$\mathsf{Kov}^{\cup}_X = \Delta^{\cup\star}_X - \Delta^{\cup}_X.$$

Первый вариант симметричных определений ковариации, Фреше-корреляции и сет-расстояния (сумма). Определение (симметричное сет-расстояние, сумма):

$$\Delta^{\Sigma}_X = \Delta^{\cup}_X + \Delta^{\cap}_X = u_X - p_X,$$

$$\Delta^{\Sigma\star}_X = \Delta^{\cup\star}_X + \Delta^{\cap\star}_X = u^{\star}_X - p^{\star}_X.$$

Определение (симметричная ковариация, сумма):

$$\mathsf{Kov}^{\Sigma}_X = \mathsf{Kov}^{\cup}_X + \mathsf{Kov}^{\cap}_X.$$

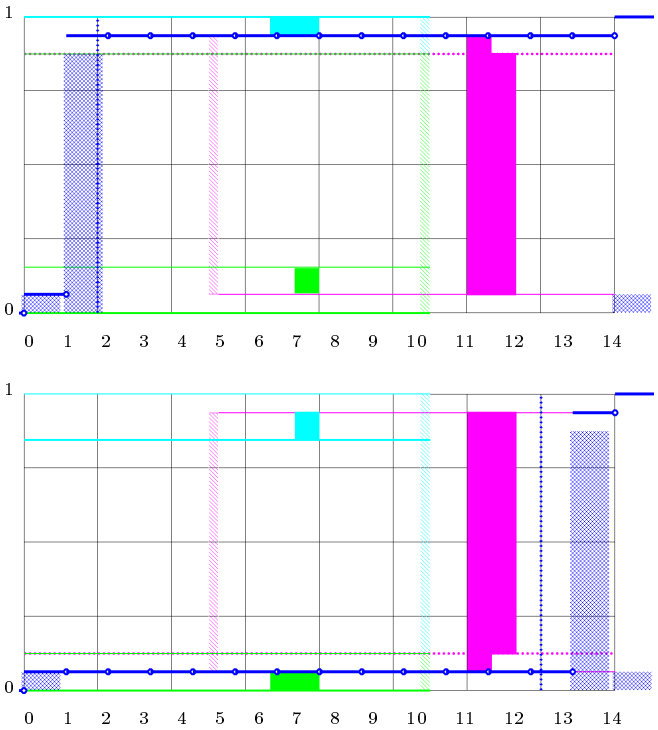


Рисунок 11: Шкалы зависимости и общности для 14-м.с. $X \subseteq \mathfrak{X}$ (фуксия — общность, аква — зависимость пересечения, лайм — зависимость объединения) и их сравнение с распределением вероятностей и ф.р. его мощностного индикатора (синий). Сверху вниз — два варианта вероятностного распределения 14-м.с. X: "сковорода" ($p_x = 1/8; p_X = 1/16; u_X = 15/16; p(\{x\}/\!\!/X) = 1/16; x \in X; \mathsf{Kov}^{\cap}_X > 0, \mathsf{Kov}^{\cup}_X < 0$) и "торт" ($p_x = 7/8; p_X = 1/16; u_X = 15/16; p(X - \{x\}/\!\!/X) = 1/16; x \in X; \mathsf{Kov}^{\cap}_X < 0, \mathsf{Kov}^{\cup}_X > 0$); см. детали в табл. 25.

	“сковорода”	“торт”
$p_x, x \in X$	1/8	7/8
шкала зависимости пересечения:		
$p^\star_X$	$(1/8)^{14}$	$(7/8)^{14} \approx 0.154$
p_X	1/16	1/16
p^-_X	0	0
p^+_X	1/8	7/8
шкала зависимости объединения:		
$u^\star_X$	≈ 0.846	≈ 1
u_X	15/16	15/16
u^-_X	1/8	7/8
u^+_X	1	1
частичные ковариации:		
$\mathsf{Kov}^\cap_X$	1/16	≈ -0.092
$\mathsf{Kov}^\cup_X$	≈ -0.092	1/16
шкала общности:		
$\langle p \rangle_X$	1/8	7/8
частичные сет-расстояния:		
$\Delta^\cap_X$	1/16	13/16
$\Delta^\cup_X$	13/16	1/16
$\Delta^{\cap\star}_X$	1/8	≈ 0.721
$\Delta^{\cup\star}_X$	≈ 0.721	1/8
сет-расстояния:		
$\Delta^{\Sigma\star}_X$	≈ 0.846	≈ 0.846
Δ^Σ_X	7/8	7/8
$\Delta^{\Gamma\star}_X$	≈ 0.600	≈ 0.600
Δ^Γ_X	≈ 0.451	≈ 0.451
$\Delta^{\min\star}_X$	1/4	1/4
$\Delta^{\min}_X$	1/8	1/8
близости:		
$1 - \Delta^\Sigma_X$	1/8	1/8
$1 - \Delta^\Gamma_X$	≈ 0.549	≈ 0.549
$1 - \Delta^{\min}_X$	7/8	7/8
ковариации:		
Kov^Σ_X	≈ -0.029	≈ -0.029
Kov^Γ_X	≈ 0.150	≈ 0.150
$\mathsf{Kov}^{\min}_X$	1/8	1/8
Фреше-корреляции:		
Kor^Σ_X	≈ -0.189	≈ -0.189
Kor^Γ_X	≈ 0.249	≈ 0.249
$\mathsf{Kor}^{\min}_X$	1/2	1/2

Таблица 25: Значения основных вероятностных показателей вероятностного распределения 14-м.с. X для двух его вариантов: “сковорода” и “торт”; см. рис. 11.

Метрическая интерпретация симметричной ковариации (сумма):

$$\mathsf{Kov}^\Sigma_X = \Delta^{\Sigma\star}_X - \Delta^\Sigma_X.$$

Второй вариант симметричных определений ковариации, Фреше-корреляции и сет-расстояния, минимум. Определение (симметричное сет-расстояние, минимум):

$$\Delta^{\min\star}_X = 2\min\{\Delta^{\cup\star}_X, \Delta^{\cap\star}_X\},$$

$$\Delta^{\min}_X = 2\min\{\Delta^\cup_X, \Delta^\cap_X\}.$$

Определение (симметричная ковариация, минимум):

$$\mathsf{Kov}^{\min}_X = 2\min\{\Delta^{\cup\star}_X, \Delta^{\cap\star}_X\} - 2\min\{\Delta^\cup_X, \Delta^\cap_X\}.$$

Метрическая интерпретация симметричной ковариации (минимум):

$$\mathsf{Kov}^{\min}_X = \Delta^{\min\star}_X - \Delta^{\min}_X.$$

Третий вариант симметричных определений ковариации, Фреше-корреляции и сет-расстояния, среднее геометрическое. Определение (симметричное сет-расстояние, среднее геометрическое):

$$\Delta^{\Gamma\star}_X = 2\sqrt{\Delta^{\cup\star}_X \Delta^{\cap\star}_X},$$

$$\Delta^\Gamma_X = 2\sqrt{\Delta^\cup_X \Delta^\cap_X}.$$

Определение (симметричная ковариация, минимум):

$$\mathsf{Kov}^\Gamma_X = 2\sqrt{\Delta^{\cup\star}_X \Delta^{\cap\star}_X} - 2\sqrt{\Delta^\cup_X \Delta^\cap_X}.$$

Метрическая интерпретация симметричной ковариации (минимум):

$$\mathsf{Kov}^\Gamma_X = \Delta^{\Gamma\star}_X - \Delta^\Gamma_X.$$

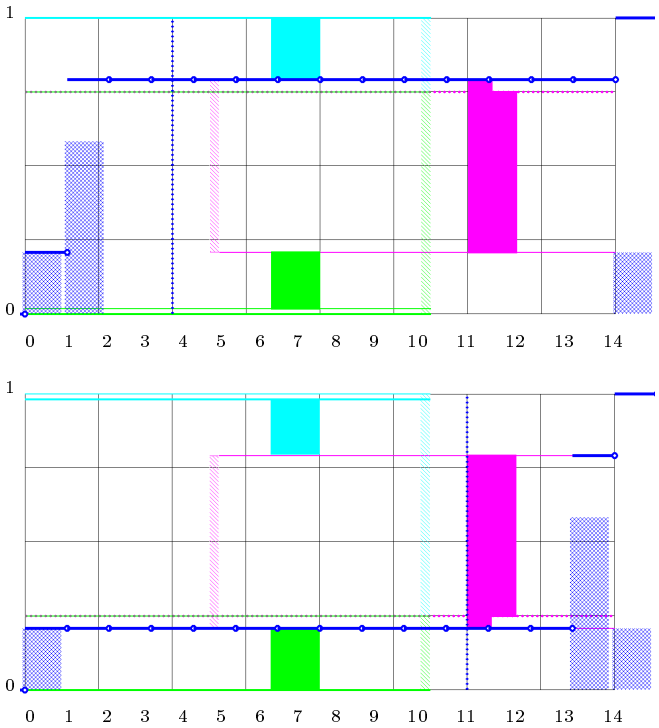


Рисунок 12: Шкалы зависимости и общности для 14-м.с. $X \subseteq \mathfrak{X}$ (фуксия — общность, аква — зависимость пересечения, лайм — зависимость объединения) и их сравнение с распределением вероятностей и ф.р. его мощностного индикатора (синий). Сверху вниз — два варианта вероятностного распределения 14-м.с. X: “сковорода” ($p_x = 1/4; p_X = 5/24; u_X = 19/24; p(\{x\} /\!\!/ X) = 1/24; x \in X; \mathsf{Kov}^\cap_X > 0, \mathsf{Kov}^\cup_X > 0$) и “торт” ($p_x = 3/4; p_X = 5/24; u_X = 19/24; p(X - \{x\} /\!\!/ X) = 1/24; x \in X; \mathsf{Kov}^\cap_X > 0, \mathsf{Kov}^\cup_X > 0$); см. детали в табл. 26.

	“сковорода”	“торт”
$p_x, x \in X$	1/4	3/4
шкала зависимости пересечения:		
$p^\star_X$	$(1/4)^{14} \approx 4 \cdot 10^{-9}$	$(3/4)^{14} \approx 0.018$
p_X	5/24	5/24
p^-_X	0	0
p^+_X	1/4	3/4
шкала зависимости объединения:		
$u^\star_X$	≈ 0.982	$\approx 1 - 4 \cdot 10^{-9}$
u_X	19/24	19/24
u^-_X	1/4	3/4
u^+_X	1	1
частичные ковариации:		
$\mathsf{Kov}^\cap_X$	5/24	≈ 0.19
$\mathsf{Kov}^\cup_X$	≈ 0.19	5/24
шкала общности:		
$\langle p \rangle_X$	1/4	3/4
частичные сет-расстояния:		
$\Delta^\cap_X$	1/24	13/24
$\Delta^\cup_X$	13/24	1/24
$\Delta^{\cap\star}_X$	1/4	≈ 0.732
$\Delta^{\cup\star}_X$	≈ 0.732	1/4
сет-расстояния:		
$\Delta^{\Sigma\star}_X$	≈ 0.982	≈ 0.982
Δ^Σ_X	7/12	7/12
$\Delta^{\Gamma\star}_X$	≈ 0.856	≈ 0.856
Δ^Γ_X	≈ 0.300	≈ 0.300
$\Delta^{\min\star}_X$	1/2	1/2
$\Delta^{\min}_X$	1/12	1/12
близости:		
$1 - \Delta^\Sigma_X$	5/12	5/12
$1 - \Delta^\Gamma_X$	≈ 0.700	≈ 0.700
$1 - \Delta^{\min}_X$	11/12	11/12
ковариации:		
Kov^Σ_X	≈ 0.398	≈ 0.398
$\mathsf{Kov}^{\min}_X$	5/12	5/12
Kov^Γ_X	≈ 0.555	≈ 0.555
Фреше-корреляции:		
Kor^Σ_X	≈ 0.406	≈ 0.406
Kor^Γ_X	≈ 0.649	≈ 0.649
$\mathsf{Kor}^{\min}_X$	5/6	5/6

Таблица 26: Значения основных вероятностных показателей вероятностного распределения 14-м.с. X для двух его вариантов: “сковорода” и “торт”; см. рис. 12.

9 Сравнение классических и эвентологических показателей общности событий (примеры: “слабый консенсус”, “все против одного”, “всё или ничего”)

9.1 Примеры для тестирования показателей общности м.с.

Для тестирования показателей событийной общности или разобщенности требуются простые примеры м.с. с простой структурой общности событий, которая имеет интуитивно предсказуемые значения показателей их общности.

9.1.1 Пример: “не больше одного против”

Рассмотрим пример сопряженного м.с. $\mathbb{X}$, структура событийной общности которого может быть коротко названа “не больше одного против” в силу того, что определяющая его вероятностная мера распределена только на тех террасных событиях, которые, либо “расположены” на диагонали (и имеют вероятность α, равномерно распределенную на все M диагональных террасных событий), либо “занумерованы” m-подмножествами, в которых *все кроме одного* события принадлежат одному и тому же категориальному множеству $\mathfrak{X}_i$ $(i = 1, \ldots, M)$ (и имеют вероятность $1-\alpha$, равномерно распределенную на все $mM(M-1)$ таких террасных событий).

Первое семейство, которое состоит из всех категориальных м.с., так как, именно, ими занумерованы диагональные террасные события, обозначим

$$\mathfrak{T}_d = \{\mathfrak{X}_i, i = 1, \ldots, M\},$$

а второе семейство обозначим

$$\mathfrak{T}_1 = \{X^j_{i,k}, i = 1, ..., M, j = 1, ..., m, k \in \{1, ..., M\} - \{i\}\},$$

где

$$X^j_{i,k} = \mathfrak{X}_i \Delta \{x^j_i, x^j_k\} = \{x^1_i, \ldots, x^j_k, \ldots, x^m_i\},\ k \neq i,$$

— m-подмножества, которыми занумерованы террасные события из семейства $\mathfrak{T}_1$.

Следовательно,

$$\mathsf{ter}_{\mathfrak{X}_i /\!\!/ \mathbb{X}} = \bigcap_{x \in \mathfrak{X}_i} x = \bigcap_{j=1}^{m} x^j_i$$

для $i = 1, \ldots, M$; а

$$\mathsf{ter}_{X^j_{i,k} /\!\!/ \mathbb{X}} = \bigcap_{x \in \mathfrak{X}^j_{i,k}} x = x^j_k \cap \left(\bigcap_{s \neq j} x^s_i \right)$$

для $i = 1, \ldots, M, j = 1, \ldots, m, k \in \{1, \ldots, M\} - \{i\}$.

Террасные события $\mathsf{ter}_{\mathfrak{X}_i /\!\!/ \mathbb{X}}$, занумерованные категориальными м.с. $\mathfrak{X}_i \in \mathfrak{T}_d$, происходят тогда, когда все без исключения эксперты относят наблюдаемое событие к i-той категории.

Террасные события $\mathsf{ter}_{X^j_{i,k} /\!\!/ \mathbb{X}}$, занумерованные м.с. $X^j_{i,k} \in \mathfrak{T}_1$, происходит тогда, когда все кроме одного j-го эксперта относят наблюдаемое событие к i-той категории, а j-й эксперт — к k-той, $k \neq i$. Иначе говоря, только j-й эксперт относит наблюдаемое событие *не к i-той категории*.

Замечание. Условие "не больше одного против" означают здесь "либо все m экспертов имеют общее мнение, либо $m-1$ эксперт имеет общее мнение". Поэтому, пример "не больше одного против" имеет смысл, когда

$$m > 2, \qquad (9.1.1.\star)$$

так как об *общем мнении* экспертов имеет смысл говорить, лишь начиная с двух экспертов: $m-1=2$.

Заметим также, что m-мерная сопряженная $\underbrace{M\times\ldots\times M}_{m}$ таблица, характеризуемая м.с. $\mathbb{X}$, содержит ненулевые клеточные вероятности либо на диагонали в M клетках, которые соответствуют террасным событиям из $\mathfrak{T}_d$, либо в $mM(M-1)$ клетке, которые соответствуют террасным событиям из $\mathfrak{T}_1$. Следовательно, вероятностное распределение м.с. $\mathbb{X}$ определяется следующими вероятностями (1-го или 2-го рода)

$$p_{X/\!/\mathbb{X}} = p(X/\!/\mathbb{X}) =$$

$$= \begin{cases} \dfrac{\alpha}{M}, & X \in \mathfrak{T}_d; \\ \dfrac{(1-\alpha)}{mM(M-1)}, & X \in \mathfrak{T}_1; \\ 0, & X \in \mathfrak{T} - \mathfrak{T}_d - \mathfrak{T}_1, \end{cases} \qquad (9.1.1.2)$$

где

$$\mathfrak{T} = \{X : X = \{x^1_{i_1}, ..., x^m_{i_m}\}, i_j = 1, ..., M, j = 1, ..., m\}$$

— семейство всех M^m подмножеств сопряженного м.с. $\mathbb{X}$, которыми занумерованы террасные события, порожденные $\mathbb{X}$.

Рассмотрим теперь i-тое категориальное m-м.с. $\mathfrak{X}_i \subset \mathbb{X}$ и его вероятностное распределение 1-го рода

$$\{p(X/\!/\mathfrak{X}_i), X \subseteq \mathfrak{X}_i\},$$

которое является *проекцией распределения* м.с. $\mathbb{X}$ на подм.с. $\mathfrak{X}_i \subseteq \mathbb{X}$, и образовано 2^m вероятностями террасных событий 1-го рода для $X \subseteq \mathfrak{X}_i$

$$p(X/\!/\mathfrak{X}_i) = \mathbf{P}\left(\mathsf{ter}(X/\!/\mathfrak{X}_i)\right) = \mathbf{P}\left(\bigcap_{x\in X} x \bigcap_{x\in\mathfrak{X}_i - X} x^c\right).$$

Вероятностное распределения сопряженного м.с. $\mathbb{X}$ определяет распределение своей проекции на категориальное м.с. $\mathfrak{X}_i$ по формулам

$$p(X/\!/\mathfrak{X}_i) = \sum_{Y\cap\mathfrak{X}_i = X} p(Y/\!/\mathbb{X})$$

для $X \subseteq \mathfrak{X}_i$, из которых следуют формулы

$$p(X/\!/\mathfrak{X}_i) =$$

$$= \begin{cases} \dfrac{\alpha}{M}, & X = \mathfrak{X}_i, \\ \dfrac{1-\alpha}{mM}, & X = \{x^j_i\}, \\ & j = 1, \ldots, m, \\ \dfrac{1-\alpha}{mM}, & X = \mathfrak{X}_i \cap X^j_{i,k}, \\ & j = 1, \ldots, m, \\ & k \in \{1, \ldots, M\} - \{i\}, \\ 1 - \dfrac{1}{M} - \dfrac{1-\alpha}{M}, & X = \emptyset, \\ 0, & \text{иначе.} \end{cases} \qquad (9.1.1.3)$$

В силу того, что

$$u_{\mathfrak{X}_i} = 1 - p(\emptyset/\!/\mathfrak{X}_i) = \frac{1}{M} + \frac{1-\alpha}{M},$$

и для $x \in \mathfrak{X}_i = \{x^1_i, \ldots, x^m_i\}$

$$p_x = \pi^j_i = \mathbf{P}(x^j_i) =$$

$$= \frac{\alpha}{M} + \frac{1-\alpha}{mM} + (m-1)\frac{1-\alpha}{mM} = \frac{1}{M}, \qquad (9.1.1.4)$$

Из (9.1.1.2) следует, что сет-расстояние $\mathfrak{X}_i$ равно

$$\Delta_{\mathfrak{X}_i} = u_{\mathfrak{X}_i} - \frac{1}{|\mathfrak{X}_i|}\sum_{x\in\mathfrak{X}_i} p_x = \frac{1-\alpha}{M}. \qquad (9.1.1.2)$$

Вспомните границы Фреше для сет-расстояния (3.4.1) и (3.4.2) и вычислите их для i-го категориального м.с. $\mathfrak{X}_i$:

$$\Delta^-_{\mathfrak{X}_i} = \max_{x\in\mathfrak{X}_i} p_x - \frac{1}{|\mathfrak{X}_i|}\sum_{x\in\mathfrak{X}_i} p_x = 0,$$

$$\Delta^+_{\mathfrak{X}_i} = \min\left\{1, \sum_{x\in\mathfrak{X}_i} p_x\right\} - \frac{1}{|\mathfrak{X}_i|}\sum_{x\in\mathfrak{X}_i} p_x =$$

$$= \min\left\{1, \frac{m}{M}\right\} - \frac{1}{M} = \begin{cases} \dfrac{m-1}{M}, & m \le M, \\ \dfrac{M-1}{M}, & m > M, \end{cases}$$

в силу того, что $\sum\limits_{x\in\mathfrak{X}_i} p_x = \dfrac{m}{M}$.

Отсюда размах возможных значений сет-расстояния i-го категориального м.с. $\mathfrak{X}_i$ равен (3.4.3)

$$\Delta^+_{\mathfrak{X}_i} - \Delta^-_{\mathfrak{X}_i} = \Delta^+_{\mathfrak{X}_i} = \begin{cases} \dfrac{m-1}{M}, & m \le M, \\ \dfrac{M-1}{M}, & m > M, \end{cases}$$

Поэтому каждое i-е категориальное м.с. $\mathfrak{X}_i$ имеет *Фреше-расстояние*

$$\mathbb{\Delta}_{\mathfrak{X}_i} = \frac{\Delta_{\mathfrak{X}_i} - \Delta^-_{\mathfrak{X}_i}}{\Delta^+_{\mathfrak{X}_i} - \Delta^-_{\mathfrak{X}_i}} = \begin{cases} \dfrac{1-\alpha}{m-1}, & m \le M, \\ \dfrac{1-\alpha}{M-1}, & m > M, \end{cases}$$

и следующие показатели общности:

$$1 - \Delta_{\mathfrak{X}_i} = 1 - \frac{1}{M}(1-\alpha)$$

— *O-близость*;

$$1 - \mathbb{\Delta}_{\mathfrak{X}_i} = \frac{\Delta^+_{\mathfrak{X}_i} - \Delta_{\mathfrak{X}_i}}{\Delta^+_{\mathfrak{X}_i} - \Delta^-_{\mathfrak{X}_i}} = \begin{cases} 1 - \dfrac{1-\alpha}{m-1}, & m \le M, \\ 1 - \dfrac{1-\alpha}{M-1}, & m > M. \end{cases}$$

— *Фреше-O-близость*.

O-кучность категориального м.с. $\mathfrak{X}_i$ равна

$$1 - \nabla_{\mathfrak{X}_i} = \frac{\langle p \rangle_{\mathfrak{X}_i}}{u_{\mathfrak{X}_i}} = \frac{1/M}{(2-\alpha)/M} = \frac{1}{2-\alpha}.$$

Верхняя и нижняя границы и размах кучности $\mathfrak{X}_i$ равны соответственно:

$$(1 - \nabla_{\mathfrak{X}_i})^+ = \frac{\langle p \rangle_{\mathfrak{X}_i}}{u^-_{\mathfrak{X}_i}} = 1,$$

$$(1 - \nabla_{\mathfrak{X}_i})^- = \frac{\langle p \rangle_{\mathfrak{X}_i}}{u^+_{\mathfrak{X}_i}} =$$

$$= \frac{\langle p \rangle_{\mathfrak{X}_i}}{\min\left\{1, \sum_{x \in \mathfrak{X}_i} p_x\right\}} = \begin{cases} \frac{1}{m}, & m \le M, \\ \frac{1}{M}, & m > M, \end{cases}$$

$$(1 - \nabla_{\mathfrak{X}_i})^+ - (1 - \nabla_{\mathfrak{X}_i})^- = \begin{cases} 1 - \frac{1}{m}, & m \le M, \\ 1 - \frac{1}{M}, & m > M. \end{cases}$$

Отсюда *Фреше-O-кучность* категориального м.с. $\mathfrak{X}_i$ равна

$$1 - \mathbb{\nabla}_{\mathfrak{X}_i} = \begin{cases} \frac{1}{m-1}\left(\frac{m}{2-\alpha} - 1\right), & m \le M, \\ \frac{1}{M-1}\left(\frac{M}{2-\alpha} - 1\right), & m > M. \end{cases}$$

По определениям *диагональная каппа (4.5.4), классическая каппа (4.5.5), каппа-O-близость (4.5.6) и каппа-Фреше-O-близость (4.5.7)* м.с. $\mathbb{X}$ равны соответственно:

$$\kappa_{\mathrm{diag}}(\mathbb{X}) = \alpha; \tag{9.1.1.4}$$

$$\kappa(\mathbb{X}) = \frac{\kappa_{\mathrm{diag}}(\mathbb{X}) - \sum\limits_{i=1}^{M}\prod\limits_{j=1}^{m}\pi_i^j}{1 - \sum\limits_{i=1}^{M}\prod\limits_{j=1}^{m}\pi_i^j} =$$

$$= \frac{\alpha - \sum\limits_{i=1}^{M}\prod\limits_{j=1}^{m}\frac{1}{M}}{1 - \sum\limits_{i=1}^{M}\prod\limits_{j=1}^{m}\frac{1}{M}} = \frac{\alpha - \frac{1}{M^m}}{1 - \frac{1}{M^m}}; \tag{9.1.1.5}$$

$$\varkappa_{\Delta}(\mathbb{X}) = \frac{1}{M}\sum_{i=1}^{M}(1 - \Delta_{\mathfrak{X}_i}) = 1 - \frac{1}{M}(1-\alpha); \tag{9.1.1.6}$$

$$\varkappa_{\mathbb{\Delta}}(\mathbb{X}) = \frac{1}{M}\sum_{i=1}^{M}(1 - \mathbb{\Delta}_{\mathfrak{X}_i}) =$$

$$= \begin{cases} 1 - \dfrac{1-\alpha}{m-1}, & m \le M, \\ 1 - \dfrac{1-\alpha}{M-1}, & m > M. \end{cases} \tag{9.1.1.7}$$

а *каппа-O-кучность и каппа-Фреше-O-кучность м.с.* $\mathbb{X}$ вычисляются аналогичными приемами по формулам (4.5.8) и (4.5.9) соответственно:

$$\varkappa_{\nabla}(\mathbb{X}) = \frac{1}{M}\sum_{i=1}^{M}(1 - \nabla_{\mathfrak{X}_i}) = \frac{1}{2-\alpha}, \tag{9.1.1.8}$$

$$\varkappa_{\mathbb{\nabla}}(\mathbb{X}) = \frac{1}{M}\sum_{i=1}^{M}(1 - \mathbb{\nabla}_{\mathfrak{X}_i}) =$$

$$= \begin{cases} \frac{1}{m-1}\left(\frac{m}{2-\alpha} - 1\right), & m \le M, \\ \frac{1}{M-1}\left(\frac{M}{2-\alpha} - 1\right), & m > M. \end{cases} \tag{9.1.1.9}$$

Вывод. Когда мнение *только одного* из экспертов всегда противоречит мнениям остальных экспертов, между собой совпадающих, диагональная каппа равна α, классическая каппа для всех α меньше (!?) диагональной, приближается к ней снизу с ростом α, а при $\alpha < M^{-m}$ показывает отрицательную (!?) общность, что противоречит интуиции; вместе с тем, эвентологические показатели — каппа-O-близость, каппа-Фреше-O-близость, каппа-O-кучность и каппа-Фреше-O-кучность для всех α больше диагональной каппа, что интуитивно оправдано, так как диагональная каппа не учитывает "неполные" общности мнений экспертов, не "попадающие" на диагональ.

9.1.2 Пример: "всё или ничего"

Рассмотрим пример м.с. $\mathbb{X}$, структура событийной общности которого может быть коротко названа "всё или ничего" в силу того, что все m экспертов либо приходят к консенсусу по каждой категории с вероятностью α/M, либо с вероятностью $1-\alpha$ среди их мнений нет и двух общих ни по одной из категорий, соответствующие отсутствию консенсуса терасные события *равновероятны*, а сумма их вероятностей равна $1-\alpha$.

Обозначим

$$\mathfrak{T} = \{\{x^1_{i_1}, \dots, x^m_{i_m}\} : i_k = 1, \dots, M, k = 1, \dots, m\}$$

совокупность всех подмножеств, которыми занумерованы *клеточные террасные события*

$$\mathsf{ter}_{\{x^1_{i_1}, \dots, x^m_{i_m}\}/\!/\mathbb{X}},$$

порожденные сопряженным м.с. $\mathbb{X}$;

$$\mathfrak{T}_d = \Big\{\{x^1_i, \dots, x^m_i\} : i = 1, \dots, M\Big\} =$$

$$= \Big\{\mathfrak{X}_i : i = 1, \dots, M\Big\} \subseteq \mathfrak{T}$$

— диагональная подсовокупность категориальных m-м.с., в которых все события имеют один и тот же категорный индекс;

$$\mathfrak{T}_0 = \Big\{\{x^1_{i_1}, \dots, x^m_{i_m}\} : i_k \neq i_s, k \neq s\Big\} \subseteq \mathfrak{T}$$

— 0-подсовокупность m-м.с., в которых события не имеют совпадающих категорных индексов $i_1, \dots, i_m$;

$$\mathfrak{T}_\star = \Big\{\{x^1_{i_1}, \dots, x^m_{i_m}\} : \exists k, s; i_k = i_s\Big\} = \mathfrak{T} - \mathfrak{T}_d - \mathfrak{T}_0$$

— 1-подсовокупность m-м.с., в которых события имеют хотя бы одну пару совпадающих категорных индексов.

Мощность 0-совокупности $\mathfrak{T}_0$ равна числу размещений из M по m при $m \leq M$ и нулю в противном случае:

$$|\mathfrak{T}_0| = \begin{cases} A^m_M = \frac{M!}{(M-m)!}, & m \leq M, \\ 0, & m > M. \end{cases}$$

Замечание. Будем считать, что

$$1 < m \leq M \qquad (9.1.2.\star)$$

в силу того, что $|\mathfrak{T}_0| = 0$ при $m > M$, а для $m = 1$ понятие общности мнений экспертов теряет смысл.

Поскольку $|\mathfrak{T}| = M^m$ — мощность всей совокупности совпадает с числом клеток в сопряженной (m, M)-таблице, $|\mathfrak{T}_d| = M$ — мощность диагональной совокупности равна числу категорий, то

$$|\mathfrak{T}_\star| = M^m - A^m_M - M$$

— мощность 1-совокупности $\mathfrak{T}_\star$ равна мощности оставшейся части.

В силу сделанных предположений вероятностное распределение сопряженного м.с. $\mathbb{X}$ имеет вид:

$$p(X/\!/\mathbb{X}) = \begin{cases} \frac{\alpha}{|\mathfrak{T}_d|} = \frac{\alpha}{M}, & X \in \mathfrak{T}_d, \\ \frac{1-\alpha}{|\mathfrak{T}_0|} = \frac{1-\alpha}{A^m_M}, & X \in \mathfrak{T}_0, \\ 0, & X \in \mathfrak{T}_\star, \end{cases} \qquad (9.1.2.1)$$

и определяет вероятностные распределения своих подмножеств, в том числе категориальных m-подм.с. $\mathfrak{X}_i \subseteq \mathbb{X}, i = 1, \dots, M$, по общим формулам для $X \subseteq \mathfrak{X}_i$:

$$p(X/\!/\mathfrak{X}_i) = \sum_{Y \cap \mathfrak{X}_i = X} p(Y/\!/\mathbb{X}), \qquad (9.1.2.2)$$

из которых следует, что

$$p(X/\!/\mathfrak{X}_i) =$$

$$= \begin{cases} \sum\limits_{\substack{Y \cap \mathfrak{X}_i = \mathfrak{X}_i \\ Y \in \mathfrak{T}_d}} p(Y/\!/\mathbb{X}) = p(\mathfrak{X}_i/\!/\mathbb{X}) = \frac{\alpha}{M}, & \\ & X = \mathfrak{X}_i, \quad (9.1.2.3) \\ \sum\limits_{\substack{Y \cap \mathfrak{X}_i = \{x^j_i\} \\ Y \in \mathfrak{T}_0}} p(Y/\!/\mathbb{X}) = \frac{1-\alpha}{M}, & X = \{x^j_i\}, (9.1.2.3') \\ & j = 1, \dots, m, \\ \sum\limits_{\substack{Y \cap \mathfrak{X}_i = \emptyset \\ Y \in \mathfrak{T}_0}} p(Y/\!/\mathbb{X}) = 1 - \frac{\alpha}{M} - m\frac{1-\alpha}{M}, & \\ & X = \emptyset, \quad (9.1.2.3'') \\ 0, & \text{иначе,} \end{cases}$$

поскольку в $(9.1.2.3')$ число слагаемых в сумме равно A^{m-1}_{M-1} а сама сумма поэтому равна

$$A^{m-1}_{M-1}\frac{1-\alpha}{A^m_M} = \frac{1-\alpha}{M},$$

а в $(9.1.2.3'')$ сумма равна

$$p(\emptyset/\!/\mathfrak{X}_i) = 1 - \sum_{X \neq \emptyset} p(X/\!/\mathfrak{X}_i) = 1 - \frac{\alpha}{M} - m\frac{1-\alpha}{M}.$$

Отсюда получаем

$$u_{\mathfrak{X}_i} = 1 - p(\emptyset/\!/\mathfrak{X}_i) = \frac{\alpha}{M} + m\frac{1-\alpha}{M}, \qquad (9.1.2.4)$$

$$\pi^j_i = \mathbf{P}(x^j_i) = \frac{\alpha}{M} + \frac{1-\alpha}{M} = \frac{1}{M}, \qquad (9.1.2.5)$$

$$\max_{x^j_i \in \mathfrak{X}_i} \pi^j_i = \frac{1}{M}. \qquad (9.1.2.6)$$

По определению сет-расстояние категориального м.с. $\mathfrak{X}_i$ равно

$$\Delta_{\mathfrak{X}_i} = u_{\mathfrak{X}_i} - \frac{1}{m}\sum_{j=1}^m \pi^j_i = (m-1)\frac{1-\alpha}{M}, \qquad (9.1.2.7)$$

а **О-близость** м.с. $\mathfrak{X}_i$ равна

$$1 - \Delta_{\mathfrak{X}_i} = 1 - (m-1)\frac{1-\alpha}{M}, \qquad (9.1.2.8)$$

Прежде, чем вычислить их границы Фреше, найдем границы Фреше для вероятности объединение $u_{\mathfrak{X}_i}$:

$$u^-_{\mathfrak{X}_i} = \max_{x^j_i \in \mathfrak{X}_i} \pi^j_i = \frac{1}{M},$$

$$u^+_{\mathfrak{X}_i} = \min\left\{1, \sum_{j=1}^{m} \pi_i^j\right\} =$$

$$= \min\left\{1, \frac{m}{M}\right\} = \begin{cases} \frac{m}{M}, & m \le M, \\ 1, & m > 1. \end{cases} = \frac{m}{M}$$

в силу предположения (9.1.2.⋆).

Границы Фреше сет-расстояния имеют вид:

$$\Delta^-_{\mathfrak{X}_i} = u^-_{\mathfrak{X}_i} - \frac{1}{m}\sum_{j=1}^{m} \pi_i^j = \frac{1}{M} - \frac{1}{M} = 0, \qquad (9.1.2.8)$$

$$\Delta^+_{\mathfrak{X}_i} = u^+_{\mathfrak{X}_i} - \frac{1}{m}\sum_{j=1}^{m} \pi_i^j = \frac{m}{M} - \frac{1}{M} = \frac{m-1}{M}. \qquad (9.1.2.9)$$

Отсюда размах возможных значений сет-расстояния категориального м.с. $\mathfrak{X}_i$ для $i = 1, \ldots, M$ равен правой границе Фреше:

$$\Delta^+_{\mathfrak{X}_i} - \Delta^-_{\mathfrak{X}_i} = \frac{m-1}{M}, \qquad (9.1.2.10)$$

а **Фреше-О-близость** м.с. $\mathfrak{X}_i$ определяется величиной

$$1 - \mathbb{\Delta}_{\mathfrak{X}_i} = \frac{\frac{m-1}{M} - \frac{m-1}{M}(1-\alpha)}{\frac{m-1}{M}} = \alpha.$$

По определениям (4.5.4), (4.5.5), (4.5.6) и (4.5.7) *диагональная каппа, классическая каппа, каппа-О-близость и каппа-Фреше-О-близость* м.с. $\mathbb{X}$ равны:

$$\kappa_{\mathrm{diag}}(\mathbb{X}) = \alpha; \qquad (9.1.2.11)$$

$$\kappa(\mathbb{X}) = \frac{\kappa_{\mathrm{diag}}(\mathbb{X}) - \sum_{i=1}^{M}\prod_{j=1}^{m}\pi_i^j}{1 - \sum_{i=1}^{M}\prod_{j=1}^{m}\pi_i^j} =$$

$$= \frac{\alpha - \frac{1}{M^{m-1}}}{1 - \frac{1}{M^{m-1}}}; \qquad (9.1.2.12)$$

$$\varkappa_{\Delta}(\mathbb{X}) = \frac{1}{M}\sum_{i=1}^{M}\left(1 - (m-1)\frac{1-\alpha}{M}\right) =$$

$$= 1 - (m-1)\frac{1-\alpha}{M}. \qquad (9.1.2.13)$$

$$\varkappa_{\mathbb{\Delta}}(\mathbb{X}) = \frac{1}{M}\sum_{i=1}^{M}(1 - \mathbb{\Delta}_{\mathfrak{X}_i}) = \alpha, \qquad (9.1.2.14)$$

а *каппа-О-кучность и каппа-Фреше-О-кучность м.с.* $\mathbb{X}$ вычисляются аналогичными приемами по формулам (4.5.8) и (4.5.9) соответственно:

$$\varkappa_{\nabla}(\mathbb{X}) = \frac{1}{M}\sum_{i=1}^{M}(1 - \nabla_{\mathfrak{X}_i}) =$$

$$= \frac{1}{\alpha + m(1-\alpha)}, \qquad (9.1.2.15)$$

$$\varkappa_{\mathbb{\nabla}}(\mathbb{X}) = \frac{1}{M}\sum_{i=1}^{M}(1 - \mathbb{\nabla}_{\mathfrak{X}_i}) =$$

$$= \frac{\alpha}{\alpha + m(1-\alpha)}. \qquad (9.1.2.16)$$

9.1.3 Пример: "слабый консенсус"

Рассмотрим пример м.с. $\mathbb{X}$, структура событийной общности которого может быть коротко названа "слабый консенсус" в силу того, что с вероятностью α все эксперты приходят к *консенсусу* по каждой категории; а с вероятностью $1-\alpha$ эксперты приходят к *слабому консенсусу*, т.е. лишь один эксперт по каждой категории имеют мнение, отличное от мнения остальных экспертов[18]. Кроме того, предполагается, что террасные события, как соответствующие консенсусу, так и соответствующие слабому консенсусу *равновероятны*, а суммы их вероятностей равны α и $1-\alpha$ соответственно.

Обозначим $\mathfrak{T}$, как и в параграфе 9.1.2 совокупность $\mathfrak{T}$ всех подмножеств $\{x^1_{i_1}, \ldots, x^m_{i_m}\}$, которыми занумерованы *клеточные террасные события*

$$\mathsf{ter}_{\{x^1_{i_1},\ldots,x^m_{i_m}\}/\!/\mathbb{X}},$$

порожденные сопряженным м.с. $\mathbb{X}$.

Обозначим также $\mathfrak{T}_d$ — совокупность всех категориальных m-подм.с. $\mathfrak{X}_i, i = 1, \ldots, M$, которыми занумерованы клеточные террасные события, соответствующие *консенсусу* мнений экспертов.

Обозначим $\mathfrak{T}_w$ — совокупность m-подм.с., которыми занумерованы *клеточные террасные события*, соответствующие *слабому консенсусу* мнений экспертов.

Предположения примера "слабый консенсус" означают, то

$$\sum_{X\in\mathfrak{T}_d} p(X/\!/\mathbb{X}) = \alpha,$$

$$\sum_{X\in\mathfrak{T}_w} p(X/\!/\mathbb{X}) = 1-\alpha,$$

$$\sum_{X\in\mathfrak{T}-\mathfrak{T}_d-\mathfrak{T}_w} p(X/\!/\mathbb{X}) = 0.$$

Напомним также, что

$$|\mathfrak{T}_d| = M,$$

и вычислим мощность совокупности $\mathfrak{T}_w$:

$$|\mathfrak{T}_w| = mM(M-1).$$

[18] Очевидно, что "слабый консенсус" может происходить, лишь когда количество экспертов не меньше трех: $m \ge 3$.

поскольку слабый консенсус по каждой i-той категории может случиться $m(M-1)$ способами, а всего категорий M.

Замечание. Пример "слабый консенсус" имеет смысл, когда

$$1 < M, \quad 1 < m, \tag{9.1.3.$\star$}$$

в силу того, что $|\mathfrak{T}_w| = 0$ при $M = 1$, а для $m = 1$ не имеет смысла понятие общность мнения экспертов.

В силу сделанных предположений вероятностное распределение сопряженного м.с. $\mathbb{X}$ имеет вид:

$$p(X/\!\!/\mathbb{X}) = \begin{cases} \dfrac{\alpha}{|\mathfrak{T}_d|} = \dfrac{\alpha}{M}, & X \in \mathfrak{T}_d, \\ \dfrac{1-\alpha}{|\mathfrak{T}_w|} = \dfrac{1-\alpha}{mM(M-1)}, & X \in \mathfrak{T}_w, \\ 0, & \text{иначе}, \end{cases} \tag{9.1.3.1}$$

и определяет вероятностные распределения своих подмножеств, в том числе категориальных m-подм.с. $\mathfrak{X}_i \subseteq \mathbb{X}, i = 1, \ldots, M$, по общим формулам для $X \subseteq \mathfrak{X}_i$:

$$p(X/\!\!/\mathfrak{X}_i) = \sum_{Y \cap \mathfrak{X}_i = X} p(Y/\!\!/\mathbb{X}), \tag{9.1.3.2}$$

из которых следует, что

$$p(X/\!\!/\mathfrak{X}_i) = \begin{cases} \sum\limits_{\substack{Y \cap \mathfrak{X}_i = \mathfrak{X}_i \\ Y \in \mathfrak{T}_d}} p(Y/\!\!/\mathbb{X}) = \dfrac{\alpha}{M}, & |X| = m, \\ \sum\limits_{\substack{Y \cap \mathfrak{X}_i = X \\ Y \in \mathfrak{T}_w}} p(Y/\!\!/\mathbb{X}) = \dfrac{1-\alpha}{mM}, & |X| = m-1, \\ \sum\limits_{\substack{Y \cap \mathfrak{X}_i = \emptyset \\ Y \in \mathfrak{T}_w}} p(Y/\!\!/\mathbb{X}) = 1 - \dfrac{1}{M}, & X = \emptyset, \\ 0, & \text{иначе}, \end{cases} \tag{9.1.3.3}$$

$$\tag{9.1.3.3$'$}$$

$$\tag{9.1.3.3$''$}$$

поскольку в (9.1.3.3$'$) число слагаемых в каждой[19] такой сумме равно $M-1$ а сама сумма поэтому равна

$$(M-1)\frac{1-\alpha}{mM(M-1)} = \frac{1-\alpha}{mM},$$

а в (9.1.3.3$''$) сумма равна дополнению до единицы суммы (9.1.3.3) и m сумм (9.1.3.3$'$):

$$p(\emptyset/\!\!/\mathfrak{X}_i) = 1 - \sum_{X \neq \emptyset} p(X/\!\!/\mathfrak{X}_i) = 1 - \frac{\alpha}{M} - m\frac{1-\alpha}{mM} = 1 - \frac{1}{M}.$$

[19] для каждого $X \subseteq \mathfrak{X}_i$, такого что $|X| = m-1$.

Отсюда получаем

$$u_{\mathfrak{X}_i} = 1 - p(\emptyset/\!\!/\mathfrak{X}_i) = \frac{1}{M}, \tag{9.1.3.4}$$

$$\pi_i^j = \mathbf{P}(x_i^j) = {}^{20} = \frac{1}{M} - \frac{1-\alpha}{mM}, \tag{9.1.3.5}$$

$$\max_{x_i^j \in \mathfrak{X}_i} \pi_i^j = \frac{1}{M} - \frac{1-\alpha}{mM}. \tag{9.1.3.6}$$

Сет-расстояние категориального м.с. $\mathfrak{X}_i$ по определению равно

$$\Delta_{\mathfrak{X}_i} = u_{\mathfrak{X}_i} - \frac{1}{m}\sum_{j=1}^{m} \pi_i^j =$$

$$= \frac{1}{M} - \frac{1}{M} + \frac{1-\alpha}{mM} = \frac{1-\alpha}{mM}, \tag{9.1.3.7}$$

а **О-близость** м.с. $\mathfrak{X}_i$ равна

$$1 - \Delta_{\mathfrak{X}_i} = 1 - \frac{1-\alpha}{mM}. \tag{9.1.3.7$'$}$$

Прежде, чем вычислить его границы Фреше, найдем границы Фреше для вероятности объединение $u_{\mathfrak{X}_i}$:

$$u^-_{\mathfrak{X}_i} = \max_{x_i^j \in \mathfrak{X}_i} \pi_i^j = \frac{1}{M} - \frac{1-\alpha}{mM},$$

$$u^+_{\mathfrak{X}_i} = \min\left\{1, \sum_{j=1}^{m} \pi_i^j\right\} = \min\left\{1, \frac{m}{M} - \frac{1-\alpha}{M}\right\} =$$

$$= \begin{cases} \dfrac{m}{M} - \dfrac{1-\alpha}{M}, & m \leq M, \\ 1, & m > M. \end{cases}$$

Границы Фреше сет-расстояния имеют вид:

$$\Delta^-_{\mathfrak{X}_i} = u^-_{\mathfrak{X}_i} - \frac{1}{m}\sum_{j=1}^{m} \pi_i^j = 0, \tag{9.1.3.8}$$

$$\Delta^+_{\mathfrak{X}_i} = u^+_{\mathfrak{X}_i} - \frac{1}{m}\sum_{j=1}^{m} \pi_i^j =$$

$$= \begin{cases} \dfrac{m-1}{M} - \dfrac{m-1}{mM}(1-\alpha), & m \leq M, \\ 1 - \dfrac{1}{M} + \dfrac{1-\alpha}{mM}, & m > M. \end{cases} \tag{9.1.3.9}$$

Отсюда

$$\Delta^+_{\mathfrak{X}_i} - \Delta_{\mathfrak{X}_i} =$$

$$= \begin{cases} \dfrac{m-1}{M} - \dfrac{1-\alpha}{M}, & m \leq M, \\ 1 - \dfrac{1}{M}, & m > M. \end{cases} \tag{9.1.3.10}$$

[20] т.к. $\mathbf{P}(x_i^j) = \frac{\alpha}{M} + (m-1)\frac{1-\alpha}{mM} = \frac{\alpha}{M} + (1 - \frac{1}{m})\frac{1-\alpha}{M} = \frac{\alpha}{M} + \frac{1-\alpha}{M} - \frac{1-\alpha}{mM} = \frac{1}{M} - \frac{1-\alpha}{mM}$.

Размах возможных значений сет-расстояния категориального м.с. $\mathfrak{X}_i$ для $i = 1, \ldots, M$ равен правой границе Фреше:

$$\Delta^+_{\mathfrak{X}_i} - \Delta^-_{\mathfrak{X}_i} = \\ = \begin{cases} \dfrac{m-1}{M} - \dfrac{1-\alpha}{M} + \dfrac{1-\alpha}{mM}, & m \le M, \\ 1 - \dfrac{1}{M} + \dfrac{1-\alpha}{mM}, & m > M. \end{cases} \quad (9.1.3.11)$$

а **Фреше-О-близость** м.с. $\mathfrak{X}_i$ определяется величиной

$$1 - \mathbb{\Delta}_{\mathfrak{X}_i} = \begin{cases} \dfrac{\frac{m-1}{M} - \frac{1-\alpha}{M}}{\frac{m-1}{M} - \frac{1-\alpha}{M} + \frac{1-\alpha}{mM}}, & m \le M, \\ \dfrac{1 - \frac{1}{M}}{1 - \frac{1}{M} + \frac{1-\alpha}{mM}}, & m > M. \end{cases}$$

$$1 - \mathbb{\Delta}_{\mathfrak{X}_i} = \begin{cases} \dfrac{m-2+\alpha}{m-2+\alpha+\frac{1-\alpha}{m}}, & m \le M, \\ \dfrac{M-1}{M-1+\frac{1-\alpha}{m}}, & m > M. \end{cases}$$

По определениям (4.5.4), (4.5.5) и (4.5.6) *диагональная каппа, классическая каппа, каппа-О-близость и каппа-Фреше-О-близость* м.с. $\mathbb{X}$ равны:

$$\kappa_{\mathrm{diag}}(\mathbb{X}) = \alpha; \quad (9.1.3.12)$$

$$\kappa(\mathbb{X}) = \frac{\kappa_{\mathrm{diag}}(\mathbb{X}) - \sum\limits_{i=1}^{M}\prod\limits_{j=1}^{m}\pi_i^j}{1 - \sum\limits_{i=1}^{M}\prod\limits_{j=1}^{m}\pi_i^j} = \\ = \frac{\alpha - M\left(\frac{1}{M} - \frac{1-\alpha}{mM}\right)^m}{1 - M\left(\frac{1}{M} - \frac{1-\alpha}{mM}\right)^m}; \quad (9.1.3.13)$$

$$\varkappa_{\Delta}(\mathbb{X}) = \frac{1}{M}\sum_{i=1}^{M}(1 - \Delta_{\mathfrak{X}_i}) = \\ = 1 - \frac{m-1}{M}(1-\alpha), \quad (9.1.3.14)$$

$$\varkappa_{\mathbb{\Delta}}(\mathbb{X}) = \frac{1}{M}\sum_{i=1}^{M}(1 - \mathbb{\Delta}_{\mathfrak{X}_i}) = \\ = \begin{cases} \dfrac{m-2+\alpha}{m-2+\alpha+\frac{1-\alpha}{m}}, & m \le M, \\ \dfrac{M-1}{M-1+\frac{1-\alpha}{m}}, & m > M, \end{cases} \quad (9.1.3.15)$$

а *каппа-О-кучность и каппа-Фреше-О-кучность м.с.* $\mathbb{X}$ вычисляются аналогичными приемами по формулам (4.5.8) и (4.5.9) соответственно:

$$\varkappa_{\nabla}(\mathbb{X}) = \frac{1}{M}\sum_{i=1}^{M}(1 - \nabla_{\mathfrak{X}_i}) = 1 - \frac{1-\alpha}{M}, \quad (9.1.3.16)$$

$$\varkappa_{\mathbb{\nabla}}(\mathbb{X}) = \frac{1}{M}\sum_{i=1}^{M}(1 - \mathbb{\nabla}_{\mathfrak{X}_i}) = \\ = \begin{cases} \left(1 - \dfrac{1-\alpha}{m}\right)\dfrac{m(m-2+\alpha)}{m(m-2+\alpha)+1-\alpha}, & m \le M, \\ \left(1 - \dfrac{1-\alpha}{m}\right)\dfrac{m(M-1)}{m(M-1)+1-\alpha}, & m > M. \end{cases} \quad (9.1.3.17)$$

10 Шкалы вероятностной зависимости и общности м.с. и их точки “отсчета”

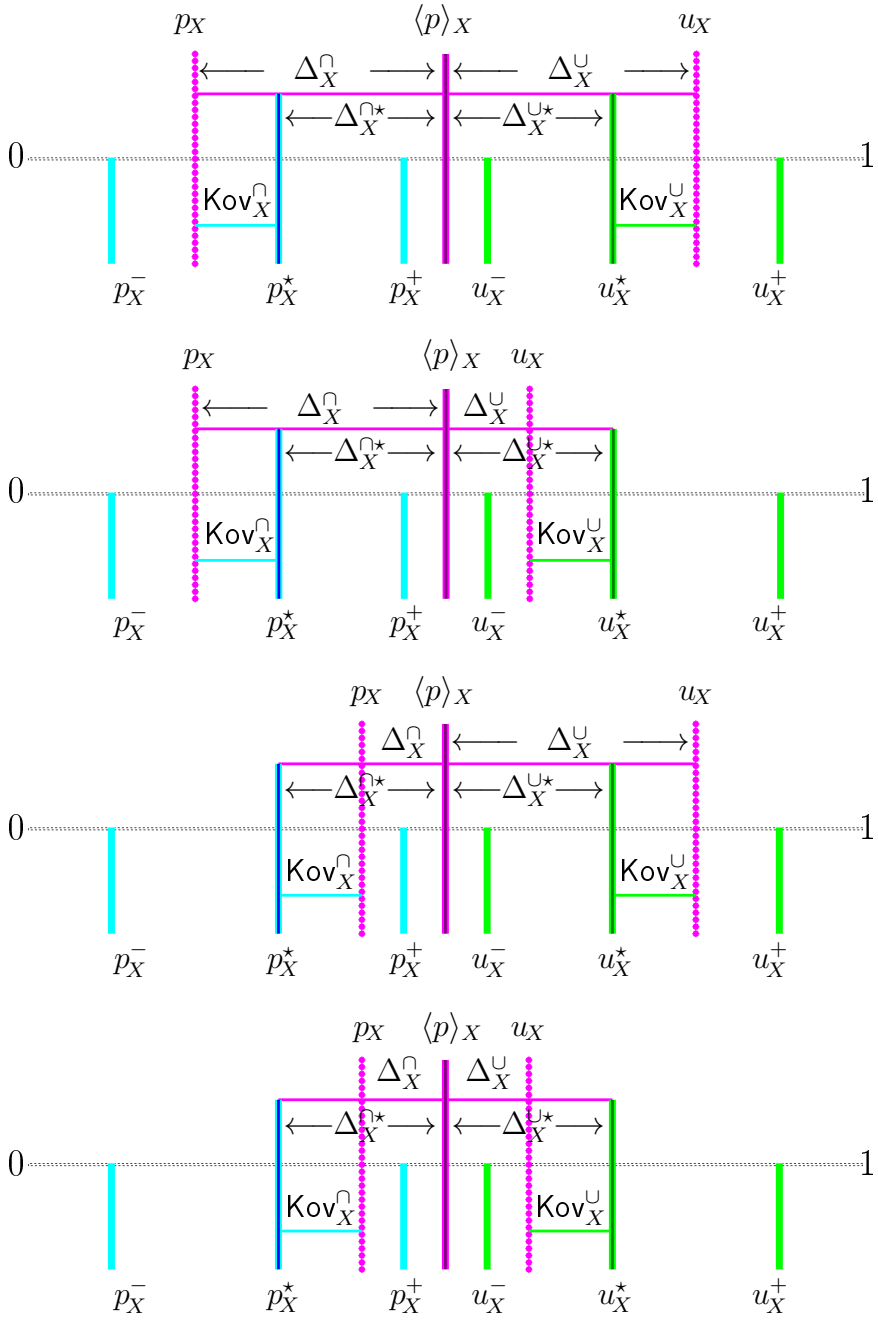


Рисунок 13: Шкалы измерения зависимости и общности м.с. $X \subseteq \mathfrak{X}$ со своими “точками отсчета”: **шкала общности для двух сеттрасстояний** по пересечению $\Delta^{\cap}_X$ и объединению $\Delta^{\cup}_X$ (фуксия) со “**средневероятной точкой отсчета**” $\langle p\rangle_X$; две **шкалы зависимости для ковариаций** по пересечению $\mathsf{Kov}^{\cap}_X$ (аква) и по объединению $\mathsf{Kov}^{\cup}_X$ (лайм) с “**независимыми точками отсчета**” $p^\star_X$ и $u^\star_X$ соответственно. Сверху вниз показаны все 4 варианта взаимного расположения вероятностей пересечения p_X и объединения u_X относительно “независимых точек отсчета” (т.е. все комбинации знаков двух ковариаций м.с. X по пересечению и по объединению): $p_X < p^\star_X,\ u^\star_X < u_X$ $(\mathsf{Kov}^{\cap}_X < 0, \mathsf{Kov}^{\cup}_X < 0)$; $p_X < p^\star_X,\ u_X < u^\star_X$ $(\mathsf{Kov}^{\cap}_X < 0, \mathsf{Kov}^{\cup}_X > 0)$; $p^\star_X < p_X,\ u^\star_X < u_X$ $(\mathsf{Kov}^{\cap}_X > 0, \mathsf{Kov}^{\cup}_X < 0)$ и $p^\star_X < p_X,\ u_X < u^\star_X$ $(\mathsf{Kov}^{\cap}_X > 0, \mathsf{Kov}^{\cup}_X > 0)$.

11 Иллюстрация свойств мер разобщенности событий на примерах некоторых распределений множеств событий

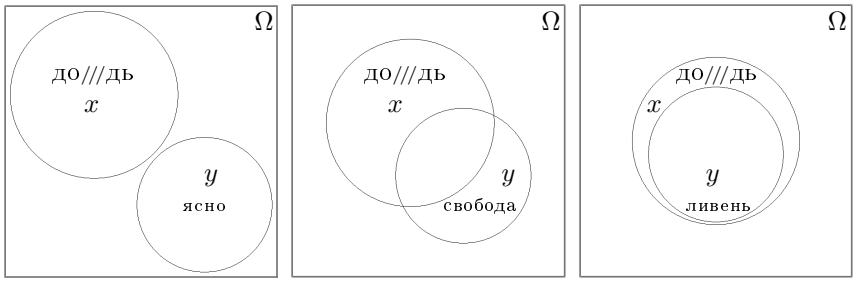


Рисунок 14: Диаграммы Венна дуплетов событий. Слева: разобщенные события “до///дь” и “ясно”; в центре независимые события “до///дь” и “свобода”; справа “до///дь” и “ливень” — события наибольшей общности.

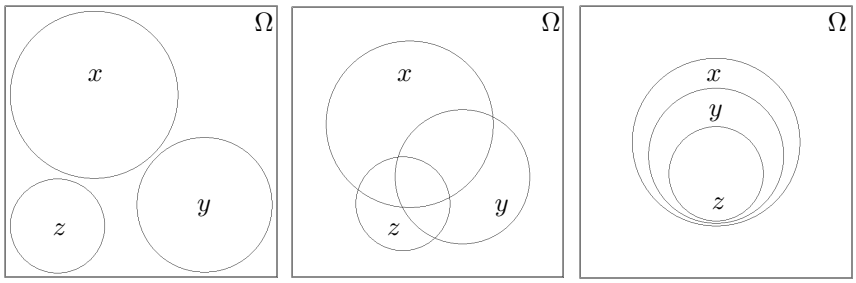
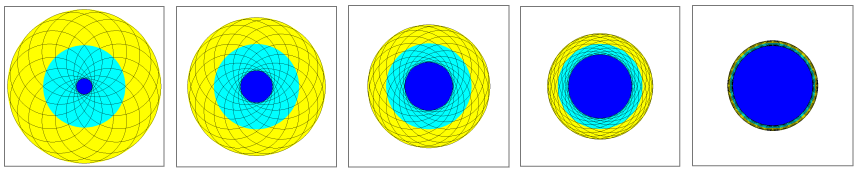
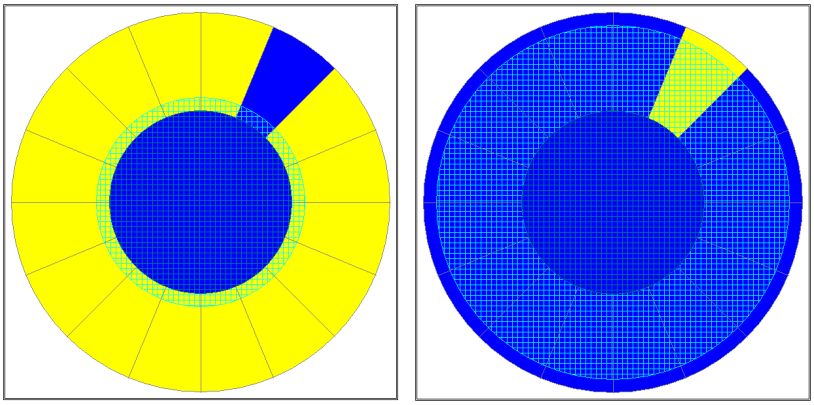


Рисунок 15: Диаграммы Венна триплета разобщенных событий (слева), триплета независимых событий (в центре) и триплета событий наибольшей общности (справа).



$$\Delta^{\cup}_{\mathfrak{X}_1} \geq \Delta^{\cup}_{\mathfrak{X}_2} \geq \Delta^{\cup}_{\mathfrak{X}_3} \geq \Delta^{\cup}_{\mathfrak{X}_4} \geq \Delta^{\cup}_{\mathfrak{X}_5}$$
$$\Delta^{\cap}_{\mathfrak{X}_1} \geq \Delta^{\cap}_{\mathfrak{X}_2} \geq \Delta^{\cap}_{\mathfrak{X}_3} \geq \Delta^{\cap}_{\mathfrak{X}_4} \geq \Delta^{\cap}_{\mathfrak{X}_5}$$

Рисунок 16: Диаграммы Венна пяти множеств из 16 событий, общность которых, измеренная обоими сет-расстояниями, возрастает слева направо; событийной “точкой отсчета” служит средневероятное событие множества из 16 событий (голубой), с которым сближается сверху объединение 16 событий (желтый) и снизу — пересечение 16 событий (синий).



$$\langle p\rangle_{x_1} - p_{x_1} = \Delta^{\cap}_{\mathfrak{x}_1} \ll \Delta^{\cup}_{\mathfrak{x}_1} = u_{x_1} - \langle p\rangle_{x_1} \qquad \langle p\rangle_{x_2} - p_{x_2} = \Delta^{\cap}_{\mathfrak{x}_2} \gg \Delta^{\cup}_{\mathfrak{x}_2} = u_{x_2} - \langle p\rangle_{x_2}$$

Рисунок 17: Диаграммы Венна двух множеств $\mathfrak{X}_1$ (“сковорода”) и $\mathfrak{X}_2$ (“торт”), каждое из 16 событий, общность которых, очевидно, велика, но измеренная сет-расстояниями относительно объединения $\Delta^{\cup}_{\mathfrak{x}_1}$ и пересечения $\Delta^{\cap}_{\mathfrak{x}_1}$ приводит к противоречивым результатам. Слева первое множество событий $\mathfrak{X}_1$ имеет высокую общность (малое сет-расстояние), измеренную сет-расстоянием относительно пересечения, но низкую общность, измеренную сет-расстоянием относительно объединения. Справа второе множество событий $\mathfrak{X}_2$ имеет высокую общность (малое сет-расстояние), измеренную сет-расстоянием относительно объединения, но низкую общность, измеренную сет-расстоянием относительно пересечения. Одно из 16-ти событий в каждом множестве показано синим цветом, их объединение — желтым; пересечение — темно-синий круг в центре; средневероятное событие, выделенное голубой решеткой, происходит со средней для каждого множества событий вероятностью.

12 Частичные (односторонние) и полные (двусторонние) показатели разобщенности м.с.

12.1 Эвентологические показатели разобщенности сопряженного м.с.

В этом параграфе дается определение показателей разобщенности сопряженного м.с., которое определяет некоторую сопряженную таблицу. Все эти эвентологические показатели конструируются тем или иным образом из двух частичных *сет-расстояний* м.с. (относительно объединения и относительно пересечения), которые обладают свойствами, аналогичными свойствам расстояния, и введены в математической эвентологии в качестве числовых характеристик разобщенности произвольных м.с.

Структура сопряженного м.с. Сопряженное м.с. $\mathbb{X}$, определяющее m-мерную сопряженную таблицу объема M^m, представимо в виде

$$\mathbb{X} = \sum_{i=1}^{M} \mathfrak{X}_i \tag{12.1.1}$$

— суммы категориальных m-м.с.

$$\mathfrak{X}_i = \{x_i^1, \ldots, x_i^m\}, \tag{12.1.2}$$

которые могут иметь произвольное вероятностное распределение размерности 2^m:

$$\{p(X/\!/\mathfrak{X}_i), X \subseteq \mathfrak{X}_i\},$$

или в виде

$$\mathbb{X} = \sum_{j=1}^{m} \mathbf{X}^j \tag{12.1.3}$$

— суммы экспертных M-м.с.

$$\mathbf{X}^j = \{x_1^j, \ldots, x_M^j\}, \tag{12.1.4}$$

которые состоят из непересекающихся событий, образующих покрытие Ω:

$$\sum_{i=1}^{M} p_{x_i^j} = 1, \tag{12.1.5}$$

где $j = 1, \ldots, m$.

12.1.1 Сумма частичных сет-расстояний м.с.

Рассматривается сопряженный показатель разобщенности м.с., основанный на сумме частичных сет-расстояний категориальных м.с., образующих данное сопряженное м.с.

Показатели разобщенности категориальных м.с. Каждое категориальное м.с. $\mathfrak{X}_i$ имеет числовую характеристику разобщенности его событий — сет-расстояние, вычисляемое как сумма частичных показателей

$$\Delta_{\mathfrak{X}_i}^{\Sigma} = \Delta_{\mathfrak{X}_i}^{\cup} + \Delta_{\mathfrak{X}_i}^{\cap} = u_{\mathfrak{X}_i} - p_{\mathfrak{X}_i}, \tag{12.1.6}$$

где

$$\Delta_{\mathfrak{X}_i}^{\cup} = u_{\mathfrak{X}_i} - \langle p \rangle_{\mathfrak{X}_i}, \tag{12.1.7}$$

$$\Delta_{\mathfrak{X}_i}^{\cap} = \langle p \rangle_{\mathfrak{X}_i} - p_{\mathfrak{X}_i} \tag{12.1.8}$$

— частичные сет-расстояния категориального м.с. $\mathfrak{X}_i$, характеризующие частичную разобщенность событий относительно объединения и пересечения соответственно.

Замечание. Сначала убедимся, что

$$\sum_{i=1}^{M} \langle p \rangle_{\mathfrak{X}_i} = \sum_{i=1}^{M} \frac{1}{m} \sum_{j=1}^{m} p_{x_i^j} =$$

$$= \frac{1}{m} \sum_{j=1}^{m} \sum_{i=1}^{M} p_{x_i^j} = 1, \tag{12.1.9}$$

в силу свойства событий (12.1.5) экспертных м.с. Отсюда получаем, что

$$\sum_{i=1}^{M} \Delta_{\mathfrak{X}_i}^{\cap} = \sum_{i=1}^{M} \langle p \rangle_{\mathfrak{X}_i} - \sum_{i=1}^{M} p_{\mathfrak{X}_i} = 1 - \kappa_{\mathrm{diag}}^{\cap}(\mathbb{X}), \tag{12.1.10}$$

где

$$\kappa_{\mathrm{diag}}^{\cap}(\mathbb{X}) = \sum_{i=1}^{M} p_{\mathfrak{X}_i} \tag{12.1.11}$$

— сумма клеточных вероятностей, расположенных на главной диагонали сопряженной таблицы, т.е. — это классическая диагональная каппа, известный простейший показатель общности событий в сопряженном м.с. $\mathbb{X}$, определяемый нами как сумма

$$p_{\mathfrak{X}_i} = \mathbf{P}\left(\bigcap_{x \in \mathfrak{X}_i} x\right) \tag{12.1.12}$$

— вероятностей пересечений всех событий в каждом категориальном м.с. $\mathfrak{X}_i, i = 1, \ldots, M$.

Аналогично (12.1.10) получаем, что

$$\sum_{i=1}^{M} \Delta_{\mathfrak{X}_i}^{\cup} = \sum_{i=1}^{M} u_{\mathfrak{X}_i} - \sum_{i=1}^{M} \langle p \rangle_{\mathfrak{X}_i} = \kappa_{\mathrm{diag}}^{\cup}(\mathbb{X}) - 1, \tag{12.1.13}$$

где

$$\kappa_{\mathrm{diag}}^{\cup}(\mathbb{X}) = \sum_{i=1}^{M} u_{\mathfrak{X}_i} \tag{12.1.14}$$

— еще одна диагональная каппа, новый показатель общности событий в сопряженном м.с. $\mathbb{X}$, определяемый нами как сумма

$$u_{\mathfrak{X}_i} = \mathbf{P}\left(\bigcup_{x \in \mathfrak{X}_i} x\right) \tag{12.1.15}$$

— вероятностей объединений всех событий в каждом категориальном м.с. $\mathfrak{X}_i, i = 1, \ldots, M$.

Сложив (12.1.10) и (12.1.14) получим величину

$$\Delta^{\Sigma}(\mathbb{X}) = \sum_{i=1}^{M} \Delta^{\Sigma}_{\mathfrak{X}_i} = \kappa^{\cup}_{\mathrm{diag}}(\mathbb{X}) - \kappa^{\cap}_{\mathrm{diag}}(\mathbb{X}), \quad (12.1.16)$$

которую может играть роль сопряженного показателя общности м.с. $\mathbb{X}$.

Лемма. *Для суммы сет-расстояний всех категориальных м.с. $\mathfrak{X}_i, i = 1, \ldots, M$, составляющих произвольное сопряженное м.с. $\mathbb{X}$, выполняется равенство:*

$$\sum_{i=1}^{M} \Delta^{\Sigma}_{\mathfrak{X}_i} = m\big(1 - \kappa^{\cap}_{\mathrm{diag}}(\mathbb{X})\big). \quad (12.1.17)$$

Доказательство.

Следствие 1. Из (12.1.17) получаем

$$\kappa^{\cap}_{\mathrm{diag}}(\mathbb{X}) = 1 - \sum_{i=1}^{M} \left(\frac{1}{m}\Delta^{\Sigma}_{\mathfrak{X}_i}\right). \quad (12.1.18)$$

Следствие 2. В силу леммы этот "новый" показатель также можно выразить через диагональную каппа относительно пересечения и сет-расстояния категориальных м.с. по формуле

$$\kappa^{\cup}_{\mathrm{diag}}(\mathbb{X}) = m\big(1 - \kappa^{\cap}_{\mathrm{diag}}(\mathbb{X})\big) + \kappa^{\cap}_{\mathrm{diag}}(\mathbb{X}) =$$

$$= 1 + \sum_{i=1}^{M} \left(1 - \frac{1}{m}\right) \Delta^{\Sigma}_{\mathfrak{X}_i}. \quad (12.1.19)$$

В формуле (12.1.16) для сопряженного показателя разобщенности м.с. $\mathbb{X}$, определяемого как сумма частичных сет-расстояний категориальных м.с., принимает участие классическая диагональная каппа $\kappa^{\cap}_{\mathrm{diag}}(\mathbb{X})$ (12.1.11), которая в данном контексте называется *диагональная каппа относительно пересечения*, и еще один показатель $\kappa^{\cup}_{\mathrm{diag}}(\mathbb{X})$ (12.1.14) — *диагональная каппа относительно объединения.*

Выводы. Данный способ определения сопряженного показателя разобщенности как суммы частичных сет-расстояний не зависит от "средневероятной точки отсчета", которая характеризует шкалу измерения общности и разобщенности, и приводит к показателям, которые можно выразить через диагональную каппа относительно пересечения, совпадающую с классическим каппа показателем. Тем самым теряются преимущества частичных сет-расстояний, которые могут учитывать не только полную общности, выражаемую классической диагональной каппа — суммой диагональных клеточных вероятностей, но и частичные общности, зависящие от клеточных вероятностей вне главной диагонали сопряженной таблицы. Таким образом, вместо суммы частичных показателей требуется найти иной способ измерения, сохраняющий информацию о частичной общности и разобщенности событий из сопряженного м.с. В качестве претендентов на место суммы частичных сет-расстояний, теряющей частичную информацию, могут быть рассмотрены две функции от частичных сет-расстояний: *минимум* и *среднее геометрическое.*

12.1.2 Минимум частичных сет-расстояний м.с.

Рассматривается сопряженный показатель разобщенности м.с., основанный на минимуме частичных сет-расстояний категориальных м.с., образующих данное сопряженное м.с.

Показатели разобщенности категориальных м.с. Каждое категориальное м.с. $\mathfrak{X}_i$ имеет числовую характеристику разобщенности его событий — сет-расстояние, вычисляемое как минимум частичных показателей

$$\Delta^{\min}_{\mathfrak{X}_i} = 2\min\left\{\Delta^{\cup}_{\mathfrak{X}_i}, \Delta^{\cap}_{\mathfrak{X}_i}\right\} \quad (12.1.6)$$

где

$$\Delta^{\cup}_{\mathfrak{X}_i} = u_{\mathfrak{X}_i} - \langle p\rangle_{\mathfrak{X}_i}, \quad (12.1.7)$$

$$\Delta^{\cap}_{\mathfrak{X}_i} = \langle p\rangle_{\mathfrak{X}_i} - p_{\mathfrak{X}_i} \quad (12.1.8)$$

— частичные сет-расстояния категориального м.с. $\mathfrak{X}_i$, характеризующие частичную разобщенность событий относительно объединения и пересечения соответственно.

12.1.3 Среднее геометрическое частичных сет-расстояний м.с.

Рассматривается сопряженный показатель разобщенности м.с., основанный на среднем геометрическом частичных сет-расстояний категориальных м.с., образующих данное сопряженное м.с.

Показатели разобщенности категориальных м.с. Каждое категориальное м.с. $\mathfrak{X}_i$ имеет числовую характеристику разобщенности его событий — сет-расстояние, вычисляемое как среднее геометрическое частичных показателей

$$\Delta^{\Gamma}_{\mathfrak{X}_i} = 2\sqrt{\Delta^{\cup}_{\mathfrak{X}_i}\Delta^{\cap}_{\mathfrak{X}_i}}, \quad (12.1.6)$$

где

$$\Delta^{\cup}_{\mathfrak{X}_i} = u_{\mathfrak{X}_i} - \langle p\rangle_{\mathfrak{X}_i}, \quad (12.1.7)$$

$$\Delta^{\cap}_{\mathfrak{X}_i} = \langle p\rangle_{\mathfrak{X}_i} - p_{\mathfrak{X}_i} \quad (12.1.8)$$

— частичные сет-расстояния категориального м.с. $\mathfrak{X}_i$, характеризующие частичную разобщенность событий относительно объединения и пересечения соответственно.

12.1.4 Нормализованное произведение частичных сет-расстояний м.с.

Рассматривается сопряженный показатель разобщенности м.с., основанный на нормализованном произведении частичных сет-расстояний категориальных м.с., образующих данное сопряженное м.с.

Показатели разобщенности категориальных м.с. Каждое категориальное м.с. $\mathfrak{X}_i$ имеет числовую характеристику разобщенности его событий — сет-расстояние, вычисляемое как нормализованное произведение частичных показателей

$$\Delta^{\Pi}_{\mathfrak{X}_i} = \left(\frac{2}{u_{\mathfrak{X}_1} - p_{\mathfrak{X}_1}}\right)^2 \Delta^{\cup}_{\mathfrak{X}_i}\Delta^{\cap}_{\mathfrak{X}_i}, \tag{12.1.6}$$

где

$$\Delta^{\cup}_{\mathfrak{X}_i} = u_{\mathfrak{X}_i} - \langle p\rangle_{\mathfrak{X}_i}, \tag{12.1.7}$$

$$\Delta^{\cap}_{\mathfrak{X}_i} = \langle p\rangle_{\mathfrak{X}_i} - p_{\mathfrak{X}_i} \tag{12.1.8}$$

— частичные сет-расстояния категориального м.с. $\mathfrak{X}_i$, характеризующие частичную разобщенность событий относительно объединения и пересечения соответственно.

12.2 Классические сопряженные таблицы

Напомним, что классическая (m, M)-задача общности мнений m экспертов по M альтернативным категориям и в G наблюдениях характеризуется *m-мерной сопряженной* $\underbrace{M\times\ldots\times M}_{m}$ *таблицей* объема M^m, определяемой *$m\times G$ таблицей наблюдений*

	1	…	j	…	m
1	$i(1,1)$	…	$i(j,1)$	…	$i(m,1)$
…	…	…	…	…	…
g	$i(1,g)$	…	$i(j,g)$	…	$i(m,g)$
…	…	…	…	…	…
G	$i(1,G)$	…	$i(j,G)$	…	$i(m,G)$

. (12.2.1)

В каждой клетке (j,g) этой таблицы наблюдений записан номер категории $i(j,g)\in\{1,\ldots,M\}$, которую выбрал j-й эксперт $(j=1,\ldots,m)$ в g-том наблюдении $(g=1,\ldots,G)$.

Для $m=2$ *сопряженная $M\times M$ таблица клеточных статистик* имеет вид

	x^1_1	…	x^1_k	…	x^1_M	Ω
x^2_1	g^{12}_{11}	…	g^{12}_{1k}	…	g^{12}_{1M}	g^2_1
…	…	…	…	…	…	…
x^2_k	g^{12}_{k1}	…	g^{12}_{kk}	…	g^{12}_{kM}	g^2_k
…	…	…	…	…	…	…
x^2_M	g^{12}_{M1}	…	g^{12}_{Mk}	…	g^{12}_{MM}	g^2_M
Ω	g^1_1	…	g^1_k	…	g^1_M	G

, (12.2.2)

где $g^{12}_{i_1i_2}$ — количество наблюдений, в которых 1-й эксперт выбирает i_1-ю, а 2-й — i_2-ю категорию; $g^1_{i_1}$ — количество наблюдений, в которых 1-й эксперт выбирает i_1-ю категорию; $g^2_{i_2}$ — количество наблюдений, в которых 2-й эксперт выбирает i_2-ю категорию $(i_1,i_2=1,\ldots,M)$.

Для $m=2$ *сопряженной $M\times M$ таблице клеточных статистик* соответствует *сопряженная $M\times M$ таблица клеточных вероятностей*, которая имеет аналогичный вид

	x^1_1	…	x^1_k	…	x^1_M	Ω
x^2_1	π^{12}_{11}	…	π^{12}_{1k}	…	π^{12}_{1M}	π^2_1
…	…	…	…	…	…	…
x^2_k	π^{12}_{k1}	…	π^{12}_{kk}	…	π^{12}_{kM}	π^2_k
…	…	…	…	…	…	…
x^2_M	π^{12}_{M1}	…	π^{12}_{Mk}	…	π^{12}_{MM}	π^2_M
Ω	π^1_1	…	π^1_k	…	π^1_M	1

, (12.2.3)

где $\pi^{12}_{i_1i_2}$ — вероятность того, что 1-й эксперт выберет i_1-ю, а 2-й — i_2-ю категорию, оценкой которой служит соответствующая клеточная статистика $\widehat{\pi}^{12}_{i_1i_2} = g^{12}_{i_1i_2}/G$ из таблицы (12.2.1); $\pi^1_{i_1}$ — вероятность того, 1-й эксперт выберет i_1-ю категорию; $\pi^2_{i_2}$ — вероятность того, 2-й эксперт выберет i_2-ю категорию, оценками которых также служат соответствующие клеточные статистики: $\widehat{\pi}^2_{i_2} = g^2_{i_2}/G$, $\widehat{\pi}^1_{i_1} = g^1_{i_1}/G$ из той же таблицы (12.2.1).

Для классической (m,M)-задачи в G наблюдениях клеточные статистики $g^{1\ldots m}_{i_1\ldots i_m}$ *m-мерной сопряженной* $\underbrace{M\times\ldots\times M}_{m}$ *таблицы* объема M^m выражаются через элементы $i(j,g)$ *$m\times G$ таблицы наблюдений* по формуле:

$$g^{1\ldots m}_{i_1\ldots i_m} = \sum_{g=1}^{G}\prod_{j=1}^{m}\mathbf{1}(i(j,g)=i_j), \tag{12.2.4}$$

где

$$\mathbf{1}(i(j,g)=i_j) = \begin{cases}1, & i(j,g)=i_j,\\ 0, & i(j,g)\neq i_j\end{cases}$$

— индикатор равенства числовых индексов $i(j,g)=i_j$, а $i_j=1,\ldots,M$ для $j=1,\ldots,m$.

12.3 Террасные сопряженные таблицы

Перейдем к *террасной задаче общности мнений экспертов*, которая обобщает классическую только в том, что позволяет категориям быть неальтернативными.

Террасная (m,M)-задача общности мнений m экспертов по M произвольным категориям и в G наблюдениях характеризуется *m-мерной террасной сопряженной* $\underbrace{2^M\times\ldots\times 2^M}_{m}$ *таблицей* объема 2^{mM}, определяемой *террасной $m\times G$ таблицей наблюдений*

	1	…	j	…	m
1	$X(1,1)$	…	$X(j,1)$	…	$X(m,1)$
…	…	…	…	…	…
g	$X(1,g)$	…	$X(j,g)$	…	$X(m,g)$
…	…	…	…	…	…
G	$X(1,G)$	…	$X(j,G)$	…	$X(m,G)$

. (12.3.1)

В каждой клетке (j,g) этой террасной таблицы наблюдений записано подмножество категорий $X(j,g)\subseteq\mathbf{X}^j_M$ экспертного M-м.с.

$$\mathbf{X}^j_M = \{x^j_1,\ldots,x^j_M\},$$

которое выбрал j-й эксперт ($j = 1, \ldots, m$) в g-том наблюдении ($g = 1, \ldots, G$).

Для $m = 2$ *террасная сопряженная* $2^M \times 2^M$ *таблица клеточных статистик* имеет вид (12.3.2), где

- $g^{12}_{\gamma_1\gamma_2}$ — количество наблюдений, в которых 1-й эксперт выбирает γ_1-е подмножество $X^1_{\gamma_1} \subseteq \mathbf{X}^1_M$, а 2-й — γ_2-е подмножество категорий $X^2_{\gamma_2} \subseteq \mathbf{X}^2_M$;
- $g^1_{\gamma_1}$ — количество наблюдений, в которых 1-й эксперт выбирает γ_1-е подмножество категорий $X^1_{\gamma_1} \subseteq \mathbf{X}^1_M$;
- $g^2_{\gamma_2}$ — количество наблюдений, в которых 2-й эксперт выбирает γ_2-е подмножество категорий $X^2_{\gamma_2} \subseteq \mathbf{X}^2_M$ ($\gamma_1, \gamma_2 = 1, \ldots, 2^M$).

Для $m = 2$ *террасной сопряженной* $2^M \times 2^M$ *таблице клеточных статистик* соответствует *террасная сопряженная* $2^M \times 2^M$ *таблица клеточных вероятностей*, которая имеет аналогичный вид (12.3.3), где

- $\pi^{12}_{\gamma_1\gamma_2} = p(X^1_{\gamma_1} + X^2_{\gamma_2} /\!\!/ \mathbf{X}^1_M + \mathbf{X}^2_M)$ — совместная вероятность 1-го рода того, что 1-й эксперт выберет γ_1-е подмножество $X^1_{\gamma_1} \subseteq \mathbf{X}^1_M$, а 2-й — γ_2-е подмножество $X^2_{\gamma_2} \subseteq \mathbf{X}^2_M$, оценкой которой служит соответствующая клеточная статистика $\widehat{\pi}^{12}_{\gamma_1\gamma_2} = g^{12}_{\gamma_1\gamma_2}/G$ из таблицы (12.3.1);
- $\pi^1_{\gamma_1} = p(X^1_{\gamma_1} /\!\!/ \mathbf{X}^1_M)$ — маргинальная вероятность 1-го рода того, 1-й эксперт выберет γ_1-е подмножество $X^1_{\gamma_1} \subseteq \mathbf{X}^1_M$;
- $\pi^2_{\gamma_2} = p(X^2_{\gamma_2} /\!\!/ \mathbf{X}^2_M)$ — маргинальная вероятность 1-го рода того, 2-й эксперт выберет γ_2-е подмножество $X^1_{\gamma_2} \subseteq \mathbf{X}^2_M$, оценками которых также служат соответствующие клеточные статистики: $\widehat{\pi}^2_{\gamma_2} = g^2_{\gamma_2}/G$, $\widehat{\pi}^1_{\gamma_1} = g^1_{\gamma_1}/G$ из той же таблицы (12.3.1).

Для террасной (m, M)-задачи в G наблюдениях клеточные статистики $g^{1\ldots m}_{\gamma_1\ldots\gamma_m}$ *m-мерной террасной сопряженной* $\underbrace{M \times \ldots \times M}_{m}$ *таблицы*[21] объема M^m выражаются через элементы $X(j,g)$ *террасной* $m \times G$ *таблицы наблюдений* (12.3.1) по формуле:

$$g^{1\ldots m}_{\gamma_1\ldots\gamma_m} = \sum_{g=1}^{G} \prod_{j=1}^{m} \mathbf{1}\left(X(j,g) = X^j_{\gamma_j}\right), \qquad (12.3.4)$$

где

$$\mathbf{1}\left(X(j,g) = X^j_{\gamma_j}\right) = \begin{cases} 1, & X(j,g) = X^j_{\gamma_j}, \\ 0, & X(j,g) \neq X^j_{\gamma_j} \end{cases}$$

— индикатор равенства подмножеств $X(j,g) = X^j_{\gamma_j} \subseteq \mathbf{X}^j_M$, а $\gamma_j = 1, \ldots, 2^M$ для $j = 1, \ldots, m$.

[21] Для $m = 2$ см. табл. (12.3.2).

При $m = 2$ террасная сопряженная таблица (12.3.1) — это таблица совместного вероятностного распределения 1-го рода двух экспертных м.с. $\mathbf{X}^1_M$ и $\mathbf{X}^2_M$, т.е. вероятностное распределение 1-го рода их объединения — м.с. $\mathbf{X}^1_M + \mathbf{X}^2_M$, у которого суммы совместных вероятностей по строкам и столбцам этой таблицы дают маргинальные вероятностные распределения 1-го рода экспертных м.с. $\mathbf{X}^1_M$ и $\mathbf{X}^2_M$ соответственно.

	ter $\left(x^1_1 /\!\!/ \mathbf{x}^1_M\right)$	...	ter $\left(x^1_{2^M} /\!\!/ \mathbf{x}^1_M\right)$	Ω
ter $\left(x^2_1 /\!\!/ \mathbf{x}^2_M\right)$	g^{12}_{11}	...	$g^{12}_{12^M}$	g^2_1
...	...	...	...	...
ter $\left(x^2_{2^M} /\!\!/ \mathbf{x}^2_M\right)$	$g^{12}_{2^M 1}$	...	$g^{12}_{2^M 2^M}$	$g^2_{2^M}$
Ω	g^1_1	...	$g^1_{2^M}$	G

,(12.3.2)

	ter $\left(x^1_1 /\!\!/ \mathbf{x}^1_M\right)$	...	ter $\left(x^1_{2^M} /\!\!/ \mathbf{x}^1_M\right)$	Ω
ter $\left(x^2_1 /\!\!/ \mathbf{x}^2_M\right)$	π^{12}_{11}	...	$\pi^{12}_{12^M}$	π^2_1
...	...	...	...	...
ter $\left(x^2_{2^M} /\!\!/ \mathbf{x}^2_M\right)$	$\pi^{12}_{2^M 1}$	...	$\pi^{12}_{2^M 2^M}$	$\pi^2_{2^M}$
Ω	π^1_1	...	$\pi^1_{2^M}$	1

,(12.3.3)

12.4 Сопряженное вероятностное распределение м.с.

Классическая задача общности мнений многих экспертов по многим категориям возникает при обработке результатов эксперимента, которые могут быть записаны в виде *классической сопряженной* (m, M)*-таблицы*. Такая таблица характеризуется соответствующим *классическим сопряженным* mM*-м.с.*, имеющим вероятностное распределение 1-го рода, которое в силу особенностей классической задачи обладает специальной структурой общности, ассоциации и зависимости событий, определяемой *альтернативными категориями*.

Обобщением классической задачи служит *террасная задача общности* мнений многих экспертов по многим категориям, которая возникает при обработке результатов *террасного эксперимента*, записываемых в виде *террасной сопряженной* (m, M)*-таблицы*. Такая таблица характеризуется соответствующим *террасным сопряженным* mM*-м.с.*, имеющим вероятностное распределение 1-го рода, вообще говоря, произвольного вида, определяемого *неальтернативными категориями*. Опираясь на особенности террасной задачи, можно сделать некоторые предположения о свойствах подобных террасных сопряженных м.с. и их вероятностных распределений, определяющих семейство распределений, рамок которого достаточно для решения произвольной террасной сопряженной задачи.

Чтобы сформулировать данные предположения, нам понадобится обозначение

$$\mathbf{X}_M = \{x_1, \ldots, x_M\} \qquad (12.4.1)$$

для *экспериментаторского* M*-м.с*, каждое событие

которого $x_i, i = 1, \ldots, M$, происходит тогда, когда *экспериментатор* предъявляет экспертам предмет, обладающий i-м свойством из M заданных в террасном эксперименте неальтернативных свойств. Таким образом, любой предмет, предъявляемый экспертам в террасном эксперименте, характеризуется подмножеством событий $X \subseteq \mathbf{X}_M$, соответствующим тому подмножеству свойств, которыми он обладает. Задачей эксперта в террасном эксперименте является оценка: "произошло или нет каждое из событий *экспериментаторского M-м.с.*". Иначе говоря, определить подмножество произошедших событий $X \subseteq \mathbf{X}_M$, что эквивалентно оценке: "произошло или нет террасное событие"

$$\mathsf{ter}(X /\!\!/ \mathbf{X}_M) = \bigcap_{x \in X} x \bigcap_{x \in \mathbf{X}_M - X} x^c, \tag{12.4.2}$$

порожденное *экспериментаторским м.с.* $\mathbf{X}_M$ и занумерованное подмножеством $X \subseteq \mathbf{X}_M$.

Вероятности

$$p(X /\!\!/ \mathbf{X}_M) = \mathbf{P}\Big(\mathsf{ter}(X /\!\!/ \mathbf{X}_M)\Big) \tag{12.4.3}$$

террасных событий (12.4.2) образуют известное экспериментатору, но неизвестное экспертам вероятностное распределение 1-го рода

$$\{p(X /\!\!/ \mathbf{X}_M), X \subseteq \mathbf{X}_M\} \tag{12.4.4}$$

экспериментаторского M-м.с. $\mathbf{X}_M$.

Обозначим, как и раньше, для $j = 1, \ldots, m$

$$\mathbf{X}^j = \{x_1^j, \ldots, x_M^j\} = \{x^j, x \in \mathbf{X}_M\} \tag{12.4.5}$$

— *j-е экспертное M-м.с.*, которое характеризует событийные оценки j-го эксперта, т.е. каждое событие $x^j \in \mathbf{X}^j$ которого происходит тогда, когда j-й эксперт оценивает, что произошло событие $x \in \mathbf{X}_M$ из *экспериментаторского M-м.с.* $\mathbf{X}_M$ (12.4.1). Обозначим также для $j = 1, \ldots, m$

$$\{p^j(X /\!\!/ \mathbf{X}^j), X \subseteq \mathbf{X}^j\} \tag{12.4.6}$$

— вероятностное распределение 1-го рода j-го экспертного M-м.с. $\mathbf{X}^j$, где

$$p^j(X /\!\!/ \mathbf{X}^j) = \mathbf{P}\Big(\mathsf{ter}(X /\!\!/ \mathbf{X}^j)\Big) =$$

$$= \mathbf{P}\left(\bigcap_{x \in X} x^j \bigcap_{x \in \mathbf{X}_M - X} (x^j)^c\right) \tag{12.4.7}$$

— вероятность соответствующего террасного события, порожденное j-м экспертным M-м.с. $\mathbf{X}^j$ и занумерованное подмножеством $X \subseteq \mathbf{X}_M$ *экспериментаторского M-м.с.* $\mathbf{X}_M$.

12.5 Террасные категориальные м.с.

Террасные категориальные м.с.

$$\mathfrak{X}_\gamma = \{\mathsf{ter}(X_\gamma^1 /\!\!/ \mathbf{X}_M^1), \ldots, \mathsf{ter}(X_\gamma^m /\!\!/ \mathbf{X}_M^m)\} \tag{12.5.1}$$

— это m-м.с., для каждого $\gamma = 1, \ldots, 2^M$ составленные из "одноименных" террасных событий $\mathsf{ter}(X_\gamma^j /\!\!/ \mathbf{X}_M^j)$, порожденных *экспертными M-м.с.*

$$\mathbf{X}_M^j = \{x_1^j, \ldots, x_M^j\} \tag{12.5.2}$$

и занумерованных их подмножествами $X_\gamma^j \subseteq \mathbf{X}_M^j, j = 1, \ldots, m$, которые "имеют одно и то же имя", т.е. соответствуют одному и тому же подмножеству $X_\gamma \subseteq \mathfrak{X}_M$ *экспериментаторского M-м.с.*

$$\mathbf{X}_M = \{x_1, \ldots, x_M\}. \tag{12.5.3}$$

13 Выборочные дисперсии и доверительные интервалы для оценок показателей общности м.с. и методы их статистической оценки

13.1 Выборочная дисперсия оценки сет-расстояния и связанных с ним характеристик м.с.

Сет-О-расстояние:

$$\Delta_X^{\cup} = u_X - \langle p \rangle_X$$

Сет-П-расстояние:

$$\Delta_X^{\cap} = \langle p \rangle_X - p_X$$

Сет-расстояние (вероятность разобщенности м.с.):

$$\Delta_X^{\Sigma} = \Delta_X^{\cup} + \Delta_X^{\cap} = u_X - p_X$$

Близость (вероятность общности м.с.)

$$\delta_X^{\Sigma} = 1 - \Delta_X^{\Sigma}$$

Задача. Требуется найти формулы для выборочных дисперсий оценок этих числовых характеристик м.с. X по его G независимым испытаниям (независимым наблюдениям за м.с. X). Например,

$$\widehat{\mathrm{Var}}(\widehat{\Delta}_X^{\Sigma}),$$

где $\widehat{\Delta}_X^{\Sigma}$ — оценка сет-расстояния Δ_X^{Σ} по G независимым испытаниям (наблюдениям за) м.с. X, и т.д.

Подзадачи. В силу того, что все перечисленные величины конструируются из следующих составляющих:

$$\langle p \rangle_X$$

— средняя вероятность,

$$u_X, u_X^-, u_X^+, u_X^\star$$

— вероятность объединения, ее нижняя и верхняя границы Фреше и и она же для тотально независимого м.с.,

$$p_X, p_X^-, p_X^+, p_X^\star,$$

— вероятность пересечения, ее нижняя и верхняя границы Фреше и и она же для тотально независимого м.с.; то сначала необходимо вывести аналогичные формулы для выборочной дисперсии оценок этих составляющих величин.

Подзадача 1. Требуется найти формулу для выборочной дисперсии оценки вероятности пересечения событий из м.с. X по G независимым испытаниям (наблюдениям за) м.с. X

$$\widehat{\mathrm{Var}}(\widehat{p}_X),$$

где

$$\widehat{p}_X = \frac{1}{G}\sum_{g=1}^{G} \mathbf{1}_{\mathsf{ter}_X}(\omega_g)$$

— оценка вероятности пересечения

$$p_X = \mathbf{P}\left(\mathsf{ter}_X\right)$$

по G независимым испытаниям (наблюдениям за) м.с. X, а

$$\mathsf{ter}_X = \bigcap_{x\in X} x$$

— террасное событие 2-го рода, занумерованное м.с. X.

Поскольку $\mathbf{1}_{\mathsf{ter}_X}$ — булева с.в., которая принимает значение 1 с вероятностью p_X, то ее выборочная дисперсия равна

$$\widehat{\mathrm{Var}}(\mathbf{1}_{\mathsf{ter}_X}) = \widehat{p}_X(1-\widehat{p}_X),$$

а выборочная дисперсия среднего арифметического G таких булевых с.в. равна

$$\widehat{\mathrm{Var}}(\widehat{p}_X) = \frac{1}{G^2}\sum_{g=1}^{G}\widehat{\mathrm{Var}}\left(\mathbf{1}_{\mathsf{ter}_X}\right) = \frac{1}{G}\widehat{p}_X(1-\widehat{p}_X), \tag{13.1.1}$$

так как испытания независимы.

Подзадача 2. Требуется найти формулу для выборочной дисперсии оценки вероятности объединения событий из м.с. X по G независимым испытаниям (наблюдениям за) м.с. X

$$\widehat{\mathrm{Var}}(\widehat{u}_X),$$

где

$$\widehat{u}_X = \frac{1}{G}\sum_{g=1}^{G} \mathbf{1}_{\mathsf{Ter}_X}(\omega_g)$$

— оценка вероятности объединения

$$u_X = \mathbf{P}\left(\mathsf{Ter}_X\right)$$

по G независимым испытаниям (наблюдениям за) м.с. X, а

$$\mathsf{Ter}_X = \bigcup_{x\in X} x$$

— террасное событие 4-го рода, занумерованное м.с. X.

Поскольку $\mathbf{1}_{\mathsf{Ter}_X}$ — булева с.в., которая принимает значение 1 с вероятностью u_X, то ее выборочная дисперсия равна

$$\widehat{\mathrm{Var}}(\mathbf{1}_{\mathsf{Ter}_X}) = \widehat{u}_X(1-\widehat{u}_X),$$

а выборочная дисперсия среднего арифметического G таких булевых с.в. равна

$$\widehat{\mathrm{Var}}(\widehat{u}_X) = \frac{1}{G^2}\sum_{g=1}^{G}\widehat{\mathrm{Var}}\left(\mathbf{1}_{\mathsf{Ter}_X}\right) =$$

$$= \frac{1}{G}\widehat{u}_X(1-\widehat{u}_X), \tag{13.1.2}$$

так как испытания независимы.

Подзадача 3. Требуется найти формулу для выборочной дисперсии оценки сет-расстояния м.с. X по G независимым испытаниям (наблюдениям за) м.с. X

$$\widehat{\mathrm{Var}}(\widehat{\Delta}_X^\Sigma),$$

где

$$\widehat{\Delta}_X^\Sigma = \widehat{\Delta}_X^\cup + \widehat{\Delta}_X^\cap = \widehat{u}_X - \widehat{\langle p\rangle}_X + \widehat{\langle p\rangle}_X - \widehat{p}_X = \widehat{u}_X - \widehat{p}_X$$

— оценка сет-расстояния

$$\Delta_X^\Sigma = \Delta_X^\cup + \Delta_X^\cap$$

по G независимым испытаниям (наблюдениям за) м.с. X.

В силу решения подзадач 1 и 2 выборочная дисперсия сет-расстояния м.с. X равна

$$\widehat{\mathrm{Var}}(\widehat{\Delta}_X^\Sigma) = \widehat{\mathrm{Var}}(\widehat{u}_X - \widehat{p}_X) =$$

$$= \widehat{\mathrm{Var}}(\widehat{u}_X) + \widehat{\mathrm{Var}}(\widehat{p}_X) - 2\widehat{\mathsf{cov}}\left(\widehat{u}_X, \widehat{p}_X\right).$$

Поскольку $\mathbf{1}_{\mathsf{Ter}_X}\mathbf{1}_{\mathsf{ter}_X} = \mathbf{1}_{\mathsf{ter}_X}$ из-за того, что $\mathsf{ter}_X \subseteq \mathsf{Ter}_X$, то ковариация булевых с.в. $\mathbf{1}_{\mathsf{Ter}_X}$ и $\mathbf{1}_{\mathsf{ter}_X}$ равна

$$\mathsf{cov}\left(\mathbf{1}_{\mathsf{Ter}_X}, \mathbf{1}_{\mathsf{ter}_X}\right) =$$

$$= \mathbf{E}\left(\mathbf{1}_{\mathsf{Ter}_X}\mathbf{1}_{\mathsf{ter}_X}\right) - \mathbf{E}\left(\mathbf{1}_{\mathsf{Ter}_X}\right)\mathbf{E}\left(\mathbf{1}_{\mathsf{ter}_X}\right) = p_X\left(1-u_X\right),$$

а выборочная ковариация их выборочных средних по G независимым испытаниям (наблюдениям за) м.с. X равна

$$\widehat{\mathsf{cov}}\left(\widehat{u}_X, \widehat{p}_X\right) = \frac{1}{G}\widehat{p}_X\left(1-\widehat{u}_X\right).$$

Отсюда выборочная дисперсия оценки сет-расстояния равна

$$\widehat{\mathrm{Var}}(\widehat{\Delta}_X^{\Sigma}) =$$

$$= \frac{1}{G}\Big(\widehat{u}_X(1-\widehat{u}_X)+\widehat{p}_X(1-\widehat{p}_X)-2\widehat{p}_X\,(1-\widehat{u}_X)\Big) =$$

$$= \frac{1}{G}\Big((1-\widehat{u}_X+\widehat{p}_X)-(1-\widehat{u}_X+\widehat{p}_X)^2\Big) =$$

$$= \frac{1}{G}(1-\widehat{\Delta}_X^{\Sigma})\widehat{\Delta}_X^{\Sigma}. \qquad (13.1.3)$$

Выборочная дисперсия (13.1.3) — это вогнутая функция оценки сет-расстояния $\widehat{\Delta}_X^{\Sigma}$, которая достигает максимума при $\widehat{\Delta}_X^{\Sigma}=1/2$, равного $\frac{1}{4G}$, и минимума, равного нулю, когда $\widehat{\Delta}_X^{\Sigma}=0$ и когда $\widehat{\Delta}_X^{\Sigma}=1$.

Из (13.1.3) следует, что выборочная дисперсия оценки близости $\delta_X^{\Sigma}=1-\Delta_X^{\Sigma}$ и сет-расстояния Δ_X^{Σ} совпадают:

$$\widehat{\mathrm{Var}}(\widehat{\delta}_X^{\Sigma}) = \frac{1}{G}(1-\widehat{\delta}_X^{\Sigma})\widehat{\delta}_X^{\Sigma} = \frac{1}{G}\widehat{\Delta}_X^{\Sigma}(1-\widehat{\Delta}_X^{\Sigma}). \qquad (13.1.4)$$

13.2 Выборочная дисперсия оценки нормализаций сет-расстояния

Сет-расстояние (вероятность разобщенности м.с.):

$$\Delta_X^{\Sigma} = \Delta_X^{\cup} + \Delta_X^{\cap} = u_X - p_X \qquad (13.2.1)$$

Близость (вероятность общности м.с.)

$$\delta_X^{\Sigma} = 1 - \Delta_X^{\Sigma} \qquad (13.2.2)$$

Фреше-нормализация сет-расстояния:

$$\lceil\Delta_X^{\Sigma}\rceil^{\pm} = \frac{\Delta_X^{\Sigma}-\Delta_X^{\Sigma-}}{\Delta_X^{\Sigma+}-\Delta_X^{\Sigma-}} \qquad (13.2.3)$$

Коэн-нормализация сет-расстояния:

$$\lceil\Delta_X^{\Sigma}\rceil^{\star} = \frac{\Delta_X^{\Sigma}-\Delta_X^{\Sigma\star}}{1-\Delta_X^{\Sigma\star}} \qquad (13.2.4)$$

Фреше-Коэн-нормализация сет-расстояния:

$$\lceil\Delta_X^{\Sigma}\rceil = \lceil\lceil\Delta_X^{\Sigma}\rceil^{\star}\rceil^{\pm} = \lceil\lceil\Delta_X^{\Sigma}\rceil^{\pm}\rceil^{\star} =$$

$$= \frac{\Delta_X^{\Sigma}-\Delta_X^{\Sigma\star}}{\Delta_X^{\Sigma+}-\Delta_X^{\Sigma\star}} \qquad (13.2.5)$$

Здесь

$$\Delta_X^{\Sigma-} = \Delta_X^{\cup-} + \Delta_X^{\cap-} = u_X^- - p_X^+ \qquad (13.2.6)$$

$$\Delta_X^{\Sigma+} = \Delta_X^{\cup+} + \Delta_X^{\cap+} = u_X^+ - p_X^- \qquad (13.2.7)$$

— нижняя и верхняя границы Фреше сет-расстояния Δ_X^{Σ}, а

$$\Delta_X^{\Sigma\star} = \Delta_X^{\cup\star} + \Delta_X^{\cap\star} = u_X^{\star} - p_X^{\star} \qquad (13.2.8)$$

— сет-расстояние м.с. X, будь его события тотально независимы.

Задача. Требуется найти формулы для выборочных дисперсий оценок этих числовых характеристик м.с. X по его G независимым испытаниям (независимым наблюдениям за м.с. X). Например,

$$\widehat{\mathrm{Var}}(\widehat{\lceil\Delta_X^{\Sigma}\rceil}), \qquad (13.2.9)$$

где $\widehat{\lceil\Delta_X^{\Sigma}\rceil}$ — оценка Фреше-Коэн-нормализации сет-расстояния Δ_X^{Σ} по G независимым испытаниям (наблюдениям за) м.с. X, и т.д.

Решение. Фреше-Коэн-нормализации сет-расстояния Δ_X^{Σ} определяется формулой

$$\lceil\Delta_X^{\Sigma}\rceil = \frac{\Delta_X^{\Sigma}-\Delta_X^{\Sigma\star}}{\Delta_X^{\Sigma+}-\Delta_X^{\Sigma\star}}, \qquad (13.2.10)$$

где

$$\Delta_X^{\Sigma\star} = u_X^{\star} - p_X^{\star} = \mathbf{P}^{\star}(\mathsf{Ter}_X - \mathsf{ter}_X) =$$

$$= 1 - \prod_{x\in X} p_x - \prod_{x\in X}(1-p_x) \qquad (13.2.11)$$

— вероятность события $\mathsf{Ter}_X - \mathsf{ter}_X$, порожденного м.с. X, будь его события тотально независимы;

$$\Delta_X^{+} = u_X^{+} - p_X^{-} \qquad (13.2.12)$$

— верхняя граница Фреше, где

$$u_X^{+} = \min\left\{1, \sum_{x\in X} p_x\right\}, \qquad (13.2.13)$$

$$p_X^{-} = \max\left\{0, 1-\sum_{x\in X}(1-p_x)\right\}. \qquad (13.2.14)$$

Отсюда

$$\widehat{\Delta_X^{\Sigma+}} =$$

$$= \min\left\{1, \sum_{x\in X}\hat{p}_x\right\} - \max\left\{0, 1-\sum_{x\in X}(1-\hat{p}_x)\right\}, \qquad (13.2.15)$$

$$\widehat{\Delta_X^{\Sigma\star}} = 1 - \prod_{x\in X}\hat{p}_x - \prod_{x\in X}(1-\hat{p}_x). \qquad (13.2.16)$$

Кроме того, оценка нижней границы Фреше сет-расстояния м.с. X имеет вид

$$\widehat{\Delta_X^{\Sigma-}} = \max_{x\in X}\hat{p}_x - \min_{x\in X}\hat{p}_x, \qquad (13.2.17)$$

так как сама нижняя граница определяется формулой:

$$\Delta_X^{\Sigma-} = u_X^- - p_X^+ = \max_{x\in X} p_x - \min_{x\in X} p_x. \qquad (13.2.18)$$

Из основанного на неравенстве (13.3.19) предположения, что выборочные оценки $\widehat{\Delta_X^{\Sigma+}}, \widehat{\Delta_X^{\Sigma-}}$ и $\widehat{\Delta_X^{\Sigma\star}}$

можно считать известными константами по отношению к выборочной оценке $\widehat{\Delta}_X^{\Sigma}$, получаем из (13.2.10) формулы для выборочных дисперсий Фреше-Коэн-нормализаций сет-расстояния и близости в следующем виде:

$$\widehat{\mathrm{Var}}(\widehat{\lceil \Delta_X^{\Sigma} \rceil}) = \widehat{\mathrm{Var}}(\widehat{\Delta}_X^{\Sigma}) \frac{1}{(\widehat{\Delta_X^{\Sigma+}} - \widehat{\Delta_X^{\Sigma\star}})^2}, \tag{13.2.19}$$

$$\widehat{\mathrm{Var}}(\widehat{\lceil \delta_X^{\Sigma} \rceil}) = \widehat{\mathrm{Var}}(\widehat{\delta}_X^{\Sigma}) \frac{1}{(\widehat{\Delta_X^{\Sigma\star}} - \widehat{\Delta_X^{\Sigma-}})^2}. \tag{13.2.20}$$

13.3 Связь между выборочными дисперсиями оценок вероятностей событий и террасных событий, порожденных м.с.

Чтобы получить статистические оценки вероятностного распределения

$$\{p(X/\!\!/\mathfrak{X}), X \subseteq \mathfrak{X}\} \tag{13.3.1}$$

некоторого м.с. $\mathfrak{X}$, где

$$p(X/\!\!/\mathfrak{X}) = \mathbf{P}\left(\mathsf{ter}(X/\!\!/\mathfrak{X})\right) \tag{13.3.2}$$

— вероятности террасных событий

$$\mathsf{ter}(X/\!\!/\mathfrak{X}) = \bigcap_{x \in X} x \bigcap_{x \in \mathfrak{X} - X} x^c, \tag{13.3.3}$$

порожденных $\mathfrak{X}$, т.е. чтобы получить оценки $\hat{p}(X/\!\!/\mathfrak{X})$ вероятностей террасных событий; а также — оценки различных вероятностных характеристик $\mathfrak{X}$, например, оценки $\hat{p}_x$ вероятностей его событий $x \in \mathfrak{X}$:

$$p_x = \mathbf{P}(x) = \sum_{x \in X \subseteq \mathfrak{X}} p(X/\!\!/\mathfrak{X}), \tag{13.3.4}$$

требуется некоторый ряд из G независимых наблюдений за его реализациями (ряд независимых испытаний этого м.с.) — террасными событиями:

$$\mathsf{ter}(X_1/\!\!/\mathfrak{X}), \ldots, \mathsf{ter}(X_g/\!\!/\mathfrak{X}), \ldots, \mathsf{ter}(X_G/\!\!/\mathfrak{X}), \tag{13.3.5}$$

который означает, что в g-том наблюдении произошло террасное событие $\mathsf{ter}(X_g/\!\!/\mathfrak{X})$, занумерованное подмножеством $X_g \subseteq \mathfrak{X}$, образованном событиями, наступившими в g-том наблюдении.

В этом параграфе нас будет интересовать связь между выборочными дисперсиями оценок

$$\hat{p}(X/\!\!/\mathfrak{X}) = \frac{1}{G} \sum_{g=1}^{G} \mathbf{1}_{\mathsf{ter}(X/\!\!/\mathfrak{X})}(\omega_g) \tag{13.3.6}$$

— вероятностей террасных событий $\mathsf{ter}(X/\!\!/\mathfrak{X}), X \subseteq \mathfrak{X}$; и оценок

$$\hat{p}_x = \frac{1}{G} \sum_{g=1}^{G} \mathbf{1}_x(\omega_g) \tag{13.3.7}$$

— вероятностей событий $x \in \mathfrak{X}$, где $\omega_g \in \mathsf{ter}(X_g/\!\!/\mathfrak{X})$.

13.4 Простейшие свойства оценок вероятностей событий и террасных событий

Статистические оценки (13.3.6) и (13.3.7) мы будем рассматривать как с.в., составленные суммами булевых с.в.

$$\mathbf{1}_{\mathsf{ter}(X/\!\!/\mathfrak{X})}(\omega_g) = \begin{cases} 1, & \omega_g \in \mathsf{ter}(X/\!\!/\mathfrak{X}), \\ 0, & \text{иначе}, \end{cases} \tag{13.3.8}$$

$$\mathbf{1}_x(\omega_g) = \begin{cases} 1, & \omega_g \in x, \\ 0, & \text{иначе}. \end{cases} \tag{13.3.9}$$

Поскольку $\mathbf{E}\mathbf{1}_{\mathsf{ter}(X/\!\!/\mathfrak{X})} = p(X/\!\!/\mathfrak{X})$, $\mathbf{E}\mathbf{1}_x = p_x$, то

$$\widehat{\mathbf{E}}\mathbf{1}_{\mathsf{ter}(X/\!\!/\mathfrak{X})} = \hat{p}(X/\!\!/\mathfrak{X}), \tag{13.3.10}$$

$$\widehat{\mathbf{E}}\mathbf{1}_x = \hat{p}_x. \tag{13.3.11}$$

Выборочная дисперсия и ковариация оценки вероятности террасного события. Так как

$$\mathrm{Var}\left(\mathbf{1}_{\mathsf{ter}(X/\!\!/\mathfrak{X})}\right) = p(X/\!\!/\mathfrak{X})(1 - p(X/\!\!/\mathfrak{X})),$$

то

$$\widehat{\mathrm{Var}}\left(\mathbf{1}_{\mathsf{ter}(X/\!\!/\mathfrak{X})}\right) = \hat{p}(X/\!\!/\mathfrak{X})(1 - \hat{p}(X/\!\!/\mathfrak{X})), \tag{13.3.12}$$

В силу того, что для $X \neq X'$

$$\mathsf{ter}(X/\!\!/\mathfrak{X}) \cap \mathsf{ter}(X'/\!\!/\mathfrak{X}) = \emptyset,$$

$$\mathsf{cov}\left(\mathbf{1}_{\mathsf{ter}(X/\!\!/\mathfrak{X})}, \mathbf{1}_{\mathsf{ter}(X'/\!\!/\mathfrak{X})}\right) =$$

$$= \begin{cases} p(X/\!\!/\mathfrak{X})(1 - p(X/\!\!/\mathfrak{X})), & X = X', \\ -p(X/\!\!/\mathfrak{X})p(X'/\!\!/\mathfrak{X}), & X \neq X', \end{cases}$$

поэтому

$$\widehat{\mathsf{cov}}\left(\mathbf{1}_{\mathsf{ter}(X/\!\!/\mathfrak{X})}, \mathbf{1}_{\mathsf{ter}(X'/\!\!/\mathfrak{X})}\right) =$$

$$= \begin{cases} \hat{p}(X/\!\!/\mathfrak{X})(1 - \hat{p}(X/\!\!/\mathfrak{X})), & X = X', \\ -\hat{p}(X/\!\!/\mathfrak{X})\hat{p}(X'/\!\!/\mathfrak{X}), & X \neq X'. \end{cases} \tag{13.3.13}$$

Выборочная дисперсия и ковариация оценки вероятности события. Аналогично, так как

$$\mathrm{Var}\left(\mathbf{1}_x\right) = p_x(1 - p_x),$$

то

$$\widehat{\mathrm{Var}}\left(\mathbf{1}_x\right) = \hat{p}_x(1 - \hat{p}_x). \tag{13.3.14}$$

В силу того, что для $x, y \in \mathfrak{X}$ парная ковариация индикаторов событий x и y вычисляется по формуле

$$\mathsf{cov}\left(\mathbf{1}_x, \mathbf{1}_y\right) = \begin{cases} p_x(1 - p_x), & x = y, \\ p_{xy} - p_x p_y = \mathsf{Kov}_{xy}, & x \neq y, \end{cases}$$

ее оценка имеет вид:

$$\widehat{\mathsf{cov}}\left(\mathbf{1}_x, \mathbf{1}_y\right) = \begin{cases} \hat{p}_x(1 - \hat{p}_x), & x = y, \\ \hat{p}_{xy} - \hat{p}_x\hat{p}_y = \widehat{\mathsf{Kov}}_{xy}, & x \neq y, \end{cases} \tag{13.3.15}$$

где

$$\hat{p}_{xy} = \frac{1}{G}\sum_{g=1}^{G} \mathbf{1}_{\mathsf{ter}_{\{x,y\}/\!/\mathfrak{X}}}(\omega_g) \tag{13.3.16}$$

— оценка вероятности 2-го рода

$$p_{xy} = \mathbf{P}\left(\mathsf{ter}_{\{x,y\}/\!/\mathfrak{X}}\right) = \mathbf{P}(x \cap y);$$

$$\mathsf{ter}_{\{x,y\}/\!/\mathfrak{X}} = x \cap y$$

— террасное событие 2-го рода, занумерованное дуплетом событий $\{x, y\} \subseteq \mathfrak{X}$;

$$\mathsf{Kov}_{xy} = p_{xy} - p_x p_y$$

— парная ковариация 2-го рода событий x и y, для дуплета событий совпадающая с парной ковариацией их индикаторов;

$$\widehat{\mathsf{Kov}}_{xy} = \hat{p}_{xy} - \hat{p}_x \hat{p}_y$$

— ее оценка, также для дуплета событий совпадающая с оценкой парной ковариации их индикаторов.

Выборочный коэффициент вариации оценок вероятности события и вероятности террасного события.

Поскольку, когда $x \in X$, вероятности событий p_x (13.3.4) существенно больше вероятностей террасных событий $p(X/\!/\mathfrak{X})$ (13.3.2) и, вообще говоря, тем больше, чем больше мощность м.с. $\mathfrak{X}$, то для сравнения их статистических отклонений от своих средних значений полезно использовать относительные показатели, например, выборочные коэффициенты вариации.

Для оценки вероятности террасного события $\mathsf{ter}(X/\!/\mathfrak{X})$ выборочный коэффициент вариации равен

$$\hat{V}_\sigma(\mathbf{1}_{\mathsf{ter}(X/\!/\mathfrak{X})}) = \frac{\hat{\sigma}(\mathbf{1}_{\mathsf{ter}(X/\!/\mathfrak{X})})}{\hat{p}(X/\!/\mathfrak{X})} =$$

$$= \sqrt{\frac{1-\hat{p}(X/\!/\mathfrak{X})}{\hat{p}(X/\!/\mathfrak{X})}}, \tag{13.3.17}$$

где

$$\hat{\sigma}(\mathbf{1}_{\mathsf{ter}(X/\!/\mathfrak{X})}) = \left(\widehat{\mathrm{Var}}(\mathbf{1}_{\mathsf{ter}(X/\!/\mathfrak{X})})\right)^{1/2} =$$

$$= \sqrt{\hat{p}(X/\!/\mathfrak{X})(1-\hat{p}(X/\!/\mathfrak{X}))}$$

— оценка среднеквадратичного отклонения индикатора террасного события $\mathsf{ter}(X/\!/\mathfrak{X})$.

Для оценки вероятности события $x \in \mathfrak{X}$ выборочный коэффициент вариации равен

$$\hat{V}_\sigma(\mathbf{1}_x) = \frac{\hat{\sigma}(\mathbf{1}_x)}{\hat{p}_x} = \sqrt{\frac{1-\hat{p}_x}{\hat{p}_x}}, \tag{13.3.18}$$

где

$$\hat{\sigma}(\mathbf{1}_x) = \left(\widehat{\mathrm{Var}}(\mathbf{1}_x)\right)^{1/2} = \sqrt{\hat{p}_x(1-\hat{p}_x)}$$

— оценка среднеквадратичного отклонения индикатора события $x \in \mathfrak{X}$.

Из (13.3.17) и (13.3.18) получаем, что, когда $x \in X$,

$$\hat{V}_\sigma(\mathbf{1}_{\mathsf{ter}(X/\!/\mathfrak{X})}) \geq \hat{V}_\sigma(\mathbf{1}_x), \tag{13.3.19}$$

поскольку $\hat{p}_x \geq \hat{p}(X/\!/\mathfrak{X})$ для $x \in X$ в силу (13.3.4).

13.5 Доверительный интервал

Допустим $\Delta_{X_1}, \ldots, \Delta_{X_G}$ — независимая выборка из нормально распределенной совокупности с ненаблюдаемыми (параметрами) средним $\mu = \mathbf{E}(\Delta_X)$ и дисперсией $\sigma^2 = \mathrm{Var}(\Delta_X)$. Пусть

$$m = \frac{1}{G}(\Delta_{X_1} + \cdots + \Delta_{X_G}) = \widehat{\Delta}_X,$$

$$S^2 = \frac{1}{G-1}\sum_{g=1}^{G}\left(\Delta_{X_g} - m\right)^2 = \widehat{\mathrm{Var}}(\widehat{\Delta}_X),$$

где $m = \widehat{\Delta}_X$ — выборочное среднее, а $S^2 = \widehat{\mathrm{Var}}(\widehat{\Delta}_X)$ — выборочная дисперсия. Тогда

$$T = \frac{m-\mu}{S/\sqrt{G}}$$

имеет t-распределение Стьюдента с $G-1$ степенью свободы. Заметим, что распределение T не зависит от значений ненаблюдаемых параметров μ и σ^2. Допустим мы хотим вычислить α доверительный интервал для μ. Тогда, обозначая t_α как α-квантиль этого распределения[22],

$$\mathbf{P}\left(-t_{\frac{1+\alpha}{2}} \leq T \leq t_{\frac{1+\alpha}{2}}\right) = 0.95,$$

(если $\alpha = 0.95$, то есть 2.5% шанс, что T будет меньше, чем $-t_{\frac{1+\alpha}{2}} = {}^{23} = t_{\frac{1-\alpha}{2}}$ и 2.5% шанс, что T будет больше, чем $+t_{\frac{1+\alpha}{2}}$. Таким образом, вероятность, что T будет между $-t_{\frac{1+\alpha}{2}}$ и $+t_{\frac{1+\alpha}{2}}$ равна 0.95.)

Следовательно

$$\mathbf{P}\left(\widehat{\Delta}_X - t_{\frac{1+\alpha}{2}}\frac{S}{\sqrt{G}} \leq \mu \leq \widehat{\Delta}_X + t_{\frac{1+\alpha}{2}}\frac{S}{\sqrt{G}}\right) = 0.95$$

и мы имеем статистическую оценку границ теоретического 95% доверительного интервала для $\mu = \mathbf{E}(\Delta_X)$.

14 Приложение

14.1 Цитаты о позициях некоторых авторов, развивающих теорию каппа-показателей, по поводу “independent raters” и “chance-corrected kappa”

Метка **(Yes; Yes)** означает, что автор считает возможным предполагать экспертов независимыми

[22] α-квантиль t_α распределения с.в. T — это такое значение, что $\mathbf{P}(T < t_\alpha) = \alpha$.

[23] в силу симметрии распределения Стьюдента.

("independent raters"), а каппа-показатель — скорректировать таким образом, чтобы "удалить из него случайность" ("chance-corrected kappa"). Метка **(?; ?)** означает, что позиция авторов по этим двум вопросам не сформулирована. Моя точка зрения **(No; No)** совпадает с позицией Adejumo et al [6, 2004, p.12] и Agresti [8, 2007, p.21].

(Yes; Yes) — **Cohen [9, 1960, p.38]**: "Thus, the judges may be clinical psychologists, the categories schizophrenic, neurotic, and brain-damaged, and the units psychological test protocols; or the judges may be social psychologists, the categories various types of leadership, and the units small groups, etc. These have in common the following conditions, which may be taken as assumptions of the coefficient of agreement to be proposed:

1. The units are *independent.*
2. The categories of the nominal scale are *independent, mutually exclusive, and exhaustive.*
3. The judges operate *independently.*"

Cohen [9, 1960, p.40]: "To the extent to which *nonchance factors* are operating in the direction of agreement, p_0 will exceed p_c; their difference, $p_0 - p_c$ represents the proportion of the cases in which beyond-chance agreement occurred and is the numerator of the coefficient. The coefficient κ is simply the proportion of *chance-expected disagreements* which do not occur, or alternatively, it is the proportion of agreement after *chance agreement is removed* from consideration".

(No; No) — **Adejumo et al [6, 2004, p.12]**: "Thus, the common statement that kappa is a "chance-corrected measure of agreement" is misleading. As a test statistic, kappa can verify that agreement exceeds chance levels. But as a measure of the level of agreement, kappa is not "chance-corrected"; ... A better case for using kappa to qualify rater agreement is that, under certain conditions, it approximates the intra-class correlation. But this too is problematic in that (1) these conditions are not always met, and (2) one could instead directly calculate the intra-class correlation".

(?; Yes) — **Gwet [11, 2014, p.32]**: "I consider a subject categorization to be random if it is not based on any known and predetermined process".

(?; No) — **Uebersax [13, 2009]**: "Kappa is not really a chance-corrected measure of agreement. Kappa is an omnibus index of agreement. It does not make distinctions among various types and sources of disagreement"; [14, 2010]: "Despite its reputation as a chance-corrected agreement measure, kappa does not correct for chance agreement".

(Yes; Yes) — **Shoukri [16, 2004, §3.3, p.36]**: "Cohen [9, 1960] proposed Kappa as a *chance-corrected* measure of agreement, to discount the observed proportion of agreement by the expected level of agreement, given the observed marginal distributions of the rater's responses under the assumption that the raters reports are *statistically independent*". **Shoukri [16, 2004, §4.4, p.62]**: "Each of the k subjects are rated by n $>$ 2 raters *independently into one of c mutually exclusive and exhaustive* nominal categories".

(Yes; Yes) — **Fleiss et al [12, 2007, p.599]**: "Suppose that each of a sample of n subjects is rated *independently* by the same two raters, with the ratings being on a categorical scale consisting of k categories"; [12, 2007, pages 602-603]: "Different opinions have been stated on the need to incorporate chance-expected agreement into the assessment of interrater reliability. Kappa is a measure of agreement with desirable properties: *the correction for chance is incorporated into this measure of agreement*";

(No; No)= **("separately"; statistically independent)** — **Agresti [8, 2007, p.21]**: "Analyzing associations is at the heart of multivariate statistical analysis"; **[8, 2007, p.260]**: "...two experts *separately* classified cases..." (not independently); **[8, 2007, p.261]**: "Many categorical scales are quite subjective, and perfect agreement is rare. Agreement is distinct from association. Strong agreement requires strong association, but strong association can exist without strong agreement. For ordering variables if observer X consistently classifies subjects one level higher than observer Y, the strength of agreement is poor even though the association is strong"; **[8, 2007, p.264]**: "The most popular index is Cohen's kappa. It compares the agreement with that expected if the ratings were *independent*"; "Kappa equals "0" when the agreement merely equals that expected under *independence*, and it equals "1" when perfect agreement occurs"; **[7, 1992, p.202]**: "The most popular measure of this type, Cohen's kappa, gives a number on a scale for which "0" indicates agreement no better than would be expected if the ratings were *statistically independent*, and "1" indicates perfect agreement".

Благодарности

Автор с благодарностью помнит о вдохновляющем вкладе всех участников научного семинара по эвентологической математике ИМиФИ СФУ в обсуждение эвентологической теории общности, ассоциации и зависимости событий.

Список литературы

[1] О. Ю. Воробьев и Е. Е. Голденок. Метрические эвентологические распределения. *Вестник КрасГУ, серия физико-математические науки*, 1:174–181, 2004.

[2] Е. Е. Голденок. *Моделирование структур зависимостей и взаимодействий случайных событий в вероятностных системах. Автореферат дисс. ... канд. физ.-мат. наук.* Красноярский государственный технический университет, Красноярск, 2002.

[3] О. Ю. Воробьев. *Эвентология.* Сибирский федеральный университет, Красноярск, 2007, 435с., https://www.academia.edu/179393/Vorobyev_O.Yu._Eventology._-_Krasnoyarsk_Siberian_Federal_University._-_2007._-_435p.

[4] О. Ю. Воробьев. Меры рассеяния м.с. по пространству элементарных исходов. *Труды XV Международ. ЭМ конференции по эвентологической математике и смежным вопросам*, Красноярск: СФУ (под ред. Олега Воробьева):54–63, 2011.

[5] О. Ю. Воробьев. Сет-дистограммные измерители разобщенности событий. *Труды XIII ФАМЭМС конференции по финансово-актуарной математике и эвентологии многомерной статистики*, Красноярск: СФУ (под ред. Олега Воробьева):95–111, 2014.

[6] A.O. Adejumo, C. Heumann, and H. Toutenburg. A review of agreement measure as a subset of association measure between raters. http://nbn-resolving.de/urn:nbn:de:bvb:19-epub-1755-8, (385):43 pages, Sonderforschungsbereich 386 der Ludwig-Maximilians-Universitat Munchen, 2004.

[7] A. Agresti. Modelling patterns of agreement and disagreement. *Statistical Methods in Medical Research*, 1:201–208, 1992.

[8] A. Agresti. *An introduction to Categorical Data Analysis.* Wiley & Sons, New Jersey, second edition edition, 2007.

[9] J. Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, XX(1):37–46, 1960.

[10] E. E. Goldenok. *Modeling dependencies and interactions structures of random events in probabilistic systems. PhD Synopsis.* Krasnoyarsk State Technical Univertsity, Krasnoyarsk, 2002.

[11] K.L. Gwet. *Handbook of inter-rater reliability. Fourth Edition. (The Definitive Guide to Measuring the Extent of Agreement Among Raters).* Advanced Analytics, LLC, Gaithersburg, MD, 2014.

[12] Fleiss J.L., Levin B., and Paik M.C. *Statistical Methods for Rates and Proportions. Third Edition.* Wiley, New York, 2007.

[13] Uebersax J.S. Kappa coefficients: A critical appraisal. *http://john-uebersax.com/stat/kappa.htm*, 2009.

[14] Uebersax J.S. The myth of chance-corrected agreement. *http://john-uebersax.com/stat/kappa2.htm*, 2010.

[15] K. Pearson. On the theory of contingency and its relation to association and normal correlation. *Draper's Company Research Memories, Biometric Series*, I, 1904.

[16] M.M. Shoukri. *Measures of interobserver agreement.* Chapman and Hall, New York, 2003.

[17] O. Yu. Vorobyev. *Eventology.* Siberian Federal University, Krasnoyarsk, Russia, 2007, 435p., https://www.academia.edu/179393/Vorobyev_O.Yu._Eventology._-_Krasnoyarsk_Siberian_Federal_University._-_2007._-_435p.

[18] 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), http://fam.conf.sfu-kras.ru/adds/XV-em2011-14January2012.pdf.

[19] O. Yu. Vorobyev. Set-*d*istogram measurers of disagreement of events. In. *Proc. of the XIII Intern. FAMEMS Conf. on Financial and Actuarial Mathematics and Eventology of Multivariate Statistics*, Krasnoyarsk, SFU (Oleg Vorobyev ed.):10p., 2014.

[20] O. Yu. Vorobyev and E. E. Goldenok. Metric eventological distributions. *Vestnik of the Krasnoyarsk State University. Mathematics & Physics Series*, 1:174–181, 2004 (in Russian, abstract in English).

Сет-*д*истограммные измерители разобщенности событий

Олег Юрьевич Воробьев
Сибирский федеральный университет
Институт математики и фундаментальной информатики
Красноярск
oleg.yu.vorobyev@gmail.com

Аннотация. *Рассматривается расширение понятия сет-расстояния N-множества событий, характеризующего его разобщенность, до $(N+1)$-семейства сет-расстояний различных порядков (сет-**д**истограммы множества событий), которое характеризует распределение разобщенности событий относительно мощности подмножеств данного множества событий и служит удобным инструментом детального измерения разобщенности множества событий. Приводятся примеры сет-**д**истограмм для нескольких типов вероятностных распределений множеств событий с разными структурами общности и разобщенности.*

Ключевые слова. *Вероятность, многомерный статистический анализ, эвентология, множество событий, общность, разобщенность, сет-расстояние, сет-**д**истограмма, теоретико-множественная операция, чердак, подвал.*

Какая ночь! Как воздух чист,
Как серебристый дремлет лист,
Как тень черна прибрежных ив,
Как безмятежно спит залив,
Как не вздохнет нигде волна,
Как тишиною грудь полна!
Полночный свет, ты тот же день:
Белей лишь блеск, чернее тень,
Лишь тоньше запах сочных трав,
Лишь ум светлей, мирнее нрав,
Да вместо страсти хочет грудь
Вот этим воздухом вздохнуть.

Афанасий Фет, 1857?

1 Введение

Время от времени вы ловите себя на том, что не можете найти слов, пребывая в полном онемении от стечения обстоятельств. Вы не в силах выразить происходящее, погружая ваш разум в мучительные поиски терминов и определений. Не отчаивайтесь, с вами всё в порядке. Такое происходит с каждым в тот момент, когда окружающая действительность беременна новым термином, когда нового термина требует стечение обстоятельств, в которые его/ее угораздило попасть. К примеру, ведь существовали *случай, риск, шансы* до существования небезызвестных кавалера де Мере, Паскаля, Ферма и Гюйгенса, "пожертвовавших" свои имена на алтарь столь благородного дела — избавления от необходимости повторять весь этот букет понятий, заменив их на краткое "теория вероятностей".

Рисунок 1: Стеклянный цикл "Объединения & Пересечения". — Валентин Толедо, Санта Моника, Калифорния, 2010.

Так и сегодня. На наших глазах зарождается новый термин, главной задачей которого станет обозначение того нового метода исчисления *событий, обстоятельств, явлений, феноменов*, которым мы ежедневно пользуемся, принимая решения. То есть, если люди все чаще и чаще станут сталкиваться со всем тем, что сегодня происходит там, где действительно исчисляют события ради принятия решений, они сразу поймут, что они вторглись во владения "эвентологии". Возможно, сей термин будет означать некую новую науку *исчисления событий*, дополняющую науку *исчисления вероятностей* — теорию вероятностей.

2 Теоретико-множественные операции над множеством событий и сопутствующие вероятности

2.1 Произвольная теоретико-множественная операция, чердак и подвал множеств событий (м.с.) и их вероятности

Пусть $(\Omega, \mathcal{A}, \mathbf{P})$ — *всеобщее вероятностное пространство* [3, 7], а $\mathfrak{X} \subseteq \mathcal{A}$ — конечное множество событий (м.с.), состоящее из $N = |\mathfrak{X}|$ событий.

© 2014 О.Ю.Воробьев

Олег Воробьев (ред.), Труды XIII ФАМЭМС'2014, Красноярск: СФУ

• *Объединение* (также *сумма* или *соединение*) *м.с.* $\mathfrak{X}$ в теории вероятностей — теоретико-множественная операция над $\mathfrak{X}$, результатом которой служит событие, составленное из элементарных исходов, при которых наступает хотя бы одно событие из данного м.с. В террасных обозначениях *объединение м.с.* $\mathfrak{X}$ имеет вид

$$\bigcup_{x\in\mathfrak{X}} x = \mathsf{Ter}_{\mathfrak{X}/\!/\mathfrak{X}} \tag{2.1.1}$$

и называется $\mathfrak{X}$-*террасным событием 4-го рода* [2], его вероятность, т.е. *вероятность объединения событий*, имеет специальное обозначение:

$$u_{\mathfrak{X}} = u_{\mathfrak{X}/\!/\mathfrak{X}} = \mathbf{P}(\mathsf{Ter}_{\mathfrak{X}/\!/\mathfrak{X}}). \tag{2.1.2}$$

• *Пересечение* (также *произведение*) *м.с.* $\mathfrak{X}$ в теории вероятностей — теоретико-множественная операция над $\mathfrak{X}$, результатом которой служит событие, состоящее из тех и только тех элементарных исходов, при которых наступают все события из данного м.с. В террасных обозначениях *пересечение м.с.* $\mathfrak{X}$ имеет вид

$$\bigcap_{x\in\mathfrak{X}} x = \mathsf{ter}_{\mathfrak{X}/\!/\mathfrak{X}}. \tag{2.1.3}$$

и называется $\mathfrak{X}$-*террасным событием 2-го рода* [2], его вероятность, т.е. *вероятность пересечения событий*, имеет специальное обозначение:

$$p_{\mathfrak{X}} = p_{\mathfrak{X}/\!/\mathfrak{X}} = \mathbf{P}(\mathsf{ter}_{\mathfrak{X}/\!/\mathfrak{X}}). \tag{2.1.4}$$

• n-*слой*[1] *м.с.* $\mathfrak{X}$ в теории вероятностей — теоретико-множественная операция над $\mathfrak{X}$, результатом которой служит событие, состоящее из тех и только тех элементарных исходов, при которых наступает ровно n событий из данного м.с.; в террасных обозначениях определяется формулой

$$\left(C^n_{\mathfrak{X}}\right)_{x\in\mathfrak{X}} x = \bigcup_{\substack{|X|=n\\ X\subseteq\mathfrak{X}}} \mathsf{ter}(X/\!/\mathfrak{X}), \tag{2.1.5}$$

где для $X\subseteq\mathfrak{X}$

$$\mathsf{ter}(X/\!/\mathfrak{X}) = \bigcap_{x\in X}\ \bigcap_{x\in\mathfrak{X}-X} x^c \tag{2.1.6}$$

— X-*террасные события 1-го рода*. Вероятности n-слоев имеют специальные обозначения:

$$\pi_n = \pi_n(\mathfrak{X}) = \mathbf{P}\left(\left(C^n_{\mathfrak{X}}\right)_{x\in\mathfrak{X}} x\right). \tag{2.1.7}$$

• n-*чердак*[2] *м.с.* $\mathfrak{X}$ в теории вероятностей — теоретико-множественная операция над $\mathfrak{X}$, которая определяется как *объединение всех* n-*пересечений* событий из $\mathfrak{X}$ для $n = 0, 1, \ldots, N$. В террасных обозначениях n-*чердак м.с.* $\mathfrak{X}$ определяется формулой

$$\overset{n}{\bigodot}_{x\in\mathfrak{X}} x = \bigcup_{\substack{|X|=n\\ X\subseteq\mathfrak{X}}} \mathsf{ter}_{X/\!/\mathfrak{X}}, \tag{2.1.8}$$

где для $X\subseteq\mathfrak{X}$

$$\mathsf{ter}_{X/\!/\mathfrak{X}} = \bigcap_{x\in X} x \tag{2.1.9}$$

— X-*террасные события 2-го рода*. Вероятности n-чердаков, т.е. *вероятности объединений* n-*пересечений событий*, имеют специальные обозначения:

$$\overset{n}{u}_{\mathfrak{X}} = \mathbf{P}\left(\overset{n}{\bigodot}_{x\in\mathfrak{X}} x\right). \tag{2.1.10}$$

Средняя вероятность n-*пересечений событий* из $\mathfrak{X}$ имеет обозначение

$$\overset{n}{\langle p\rangle}_{\mathfrak{X}} = \frac{1}{C^n_N} \sum_{\substack{|X|=n\\ X\subseteq\mathfrak{X}}} p_{X/\!/\mathfrak{X}}, \tag{2.1.11}$$

которое не следует путать с обозначением *вероятности пересечений* n-*объединений* (2.1.14).

• n-*подвал*[3] *м.с.* $\mathfrak{X}$ в теории вероятностей — теоретико-множественная операция над $\mathfrak{X}$, которая определяется как *пересечение всех* n-*объединений* событий из $\mathfrak{X}$ для $n = 0, 1, \ldots, N$. В террасных обозначениях n-*подвал м.с.* $\mathfrak{X}$ определяется формулой

$$\overset{n}{\boxdot}_{x\in\mathfrak{X}} x = \bigcap_{\substack{|X|=n\\ X\subseteq\mathfrak{X}}} \mathsf{Ter}_{X/\!/\mathfrak{X}}. \tag{2.1.12}$$

где для $X\subseteq\mathfrak{X}$

$$\mathsf{Ter}_{X/\!/\mathfrak{X}} = \bigcup_{x\in X} x \tag{2.1.13}$$

— X-*террасные события 4-го рода*. Вероятности n-подвалов, т.е. *вероятности пересечений* n-*объединений событий*, имеют специальные обозначения:

$$\overset{n}{p}_{\mathfrak{X}} = \mathbf{P}\left(\overset{n}{\boxdot}_{x\in\mathfrak{X}} x\right). \tag{2.1.14}$$

Средняя вероятность n-*объединений событий* из $\mathfrak{X}$ имеет обозначение

$$\overset{n}{\langle u\rangle}_{\mathfrak{X}} = \frac{1}{C^n_N} \sum_{\substack{|X|=n\\ X\subseteq\mathfrak{X}}} u_{X/\!/\mathfrak{X}}, \tag{2.1.15}$$

которое не следует путать с обозначением *вероятности объединений* n-*пересечений* (2.1.10).

[1] англ. *n-layer*.

[2] англ. *n-attic*; то, что раньше называлось n-куполом [5] (а затем на короткие времена n-иерархией и n-голосованием).

[3] англ. *n-cellar*.

• *Произвольная теоретико-множественная* $\mathcal{O}$-*операция над* N-*м.с.* $\mathfrak{X}$ определяется [1] формулой

$$\underset{x\in\mathfrak{X}}{\mathcal{O}} x = \sum_{X\in\mathcal{O}} \mathsf{ter}(X/\!\!/\mathfrak{X}), \tag{2.1.16}$$

где $\mathcal{O} \subseteq 2^{\mathfrak{X}}$ — произвольная совокупность подмножеств $\mathfrak{X}$, характеризующая данную теоретико-множественную операцию. Иногда используется более подробное обозначение [5] для произвольной теоретико-множественной $\mathcal{O}$-операции (2.1.16):

$$\mathsf{terop}\left(\mathcal{O}/\!\!/2^{\mathfrak{X}}\right) = \underset{x\in\mathfrak{X}}{\mathcal{O}} x. \tag{2.1.17}$$

В частности, произвольная $\mathcal{O}$-операция становится *объединением* при

$$\mathcal{O} = \{X \subseteq \mathfrak{X} : |X| \geqslant 1\}, \tag{2.1.18}$$

пересечением при

$$\mathcal{O} = \{X \subseteq \mathfrak{X} : |X| \geqslant |\mathfrak{X}|\}, \tag{2.1.19}$$

n-*слоем* при

$$\mathcal{O} = \{X \subseteq \mathfrak{X} : |X| = n\}, \tag{2.1.20}$$

n-*чердаком* при

$$\mathcal{O} = \{X \subseteq \mathfrak{X} : |X| \geqslant n\}, \tag{2.1.21}$$

n-*подвалом* при[4]

$$\mathcal{O} = \{X \subseteq \mathfrak{X} : |X| \geqslant N + 1 - n\}. \tag{2.1.22}$$

2.2 Свойства чердаков множества событий

Лемма 1 ***(характеризации*** n***-чердака м.с. мощностными индикаторами).*** Характеризация n-*чердака м.с.* $\mathfrak{X}$ через его *мощностной индикатор*[5] и мощностной индикатор его М-дополнения $\mathfrak{X}^{(c)} = \{x^c : x \in \mathfrak{X}\}$ имеют вид:

$$\overset{n}{\underset{x\in\mathfrak{X}}{\odot}} x = \begin{cases} \{\omega : \Upsilon_{\mathfrak{X}}(\omega) \geqslant n\}, \\ \\ \{\omega : \Upsilon_{\mathfrak{X}^{(c)}}(\omega) < N - n + 1\}. \end{cases} \tag{2.2.1}$$

Доказательство. Из определения (2.1.5) и формул обращения между террасными событиями 1-го и 2-го рода [2]

$$\mathsf{ter}_{X/\!\!/\mathfrak{X}} = \sum_{X\subseteq Y} \mathsf{ter}(X/\!\!/\mathfrak{X}),$$

где

$$\mathsf{ter}(X/\!\!/\mathfrak{X}) = \bigcap_{x\in X} x \bigcap_{x\in\mathfrak{X}-X} x^c$$

— X-террасные события 1-го рода, $X \subseteq \mathfrak{X}$, порожденные м.с. $\mathfrak{X}$, получаем то, что требуется:

$$\overset{n}{\underset{x\in\mathfrak{X}}{\odot}} x = \sum_{\substack{|X|\geqslant n \\ X\subseteq\mathfrak{X}}} \mathsf{ter}(X/\!\!/\mathfrak{X}) = \{\omega : \Upsilon_{\mathfrak{X}}(\omega) \geqslant n\}.$$

Лемма доказана.

Следствие 1. В более подробных обозначениях [5]

$$\overset{n}{\underset{x\in\mathfrak{X}}{\odot}} x = \mathsf{terop}\left(C^{n+}_{\mathfrak{X}}/\!\!/2^{\mathfrak{X}}\right), \tag{2.2.2}$$

где

$$C^{n+}_{\mathfrak{X}} = \{X \subseteq \mathfrak{X} : |X| \geqslant n\} \subseteq 2^{\mathfrak{X}}$$

— совокупность подмножеств м.с. $\mathfrak{X}$, содержащих не менее n событий, характеризующая операцию n-чердака $\mathfrak{X}$.

Замечание 1. События $\overset{n}{\underset{x\in\mathfrak{X}}{\odot}} x \in \mathcal{A}$, определяемые для каждого $n = 0, 1, \ldots, N$ как результат теоретико-множественной операции n-*чердака м.с.* $\mathfrak{X}$, т.е. — формулами (2.1.5), (2.2.1) и (2.2.2), также называются n-*чердаками* $\mathfrak{X}$.

Замечание 2 ***(монотонное убывание чердаков по включению).*** Из характеризации (2.2.1) ясно, что

$$\overset{0}{\underset{x\in\mathfrak{X}}{\odot}} x \supseteq \overset{1}{\underset{x\in\mathfrak{X}}{\odot}} x \supseteq \ldots \supseteq \overset{N}{\underset{x\in\mathfrak{X}}{\odot}} x \tag{2.2.2}$$

— n-чердаки м.с. $\mathfrak{X}$ убывают по включению при росте $n = 0, 1, \ldots, N$.

Замечание 3 ***(частные чердаки).***

$$\overset{n}{\underset{x\in\mathfrak{X}}{\odot}} x = \begin{cases} \Omega, & n = 0, \\ \\ \bigcup\limits_{x\in\mathfrak{X}} x, & n = 1, \\ \ldots, & \ldots, \\ \\ \bigcap\limits_{x\in\mathfrak{X}} x, & n = |\mathfrak{X}|, \\ \\ \emptyset, & n > |\mathfrak{X}|. \end{cases}$$

Лемма 2 ***(расширенное правило де Моргана для чердаков м.с.).*** *При любом конечном* N*-м.с.* $\mathfrak{X}$ *и любом* $n = 0, 1, \ldots, N$ *для* n*-чердаков* $\mathfrak{X}$ *справедливы следующие расширенные правила де Моргана:*

$$\overset{n}{\underset{x\in\mathfrak{X}}{\odot}} x + \overset{N+1-n}{\underset{x\in\mathfrak{X}}{\odot}} x^c = \Omega, \tag{2.2.3}$$

[4] Доказательство формул (2.1.21) и (2.1.22) см. ниже: (2.2.1) и (2.3.4).

[5] *Мощностной индикатор конечного м.с.* $\mathfrak{X}$ определяется как с.в., равная сумме индикаторов всех событий из $\mathfrak{X}$:

$$\Upsilon_{\mathfrak{X}}(\omega) = \sum_{x\in\mathfrak{X}} \mathbf{1}_x(\omega).$$

при $n=1$:

$$\bigodot_{x\in\mathfrak{X}}^{1} x + \bigodot_{x\in\mathfrak{X}}^{N} x^c = \Omega$$

эквивалентные привычному правилу:

$$\bigcup_{x\in\mathfrak{X}} x + \bigcap_{x\in\mathfrak{X}} x^c = \Omega.$$

Доказательство следует из двух соотношений для террасных событий [2]:

$$\sum_{X\subseteq\mathfrak{X}} \mathsf{ter}(X/\!\!/\mathfrak{X}) = \Omega, \tag{2.2.4}$$

$$\mathsf{ter}(X/\!\!/\mathfrak{X}) = \mathsf{ter}\left((X^c)^{(c)}/\!\!/\mathfrak{X}^{(c)}\right), \tag{2.2.5}$$

где $X\subseteq\mathfrak{X}$, $\mathfrak{X}^{(c)}=\{x^c : x\in\mathfrak{X}\}$ — М-дополнение $\mathfrak{X}$, а $(X^c)^{(c)}=\{x^c : x\in X^c\}\subseteq\mathfrak{X}^{(c)}$ — М-дополнение м.с. $X^c=\mathfrak{X}-X$.

Действительно, из определения операции n-чердака $\mathfrak{X}$ и того, что $|X|+|X^c|=N$ для $X\subseteq\mathfrak{X}$, следует:

$$\bigodot_{x\in\mathfrak{X}}^{n} x = \sum_{\substack{|X|\geqslant n\\ X\subseteq\mathfrak{X}}} \mathsf{ter}(X/\!\!/\mathfrak{X}) = \sum_{\substack{|X^c|\leqslant N-n\\ X\subseteq\mathfrak{X}}} \mathsf{ter}(X/\!\!/\mathfrak{X}) = {}^{6}$$

$$= \sum_{\substack{|X^c|\leqslant N-n\\ X\subseteq\mathfrak{X}}} \mathsf{ter}((X^c)^{(c)}/\!\!/\mathfrak{X}^{(c)}) = {}^{7}$$

$$= \Omega - \sum_{\substack{|X^c|> N-n\\ X\subseteq\mathfrak{X}}} \mathsf{ter}((X^c)^{(c)}/\!\!/\mathfrak{X}^{(c)}) = {}^{8}$$

$$= \Omega - \sum_{\substack{|X^c|\geqslant N-n+1\\ (X^c)^{(c)}\subseteq\mathfrak{X}^{(c)}}} \mathsf{ter}((X^c)^{(c)}/\!\!/\mathfrak{X}^{(c)}) = {}^{9}$$

$$= \Omega - \bigodot_{x^c\in\mathfrak{X}^{(c)}}^{N-n+1} x^c = \Omega - \bigodot_{x\in\mathfrak{X}}^{N-n+1} x^c.$$

Лемма доказана.

2.3 Свойства подвалов множества событий

Лемма 3 *(связь между подвалами и чердаками м.с.). n-подвал м.с. $\mathfrak{X}$ совпадает с его m-чердаком*:

$$\boxdot_{x\in\mathfrak{X}}^{n} x = \bigodot_{x\in\mathfrak{X}}^{m} x, \tag{2.3.1}$$

если $m+n=N+1,\ \ m\geqslant 0, n\geqslant 0$.

Доказательство. В силу определений подвала (2.1.9) и чердака (2.1.5) утверждение леммы (2.3.1) означает, что

$$\bigcap_{\substack{|X|=n\\ X\subseteq\mathfrak{X}}} \mathsf{Ter}_{X/\!\!/\mathfrak{X}} = \bigcup_{\substack{|X|=N+1-n\\ X\subseteq\mathfrak{X}}} \mathsf{ter}_{X/\!\!/\mathfrak{X}} \tag{2.3.2}$$

для $n=0,1,\ldots,N$. Дополнение события из левой части (2.3.2) по правилу де Моргана равно

$$\left(\bigcap_{\substack{|X|=n\\ X\subseteq\mathfrak{X}}} \mathsf{Ter}_{X/\!\!/\mathfrak{X}}\right)^c = \bigcup_{\substack{|X|=n\\ X\subseteq\mathfrak{X}}} \left(\mathsf{Ter}_{X/\!\!/\mathfrak{X}}\right)^c. \tag{2.3.3}$$

Дополнения террасных событий 4-го рода (2.1.10) из правой части (2.3.3) по тем же причинам равно

$$\left(\mathsf{Ter}_{X/\!\!/\mathfrak{X}}\right)^c = \bigcap_{x\in X} x^c = \bigcap_{x^c\in X^{(c)}} x^c = \mathsf{ter}_{X^{(c)}/\!\!/\mathfrak{X}^{(c)}}$$

— террасному событию 2-го рода, порожденному $\mathfrak{X}^{(c)}$, М-дополнением м.с. $\mathfrak{X}$. Отсюда правая часть (2.3.3) равна

$$\bigcup_{\substack{|X|=n\\ X\subseteq\mathfrak{X}}} \mathsf{ter}_{X^{(c)}/\!\!/\mathfrak{X}^{(c)}} = \bigcup_{\substack{|X^{(c)}|=n\\ X^{(c)}\subseteq\mathfrak{X}^{(c)}}} \mathsf{ter}_{X^{(c)}/\!\!/\mathfrak{X}^{(c)}} = {}^{10}$$

$$= \bigodot_{x^c\in\mathfrak{X}^{(c)}}^{n} x^c = \bigodot_{x\in\mathfrak{X}}^{n} x^c = {}^{11} = \Omega - \bigodot_{x\in\mathfrak{X}}^{N+1-n} x = {}^{12}$$

$$= \Omega - \bigcup_{\substack{|X|=N+1-n\\ X\subseteq\mathfrak{X}}} \mathsf{ter}_{X/\!\!/\mathfrak{X}}$$

— дополнению события из правой части (2.3.2), что доказывает лемму.

Лемма 4 *(характеризация n-подвала м.с. мощностными индикаторами).* Характеризации *n-подвала* м.с. $\mathfrak{X}$ через его *мощностной индикатор*[13] и мощностной индикатор его М-дополнения $\mathfrak{X}^{(c)}=\{x^c : x\in\mathfrak{X}\}$ имеют вид:

$$\boxdot_{x\in\mathfrak{X}}^{n} x = \begin{cases} \{\omega : \Upsilon_{\mathfrak{X}}(\omega)\geqslant N+1-n\}, \\ \\ \{\omega : \Upsilon_{\mathfrak{X}^{(c)}}(\omega) < n\}. \end{cases} \tag{2.3.4}$$

Доказательство следует из леммы 3 и того, что $\Upsilon_{\mathfrak{X}}(\omega)+\Upsilon_{\mathfrak{X}^{(c)}}(\omega)=N$ для каждого элементарного исхода $\omega\in\Omega$.

Следствие 2. В более подробных обозначениях [5]

$$\boxdot_{x\in\mathfrak{X}}^{n} x = \mathsf{terop}\left(C_{\mathfrak{X}}^{(N+1-n)+}/\!\!/2^{\mathfrak{X}}\right), \tag{2.3.5}$$

где

$$C_{\mathfrak{X}}^{(N+1-n)+} = \{X\subseteq\mathfrak{X} : |X|\geqslant N+1-n\}\subseteq 2^{\mathfrak{X}}$$

— совокупность подмножеств м.с. $\mathfrak{X}$, содержащих не менее $N+1-n$ событий, характеризующая операцию n-подвала $\mathfrak{X}$.

[6] в силу (2.3.5).
[7] в силу (2.3.4).
[8] из-за перехода к нестрогому неравенству.
[9] по определению чердака м.с. (2.2.1)
[10] по определению n-чердака (2.1.3) м.с. $\mathfrak{X}^{(c)}$.
[11] по расширенному правилу де Моргана для чердаков (2.2.3).
[12] по определению $(N+1-n)$-чердака (2.1.3) м.с. $\mathfrak{X}$.
[13] см. сноску 5.

Следствие 3 ***(объединение n-пересечений совпадает с пересечением $N+1-n$ объединений).*** Из лемм 1 и 4 получаем для $n = 0, 1, \ldots, N$ *тождество дополнительности для чердаков и подвалов*:

$$\overset{n}{\underset{x\in\mathfrak{X}}{\odot}} x = \overset{N+1-n}{\underset{x\in\mathfrak{X}}{\boxdot}} x, \tag{2.3.6}$$

парафраз которого на языке операций объединения и пересечения имеет вид: "*объединение n-пересечений совпадает с пересечением $(N+1-n)$-объединений*"[14].

Замечание 4. События $\overset{n}{\underset{x\in\mathfrak{X}}{\boxdot}} x \in \mathcal{A}$, определяемые для каждого $n = 0, 1, \ldots, N$ как результат теоретико-множественной операции *n-подвала м.с.* $\mathfrak{X}$, т.е. — формулами (2.1.9), (2.3.1) и (2.3.2), также называются *n-подвалами* $\mathfrak{X}$.

Замечание 5 ***(монотонное возрастание подвалов по включению).*** Из характеризации (2.3.4) ясно, что

$$\overset{0}{\underset{x\in\mathfrak{X}}{\boxdot}} x \subseteq \overset{1}{\underset{x\in\mathfrak{X}}{\boxdot}} x \subseteq \ldots \subseteq \overset{N}{\underset{x\in\mathfrak{X}}{\boxdot}} x \tag{2.3.7}$$

— n-подвалы м.с. $\mathfrak{X}$ возрастают по включению при росте $n = 0, 1, \ldots, N$.

Замечание 6 ***(частные подвалы).***

$$\overset{n}{\underset{x\in\mathfrak{X}}{\boxdot}} x = \begin{cases} \emptyset, & n = 0, \\ \bigcap\limits_{x\in\mathfrak{X}} x, & n = 1, \\ \ldots, & \ldots, \\ \bigcup\limits_{x\in\mathfrak{X}} x, & n = N, \\ \Omega, & n > N. \end{cases}$$

Лемма 5 ***(расширенное правило де Моргана для подвалов м.с.).*** *При любом конечном N-м.с. $\mathfrak{X}$ и любом $n = 0, 1, \ldots, N$ для n-подвалов $\mathfrak{X}$ справедливы следующие расширенные правила де Моргана:*

$$\overset{n}{\underset{x\in\mathfrak{X}}{\boxdot}} x + \overset{N+1-n}{\underset{x\in\mathfrak{X}}{\boxdot}} x^c = \Omega, \tag{2.3.8}$$

при $n = 1$:

$$\overset{1}{\underset{x\in\mathfrak{X}}{\boxdot}} x + \overset{N}{\underset{x\in\mathfrak{X}}{\boxdot}} x^c = \Omega$$

эквивалентные привычному правилу:

$$\bigcap_{x\in\mathfrak{X}} x + \bigcup_{x\in\mathfrak{X}} x^c = \Omega.$$

Доказательство аналогично доказательству расширенного правила де Моргана для чердаков (лемма 2) и сразу следует из леммы 3.

[14] Например, в триплете событий $\mathfrak{X}$ "*объединение парных пересечений совпадает с пересечением парных объединений*", так как 2+2=4 ;)

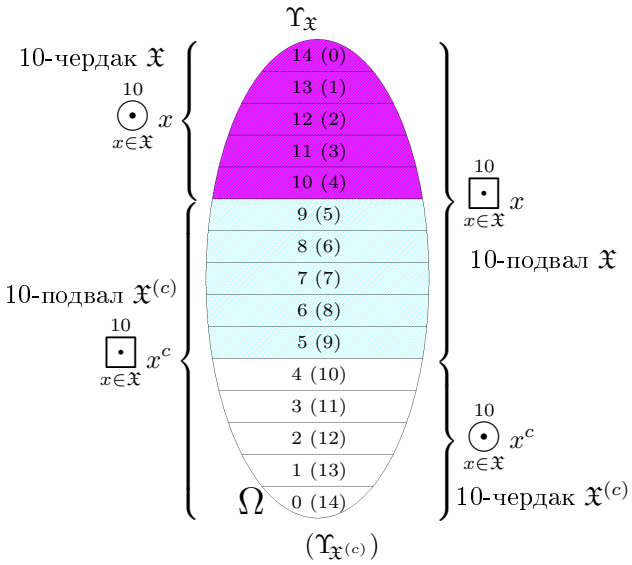


Рисунок 2: Диаграмма Венна n-чердаков и n-подвалов 14-м.с. $\mathfrak{X}$ и его М-дополнения $\mathfrak{X}^{(c)}$ ($n = 10$), на которой пространство элементарных исходов Ω изображено в виде эллипса; 10-чердак $\mathfrak{X}$ (фуксия сплошная), 10-подвал $\mathfrak{X}$ (аква-косая). Показаны номера террасных слоев м.с. и его М-дополнения (в скобках).

3 Набор сет-расстояний м.с.

3.1 Расширение понятия сет-расстояния м.с.

Сет-расстояние м.с. $\mathfrak{X}$ определяется [6] как функционал от двух отклонений точки отсчета общности (которой служит средняя вероятность событий) от вероятности пересечения и от вероятности объединения событий из этого м.с.

$$\Delta_{\mathfrak{X}} = \min\left\{\Delta^{\cap}_{\mathfrak{X}}, \Delta^{\cup}_{\mathfrak{X}}\right\}, \tag{3.1}$$

n-чердак $\mathfrak{X}$
$\overset{n}{\underset{x\in\mathfrak{X}}{\odot}} x = \begin{cases} \{\omega : \Upsilon_{\mathfrak{X}}(\omega) \geqslant n\}, \\ \{\omega : \Upsilon_{\mathfrak{X}^{(c)}}(\omega) < N - n + 1\}. \end{cases}$ (2.2.1)
n-подвал $\mathfrak{X}$
$\overset{n}{\underset{x\in\mathfrak{X}}{\boxdot}} x = \begin{cases} \{\omega : \Upsilon_{\mathfrak{X}}(\omega) \geqslant N + 1 - n\}, \\ \{\omega : \Upsilon_{\mathfrak{X}^{(c)}}(\omega) < n\}. \end{cases}$ (2.3.4)
n-чердак $\mathfrak{X}^{(c)}$
$\overset{n}{\underset{x\in\mathfrak{X}}{\odot}} x^c = \begin{cases} \{\omega : \Upsilon_{\mathfrak{X}^{(c)}}(\omega) \geqslant n\}, \\ \{\omega : \Upsilon_{\mathfrak{X}}(\omega) < N - n + 1\}. \end{cases}$ (2.2.1′)
n-подвал $\mathfrak{X}^{(c)}$
$\overset{n}{\underset{x\in\mathfrak{X}}{\boxdot}} x^c = \begin{cases} \{\omega : \Upsilon_{\mathfrak{X}^{(c)}}(\omega) \geqslant N + 1 - n\}, \\ \{\omega : \Upsilon_{\mathfrak{X}}(\omega) < n\}. \end{cases}$ (2.3.4′)

Таблица 1: Характеризации n-чердаков и n-подвалов м.с. $\mathfrak{X}$ и его М-дополнения $\mathfrak{X}^{(c)}$ их мощностными индикаторами (леммы 1 и 4), $n = 0, 1, \ldots, N$.

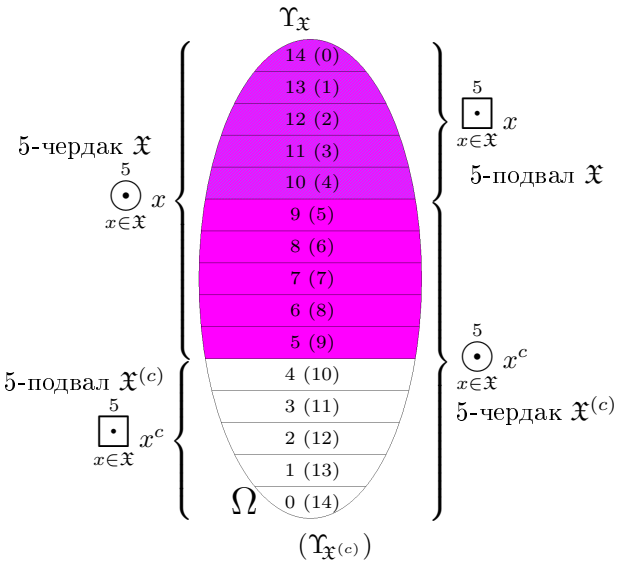


Рисунок 3: Диаграмма Венна n-чердаков и n-подвалов 14-м.с. $\mathfrak{X}$ и его М-дополнения $\mathfrak{X}^{(c)}$ ($n=5$), на которой пространство элементарных исходов Ω изображено в виде эллипса; 5-чердак $\mathfrak{X}$ (фуксия сплошная), 5-подвал $\mathfrak{X}$ (аква-косая). Показаны номера террасных слоев м.с. и его М-дополнения (в скобках).

где

$$\Delta^{\cap}_{\mathfrak{X}} = \langle p\rangle_{\mathfrak{X}} - p_{\mathfrak{X}}$$

— сет-расстояние м.с. $\mathfrak{X}$ относительно пересечения [4],

$$\Delta^{\cup}_{\mathfrak{X}} = u_{\mathfrak{X}} - \langle p\rangle_{\mathfrak{X}}$$

— сет-расстояние м.с. $\mathfrak{X}$ относительно объединения [4], а

$$\langle p\rangle_{\mathfrak{X}} = \frac{1}{|\mathfrak{X}|}\sum_{x\in\mathfrak{X}} p_x \tag{3.2}$$

— средняя вероятность событий из $\mathfrak{X}$.

Это определение можно следующим образом расширить, чтобы получить N-набор сет-расстояний. Объединение всех событий из данного N-м.с. происходит тогда, когда происходит *хотя бы одно событие* из этого N-м.с. Если объединение событий в этом определении заменить событием, которое происходит тогда, когда происходят *хотя бы n событий* из данного N-м.с., т.е. — заменить объединением n-пересечений событий из $\mathfrak{X}$, или n-чердаком $\mathfrak{X}$:

$$\overset{n}{\odot}_{x\in\mathfrak{X}} x = \bigcup_{\substack{|X|=n\\ X\subseteq\mathfrak{X}}} \mathsf{ter}_{X/\!/\mathfrak{X}},$$

а новую точку n-отсчета определить как среднюю вероятность n-пересечений:

$$\overset{n}{\langle p\rangle}_{\mathfrak{X}} = \frac{1}{C^n_N}\sum_{|X|=n} p_{X/\!/\mathfrak{X}}, \tag{3.3}$$

то прежний способ приводит к расширенному определению n-сет-расстояния м.с. для $n=0,1,\ldots,N$:

$$\overset{n}{\Delta}_{\mathfrak{X}} = \min\left\{\overset{n}{\Delta}^{\cap}_{\mathfrak{X}}, \overset{n}{\Delta}^{\cup}_{\mathfrak{X}}\right\}, \tag{3.4}$$

где

$$\overset{n}{\Delta}^{\cap}_{\mathfrak{X}} = \overset{n}{\langle p\rangle}_{\mathfrak{X}} - p_{\mathfrak{X}}$$

— n-сет-расстояние м.с. $\mathfrak{X}$ относительно пересечения,

$$\overset{n}{\Delta}^{\cup}_{\mathfrak{X}} = \overset{n}{u}_{\mathfrak{X}} - \overset{n}{\langle p\rangle}_{\mathfrak{X}}$$

— n-сет-расстояние м.с. $\mathfrak{X}$ относительно объединения n пересечений (n-чердака), а

$$\overset{n}{u}_{\mathfrak{X}} = \mathbf{P}\left(\overset{n}{\odot}_{x\in\mathfrak{X}} x\right) \tag{3.5}$$

— вероятность n-чердака[15] $\mathfrak{X}$.

Замечание 7 *(0-сет-расстояние м.с.).* Отметим, что при $n=0$ 0-сет-расстояние $\mathfrak{X}$ равно

$$\overset{0}{\Delta}_{\mathfrak{X}} = \min\left\{\overset{0}{\Delta}^{\cap}_{\mathfrak{X}}, \overset{0}{\Delta}^{\cup}_{\mathfrak{X}}\right\} = 0, \tag{3.6}$$

так как

$$\overset{0}{\Delta}^{\cap}_{\mathfrak{X}} = \overset{0}{\langle p\rangle}_{\mathfrak{X}} - p_{\mathfrak{X}} = 1 - p_{\mathfrak{X}}$$

— 0-сет-расстояние м.с. $\mathfrak{X}$ относительно пересечения;

$$\overset{0}{\Delta}^{\cup}_{\mathfrak{X}} = \overset{0}{u}_{\mathfrak{X}} - \overset{0}{\langle p\rangle}_{\mathfrak{X}} = 0$$

— 0-сет-расстояние м.с. $\mathfrak{X}$ относительно объединения;

$$\overset{0}{u}_{\mathfrak{X}} = \mathbf{P}\left(\overset{0}{\odot}_{x\in\mathfrak{X}} x\right) = \mathbf{P}(\Omega) = 1 \tag{3.7}$$

— вероятность 0-чердака $\mathfrak{X}$, а

$$\overset{0}{\langle p\rangle}_{\mathfrak{X}} = \frac{1}{C^0_N}\sum_{|X|=0} p_{X/\!/\mathfrak{X}} = p_{\emptyset} = 1 \tag{3.8}$$

— средняя вероятность 0-пересечений событий из $\mathfrak{X}$.

Замечание 8 *(вероятности чердаков м.с.).*

$$\overset{n}{u}_{\mathfrak{X}} = \begin{cases} 1, & n=0,\\ u_{\mathfrak{X}}, & n=1,\\ \ldots, & \ldots,\\ p_{\mathfrak{X}}, & n=|\mathfrak{X}|,\\ 0, & n>|\mathfrak{X}|. \end{cases}$$

В итоге понятие сет-расстояния $\Delta_{\mathfrak{X}}$ м.с. $\mathfrak{X}$ расширено до понятия N-набора сет-расстояний этого м.с.

$$\left\{\overset{0}{\Delta}_{\mathfrak{X}}, \overset{1}{\Delta}_{\mathfrak{X}}, \ldots, \overset{N}{\Delta}_{\mathfrak{X}}\right\}, \tag{3.9}$$

[15] Не путать с похожим обозначением

$$\overset{n}{\langle u\rangle}_{\mathfrak{X}} = \frac{1}{C^n_N}\sum_{\substack{|X|=n\\ X\subseteq\mathfrak{X}}} u_X$$

для средней вероятности n-объединений событий из $\mathfrak{X}$, см. (2.1.12) и (4.18).

в котором $\overset{0}{\Delta}_{\mathfrak{X}}=0$, а $\overset{1}{\Delta}_{\mathfrak{X}}=\Delta_{\mathfrak{X}}$ — это сет-расстояние в узком смысле.

Каждое n-сет-расстояние $\overset{n}{\Delta}_{\mathfrak{X}}$ из набора (3.9) может быть нормализовано[16]:

$$\overset{n}{\Delta}'_{\mathfrak{X}}=\frac{\overset{n}{\Delta}_{\mathfrak{X}}}{\overset{n}{\Delta}^{\cup}_{\mathfrak{X}}+\overset{n}{\Delta}^{\cap}_{\mathfrak{X}}}=\frac{\overset{n}{\Delta}_{\mathfrak{X}}}{\overset{n}{u}_{\mathfrak{X}}-\overset{n}{p}_{\mathfrak{X}}}, \tag{3.10}$$

чтобы получить в результате N-набор *относительных сет-расстояний* $\mathfrak{X}$

$$\left\{\overset{0}{\Delta}'_{\mathfrak{X}},\overset{1}{\Delta}'_{\mathfrak{X}},\ldots,\overset{N}{\Delta}'_{\mathfrak{X}}\right\}, \tag{3.9$'$}$$

N-набор сет-расстояний (3.9) в отличие от набора (3.9$'$) будет иногда называться набором *абсолютных* сет-расстояний. $\mathfrak{X}$,

Лемма 6 ***(формулы обращения Мебиуса между средними вероятностями пересечений событий из N-м.с. и вероятностями значений его мощностного индикатора).*** Пусть

$$\{\pi_0,\pi_1,\ldots,\pi_N\} \tag{3.11}$$

— вероятностное распределение мощностного индикатора $\Upsilon_{\mathfrak{X}}$ N-м.с. $\mathfrak{X}$, а

$$\left\{\overset{0}{\langle p\rangle}_{\mathfrak{X}},\ldots,\overset{N}{\langle p\rangle}_{\mathfrak{X}}\right\} \tag{3.12}$$

— набор средних вероятностей пересечений событий из $\mathfrak{X}$, где для $n=0,1,\ldots,N$

$$\pi_n=\mathbf{P}(\Upsilon_{\mathfrak{X}}=n)=\sum_{|X|=n}p(X/\!\!/\mathfrak{X}), \tag{3.13}$$

$$\overset{n}{\langle p\rangle}_{\mathfrak{X}}=\frac{1}{C^n_N}\sum_{|X|=n}p_{X/\!\!/\mathfrak{X}}. \tag{3.14}$$

Тогда между функциями целого аргумента (3.11) и (3.12) для $n=0,\ldots,N$ справедливы формулы обращения[17]

$$\overset{n}{\langle p\rangle}_{\mathfrak{X}}=\frac{1}{C^n_N}\sum_{m=n}^{N}\pi_m C^n_m, \tag{3.15}$$

$$\pi_n=C^n_N\sum_{m=n}^{N}(-1)^{m-n}C^{m-n}_{N-n}\overset{m}{\langle p\rangle}_{\mathfrak{X}}. \tag{3.16}$$

[16] При этом неопределенность $0/0$ считается нулем.

[17] При вычислении средних вероятностей n-пересечений $\overset{n}{\langle p\rangle}_{\mathfrak{X}}$ по формуле (3.15) удобно пользоваться следующей формулой:

$$\frac{C^n_m}{C^n_N}=\begin{cases}\prod\limits_{k=m+1}^{N}\left(1-\frac{n}{k}\right), & m<N;\\ 1, & m\geqslant N,\end{cases}$$

которую проще вычислять рекуррентно:

$$\frac{C^n_m}{C^n_N}=\frac{C^n_{m+1}}{C^n_N}\left(1-\frac{n}{m+1}\right)$$

для $m=n,\ldots,N-1$ в обратном порядке, начиная с $m=N-1$.

Доказательство. Во-первых, так как

$$\sum_{|X|=n}p_{X/\!\!/\mathfrak{X}}=^{18}=\sum_{|X|=n}\sum_{X\subseteq Y}p(Y/\!\!/\mathfrak{X})=^{19}=$$

$$=\sum_{|Y|=n}^{N}p(Y/\!\!/\mathfrak{X})\sum_{\substack{|X|=n\\X\subseteq Y}}1=\sum_{|Y|=n}^{N}p(Y/\!\!/\mathfrak{X})C^n_{|Y|}=$$

$$=\sum_{m=n}^{N}C^n_m\sum_{|Y|=m}p(Y/\!\!/\mathfrak{X})=^{20}=\sum_{m=n}^{N}\pi_m C^n_m,$$

то в силу (3.14) первая формула (3.15) доказана.

Во-вторых, просуммировав по всем $X\in C^n_{\mathfrak{X}}$ обе части формулы обращения Мебиуса

$$p(X/\!\!/\mathfrak{X})=\sum_{X\subseteq Y}(-1)^{|Y|-|X|}p_{Y/\!\!/\mathfrak{X}},$$

которая для $X\subseteq\mathfrak{X}$ выражает вероятностное распределение 1-го рода через вероятностное распределение 2-го рода, получим вторую формулу (3.16):

$$\pi_n=\sum_{X\in C^n_{\mathfrak{X}}}\sum_{X\subseteq Y}(-1)^{|Y|-|X|}p_{Y/\!\!/\mathfrak{X}}=^{21}$$

$$=\sum_{Y\in C^{n+}_{\mathfrak{X}}}(-1)^{|Y|-n}p_{Y/\!\!/\mathfrak{X}}\sum_{\substack{|X|=n\\X\subseteq Y}}1=^{22}$$

$$=\sum_{Y\in C^{n+}_{\mathfrak{X}}}(-1)^{|Y|-n}C^n_{|Y|}p_{Y/\!\!/\mathfrak{X}}=^{23}$$

$$=\sum_{m=n}^{N}(-1)^{m-n}C^n_m\sum_{Y\in C^m_{\mathfrak{X}}}p_{Y/\!\!/\mathfrak{X}}=^{24}$$

$$=\sum_{m=n}^{N}(-1)^{m-n}C^n_m C^m_N\overset{m}{\langle p\rangle}_{\mathfrak{X}}=^{25}$$

$$=C^n_N\sum_{m=n}^{N}(-1)^{m-n}C^{m-n}_{N-n}\overset{m}{\langle p\rangle}_{\mathfrak{X}}.$$

Лемма доказана.

Следствие 4 ***(формулы обращения Мебиуса между средними вероятностями объединений событий из N-м.с. и вероятностями значений его мощностного индикатора).*** Из леммы 6 для средних вероятностей *пересечений* событий, примененной к М-дополнению $\mathfrak{X}^{(c)}$ N-м.с. $\mathfrak{X}$, следуют аналогичные формулы обращения Мебиуса

[18] обращение Мебиуса.

[19] перемена порядка суммирования.

[20] формула (3.7).

[21] Перемена порядка суммирования.

[22] Число n-подмножеств Y равно $C^n_{|Y|}$.

[23] Замена множественного индекса суммирования целочисленным.

[24] В силу определения средних вероятностей m-пересечений.

[25] В силу тождества комбинаторных формул.

между средними вероятностями *объединений* событий из $\mathfrak{X}$ и масс-вероятностями его мощностного индикатора:

$$1-\overset{n}{\langle u\rangle}_{\mathfrak{X}}=\frac{1}{C_N^n}\sum_{m=n}^{N}\pi_{N-m}C_m^n, \tag{3.15}$$

$$\pi_n=^{26}=C_N^n\sum_{m=0}^{n}(-1)^m C_n^m\left(1-\overset{m+N-n}{\langle u\rangle}_{\mathfrak{X}}\right) \tag{3.16}$$

в силу того, что для $n=0,1,\ldots,N$

$$\pi_n=\pi^{(c)}_{N-n},\quad \overset{n}{\langle p\rangle}_{\mathfrak{X}^{(c)}}=1-\overset{n}{\langle u\rangle}_{\mathfrak{X}}\ .$$

Следствие 5 ***(средняя вероятность** 0**-пересечений событий из** N**-м.с.).*** Для $n=0$ средняя вероятность 0-пересечений[27] событий $x\in\mathfrak{X}$, т.е. "средняя" вероятность Ω, имеет смысл среднего

$$\overset{0}{\langle p\rangle}_{\mathfrak{X}}=\mathbf{E}\,(1)=1$$

вырожденной случайной величины, которая равна единице при любом элементарном исходе (см. доказательство (3.8)).

Следствие 6 ***(средняя вероятность** 1**-пересечений событий из** N**-м.с.).*** Для $n=1$ средняя вероятность 1-пересечений событий $x\in\mathfrak{X}$, т.е. средняя вероятность событий из $\mathfrak{X}$, имеет смысл среднего

$$\overset{1}{\langle p\rangle}_{\mathfrak{X}}=\frac{1}{|\mathfrak{X}|}\sum_{x\in\mathfrak{X}}p_x=\mathbf{E}\left(\frac{\Upsilon_{\mathfrak{X}}}{|\mathfrak{X}|}\right)$$

случайной величины $\Upsilon_{\mathfrak{X}}/|\mathfrak{X}|$ (с множеством значений $\{|Y|/|\mathfrak{X}|, Y\subseteq\mathfrak{X}\}$) — *доли событий* из м.с. $\mathfrak{X}$, наступающих при произвольном элементарном исходе.

Доказательство.

$$\overset{1}{\langle p\rangle}_{\mathfrak{X}}=\langle p\rangle_{\mathfrak{X}}=\frac{1}{|\mathfrak{X}|}\sum_{x\in\mathfrak{X}}p_x=$$

$$=\sum_{m=1}^{N}\pi_m\frac{m}{N}=\sum_{m=0}^{N}\pi_m\frac{m}{N}=\mathbf{E}\left(\frac{\Upsilon_{\mathfrak{X}}}{|\mathfrak{X}|}\right).$$

Следствие 7 ***(средняя вероятность** n**-пересечений событий из** N**-м.с. как среднее с.в.).*** Для произвольного $n=0,1,\ldots,N$ средняя n-вероятность событий $x\in\mathfrak{X}$ имеет смысл среднего

$$\overset{n}{\langle p\rangle}_{\mathfrak{X}}=\mathbf{E}\left(\frac{C^n_{\Upsilon_{\mathfrak{X}}}}{C_N^n}\right) \tag{3.17}$$

случайной величины[28]

$$\frac{C^n_{\Upsilon_{\mathfrak{X}}}}{C_N^n}=\frac{1}{C_N^n}\frac{\Upsilon_{\mathfrak{X}}(\Upsilon_{\mathfrak{X}}-1)\cdots(\Upsilon_{\mathfrak{X}}-n+1)}{n!},$$

вероятностное распределение которой определяется $(N-n+2)$-набором значений

$$\left\{0;\ \frac{C_n^n}{C_N^n},\ldots,\frac{C_N^n}{C_N^n}\right\},$$

включающим нулевое значение, и соответствующим набором вероятностей:

$$\left\{\sum_{i=0}^{n-1}\pi_i;\ \pi_n,\ldots,\pi_N\right\},$$

включающим вероятность нулевого значения с.в. $C^n_{\Upsilon_{\mathfrak{X}}}/C_N^n$:

$$\mathbf{P}\left(C^n_{\Upsilon_{\mathfrak{X}}}=0\right)=\mathbf{P}\left(\Upsilon_{\mathfrak{X}}<n\right)=\sum_{i=0}^{n-1}\pi_i.$$

Доказательство. Из леммы 6

$$\overset{n}{\langle p\rangle}_{\mathfrak{X}}=\sum_{m=n}^{N}\pi_m\frac{C_m^n}{C_N^n}.$$

В силу соглашения для $m=0,1,\ldots,N$:

$$C_m^n=\begin{cases}\dfrac{m!}{n!(m-n)!}, & 0\leqslant n\leqslant m,\\[2ex] 0, & \text{иначе,}\end{cases}$$

получаем требуемое утверждение, так как

$$\sum_{m=n}^{N}\pi_m\frac{C_m^n}{C_N^n}=\sum_{m=0}^{N}\pi_m\frac{C_m^n}{C_N^n}=\mathbf{E}\left(\frac{C^n_{\Upsilon_{\mathfrak{X}}}}{C^n_{|\mathfrak{X}|}}\right).$$

Замечание 9. В силу леммы 6 и ее следствий для каждого $n=0,1,\ldots,N$ средние вероятности n-пересечений событий из N-м.с. $\mathfrak{X}$, $\overset{n}{\langle p\rangle}_{\mathfrak{X}}$, полностью определяются вероятностным распределением его мощностного индикатора $\Upsilon_{\mathfrak{X}}$.

4 Набор сет-расстояний М-дополнения м.с.

4.1 Расширение понятия сет-расстояния М-дополнения м.с.

Сет-расстояние М-дополнения $\mathfrak{X}^{(c)}=\{x^c : x\in\mathfrak{X}\}$ м.с. $\mathfrak{X}$ определяется аналогично как функционал от двух отклонений точки отсчета общности (которой

[26] Эта формула может быть записана также в виде:

$$\pi_n=\delta(0,n)-C_N^n\sum_{m=0}^{n}(-1)^m C_n^m\ \overset{m+N-n}{\langle u\rangle}_{\mathfrak{X}}\ . \tag{3.16′}$$

[27] Под *0-пересечением событий* здесь понимается *пересечение пустого м.с.*, которое по определению полагается равным всему пространству элементарных исходов: $\bigcap\limits_{x\in\emptyset}x=\Omega$.

[28] *доли* n*-подмножеств событий* $\mathfrak{X}$*, наступающих* среди всех n-подмножеств событий из $\mathfrak{X}$ при произвольном элементарном исходе.

служит средняя вероятность дополнений событий) от вероятности пересечения и от вероятности объединения дополнений событий из $\mathfrak{X}$

$$\Delta_{\mathfrak{X}^{(c)}} = \min\left\{\Delta^{\cap}_{\mathfrak{X}^{(c)}}, \Delta^{\cup}_{\mathfrak{X}^{(c)}}\right\}, \tag{4.1}$$

где

$$\Delta^{\cap}_{\mathfrak{X}^{(c)}} = \langle p\rangle_{\mathfrak{X}^{(c)}} - p_{\mathfrak{X}^{(c)}}$$

— сет-расстояние м.с. $\mathfrak{X}^{(c)}$ относительно пересечения,

$$\Delta^{\cup}_{\mathfrak{X}^{(c)}} = u_{\mathfrak{X}^{(c)}} - \langle p\rangle_{\mathfrak{X}^{(c)}}$$

— сет-расстояние м.с. $\mathfrak{X}^{(c)}$ относительно объединения, а

$$\langle p\rangle_{\mathfrak{X}^{(c)}} = \frac{1}{|\mathfrak{X}|}\sum_{x\in\mathfrak{X}}(1-p_x) = 1 - \langle p\rangle_{\mathfrak{X}} \tag{4.2}$$

— средняя вероятность дополнений событий из $\mathfrak{X}$.

Это определение можно следующим образом расширить, чтобы получить N-набор сет-расстояний М-дополнения $\mathfrak{X}^{(c)}$. Объединение всех событий из данного N-м.с. $\mathfrak{X}^{(c)}$ происходит тогда, когда *НЕ происходит хотя бы одно событие* из $\mathfrak{X}$. Если объединение событий в этом определении заменить событием, которое происходит тогда, когда *НЕ происходят хотя бы n событий* из $\mathfrak{X}$, т.е. — заменить объединением n-пересечений дополнений событий из $\mathfrak{X}$, или n-чердаком $\mathfrak{X}^{(c)}$:

$$\bigodot_{x^c\in\mathfrak{X}^{(c)}}^{n} x^c = \bigcup_{\substack{|X^{(c)}|=n\\ X^{(c)}\subseteq\mathfrak{X}^{(c)}}} \mathsf{ter}_{X^{(c)}/\!/\mathfrak{X}^{(c)}},$$

а новую точку n-отсчета определить как среднюю вероятность n-пересечений дополнений событий из $\mathfrak{X}$:

$$\overset{n}{\langle p\rangle}_{\mathfrak{X}^{(c)}} = \frac{1}{C^n_N}\sum_{|X^{(c)}|=n} p_{X^{(c)}/\!/\mathfrak{X}^{(c)}}, \tag{4.3}$$

то прежний способ приводит к расширенному определению n-сет-расстояния М-дополнения $\mathfrak{X}^{(c)}$ м.с. $\mathfrak{X}$ для $n = 1,\ldots,N$:

$$\overset{n}{\Delta}_{\mathfrak{X}^{(c)}} = \min\left\{\overset{n}{\Delta}^{\cap}_{\mathfrak{X}^{(c)}}, \overset{n}{\Delta}^{\cup}_{\mathfrak{X}^{(c)}}\right\}, \tag{4.4}$$

где

$$\overset{n}{\Delta}^{\cap}_{\mathfrak{X}^{(c)}} = \overset{n}{\langle p\rangle}_{\mathfrak{X}^{(c)}} - p_{\mathfrak{X}^{(c)}}$$

— n-сет-расстояние М-дополнения $\mathfrak{X}^{(c)}$ м.с. $\mathfrak{X}$ относительно пересечения,

$$\overset{n}{\Delta}^{\cup}_{\mathfrak{X}^{(c)}} = \overset{n}{u}_{\mathfrak{X}^{(c)}} - \overset{n}{\langle p\rangle}_{\mathfrak{X}^{(c)}}$$

— n-сет-расстояние М-дополнения $\mathfrak{X}^{(c)}$ м.с. $\mathfrak{X}$ относительно объединения n пересечений (n-чердака), а

$$\overset{n}{u}_{\mathfrak{X}^{(c)}} = \mathbf{P}\left(\bigodot_{x^c\in\mathfrak{X}^{(c)}}^{n} x^c\right) \tag{4.5}$$

— вероятность n-чердака $\mathfrak{X}^{(c)}$.

Замечание 10 ***(0-сет-расстояние М-дополнения м.с.).*** Отметим, что при $n = 0$ 0-сет-расстояние $\mathfrak{X}^{(c)}$ равно

$$\overset{0}{\Delta}_{\mathfrak{X}^{(c)}} = \min\left\{\overset{0}{\Delta}^{\cap}_{\mathfrak{X}^{(c)}}, \overset{0}{\Delta}^{\cup}_{\mathfrak{X}^{(c)}}\right\} = 0, \tag{4.6}$$

так как

$$\overset{0}{\Delta}^{\cap}_{\mathfrak{X}^{(c)}} = \overset{0}{\langle p\rangle}_{\mathfrak{X}^{(c)}} - p_{\mathfrak{X}^{(c)}} = 1 - p_{\mathfrak{X}^{(c)}}$$

— 0-сет-расстояние $\mathfrak{X}^{(c)}$ относительно пересечения;

$$\overset{0}{\Delta}^{\cup}_{\mathfrak{X}^{(c)}} = \overset{0}{u}_{\mathfrak{X}^{(c)}} - \overset{0}{\langle p\rangle}_{\mathfrak{X}^{(c)}} = 0$$

— 0-сет-расстояние $\mathfrak{X}^{(c)}$ относительно объединения;

$$\overset{0}{u}_{\mathfrak{X}^{(c)}} = \mathbf{P}\left(\bigodot_{x^c\in\mathfrak{X}^{(c)}}^{0} x^c\right) = \mathbf{P}(\Omega) = 1 \tag{4.7}$$

— вероятность 0-чердака $\mathfrak{X}^{(c)}$, а

$$\overset{0}{\langle p\rangle}_{\mathfrak{X}^{(c)}} = \frac{1}{C^0_N}\sum_{|X^{(c)}|=0} p_{X^{(c)}/\!/\mathfrak{X}^{(c)}} = p_{\emptyset} = 1 \tag{4.8}$$

— средняя вероятность 0-пересечений событий из $\mathfrak{X}^{(c)}$.

Замечание 11 ***(частные вероятности чердаков М-дополнения м.с.).***

$$\overset{n}{u}_{\mathfrak{X}^{(c)}} = \begin{cases} 1, & n = 0,\\ u_{\mathfrak{X}^{(c)}} = 1 - p_{\mathfrak{X}}, & n = 1,\\ \ldots, & \ldots,\\ p_{\mathfrak{X}^{(c)}} = 1 - u_{\mathfrak{X}}, & n = |\mathfrak{X}|,\\ 0, & n > |\mathfrak{X}|. \end{cases}$$

В итоге понятие сет-расстояния $\Delta_{\mathfrak{X}^{(c)}}$ м.с. $\mathfrak{X}^{(c)}$ расширено до понятия N-набора сет-расстояний этого М-дополнения:

$$\left\{\overset{0}{\Delta}_{\mathfrak{X}^{(c)}}, \overset{1}{\Delta}_{\mathfrak{X}^{(c)}}, \ldots, \overset{N}{\Delta}_{\mathfrak{X}^{(c)}}\right\}, \tag{4.9}$$

в котором $\overset{0}{\Delta}_{\mathfrak{X}^{(c)}} = 0$, а $\overset{1}{\Delta}_{\mathfrak{X}^{(c)}} = \Delta_{\mathfrak{X}^{(c)}}$ — это его сет-расстояние в узком смысле.

Лемма 7 ***(сумма вероятностей 2-го рода М-дополнения м.с. по сет-индексам фиксированной мощности).*** Пусть

$$\{\pi^{(c)}_0, \pi^{(c)}_1, \ldots, \pi^{(c)}_N\}$$

— вероятностное распределение мощностного индикатора $\Upsilon_{\mathfrak{X}^{(c)}}$ М-дополнения к N-м.с. $\mathfrak{X}$, где для $m = 0, 1, \ldots, N$

$$\pi^{(c)}_m = \mathbf{P}(\Upsilon_{\mathfrak{X}^{(c)}} = m) = \sum_{|Y^{(c)}|=m} p(Y^{(c)}/\!/\mathfrak{X}^{(c)}). \tag{4.10}$$

Тогда для $n = 0, 1, \ldots, N$

$$\sum_{|X^{(c)}|=n} p_{X^{(c)}/\!/\mathfrak{X}^{(c)}} = \sum_{m=n}^{N} \pi_m^{(c)} C_m^n. \tag{4.11}$$

Доказательство аналогично доказательству леммы 6 (3.11).

Следствие 8 *(средние вероятности n-пересечений событий из M-дополнения N-м.с.).* Поскольку

$$p_{X^{(c)}/\!/\mathfrak{X}^{(c)}} = \sum_{X^{(c)} \subseteq Y^{(c)}} p(Y^{(c)}/\!/\mathfrak{X}^{(c)})$$

— вероятность 2-го рода, то средняя вероятность n-пересечений событий из $\mathfrak{X}^{(c)}$ в силу леммы 3 равна

$$\overset{n}{\langle p \rangle}_{\mathfrak{X}^{(c)}} = \frac{1}{C_N^n} \sum_{|X^{(c)}|=n} p_{X^{(c)}/\!/\mathfrak{X}^{(c)}} =$$

$$= \sum_{m=n}^{N} \pi_m^{(c)} \frac{C_m^n}{C_N^n} = \sum_{m=n}^{N} \pi_{N-m} \frac{C_m^n}{C_N^n}. \tag{4.12}$$

Следствие 9 *(средняя вероятность 0-пересечений событий из M-дополнения N-м.с.).* Для $n = 0$ средняя вероятность 0-пересечений событий $x^c \in \mathfrak{X}^{(c)}$, т.е. "средняя" вероятность Ω, имеет смысл среднего

$$\overset{0}{\langle p \rangle}_{\mathfrak{X}^{(c)}} = \mathbf{E}\,(1) = 1$$

вырожденной случайной величины, которая равна единице при любом элементарном исходе (см. доказательство (4.8)).

Следствие 10 *(средняя вероятность 1-пересечений событий M-дополнения N-м.с.).* Для $n = 1$ средняя вероятность 1-пересечений событий $x^c \in \mathfrak{X}^{(c)}$, т.е. средняя вероятность событий из $\mathfrak{X}^{(c)}$, имеет смысл среднего

$$\overset{1}{\langle p \rangle}_{\mathfrak{X}^{(c)}} = \frac{1}{|\mathfrak{X}^{(c)}|} \sum_{x^c \in \mathfrak{X}^{(c)}} (1 - p_x) = \mathbf{E}\left(\frac{\Upsilon_{\mathfrak{X}^{(c)}}}{|\mathfrak{X}^{(c)}|}\right)$$

случайной величины $\Upsilon_{\mathfrak{X}^{(c)}}/|\mathfrak{X}^{(c)}|$ — *доли событий* из $\mathfrak{X}^{(c)}$, наступающих при произвольном элементарном исходе, иначе говоря, *доли НЕнаступающих событий* в $\mathfrak{X}$.

Доказательство аналогично доказательству следствия 8 на стр. 104.

Следствие 11 *(средняя вероятность n-пересечений событий из M-дополнения N-м.с. как среднее с.в.).* Для произвольного $n = 0, 1, \ldots, N$ средняя n-вероятность событий $x \in \mathfrak{X}$ имеет смысл среднего

$$\overset{n}{\langle p \rangle}_{\mathfrak{X}^{(c)}} = \mathbf{E}\left(\frac{C^n_{\Upsilon_{\mathfrak{X}^{(c)}}}}{C_N^n}\right) \tag{4.13}$$

случайной величины[29]

$$\frac{C^n_{\Upsilon_{\mathfrak{X}^{(c)}}}}{C_N^n} = \frac{1}{C_N^n} \frac{\Upsilon_{\mathfrak{X}^{(c)}}(\Upsilon_{\mathfrak{X}^{(c)}} - 1) \cdots (\Upsilon_{\mathfrak{X}^{(c)}} - n + 1)}{n!},$$

вероятностное распределение которой определяется $(N - n + 2)$-набором значений

$$\left\{0;\ \frac{C_n^n}{C_N^n}, \ldots, \frac{C_N^n}{C_N^n}\right\},$$

включающим нулевое значение, и соответствующим набором вероятностей:

$$\left\{\sum_{i=0}^{n-1} \pi_i^{(c)};\ \pi_n^{(c)} \ldots, \pi_N^{(c)}\right\} =$$

$$= \left\{\sum_{i=N+1-n}^{N} \pi_i;\ \pi_{N-n}, \ldots, \pi_0\right\},$$

включающим вероятность нулевого значения с.в. $C^n_{\Upsilon_{\mathfrak{X}^{(c)}}}/C_N^n$:

$$\mathbf{P}\left(C^n_{\Upsilon_{\mathfrak{X}^{(c)}}} = 0\right) = \mathbf{P}\left(\Upsilon_{\mathfrak{X}^{(c)}} < n\right) =$$

$$= \sum_{i=0}^{n-1} \pi_i^{(c)} = {}^{30} = \sum_{i=N+1-n}^{N} \pi_i.$$

Доказательство аналогично доказательству следствия 8 на стр. 104.

Замечание 12. В силу леммы 7 и ее следствий для каждого $n = 0, 1, \ldots, N$ средние вероятности n-пересечений событий из M-дополнения $\mathfrak{X}^{(c)}$, $\overset{n}{\langle p \rangle}_{\mathfrak{X}^{(c)}}$, полностью определяются вероятностным распределением его мощностного индикатора $\Upsilon_{\mathfrak{X}^{(c)}}$.

Замечание 13. Поскольку для любых $\omega \in \Omega$

$$\Upsilon_{\mathfrak{X}}(\omega) + \Upsilon_{\mathfrak{X}^{(c)}}(\omega) = N,$$

вероятности значений мощностных индикаторов $\Upsilon_{\mathfrak{X}}$ и $\Upsilon_{\mathfrak{X}^{(c)}}$ связаны соотношениями дополнительности для $m = 0, 1, \ldots, N$:

$$\mathbf{P}(\Upsilon_{\mathfrak{X}^{(c)}} = N - m) = \mathbf{P}(\Upsilon_{\mathfrak{X}} = m),$$

т.е.

$$\pi^c_{N-m} = \pi_m,$$

а в итоге:

$$\{\pi_0^{(c)}, \pi_1^{(c)}, \ldots, \pi_N^{(c)}\} = \{\pi_N, \pi_{N-1}, \ldots, \pi_0\}. \tag{4.14}$$

Замечание 14. В силу обращения Мебиуса для вероятностей 2-го и 4-го родов имеем следующую формулу

$$p_{X^{(c)}} = \mathbf{P}\left(\bigcap_{x \in X} x^c\right) =$$

[29] *доли n-подмножеств событий $\mathfrak{X}$, наступающих* среди всех n-подмножеств событий из $\mathfrak{X}$ при произвольном элементарном исходе.

[30] в силу замечания (4.14).

$$= 1 - \mathbf{P}\left(\bigcup_{x\in X} x\right) = 1 - u_X \tag{4.15}$$

для вероятности n-пересечений событий из М-дополнений $X^{(c)} = \{x^c : x \in X\}$ — подмножества М-дополнения $\mathfrak{X}^{(c)}$, где $X \subseteq \mathfrak{X}$, а $X^{(c)} \subseteq \mathfrak{X}^{(c)}$ соответственно.

Из (4.15) следуют аналогичные формулы

$$\overset{n}{\langle p\rangle}_{\mathfrak{X}^{(c)}} = \frac{1}{C_N^n} \sum_{\substack{|X^{(c)}|=n \\ X^{(c)}\subseteq\mathfrak{X}^{(c)}}} p_{X^{(c)}} =$$

$$= \frac{1}{C_N^n} \sum_{\substack{|X|=n \\ X\subseteq\mathfrak{X}}} (1 - u_X) = 1 - \overset{n}{\langle u\rangle}_{\mathfrak{X}} \tag{4.16}$$

для средних вероятностей n-пересечений событий из М-дополнения $\mathfrak{X}^{(c)}$, которые дополняют средние вероятности n-объединений событий из $\mathfrak{X}$:

$$\overset{n}{\langle u\rangle}_{\mathfrak{X}} = \frac{1}{C_N^n} \sum_{\substack{|X|=n \\ X\subseteq\mathfrak{X}}} u_X \tag{4.17}$$

при $n = 0, 1, \ldots, N$.

Замечание 15. Из (4.12) и (4.16) получаем формулу

$$1 - \overset{n}{\langle u\rangle}_{\mathfrak{X}} = \sum_{m=n}^{N} \pi_{N-m} \frac{C_m^n}{C_N^n}, \tag{4.18}$$

выражающую среднюю вероятность n-объединений через вероятностное распределение м.с. $\mathfrak{X}$ для $n = 0, 1, \ldots, N$.

5 Толкование набора сет-расстояний м.с. на языке ф.р. его мощностного индикатора

5.1 Чердак м.с., связанные с ним ф.р. и мощностные индикаторы

Вероятности n-чердаков м.с. $\mathfrak{X}$ связаны со значениями ф.р. $F_{\Upsilon_{\mathfrak{X}}}$ его мощностного индикатора $\Upsilon_{\mathfrak{X}}$ формулой

$$\overset{n}{u}_{\mathfrak{X}} = \mathbf{P}(\Upsilon_{\mathfrak{X}} \geqslant n) =$$

$$= 1 - \mathbf{P}(\Upsilon_{\mathfrak{X}} < n) = 1 - F_{\Upsilon_{\mathfrak{X}}}(n). \tag{5.1.1}$$

для $n = 0, 1, \ldots, N, N+1$ Отсюда ф.р. с.в. $\Upsilon_{\mathfrak{X}}$ определяется для $r \in \mathbb{R}$ вероятностями n-чердаков $\mathfrak{X}$ по формулам

$$F_{\Upsilon_{\mathfrak{X}}}(r) = \begin{cases} 0, & r \leqslant 0, \\ \ldots, & \ldots, \\ 1 - \overset{n}{u}_{\mathfrak{X}} = \overset{N+1-n}{p}_{\mathfrak{X}^{(c)}}, & n-1 < r \leqslant n, \\ & n = 1, \ldots, N, \\ \ldots, & \ldots, \\ 1, & N < r. \end{cases} \tag{5.1.2}$$

а ф.р. с.в. $\Upsilon_{\mathfrak{X}^{(c)}} = N - \Upsilon_{\mathfrak{X}}$, мощностного индикатора М-дополнения $\mathfrak{X}^{(c)}$, определяется для $r \in \mathbb{R}$ вероятностями n-чердаков $\mathfrak{X}$ по формулам

$$F_{\Upsilon_{\mathfrak{X}^{(c)}}}(r) = \begin{cases} 0, & r \leqslant 0, \\ \ldots, & \ldots, \\ 1 - \overset{n}{p}_{\mathfrak{X}^{(c)}} = \overset{N+1-n}{u}_{\mathfrak{X}}, & n-1 < r \leqslant n, \\ & n = 1, \ldots, N, \\ \ldots, & \ldots, \\ 1, & N < r. \end{cases} \tag{5.1.3}$$

Определим на $\mathbb{R}$ *функцию средних вероятностей пересечений (ф.с.в.п.)*

$$F_{\mathfrak{X}}^{\cap}(r) = \begin{cases} 0, & r \leqslant 0, \\ \ldots, & \ldots, \\ 1 - \overset{n}{\langle p\rangle}_{\mathfrak{X}}, & n-1 < r \leqslant n, \\ & n = 1, \ldots, N, \\ \ldots, & \ldots, \\ 1, & N < r. \end{cases} \tag{5.1.4}$$

и *функцию средних вероятностей объединений (ф.с.в.о.)*

$$F_{\mathfrak{X}}^{\cup}(r) = \begin{cases} 0, & r \leqslant 0, \\ \ldots, & \ldots, \\ 1 - \overset{N+1-n}{\langle u\rangle}_{\mathfrak{X}}, & n-1 < r \leqslant n, \\ & n = 1, \ldots, N, \\ \ldots, & \ldots, \\ 1, & N < r. \end{cases} \tag{5.1.5}$$

5.2 Подвал м.с., связанные с ним ф.р. и мощностные индикаторы

Вероятности n-подвалов м.с. $\mathfrak{X}$ связаны со значениями ф.р. $F_{\Upsilon_{\mathfrak{X}}}$ его мощностного индикатора $\Upsilon_{\mathfrak{X}}$ формулой

$$\overset{n}{p}_{\mathfrak{X}} = \mathbf{P}(\Upsilon_{\mathfrak{X}} \geqslant N + 1 - n) =$$

$$= 1 - \mathbf{P}(\Upsilon_{\mathfrak{X}} < N + 1 - n) = 1 - F_{\Upsilon_{\mathfrak{X}}}(N + 1 - n). \tag{5.2.1}$$

для $n = 0, 1, \ldots, N, N+1$ Отсюда ф.р. с.в. $\Upsilon_{\mathfrak{X}}$ определяется для $r \in \mathbb{R}$ вероятностями n-подвалов $\mathfrak{X}$ по формулам

$$F_{\Upsilon_{\mathfrak{X}}}(r) = \begin{cases} 0, & r \leqslant 0, \\ \ldots, & \ldots, \\ 1 - \overset{N+1-n}{p}_{\mathfrak{X}} = \overset{n}{u}_{\mathfrak{X}^{(c)}}, & n-1 < r \leqslant n, \\ & n = 1, \ldots, N, \\ \ldots, & \ldots, \\ 1, & N < r. \end{cases} \tag{5.2.2}$$

а ф.р. с.в. $\Upsilon_{\mathfrak{X}^{(c)}} = N - \Upsilon_{\mathfrak{X}}$, мощностного индикатора М-дополнения $\mathfrak{X}^{(c)}$, определяется для $r \in \mathbb{R}$ вероят-

ностями n-подвалов $\mathfrak{X}$ по формулам

$$F_{\Upsilon_{\mathfrak{X}^{(c)}}}(r)=\begin{cases}0, & r\leqslant 0,\\ \dots, & \dots,\\ 1-\overset{N+1-n}{u}_{\mathfrak{X}^{(c)}}=\overset{n}{p}_{\mathfrak{X}}, & n-1<r\leqslant n,\\ & n=1,\dots,N,\\ \dots, & \dots,\\ 1, & N<r.\end{cases}\tag{5.2.3}$$

Определим на $\mathbb{R}$ *функцию средних вероятностей пересечений*

$$F^{\cap}_{\mathfrak{X}^{(c)}}(r)=\begin{cases}0, & r\leqslant 0,\\ \dots, & \dots,\\ 1-\overset{n}{\langle p\rangle}_{\mathfrak{X}^{(c)}}, & n-1<r\leqslant n,\\ & n=1,\dots,N,\\ \dots, & \dots,\\ 1, & N<r.\end{cases}\tag{5.2.4}$$

и *функцию средних вероятностей объединений*

$$F^{\cup}_{\mathfrak{X}^{(c)}}(r)=\begin{cases}0, & r\leqslant 0,\\ \dots, & \dots,\\ 1-\overset{N+1-n}{\langle u\rangle}_{\mathfrak{X}^{(c)}}, & n-1<r\leqslant n,\\ & n=1,\dots,N,\\ \dots, & \dots,\\ 1, & N<r.\end{cases}\tag{5.2.5}$$

Лемма 8 ***(неравенства для средних вероятностей n-пересечений и n-объединений).*** *Средние вероятности n-пересечений событий из м.с. $\mathfrak{X}$ образуют невозрастающий N-набор:*

$$\overset{0}{\langle p\rangle}_{\mathfrak{X}}\geqslant\overset{1}{\langle p\rangle}_{\mathfrak{X}}\geqslant\dots\geqslant\overset{N}{\langle p\rangle}_{\mathfrak{X}},\tag{5.2.6}$$

а средние вероятности n-объединений событий[31] *из м.с. $\mathfrak{X}$ образуют неубывающий N-набор:*

$$\overset{0}{\langle u\rangle}_{\mathfrak{X}}\leqslant\overset{1}{\langle u\rangle}_{\mathfrak{X}}\leqslant\dots\leqslant\overset{N}{\langle u\rangle}_{\mathfrak{X}}.\tag{5.2.6$'$}$$

Доказательство. Для каждого $n=0,1,\dots,N-1$ при фиксированном $X_{n+1}\subseteq\mathfrak{X}$ имеем $n+1$ неравенство вида

$$p_{X_{n+1}}\leqslant p_{X_n},$$

когда $X_n\subset X_{n+1}$. Просуммировав обе части этих неравенств по всем $X_n\subset X_{n+1}$, получим

$$(n+1)p_{X_{n+1}}\leqslant\sum_{X_n\subset X_{n+1}}p_{X_n}.$$

Еще раз просуммировав обе части полученных неравенств теперь уже по всем $X_{n+1}\subset\mathfrak{X}$, получим

$$(n+1)\sum_{X_{n+1}\subset\mathfrak{X}}p_{X_{n+1}}\leqslant\sum_{X_{n+1}\subset\mathfrak{X}}\sum_{X_n\subset X_{n+1}}p_{X_n}.$$

Переменив порядок суммирования в правой сумме, получим

$$\sum_{X_{n+1}\subset\mathfrak{X}}\sum_{X_n\subset X_{n+1}}p_{X_n}=$$

$$=\sum_{X_n\subset\mathfrak{X}}\sum_{X_n\subset X_{n+1}}p_{X_n}=(N-n)\sum_{X_n\subset\mathfrak{X}}p_{X_n}.$$

Поделив обе части неравенства на $(n+1)C_N^{n+1}$, имеем

$$\frac{1}{C_N^{n+1}}\sum_{X_{n+1}\subset\mathfrak{X}}p_{X_{n+1}}\leqslant\frac{(N-n)}{(n+1)C_N^{n+1}}\sum_{X_n\subset\mathfrak{X}}p_{X_n}.$$

В итоге получаем неравенство

$$\frac{1}{C_N^{n+1}}\sum_{X_{n+1}\subset\mathfrak{X}}p_{X_{n+1}}\leqslant\frac{1}{C_N^{n}}\sum_{X_n\subset\mathfrak{X}}p_{X_n},$$

которое представляет из себя требуемое неравенство для средних вероятностей $(n+1)$- и n-пересечений:

$$\overset{n+1}{\langle p\rangle}_{\mathfrak{X}}\leqslant\overset{n}{\langle p\rangle}_{\mathfrak{X}},$$

где, напомним, $n=0,1,\dots,N-1$. Неравенства (5.2.6) доказаны.

Из этих неравенств, примененных к М-дополнению $\mathfrak{X}^{(c)}$ м.с. $\mathfrak{X}$:

$$\overset{0}{\langle p\rangle}_{\mathfrak{X}^{(c)}}\geqslant\overset{1}{\langle p\rangle}_{\mathfrak{X}^{(c)}}\geqslant\dots\geqslant\overset{N}{\langle p\rangle}_{\mathfrak{X}^{(c)}},$$

получаем неравенства для средних вероятностей n-объединений (5.2.6$'$). Лемма доказана.

6 Сет-*дистограмма* множества событий

Определение 1 ***(сет-дистограмма м.с.).*** *Сет-**дистограммой*** м.с. $\mathfrak{X}$ называется графическое изображение набора его сет-расстояний (3.9):

$$\left\{\overset{0}{\Delta}_{\mathfrak{X}},\overset{1}{\Delta}_{\mathfrak{X}},\dots,\overset{N}{\Delta}_{\mathfrak{X}}\right\},$$

в виде диаграммы (см. рис. 4). Графическое изображение набора его относительных сет-расстояний (3.9$'$):

$$\left\{\overset{0}{\Delta}'_{\mathfrak{X}},\overset{1}{\Delta}'_{\mathfrak{X}},\dots,\overset{N}{\Delta}'_{\mathfrak{X}}\right\},$$

в виде диаграммы (см. рис. 4) называется *относительной сет-**дистограммой*** м.с. $\mathfrak{X}$.

[31] Под *0-объединением событий* здесь понимается *объединение пустого м.с.*, которое по определению полагается равным невозможному событию: $\bigcup\limits_{x\in\emptyset}x=\emptyset$.

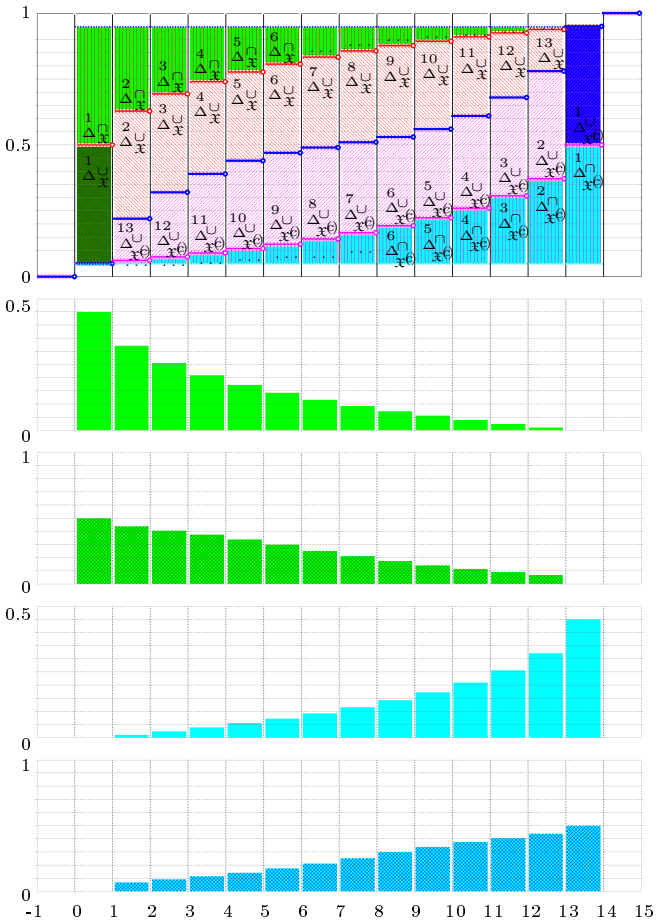


Рисунок 4: Сет-дистограммы, средние и полное сет-расстояние м.с. $\mathfrak{X}$, значения ф.р. мощностного индикатора которого приведены в Табл. 2.

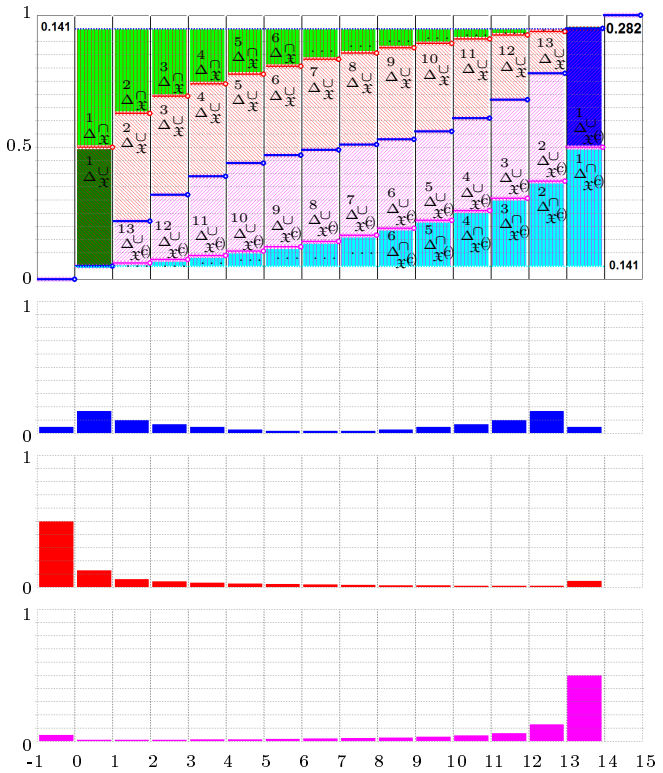


Рисунок 5: Ф.р. и две сопровождающие функции мощностного индикатора м.с. $\mathfrak{X}$ (значения ф.р. мощностного индикатора приведены в Табл. 2), а также масс-вероятности ф.р. (синий) и двух сопровождающих функций (красный и фуксия).

На рисунках 4 и 6: вверху графики ф.р. $F_{\Upsilon_{\mathfrak{X}}}$ (ступенчатый синий), функции средних вероятностей пересечения $F^{\cap}_{\Upsilon_{\mathfrak{X}}}$ (ступенчатый красный) и функции средних вероятностей объединения $F^{\cup}_{\Upsilon_{\mathfrak{X}}}$ (ступенчатый фуксия) мощностных индикаторов $\Upsilon_{\mathfrak{X}}$ 14-м.с. $\mathfrak{X}$; наборы сет-расстояний $\overset{n}{\Delta}{}^{\cap}_{\mathfrak{X}}$ и $\overset{n}{\Delta}{}^{\cup}_{\mathfrak{X}}$ м.с. $\mathfrak{X}$ (лайм и зеленый), а также $\overset{n}{\Delta}{}^{\cap}_{\mathfrak{X}^{\theta}}$ и $\overset{n}{\Delta}{}^{\cup}_{\mathfrak{X}^{\theta}}$ его M-дополнения $\mathfrak{X}^{(c)}$ (аква и синий); абсолютные и относительные сет-*д*истограммы м.с. $\mathfrak{X}$ (лайм и заштрихованный лайм, в центре) и его M-дополнения $\mathfrak{X}^{(c)}$ (аква и заштрихованная аква, внизу). Слева вверху указано значение среднего сет-расстояния $\langle\Delta_{\mathfrak{X}}\rangle$ м.с. $\mathfrak{X}$, справа внизу значение среднего сет-расстояния $\langle\Delta_{\mathfrak{X}^{(c)}}\rangle$ его M-дополнения $\mathfrak{X}^{(c)}$, справа вверху указано полное сет-расстояние $\mathfrak{X}$. Значения ф.р. мощностного индикатора $\Upsilon_{\mathfrak{X}}$ м.с. $\mathfrak{X}$ на рис. 4 приведены в таблице 2.

На рисунках 5 и 7: вверху: функция распределения $F_{\mathfrak{X}}$ (синий; значения ф.р. приведены в Табл. 2), функция средних вероятностей пересечений $F^{\cap}_{\mathfrak{X}}$ (красный) и функция средних вероятностей объединений $F^{\cup}_{\mathfrak{X}}$ (фуксия) мощностного индикатора $\Upsilon_{\mathfrak{X}}$; ниже *в другом масштабе*: масс-вероятности ф.р. $F_{\mathfrak{X}}$ (синий), а также масс-вероятности его функции средних вероятностей пересечений $F^{\cap}_{\mathfrak{X}}$ (красный) и функции средних вероятностей объединений $F^{\cup}_{\mathfrak{X}}$ (фуксия). Значения ф.р. мощностного индикатора $\Upsilon_{\mathfrak{X}}$ м.с. $\mathfrak{X}$ на рис. 5 приведены в той же таблице 2.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	0.00	0.05	0.22	0.32	0.39	0.44	0.47	0.49	0.51	0.53	0.56	0.61	0.68	0.78	0.95

Таблица 2: Значения ф.р. мощностного индикатора 14-м.с. на рис. 4.

7 Среднее и полное сет-расстояния множества событий

Чтобы иметь возможность измерять одним показателем разобщенность событий из N-м.с. $\mathfrak{X}$, можно воспользоваться *средним сет-расстоянием* $\mathfrak{X}$, которое определяется как

$$\langle\Delta_{\mathfrak{X}}\rangle = \frac{1}{N}\sum_{n=1}^{N} \overset{n}{\Delta}_{\mathfrak{X}}, \tag{7.1}$$

среднее его n-сет-расстояний ($n = 1, \ldots, N$); а также *полным сет-расстоянием* $\mathfrak{X}$, которое определяется как

$$\Delta_{\mathfrak{X}} = \langle\Delta_{\mathfrak{X}}\rangle + \langle\Delta_{\mathfrak{X}^{(c)}}\rangle, \tag{7.2}$$

сумма средних сет-расстояний $\mathfrak{X}$ *и его M-дополнения* $\mathfrak{X}^{(c)}$, и учитывает не только разобщенность событий из $\mathfrak{X}$, но и разобщенность их дополнений из $\mathfrak{X}^{(c)}$.

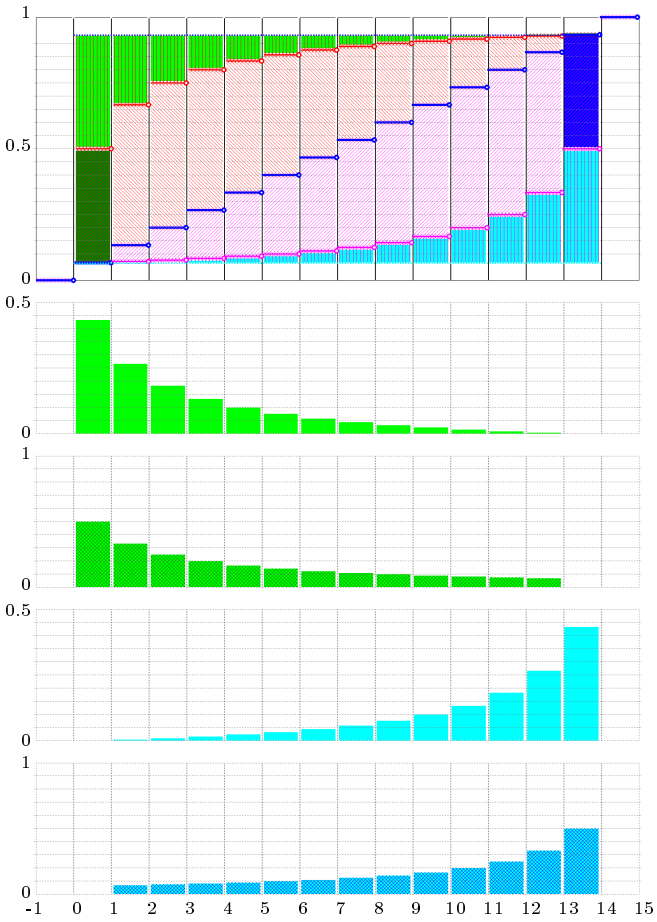


Рисунок 6: Все значения мощностного индикатора 14-м.с. $\mathfrak{X}$ равновероятны и равны 1/15, а сопровождающие ф.р. вычисляются по формулам (5.1.4) и (5.1.5) и равны соответственно: $n/(n+1)$ и $1/(15-n), n = 0, 1, \ldots, 14$.

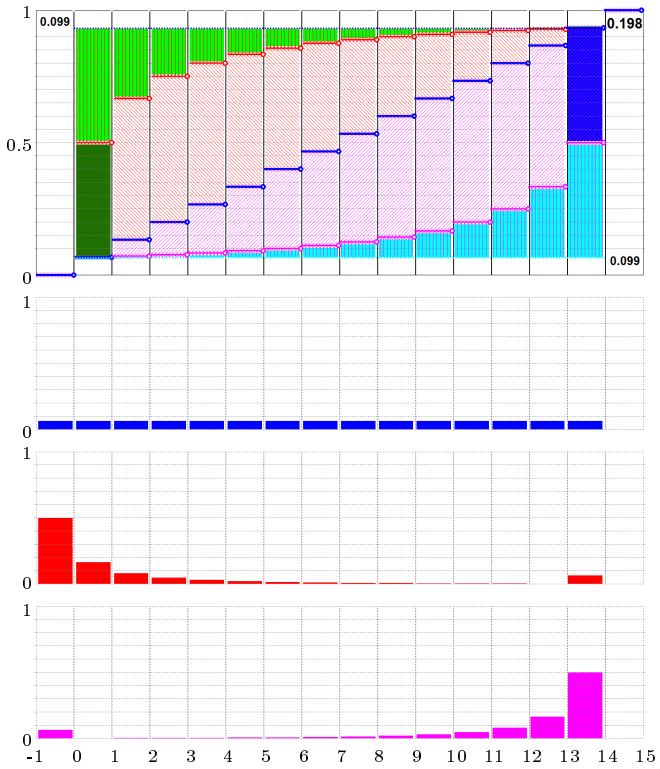


Рисунок 7: Все значения мощностного индикатора 14-м.с. $\mathfrak{X}$ равновероятны и равны 1/15, а сопровождающие ф.р. вычисляются по формулам (5.1.4) и (5.1.5) и равны соответственно: $n/(n+1)$ и $1/(15-n), n = 0, 1, \ldots, 14$.

8 Вспомогательные утверждения

Утверждение 1 *(существование N-м.с. с заданным мощностным индикатором). Пусть F_ξ — ф.р. дискретной с.в. ξ, принимающей значения из $\{0, 1, \ldots, N\}$, а $(\Omega, \mathcal{A}, \mathbf{P})$ — всеобщее вероятностное пространство. Тогда существует N-м.с. $\mathfrak{X} \subset \mathcal{A}$, ф.р. $F_{\Upsilon_\mathfrak{X}}$ мощностного индикатора $\Upsilon_\mathfrak{X}$ которого совпадает с ней на $\mathbb{R}$:*

$$F_{\Upsilon_\mathfrak{X}} = F_\xi. \tag{8.1}$$

Доказательство. Будем считать, что для $r \in \mathbb{R}$ ф.р. с.в. ξ определяется формулой

$$F_\xi(r) = \mathbf{P}(\xi < r), \tag{8.2}$$

тогда вероятности значений с.в. ξ

$$\pi_n = \mathbf{P}(\xi = n) \tag{8.3}$$

связаны с ф.р. для $n = 0, 1, \ldots, N$ соотношением

$$F_\xi(r) = \sum_{n<r} \pi_n. \tag{8.4}$$

Для доказательства утверждения (8.1) достаточно показать, что мощностной индикатор $\Upsilon_\mathfrak{X}$ N-множества *вложенных* событий

$$\mathfrak{X} = \{x_1, \ldots, x_N\}, \tag{8.5}$$

где

$$x_1 \supseteq x_2 \supseteq \ldots \supseteq x_N, \tag{8.6}$$

а вероятности событий $x_n \in \mathfrak{X}$ определяются вероятностями значений с.в. ξ (8.3) по формулам:

$$p_{x_n} = \sum_{i=n}^{N} \pi_i \tag{8.7}$$

для $n = 1, \ldots, N$, имеет ф.р. $F_{\Upsilon_\mathfrak{X}}$, совпадающую с F_ξ.

Заметим, что

$$\{\omega : \Upsilon_\mathfrak{X}(\omega) = n\} = \begin{cases} \Omega - x_1, & n = 0, \\ x_n - x_{n+1}, & n = 1, \ldots, N-1, \\ x_N, & n = N. \end{cases}$$

Отсюда, а также в силу вложенности событий из $\mathfrak{X}$

и формул (8.7) получаем, что

$$\mathbf{P}(\Upsilon_{\mathfrak{X}}=n)=\begin{cases}\pi_0=1-p_{x_1}=1-\sum\limits_{i=1}^{N}\pi_i,\ n=0,\\ \pi_n=p_{x_n}-p_{x_{n+1}},\ n=1,\ldots,N-1,\\ \pi_N=p_{x_N},\qquad\qquad n=N.\end{cases}$$

— вероятности всех значений с.в. ξ и $\Upsilon_{\mathfrak{X}}$ совпадают, следовательно совпадают и их ф.р.

Утверждение 2 *(о мощностном индикаторе M-суммы[32] нескольких N-м.с.). Пусть N-м.с.*

$$\mathfrak{X}=\left(\sum_{l=1}^{L}\right)\mathfrak{X}^l$$

— равно M-сумме N-м.с.

$$\mathfrak{X}^l=\mathfrak{X}(\cap)W^l=\{x\cap W^l:x\in\mathfrak{X}\},$$

M-пересечений м.с. $\mathfrak{X}$ с событиями W^l, которые образуют разбиение пространства элементарных исходов

$$\Omega=W^1+\ldots+W^L=\sum_{l=1}^{L}W^l.$$

Тогда для $\omega\in\Omega$

$$\Upsilon_{\mathfrak{X}}(\omega)=\sum_{l=1}^{L}\Upsilon_{\mathfrak{X}^l}(\omega).$$

Лемма 9 *(пересечение всех n-пересечений событий из N-м.с.). Пусть $\mathfrak{X}$ — N-множество событий, тогда для $n=1,2,\ldots,N$*

$$\bigcap_{\substack{X\subseteq\mathfrak{X}\\|X|=n}}\mathsf{ter}_{X/\!/\mathfrak{X}}=\mathsf{ter}_{\mathfrak{X}/\!/\mathfrak{X}}.\tag{8.3}$$

Доказательство. Утверждение леммы следует из соотношения для двух террасных событий 2-го рода

$$\mathsf{ter}_{X/\!/\mathfrak{X}}\cap\mathsf{ter}_{X/\!/\mathfrak{X}}=\mathsf{ter}_{X\cup Y/\!/\mathfrak{X}},$$

которое легко обобщается на любой набор террасных событий 2-го рода, занумерованных подмножествами из совокупности $\mathcal{B}\subseteq 2^{\mathfrak{X}}$:

$$\bigcap_{X\in\mathcal{B}}\mathsf{ter}_{X/\!/\mathfrak{X}}=\mathsf{ter}_{\bigcup_{X\in\mathcal{B}}X/\!/\mathfrak{X}},\tag{8.4}$$

поскольку[33]

$$\bigcap_{w\in\overset{n}{\mathfrak{X}}{}^{\cap}}w=\bigcap_{\substack{|X|=n\\X\subseteq\mathfrak{X}}}\mathsf{ter}_{X/\!/\mathfrak{X}}=\mathsf{ter}_{\mathfrak{X}/\!/\mathfrak{X}},$$

так как

$$\bigcup_{\substack{|X|=n\\X\subseteq\mathfrak{X}}}X=\mathfrak{X}.$$

$\star\star\star$

Множество n-пересечений $\overset{n}{\mathfrak{X}}{}^{\cap}$ событий из N-м.с. $\mathfrak{X}$ имеет мощность

$$\left|\overset{n}{\mathfrak{X}}{}^{\cap}\right|=C_N^n,$$

и разбивает Ω на $2^{C_N^n}$ террасных событий, из которых, поскольку $\overset{n}{\mathfrak{X}}{}^{\cap}$ порождено N-м.с. $\mathfrak{X}$, только

$$1+\sum_{m=n}^{N}C_N^m$$

могут быть отличны от невозможного события. Причем террасные события

$$\mathsf{ter}\left(\overset{n}{W}/\!/\overset{n}{\mathfrak{X}}{}^{\cap}\right)=\bigcap_{w\in\overset{n}{W}}w\bigcap_{w\in\overset{n}{\mathfrak{X}}{}^{\cap}-\overset{n}{W}}w^c,$$

порожденные C_N^n-м.с. $\overset{n}{\mathfrak{X}}{}^{\cap}$, совпадают с "соответствующими" террасными событиями

$$\mathsf{ter}(X/\!/\mathfrak{X})=\bigcap_{x\in X}x\bigcap_{x\in\mathfrak{X}-X}x^c,$$

порожденными N-м.с. $\mathfrak{X}$. Правило соответствия определяется леммой 10.

Лемма 10 *(правило соответствия между террасными событиями, порожденными N-м.с. $\mathfrak{X}$ и C_N^n-м.с. $\overset{n}{\mathfrak{X}}{}^{\cap}$). Пусть $\mathfrak{X}$ — N-м.с., а $\overset{n}{\mathfrak{X}}{}^{\cap}$ — C_N^n-множество n-пересечений событий из $\mathfrak{X}$, тогда между террасными событиями, порожденными этими м.с. существует следующее соответствие для каждого $X\subseteq\mathfrak{X}$ и $n=0,1,\ldots,N$:*

$$\mathsf{ter}\left(\overset{n}{W}(X)/\!/\overset{n}{\mathfrak{X}}{}^{\cap}\right)=\begin{cases}\Omega, & n=0;\\ \mathsf{ter}(X/\!/\mathfrak{X}), & |X|\geqslant n,\\ & n=1,\ldots,N;\\ \sum\limits_{|X|<n}\mathsf{ter}(X/\!/\mathfrak{X}), & |X|<n,\\ & n=1,\ldots,N,\end{cases}$$

где

$$\overset{n}{W}(X)=\begin{cases}\left\{\bigcap\limits_{x\in X_n}x:X_n\subseteq C_X^n\right\}, & |X|\geqslant n;\\ \emptyset, & |X|<n\end{cases}$$

— подмножество $\overset{n}{\mathfrak{X}}{}^{\cap}$, а

$$C_X^n=\{X\subseteq\mathfrak{X}:|X|=n\}\subseteq 2^{\mathfrak{X}}$$

[32] *M-сумма* и *M-пересечение* — сумма и пересечение нескольких м.с. по Минковскому.

[33] Здесь роль совокупности $\mathcal{B}$ играет так называемый n-слой

$$C_{\mathfrak{X}}^n=\{X\subseteq\mathfrak{X}:|X|=n\}\subseteq 2^{\mathfrak{X}}$$

для $n=0,1,\ldots,N$.

— *n-слой подмножеств м.с.* $\mathfrak{X}$.

Лемма 11.

$$\sum_{m=n}^{N} C_m^n = C_{N+1}^{n+1}. \tag{8.5}$$

Доказательство. Вычислим сумму

$$\sum_{m=n}^{N} C_m^n + \sum_{m=n-1}^{N} C_m^{n-1} = \sum_{m=n}^{N} C_m^n + \sum_{m=n}^{N} C_m^{n-1} + 1 =$$

$$= \sum_{m=n}^{N} (C_m^n + C_m^{n-1}) + 1 = \sum_{m=n}^{N} C_{m+1}^n + 1 =$$

$$= C_{N+1}^n + \sum_{m=n}^{N} C_m^n.$$

Таким образом

$$\sum_{m=n-1}^{N} C_m^{n-1} = C_{N+1}^n,$$

или

$$\sum_{m=n}^{N} C_m^n = C_{N+1}^{n+1}.$$

Замечание 16. Пусть $\mathfrak{X}$ — N-м.с., мощностной индикатор которого $\Upsilon_{\mathfrak{X}}$ принимает свои значения $0, 1, \ldots, N$ с равными вероятностями:

$$\pi_m = \mathbf{P}(\Upsilon_{\mathfrak{X}} = m) = \frac{1}{N+1}$$

для $m = 0, 1, \ldots, N$. Тогда средняя вероятность n-пересечений событий из $\mathfrak{X}$ в силу следствия 6 определяется суммой биномиальных коэффициентов по нижнему индексу:

$$\sum_{m=n}^{N} C_m^n \tag{8.6}$$

и в силу леммы 11 равна:

$$\overset{n}{\langle p \rangle}_{\mathfrak{X}} = \frac{1}{(N+1)C_N^n} \sum_{m=n}^{N} C_m^n = \frac{1}{n+1}. \tag{8.7}$$

Замечание 17 ***(средние вероятности n-пересечений событий из N-м.с. с биномиальным распределением мощностного индикатора).*** Допустим мощностной индикатор $\Upsilon_{\mathfrak{X}}$ N-м.с. $\mathfrak{X}$ имеет биномиальное (N, p)-распределение. Тогда

$$\overset{n}{\langle p \rangle}_{\mathfrak{X}} = p^n. \tag{3.12′}$$

Действительно, так как по условию

$$\pi_m = \mathbf{P}\left(\Upsilon_{\mathfrak{X}} = m\right) = C_N^m p^m (1-p)^{N-m},$$

то из (3.12) получаем:

$$\overset{n}{\langle p \rangle}_{\mathfrak{X}} = \sum_{m=n}^{N} C_N^m p^m (1-p)^{N-m} \frac{C_m^n}{C_N^n} =$$

$$= p^n \sum_{m=n}^{N} C_{N-n}^{N-m} p^{m-n} (1-p)^{N-m} = p^n.$$

Замечание 18. Пусть $\mathfrak{X}_N$ — N-м.с., мощностной индикатор которого $\Upsilon_{\mathfrak{X}_N}$ имеет биномиальное распределение параметрами $(N, \lambda/N)$:

$$\pi_m(N) = \mathbf{P}(\Upsilon_{\mathfrak{X}_N} = m) = C_N^m \left(\frac{\lambda}{N}\right)^m \left(1 - \frac{\lambda}{N}\right)^{N-m}.$$

для $m = 0, 1, \ldots, N$. Тогда средние вероятности n-пересечений событий из $\mathfrak{X}_N$ в силу замечания 9 вычисляются по формулам:

$$\overset{n}{\langle p \rangle}_{\mathfrak{X}_N} = \left(\frac{\lambda}{N}\right)^n \tag{8.9}$$

для $n = 0, 1, \ldots, N$.

При $N \to \infty$ такое $(N, \lambda/N)$-биномиальное распределение мощностного индикатора N-м.с. $\mathfrak{X}_N$ стремится к распределению Пуассона с параметром λ:

$$\pi_m = \mathbf{P}(\Upsilon_{\mathfrak{X}} = m) = \frac{\lambda^m}{m!} e^{-\lambda}.$$

для $m = 0, 1, \ldots$; а предел средних вероятностей (8.9) имеет вид

$$\lim_{N\to\infty} \overset{n}{\langle p \rangle}_{\mathfrak{X}_N} = \begin{cases} 1, & n = 0, \\ 0, & \text{иначе.} \end{cases} \tag{8.10}$$

9 Обсуждение и перспективы

А. Сет-*дистограммы*, средние и полные сет-расстояния м.с. В работе введены новые инструменты измерения общности и разобщенности событий, образующих некоторое м.с., которые названы *сет-дистограммой*, *средним* и *полным сет-расстоянием* этого м.с. Общность и разобщенность событий — это характеристика м.с. относительно новых точек отсчета — средних вероятностей пересечений событий из м.с.

Сет-дистограмма показывает распределение показателей разобщенности событий из м.с. относительно мощности их подмножеств; полностью определяется ф.р. мощностного индикатора м.с. и удобна для измерения разобщенности событий, а также для демонстрации степени разобщенности событий, образующих подмножества фиксированной мощности данного м.с.; удачно дополняет традиционные средства вероятностного описания м.с., такие, как ф.р., масс-функция вероятностей и гистограмма его мощностного индикатора, показывая детальное распределение степени разобщенности событий относительно мощности их подмножеств.

Среднее сет-расстояние м.с. служит единым числовым показателем разобщенности м.с., который

предоставляет возможность сравнивать степени разобщенности различных м.с. между собой. Сумма средних сет-расстояний м.с. и его М-дополнения названа *полным сет-расстоянием* м.с., которое учитывает не только разобщенность его событий, но и разобщенность их дополнений.

Б. Эквивалентность событий с разными именами. Помимо введения нового инструмента измерения общности и разобщенности событий, образующих некоторое м.с., в этой работе необходимо выделить еще один результат, который в силу своей фундаментальности для науки исчисления событий вполне мог бы остаться незамеченным фоном. Речь идет о леммах 1 и 4, которые могут быть переформулированы с использованием до сих пор ускользающего от безупречного определения понятия "имени события" (см., например, [3, 7]) таким образом, чтобы способствовать приближению понимания этого понятия к безупречности.

Леммы 1 и 4 (эквивалентность событий, имеющих различные имена). *Пусть* $\mathfrak{X}$ *—* N*-м.с., тогда*

$$\bigodot_{x\in\mathfrak{X}}^{n} x = \begin{cases} \{\omega : \Upsilon_{\mathfrak{X}}(\omega) \geqslant n\}, \\ \{\omega : \Upsilon_{\mathfrak{X}^{(c)}}(\omega) < N-n+1\} \end{cases} \tag{2.2.1}$$

— события, которые имеют имена: "n***-чердак м.с.** $\mathfrak{X}$*", "**объединение** n***-пересечений событий из м.с.** $\mathfrak{X}$*", "**событие, происходящее, когда происходит не менее** n **событий из м.с.** $\mathfrak{X}$*," эквивалентны;*

$$\boxdot_{x\in\mathfrak{X}}^{n} x = \begin{cases} \{\omega : \Upsilon_{\mathfrak{X}}(\omega) \geqslant N+1-n\}, \\ \{\omega : \Upsilon_{\mathfrak{X}^{(c)}}(\omega) < n\} \end{cases} \tag{2.3.4}$$

— события, которые имеют имена: "n***-подвал м.с.** $\mathfrak{X}$*", "**пересечение** n***-объединений событий из м.с.** $\mathfrak{X}$*", "**событие, происходящее, когда НЕ происходит более** $n-1$ **события из м.с.** $\mathfrak{X}$*," эквивалентны.*

Если первые два имени событий в этих утверждениях эквивалентны по определению, то эквивалентность их третьему не очевидна и формально требует доказательств, которые приведены выше. Возможно кому-то эти утверждения покажутся не слишком глубокими (особенно лемма 1). Однако весьма ценную для эвентологии роль эти леммы играют в качестве наглядных примеров того, что *любое утверждение об эквивалентности чего бы-то ни было можно рассматривать как утверждение об эквивалентности событий, имеющих различные имена.*

В. "Производные" множества пересечений и "производные" множества объединений событий из м.с. Каждое N-м.с. $\mathfrak{X}$ определяет "производные" *множества* n*-пересечений* и *множества* n*-объединений* событий из $\mathfrak{X}$ ($n = 0, 1, \ldots, N$), вероятностные распределения которых определяются вероятностным распределением самого $\mathfrak{X}$. Характеристики этих "производных" м.с. детализируют характеристики $\mathfrak{X}$. Например, их сет-расстояния при $n = 0, 1, \ldots, N$ образуют *сет-**дистограммы*** $\mathfrak{X}$ и его М-дополнения $\mathfrak{X}^{(c)}$ соответственно.

Объединение всех событий из "производного" *множества* n*-пересечений* событий из $\mathfrak{X}$ названо в работе n*-чердаком* м.с. $\mathfrak{X}$ и получило специальное обозначение: $\bigodot_{x\in\mathfrak{X}}^{n} x$, а пересечение всех событий из "производного" *множества* n*-объединений* событий из $\mathfrak{X}$ — n*-подвалом* м.с. $\mathfrak{X}$ и обозначено $\boxdot_{x\in\mathfrak{X}}^{n} x$.

Г. Расширение понятия средневероятного события. Средневероятные события N-м.с. можно определять последовательно: для самого N-множества событий (унарных пересечений событий), для C_N^2-множества парных пересечений событий из N-м.с., для C_N^3-множества тройных пересечений событий и т.д. вплоть до множества из одного элемента — полного пересечения м.с., чтобы получить в результате невозрастающий набор средневероятных событий из данного N-м.с. различных порядков $n = 0, 1, \ldots, N$.

Благодарности

Статья написана недолгими июльскими ночами 2014, каждой из которых в отдельности автор исключительно благодарен за элегантную неотвратимость. Особая благодарность Валентину Толедо[34]: его стеклянный цикл "Объединения и пересечения" оставил автора в ночном июле в который раз наедине с пронзающей бесконечность красотой теоретико-множественных операций.

Список литературы

[1] О. Ю. Воробьев. *Сет-суммирование*. Наука, Новосибирск, 1993, 137с.

[2] О. Ю. Воробьев. *Эвентология*. Сибирский федеральный университет, Красноярск, 2007, 435с., https://www.academia.edu/179393/Vorobyev_O.Yu._Eventology._-_Krasnoyarsk_Siberian_Federal_University._-_2007._-_435p.

[3] О. Ю. Воробьев. Начала аксиоматизации эвентологии. *Труды VIII Междунар. ФАМ конференции по финансово-актуарной математике и смежным вопросам*, Красноярск: СФУ (под ред. Олега Воробьева), 1:77–82, 2009.

[4] О. Ю. Воробьев. Меры рассеяния м.с. по пространству элементарных исходов. *Труды XV Междунар. ЭМ конференции по эвентологической математике и смежным вопросам*, Красноярск: СФУ (под ред. Олега Воробьева):54–63, 2011.

[5] О. Ю. Воробьев. Событийные средние в эвентологии, их асимптотические свойства, интерпретация и визуализация. *Труды XVI Междунар. ЭМ конференции по эвентологической математике и смежным вопросам*, Красноярск: СФУ (под ред. Олега Воробьева):50–56, 2012.

[6] О. Ю. Воробьев. Вне привычных измерений статистических взаимосвязей: эвентологически пересмотренный подход к измерению общности, ассоциации и статистической зависимости множеств событий. В *Трудах XIII ФАМЭМС конф. по финансово-актуарной математике и эвентологии многомерной статистики*, Красноярск, СФУ (под ред. Олега Воробьева):44–94, 2014.

[7] O. Yu. Vorobyev. On elements of axiomatizing eventology. *Journal of Siberian Federal University. Mathematics & Physics*, 3(2):157–164, 2010.

[34] *Валентин Толедо (р. 1970, Valentin Toledo)* — испанский художник; проживает в Лос-Анджелесе, Калифорния; его шедевр из стеклянного цикла: "Объединения & Пересечения" (янв. 2010, Санта Моника, Калифорния).

К обоснованиям эвентологического метода

Олег Юрьевич Воробьев
Институт математики и фундаментальной информатики
Сибирский федеральный университет
Красноярск
email: oleg.yu.vorobyev@gmail.com
url: sfu-kras.academia.edu/OlegVorobyev

Аннотация. *Обсуждаются некоторые основания эвентологического метода. Вводятся новые понятия всеобщего K-события (всеобщего колмогоровского события) и со́-бытия (имени всеобщего K-события). Формулируются две новые эвентологические аксиомы* ***со́-бытия*** *и* ***множества со́-бытий****, из которых следуют предложенные ранее аксиомы VI и VII [2, 4]. Вводятся новые понятия сет-произведения и упорядоченного сет-произведения множеств со́-бытий, которые существенно отличаются от привычных понятий прямого и декартова произведений.*

Ключевые слова. *Эвентология, вероятность, событие, всеобщее K-событие, со́-бытие, аксиома со́-бытия, аксиома множества со́-бытий, сет-произведение множеств.*

1 Эвентологические прелиминарии теории всеобщих K-событий и со́-бытий

Любое *всеобщее колмогоровское событие (K-событие)*

$$x^{\mho} \in \mathcal{A}^{\mho}$$

из сигма-алгебры *всеобщего вероятностного пространства* $(\Omega, \mathcal{A}^{\mho}, \mathbf{P})$, происходящее с ненулевой вероятностью, — это всегда *бесконечное* множество всеобщих элементарных исходов $x^{\mho} \subseteq \Omega$. Поэтому субъект исключительно в силу ограниченности его событийных знаний и опыта не способен наблюдать всеобщее K-событие *непосредственно (прямо)*[1], но лишь *опосредовано*, всякий раз оказываясь вынужденным проверять значение $\mathfrak{n}_x(\omega)$ одного из припасенных им для всеобщего K-события $x^{\mho}$ эквивалентных *обобщенных характеризующих высказываний*[2] $\mathfrak{n}_x \in \mathfrak{n}(x^{\mho})$, которыми он определяет наступление всеобщего K-события

$$x^{\mho} = \{\omega \in \Omega : \mathfrak{n}_x(\omega) = \mathfrak{true}\},$$

чтобы понять, выполняется ли отношение принадлежности

$$\omega \in x^{\mho}$$

для очередного наступившего всеобщего элементарного исхода $\omega \in \Omega$.

Итак, вместо *прямого* наблюдения всеобщего K-события $x^{\mho} \in \mathcal{A}^{\mho}$ субъекту $\mu \in \mathfrak{M}$ доступно его *опосредованное* наблюдение — наблюдение значений *характеризующего высказывания* $\mathfrak{n}_x \in \mathfrak{n}(x^{\mho})$, которое в математической эвентологии [1] называется *именем всеобщего K-события* $x^{\mho}$ и обозначается более коротко: *со́-бытие*[3] $x = \mathfrak{n}_x$. Из имен $x \in \mathcal{A}_\mu$, которыми субъект называет всеобщие K-события $x^{\mho}$, образуется *именная сигма-алгебра* $\mathcal{A}_\mu$ и складывается *со́∼бытие* субъекта $\mu \in \mathfrak{M}$. Поскольку только *именами всеобщих K-событий* из его собственной *именной сигма-алгебры* $\mathcal{A}_\mu$ субъект способен манипулировать *непосредственно* в качестве его *со́-бытий* в рамках *субъектно-именного вероятностного пространства* $(\Omega, \mathcal{A}_\mu, \mathbf{P})$.

Таким образом, каждый субъект $\mu \in \mathfrak{M}$ *опосредовано* наблюдает *всеобщее K-событие* $x^{\mho} \in \mathcal{A}^{\mho}$, наблюдая за значением одного из его имен, *со́-бытием*

$$x \in \mathfrak{name}_\mu(x^{\mho}) \subset \mathcal{A}_\mu,$$

из множества

$$\mathfrak{name}_\mu(x^{\mho}) = \{x \in \mathcal{A}_\mu : \mathfrak{event}_\mu(x) = x^{\mho}\},$$

составленного из всех имен всеобщего K-события $x^{\mho}$, которые данный субъект имеет в своем *именном тезаурусе* $\mathcal{A}_\mu$ для *опосредованно-именного* наблюдения $x^{\mho}$.

© 2014 О.Ю.Воробьев

Олег Воробьев (ред.), Труды XIII ФАМЭМС'2014, Красноярск: СФУ

[1] Для *прямого* наблюдения всеобщего K-события $x^{\mho}$ субъект должен иметь в своем распоряжении бесконечный перечень всех составляющих $x^{\mho}$ всеобщих элементарных исходов $\omega \in x^{\mho}$.

[2] Под *обобщенными характеризующими высказываниями* здесь могут пониматься логические высказывания в логиках любого порядка, т.е. простые и сложные логические высказывания, предикаты, а также любые булевозначные функции, определенные на тех или иных интерпретационных пространствах.

[3] ***Имя всеобщего K-события (со́-бытие)*** **относится к** ***всеобщему K-событию*** **так же, как** ***высказывание*** **относится к** ***суждению*** (любое высказывание суждения предполагает, что рядом существуют другие эквивалентные высказывания того же суждения); а из совокупности *со́-бытий (характеристических высказываний)* складывается *со́∼бытие* субъекта $\mu \in \mathfrak{M}$ — его собственное *множество со́-бытий*, характеризующее данного субъекта.

В итоге каждый субъект наблюдает и манипулирует всеобщими K-событиями из всеобщего вероятностного пространства $(\Omega, \mathcal{A}^{\mho}, \mathbf{P})$, манипулируя лишь их именами (со́-бытиями), т.е. *опосредовано* через его *субъектно-именное вероятностное пространство* $(\Omega, \mathcal{A}_{\mu}, \mathbf{P})$, которое можно считать субъектно-именной μ-реализацией всеобщего вероятностного пространства, или *частным вероятностным пространством*.

Вероятность определяется на субъектно-именной сигма-алгебре $\mathcal{A}$ соотношением:

$$\mathbf{P}(x) = \mathbf{P}(\mathfrak{event}_{\mu}(x))$$

для $x \in \mathcal{A}$.

1.1 Всеобщее вероятностное пространство и его субъектно-именные реализации

Рассмотрим *всеобщее вероятностное пространство* $(\Omega, \mathcal{A}^{\mho}, \mathbf{P})$ и одну из его *субъектно-именных реализаций* — частное вероятностное пространство $(\Omega, \mathcal{A}, \mathbf{P})$. Элементами сигма-алгебры $\mathcal{A}^{\mho}$ служат *всеобщие колмогоровские события (K-события)* $x^{\mho} \in \mathcal{A}^{\mho}$, а элементами сигма-алгебры $\mathcal{A}$ — *со́-бытия* $x \in \mathcal{A}$, которые служат *именами всеобщих K-событий* $x^{\mho}$. Между этими сигма-алгебрами установлено следующее соответствие.

Одному имени всеобщего K-события соответствует одно всеобщего K-событие:

$$\mathfrak{event} : \mathcal{A} \to \mathcal{A}^{\mho},$$

$$\mathfrak{event}(x) = x^{\mho}.$$

Одному всеобщему K-событию соответствует множество эквивалентных имен всеобщих K-событий:

$$\mathfrak{name} : \mathcal{A}^{\mho} \to 2^{\mathcal{A}},$$

$$\mathfrak{name}(x^{\mho}) = \{x \in \mathcal{A} : \mathfrak{event}(x) = x^{\mho}\}.$$

K-событие — это единственное значение событийного отображения $x^{\mho} = \mathfrak{event}(x)$ его имени (характеризующего высказывания) $x \in \mathcal{A}$:

$$\mathfrak{e} : \mathcal{A} \to \mathcal{A}^{\mho},$$

$$\mathfrak{e}(x) = x^{\mho}.$$

Имя всеобщего K-события — это одно из возможно многих значений именного отображения $x \in \mathfrak{name}(x^{\mho})$ всеобщего K-события $x^{\mho} \in \mathcal{A}^{\mho}$:

$$\mathfrak{n} : \mathcal{A}^{\mho} \to 2^{\mathcal{A}},$$

$$\mathfrak{n}(x^{\mho}) = \{x \in \mathcal{A} : \mathfrak{e}(x) = x^{\mho}\}.$$

Именное отображение $\mathfrak{n}$ сохраняет теоретико-множественные операции по Минковскому:

$$\mathfrak{n}(x^{\mho} \star y^{\mho}) = \mathfrak{n}(x^{\mho})(\star)\,\mathfrak{n}(y^{\mho}) =$$

$$= \{x \star y \in \mathcal{A} : \mathfrak{e}(x) = x^{\mho}; \mathfrak{e}(y) = y^{\mho}\}.$$

Событийное отображение $\mathfrak{e}$ сохраняет обычные теоретико-множественные операции над со́-бытиями:

$$\mathfrak{e}(x \star y) = \mathfrak{e}(x) \star \mathfrak{e}(y).$$

Итак, алгебра всеобщих K-событий $\mathcal{A}^{\mho}$ и алгебра их имен (субъектно-именная алгебра) $\mathcal{A}$ связаны взаимно-обратными именным, $\mathfrak{name}$, и событийным, $\mathfrak{event}$, отображениями, сохраняющими теоретико-множественные операции над всеобщими K-событиями и теоретико-множественные операции по Минковскому над со́-бытиями (логические операции над именами всеобщих K-событий): теоретико-множественные операции над всеобщими K-событиями $x^{\mho} \in \mathcal{A}^{\mho}$ служат образами или прообразами логических операций по Минковскому над именами этих всеобщих K-событий

$$x \in \mathfrak{n}(x^{\mho}) \subseteq \mathcal{A}.$$

Иначе говоря, алгебра всеобщих K-событий $\mathcal{A}^{\mho}$ — это фактор-алгебра субъектно-именной алгебры $\mathcal{A}$ относительно отношения эквивалентности, которое определяется событийным отображением $\mathfrak{event}$: два со́-бытия $x, y \in \mathcal{A}$ эквивалентны, $x \sim y$, т.т.т., когда

$$\mathfrak{event}(x) = \mathfrak{event}(y),$$

т.е. когда они служат *различными* именами одного и того же всеобщего K-события $x^{\mho} = \mathfrak{event}(x)$.

Замечание 1 ***(изоморфизм теоретико-множественных операций над K-событиями и логических операций над со́-бытиями)***. Поскольку субъект, всегда принимает решения в рамках своей субъектно-именной реализации $(\Omega, \mathcal{A}, \mathbf{P})$ всеобщего вероятностного пространства $(\Omega, \mathcal{A}^{\mho}, \mathbf{P})$, то ему приходится использовать в качестве эвентологической модели интересующей его ситуации некоторое конечное множество *со́-бытий*, т.е. конечное множество *имен всеобщих K-событий* $\mathfrak{X} \subset \mathcal{A}$. На алгебре *со́-бытий (имен всеобщих K-событий)*, т.е. на алгебре *характеристических высказываний*, определены логические операции, которые изоморфны теоретико-множественным операциям над соответствующими всеобщими K-событиями. Этот изоморфизм позволяет субъекту обращаться с *со́-бытиями* так же, как с *всеобщими K-событиями*, считая, например, что они происходят (когда высказывания истинны) или не происходят (когда высказывания ложны), и формально используя для манипулирования ими теоретико-множественные операции, которые лишь обозначают изоморфные им логические операции.

Замечание 2 ***(множеству со́-бытий соответствует мультимножество K-событий)***. Из свойств именного отображения $\mathfrak{event}$ следует, что для произвольного $N \in \mathbb{N}_0$ любому *N-множеству* $\mathfrak{X} \in \mathcal{A}$ со́-

бытий из сигма-алгебры $\mathcal{A}$ каждого частного вероятностного пространства $(\Omega, \mathcal{A}, \mathbf{P})$ соответствует N-*мультимножество*[4]

$$\mathfrak{X}^{\mho} = \left\{ \left(x^{\mho}\right)^{m_x} : x \in \mathfrak{X}, \sum_{x\in\mathfrak{X}} m_x = N \right\} \tag{1.1}$$

всеобщих K-событий из сигма-алгебры всеобщего вероятностного пространства $(\Omega, \mathcal{A}^{\mho}, \mathbf{P})$, где

$$x^{\mho} = \mathfrak{events}(x) \in \mathcal{A}^{\mho}$$

для $x \in \mathfrak{X}$. Отсюда в силу предлагаемого построения, некоторые (или даже все) всеобщие K-события, скрывающиеся под именами из $\mathfrak{X}$, могут совпадать. Таким образом, оперируя *множеством со́-бытий*, имен всеобщих K-событий (которые, разумеется, все различны), субъект получает возможность манипулировать не только *множествами всеобщих K-событий*, но и *мультимножествами всеобщих K-событий*. Эта возможность оказывается удобной во многих эвентологических приложениях, поскольку не только упрощает анализ, но и расширяет диапазон событийных исследований, позволяя, например, без ограничений предполагать наличие любой структуры зависимости у рассматриваемого множества всеобщих K-событий.

1.2 Аксиома со́-бытия

Учитывая выше изложенные предопределения и, в особенности, помня, что *со́-бытие* $x \in \mathcal{A}$ в некотором частном вероятностном пространстве $(\Omega, \mathcal{A}, \mathbf{P})$ определяется, как *имя K-события* $x^{\mho} \in \mathcal{A}^{\mho}$ из всеобщего вероятностного пространства $(\Omega, \mathcal{A}^{\mho}, \mathbf{P})$, сформулируем аксиому со́-бытия, из которой следует аксиома VI, ранее предложенная мною в [2, 2009] и [4, 2010].

Аксиома VI (аксиома со́-бытия). Для каждого $N \in \mathbb{N}_0$ и каждого N-мультимножества $\mathfrak{X}^{\mho}$ всеобщих K-событий из сигма-алгебры $\mathcal{A}^{\mho}$ всеобщего вероятностного пространства $(\Omega, \mathcal{A}^{\mho}, \mathbf{P})$ найдется частное вероятностное пространство $(\Omega, \mathcal{A}, \mathbf{P})$ и именующее $\mathfrak{X}^{\mho}$ N-множество

$$\mathfrak{X} = \{x \in \mathcal{A} : x \in \mathfrak{name}(x^{\mho}), x^{\mho} \in \mathfrak{X}^{\mho}\} \tag{1.2}$$

со́-бытий из сигма-алгебры $\mathcal{A}$ этого пространства.

Замечание 3 *(формулы обращения между N-множеством со́-бытий и N-мультимножеством всеобщих K-событий).* Формула (1.1), определяющая N-мультимножество всеобщих K-событий $\mathfrak{X}^{\mho}$ через N-множество со́-бытий $\mathfrak{X}$, прозрачно следует из предопределений. Справедливость обратной формулы (1.2), определяющей N-множество со́-событий $\mathfrak{X}$ через N-мультимножество K-событий $\mathfrak{X}^{\mho}$, следует из Аксиомы VI.

[4] Если все кратности элементов мультимножества $\mathfrak{X}^{\mho}$ равны единице: $m_x = 1, x \in \mathfrak{X}$, то мультимножество "превращается" в множество.

1.3 Аксиома множества со́-бытий

Сформулируем теперь аксиому множества со́-бытий, из которой следует аксиома VII, ранее предложенная мною в [2, 2009] и [4, 2010].

Определение 1 *(сет-симплекс).* Обозначим

$$\boldsymbol{a} = \boldsymbol{a}(\mathfrak{A}) = \{a(A/\!/\mathfrak{A}) : A \subseteq \mathfrak{A}\} \tag{1.3}$$

— произвольное *частично упорядоченное множество*[5] из 2^N неотрицательных чисел $a(A/\!/\mathfrak{A}) \geq 0$, которые "занумерованы" подмножествами *неупорядоченного* N-множества $\mathfrak{A}$ и сумма которых равна единице: $\sum_{A\subseteq\mathfrak{A}} a(A/\!/\mathfrak{A}) = 1$. Совокупность всех таких чу-множеств чисел обозначается

$$\mathcal{S}_{\mathfrak{A}}^{\left(2^N\right)} = \left\{ \boldsymbol{a} : a(A/\!/\mathfrak{A}) \geq 0, \sum_{A\subseteq\mathfrak{A}} a(A/\!/\mathfrak{A}) = 1 \right\} \tag{1.4}$$

и называется 2^N-вершинным *сет-симплексом*, порожденным N-множеством $\mathfrak{A}$.

Аксиома VII (аксиома множества со́-бытий). Для каждого $N \in \mathbb{N}_0$, каждого произвольного N-множества $\mathfrak{A}$, нумерующего своими подмножествами элементы сет-симплекса $\mathcal{S}_{\mathfrak{A}}^{\left(2^N\right)}$, и любого элемента $\boldsymbol{a} = \{a(A/\!/\mathfrak{A}), A \subseteq \mathfrak{A}\} \in \mathcal{S}_{\mathfrak{A}}^{\left(2^N\right)}$ найдется N-множество со́-событий $\mathfrak{X}$ из сигма-алгебры $\mathcal{A}$ некоторого частного вероятностного пространства $(\Omega, \mathcal{A}, \mathbf{P})$, такое, что его вероятностное распределение $\boldsymbol{p}(\mathfrak{X}) = \{p(X/\!/\mathfrak{X}),\ X \subseteq \mathfrak{X}\}$ совпадает с $\boldsymbol{a}$:

$$\boldsymbol{p}(\mathfrak{X}) = \boldsymbol{a}. \tag{1.5}$$

Замечание 4 *(о сет-симплексе вероятностных распределений множеств со́-бытий).* Из аксиомы VII следует, что для любого 2^N-вершинного сет-симплекса $\mathcal{S}_{\mathfrak{A}}^{\left(2^N\right)}$ найдется такое частное вероятностное пространство $(\Omega, \mathcal{A}, \mathbf{P})$, сигма-алгебра которого содержит совокупность множеств со́-бытий, вероятностные распределения которых заполняют весь сет-симплекс $\mathcal{S}_{\mathfrak{A}}^{\left(2^N\right)}$.

2 Теоретико-множественные прелиминарии террасного метода

Рассмотрим всеобщее вероятностное пространство $(\Omega, \mathcal{A}^{\mho}, \mathbf{P})$, в котором алгебра всеобщих K-событий $\mathcal{A}^{\mho}$ удовлетворяет стандартной системе аксиом ZFC (аксиомы Цермело-Френкеля с аксиомой выбора) теории множеств [3] и принимает определение Куратовского упорядоченной пары

$$\langle x, y\rangle = \{\{x\}, \{x, y\}\}.$$

[5] Для термина *частично упорядоченное множество* можно использовать сокращение *чу-множество*, или *по-множество* по аналогии с английским сокращением *poset* для *partially ordered set*.

Напомним, что *порядковое число*[6] — это любое *транзитивное множество*[7], вполне упорядоченное отношением принадлежности. $\emptyset$ — это порядковое число; если α — порядковое число, то $\alpha \cup \{\alpha\}$, записываемое как $\alpha + 1$, также порядковое. Это обеспечивает представление фон Неймана натуральных чисел как конечных порядковых чисел

$$n = \{0, 1, \ldots, n-1\}.$$

Мощность множества X, обозначаемая $|X|$, — это минимальное порядковое число α, такое, что существует инъективное отображение из X в α и наоборот. Таким образом, $|X| \leqslant |Y|$ имеет место, если существует инъекция из X в Y; если вдобавок не существует инъекции из Y в X, то $|X| < |Y|$.

2.1 Обозначения и аббревиатуры террасных множеств

Пусть

$$\mathfrak{S} = \{\mathfrak{X}_1, \ldots, \mathfrak{X}_n\}$$

— множество, состоящее из n конечных подмножеств базового множества $\mathcal{A}$: $\mathfrak{X}_1 \subset \mathcal{A}, \ldots \mathfrak{X}_n \subset \mathcal{A}$, которое порождает террасное разбиение $\mathcal{A}$ на террасные множества 1-го рода:

$$\mathcal{A} = \sum_{S \subseteq \mathfrak{S}} \mathsf{ter}(S /\!\!/ \mathfrak{S}),$$

где

$$\mathsf{ter}(S /\!\!/ \mathfrak{S}) = \bigcap_{X \in S} X \bigcap_{X \in \mathfrak{S} - S} (\mathcal{A} - X) \subseteq \mathcal{A}.$$

— *S-террасное множество 1-го рода*, порожденное множеством $\mathfrak{S}$.

Обозначения для мощностей террасных множеств 1-го рода:

$$\mathtt{t}_S = |\mathsf{ter}(S /\!\!/ \mathfrak{S})|.$$

Аббревиатура для террасных множеств 1-го рода и их мощностей, например,

$$\mathtt{t}_1 = \mathtt{t}_{\{\mathfrak{X}_1\}} = |\mathsf{ter}(\{\mathfrak{X}_1\} /\!\!/ \mathfrak{S})| = |\mathtt{ter}_1|,$$

$$\mathtt{t}_{12} = \mathtt{t}_{\{\mathfrak{X}_1, \mathfrak{X}_2\}} = |\mathsf{ter}(\{\mathfrak{X}_1, \mathfrak{X}_2\} /\!\!/ \mathfrak{S})| = |\mathtt{ter}_{12}|,$$

$$\mathtt{t}_{123} = \mathtt{t}_{\{\mathfrak{X}_1, \mathfrak{X}_2, \mathfrak{X}_3\}} = |\mathsf{ter}(\{\mathfrak{X}_1, \mathfrak{X}_2, \mathfrak{X}_3\} /\!\!/ \mathfrak{S})| = |\mathtt{ter}_{123}|.$$

[6] *Ординальное (порядковое) число* — число, обозначающее относительную позицию в последовательности, такую как "первая", "вторая", "третья"; иногда сокращается до: *ординал*. *Кардинальное число* — число, обозначающее количество, а не порядок элементов множества; число элементов *конечного* множества; иногда сокращается до: *кардинал*.

[7] *Транзитивное множество* — это такое множество A, что "всякий раз, когда $x \in A$, и $y \in x$, то $y \in A$"; или, эквивалентно: "всякий раз, когда $x \in A$, и x не является *ур-элементом (атомом, индивидуалом)*, то x — подмножество A: $x \subseteq A$".

Обозначения для мощностей террасных множеств 2-го рода:

$$\mathtt{n}_S = |\mathsf{ter}_{S /\!\!/ \mathfrak{S}}| = \left| \bigcap_{\mathfrak{X} \in S} \mathfrak{X} \right|.$$

Аббревиатура для мощностей террасных множеств 2-го рода, например,

$$\mathtt{n}_1 = \mathtt{n}_{\{\mathfrak{X}_1\}} = |\mathfrak{X}_1|,$$

$$\mathtt{n}_{12} = \mathtt{n}_{\{\mathfrak{X}_1, \mathfrak{X}_2\}} = |\mathfrak{X}_1 \cap \mathfrak{X}_2|,$$

$$\mathtt{n}_{123} = \mathtt{n}_{\{\mathfrak{X}_1, \mathfrak{X}_2, \mathfrak{X}_3\}} = |\mathfrak{X}_1 \cap \mathfrak{X}_2 \cap \mathfrak{X}_3|.$$

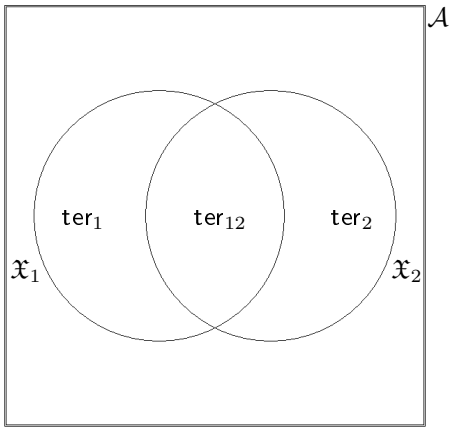


Рисунок 1: Диаграмма Венна двух конечных множеств со́-бытий $\mathfrak{X}_1$ и $\mathfrak{X}_2$ из субъектно-именной сигма-алгебры $\mathcal{A}$, порождающих разбиение $\mathcal{A}$ на четыре террасных множества, из которых обозначены три конечных террасных множества.

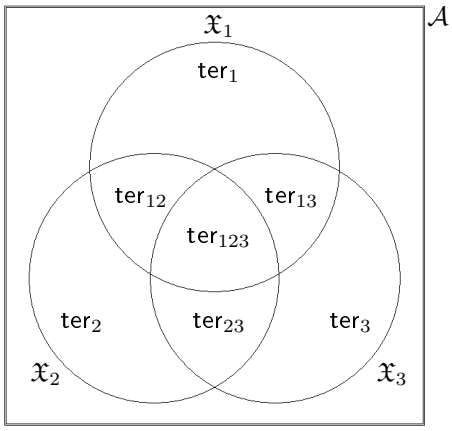


Рисунок 2: Диаграмма Венна трех конечных множеств со́-бытий $\mathfrak{X}_1, \mathfrak{X}_2$ и $\mathfrak{X}_3$ из субъектно-именной сигма-алгебры $\mathcal{A}$, порождающих разбиение $\mathcal{A}$ на восемь террасных множеств, из которых обозначены семь конечных террасных множеств.

Формулы обращения Мебиуса, связывающие мощности террасных множеств 1-го и 2-го рода для $S \subseteq \mathfrak{S}$:

$$\begin{cases} \mathtt{n}_S = \sum\limits_{S \subseteq T \subseteq \mathfrak{S}} \mathtt{t}_T, \\ \mathtt{t}_S = \sum\limits_{S \subseteq T \subseteq \mathfrak{S}} \mathtt{n}_T (-1)^{|T-S|}. \end{cases}$$

В частности, для триплета:

$$\begin{cases} \mathtt{n}_1 = \mathtt{t}_1 + \mathtt{t}_{12} + \mathtt{t}_{13} + \mathtt{t}_{123}, \\ \mathtt{n}_2 = \mathtt{t}_2 + \mathtt{t}_{12} + \mathtt{t}_{23} + \mathtt{t}_{123}, \\ \mathtt{n}_3 = \mathtt{t}_3 + \mathtt{t}_{13} + \mathtt{t}_{23} + \mathtt{t}_{123}, \\ \mathtt{n}_{12} = \mathtt{t}_{12} + \mathtt{t}_{123}, \\ \mathtt{n}_{13} = \mathtt{t}_{13} + \mathtt{t}_{123}, \\ \mathtt{n}_{23} = \mathtt{t}_{23} + \mathtt{t}_{123}, \\ \mathtt{n}_{123} = \mathtt{t}_{123}; \end{cases}$$

$$\begin{cases} \mathtt{t}_1 = \mathtt{n}_1 - \mathtt{n}_{12} - \mathtt{n}_{13} + \mathtt{n}_{123}, \\ \mathtt{t}_2 = \mathtt{n}_2 - \mathtt{n}_{12} - \mathtt{n}_{23} + \mathtt{n}_{123}, \\ \mathtt{t}_3 = \mathtt{n}_3 - \mathtt{n}_{13} - \mathtt{n}_{23} + \mathtt{n}_{123}, \\ \mathtt{t}_{12} = \mathtt{n}_{12} - \mathtt{n}_{123}, \\ \mathtt{t}_{13} = \mathtt{n}_{13} - \mathtt{n}_{123}, \\ \mathtt{t}_{23} = \mathtt{n}_{23} - \mathtt{n}_{123}, \\ \mathtt{t}_{123} = \mathtt{n}_{123}. \end{cases}$$

2.2 Обозначения и аббревиатуры террасных со́-бытий

Пусть $\mathfrak{X} \subset \mathcal{A}$ — конечное множество со́-бытий, порождающее разбиение[8]

$$\Omega = \sum_{X \subseteq \mathfrak{X}} \mathsf{ter}(X /\!\!/ \mathfrak{X})$$

на террасные со́-бытия

$$\mathsf{ter}(X) = \mathsf{ter}(X /\!\!/ \mathfrak{X}) = \bigcap_{x \in X} x \bigcap_{x \in \mathfrak{X} - X} x^c \subseteq \Omega.$$

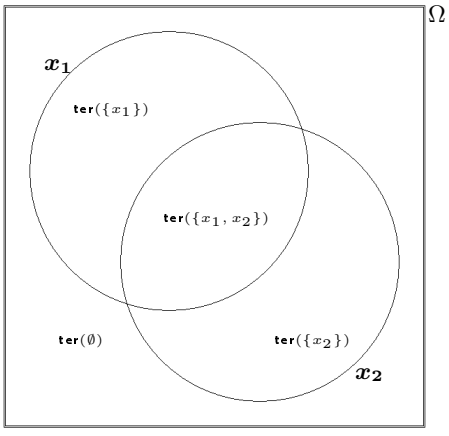


Рисунок 3: Диаграмма Венна дуплета со́-бытий $\mathfrak{X} = \{x_1, x_2\}$ из субъектно-именной сигма-алгебры $\mathcal{A}$, порождающего разбиение Ω на четыре террасных со́-бытия $\mathsf{ter}(X), X \subseteq \mathfrak{X}$.

[8] см. рис. 3 для дуплета со́-бытий и рис. 4 для триплета со́-бытий.

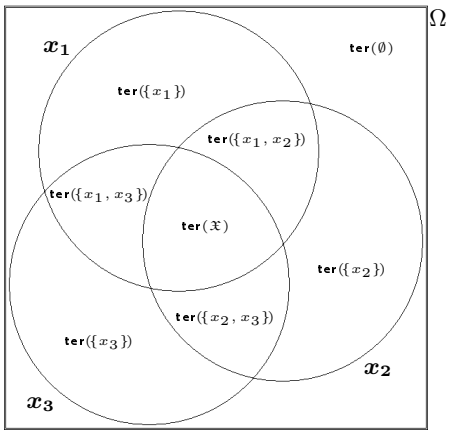
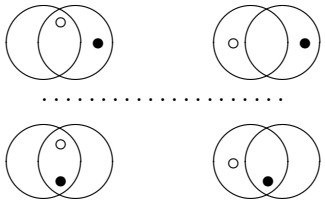


Рисунок 4: Диаграмма Венна триплета со́-бытий $\mathfrak{X} = \{x_1, x_2, x_3\}$ из субъектно-именной сигма-алгебры $\mathcal{A}$, порождающих разбиение Ω на восемь террасных со́-бытий $\mathsf{ter}(X), X \subseteq \mathfrak{X}$.

$|(\mathfrak{X}_1 \cap \mathfrak{X}_2) \otimes (\mathfrak{X}_2 \setminus \mathfrak{X}_1)| = |\mathfrak{X}_1 \cap \mathfrak{X}_2| \cdot |\mathfrak{X}_2 \setminus \mathfrak{X}_1|$ | $|(\mathfrak{X}_1 \setminus \mathfrak{X}_2) \otimes (\mathfrak{X}_2 \setminus \mathfrak{X}_1)| = |\mathfrak{X}_1 \setminus \mathfrak{X}_2| \cdot |\mathfrak{X}_2 \setminus \mathfrak{X}_1|$



$|(\mathfrak{X}_1 \cap \mathfrak{X}_2)^{\otimes 2}| = C^2_{|\mathfrak{X}_1 \cap \mathfrak{X}_2|} = |\mathfrak{X}_1 \cap \mathfrak{X}_2| \cdot (|\mathfrak{X}_1 \cap \mathfrak{X}_2| - 1)/2$ | $|(\mathfrak{X}_1 \setminus \mathfrak{X}_2) \otimes (\mathfrak{X}_1 \cap \mathfrak{X}_2)| = |\mathfrak{X}_1 \setminus \mathfrak{X}_2| \cdot |\mathfrak{X}_1 \cap \mathfrak{X}_2|$

Рисунок 5: Перечисление 4 непересекающихся слагаемых сет-произведения дуплета $\mathfrak{S} = \{\mathfrak{X}_1, \mathfrak{X}_2\}$, составленного из 2-множеств $\{\circ, \bullet\}$, содержащих по одному элементу из каждого множества дуплета: $\circ \in \mathfrak{X}_1$ и $\bullet \in \mathfrak{X}_2$, выбранных из террасных множеств $\mathsf{ter}(S /\!\!/ \mathfrak{S}), S \neq \emptyset$, порожденных $\mathfrak{S}$.

3 Сет-произведения

Определение 2 ***(сет-произведение как множество множеств).*** Для n множеств $\mathfrak{X}_1, \mathfrak{X}_2, \ldots, \mathfrak{X}_n$, образующих множество

$$\mathfrak{S} = \{\mathfrak{X}_1, \mathfrak{X}_2, \ldots, \mathfrak{X}_n\},$$

их *(неупорядоченное) сет-произведение* определяется как множество

$$\bigotimes_{\mathfrak{X} \in \mathfrak{S}} \mathfrak{X} = \bigotimes_{i=1}^{n} \mathfrak{X}_i = \mathfrak{X}_1 \otimes \ldots \otimes \mathfrak{X}_n =$$

$$= \Big\{ \{x_1, \ldots, x_n\} : x_1 \in \mathfrak{X}_1, \ldots, x_n \in \mathfrak{X}_n \Big\},$$

состоящее из *n-множеств*[9]

$$\{x_1, \ldots, x_n\} \in \bigotimes_{\mathfrak{X} \in \mathfrak{S}} \mathfrak{X},$$

которые содержат точно по одному элементу x_i из каждого множества $\mathfrak{X}_i$ $(1 \leqslant i \leqslant n)$. При $\mathfrak{S} = \emptyset$ полагается, что

$$\bigotimes_{\mathfrak{X} \in \emptyset} \mathfrak{X} = \{\emptyset\}.$$

[9] Разумеется, неупорядоченных.

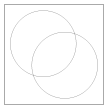
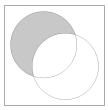
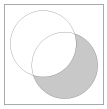
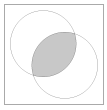
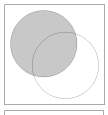
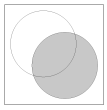
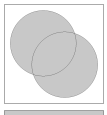
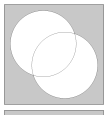
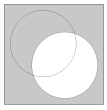
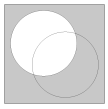
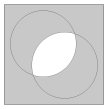
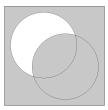
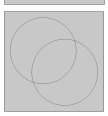
	$\varnothing$	$\emptyset$ Impossible сó-event Невозможное сó-бытие	False, Contradiction Ложь, Противоречие
	$\{\{x_1\}\}$	$x_1 \cap x_2^c$ $x_1 \setminus x_2$ Direct difference of сó-events Прямая разность сó-бытий	$x_1 \not\to x_2$ "$x_1 \not\supset x_2$" Direct nonimplication Прямая антиимпликация
	$\{\{x_2\}\}$	$x_1^c \cap x_2$ x_1 / x_2 Converse difference of сó-events Обратная разность сó-бытий	$x_1 \not\leftarrow x_2$ "$x_1 \not\subset x_2$" Converse nonimplication Обратная антиимпликация
	$\{\{x_1, x_2\}\}$	$x_1 \cap x_2$ Intersection of сó-events	$x_1 \& x_2$, $x_1 \wedge x_2$ x_1 AND x_2 Conjunction Конъюнкция
	$\{\{x_1\}, \{x_2\}\}$	$x_1 \Delta x_2$ Symmetric difference of 'o-events Симметрическая разность сó-бытий	$x_1 \not\leftrightarrow x_2, x_1 \not\equiv x_2$ x_1 XOR x_2 Exclusive disjunction Исключительная дизъюнкция
	$\{\{x_1\}, \{x_1, x_2\}\}$	x_1 Сó-events x_1 Сó-бытие x_1	x_1 Proposition x_1 Утверждение x_1
	$\{\{x_2\}, \{x_1, x_2\}\}$	x_2 Сó-events x_2 Сó-бытие x_2	x_2 Proposition x_2 Утверждение x_2
	$\{\{x_1\}, \{x_2\}, \{x_1, x_2\}\}$	$x_1 \cup x_2$ Union of сó-events Объединение сó-бытий	$x_1 \vee x_2$, $x_1 \cup x_2$ x_1 OR x_2 Disjunction Дизъюнкция
	$\{\emptyset\}$	$x_1^c \cap x_2^c$ Nonunion of сó-events Антиобъединение сó-бытий	$x_1 \downarrow x_2$, x_1 NOR x_2 Pierce arrow Стрелка Пирса Антидизъюнкция
	$\{\emptyset, \{x_1\}\}$	x_2^c Complement of the сó-event x_2 Дополнение сó-бытия x_2	$\neg x_2$ Negation of x_2 Отрицание x_2
	$\{\emptyset, \{x_2\}\}$	x_1^c Complement of the сó-event x_1 Дополнение сó-бытия x_1	$\neg x_1$ Negation of x_1 Отрицание x_1
	$\{\emptyset, \{x_1, x_2\}\}$	$x_1 \Delta x_2^c$ Equivalence of сó-events Эквивалентность сó-бытий	$x_1 \leftrightarrow x_2$, $x_1 \equiv x_2$ x_1 IFF x_2 If and only if Если и только если Эквивалентность
	$\{\emptyset, \{x_1\}, \{x_2\}\}$	$x_1^c \cup x_2^c$ Nonintersection of сó-events Антипересечение сó-бытий	$x_1 \uparrow x_2$, x_1 NAND x_2 Sheffer stroke Штрих Шеффера Антиконъюнкция
	$\{\emptyset, \{x_1\}, \{x_1, x_2\}\}$	$x_1 \cup x_2^c$, $x_1 \not\setminus x_2$ Direct nondifference of сó-events Прямая антиразность сó-бытий	$x_1 \leftarrow x_2$ "$x_1 \subset x_2$" Converse implication Обратная импликация
	$\{\emptyset, \{x_2\}, \{x_1, x_2\}\}$	$x_1^c \cup x_2$, $x_1 \not\neq x_2$ Converse nondifference of сó-events Обратная антиразность сó-бытий	$x_1 \to x_2$ "$x_1 \supset x_2$" Direct implication Прямая импликация
	$\{\emptyset, \{x_1\}, \{x_2\}, \{x_1, x_2\}\}$	Ω Sure, or certain сó-event Достоверное сó-бытие	True, Tautology Истина, Тавтология

Рисунок 6: Диаграммы Венна (1 столбец), генераторы $\mathcal{O}$-операций (2 столбец), 16 бинарных теоретико-множественных операций (3 столбец) и соответствующие им бинарные логические функции (4 столбец) над дуплетом сó-бытий $\{x_1, x_2\}$. Логические обозначения, "вступающие в противоречие с теоретико-множественными", выделены красным.

Определение 3 ***(упорядоченное сет-произведение как множество упорядоченных множеств).*** Для n множеств $\mathfrak{X}_1, \mathfrak{X}_2, \ldots, \mathfrak{X}_n$ их *упорядоченное сет-произведение* определяется как множество

$$\bigtimes_{i=1}^{n} \mathfrak{X}_i = \mathfrak{X}_1 \otimes \ldots \otimes \mathfrak{X}_n =$$

$$= \Big\{\{x_1, \ldots, x_n\} : x_1 \in \mathfrak{X}_1, \ldots, x_n \in \mathfrak{X}_n\Big\},$$

состоящее из *упорядоченных n-множеств*

$$\{x_1, \ldots, x_n\} \in \bigtimes_{i=1}^{n} \mathfrak{X}_i,$$

которые содержат точно по одному элементу x_i из каждого множества $\mathfrak{X}_i$ $(1 \leqslant i \leqslant n)$. Упорядоченное сет-произведение пустой совокупности множеств полагается равным

$$\bigtimes_{i \in \emptyset} \mathfrak{X}_i = \{\emptyset\}$$

Определение 4 ***(прямое произведение как множество упорядоченных мультимножеств).*** Для n множеств $\mathfrak{X}_1, \mathfrak{X}_2, \ldots, \mathfrak{X}_n$ их *прямое произведение* определяется как множество

$$\prod_{i=1}^{n} \mathfrak{X}_i = \mathfrak{X}_1 \cdot \ldots \cdot \mathfrak{X}_n,$$

состоящее из *упорядоченных n-мультимножеств*

$$\{x_1^{m_1}, \ldots, x_n^{m_n}\} \in \prod_{i=1}^{n} \mathfrak{X}_i,$$

которые содержат точно по m_i элементов x_i из каждого множества $\mathfrak{X}_i$ $(1 \leqslant i \leqslant n)$, где $m_i = 0, \ldots, n,\ i = 1, \ldots, n$ и

$$\sum_{i=1}^{n} m_i = n.$$

Прямое произведение пустой совокупности множеств полагается равным

$$\prod_{i \in \emptyset} \mathfrak{X}_i = \{\emptyset\}.$$

Лемма 1 ***(мощность сет-произведений и прямого произведения непересекающихся множеств).*** *Для n непересекающихся конечных множеств $\mathfrak{X}_1, \mathfrak{X}_2, \ldots, \mathfrak{X}_n$ мощности их сет-произведения, упорядоченного сет-произведения и прямого произведения совпадают и равны:*

$$\left|\bigotimes_{i=1}^{n} \mathfrak{X}_i\right| = \left|\bigtimes_{i=1}^{n} \mathfrak{X}_i\right| = \left|\prod_{i=1}^{n} \mathfrak{X}_i\right| = \prod_{i=1}^{n} |\mathfrak{X}_i|.$$

Доказательство очевидно.

Лемма 2 *(мощность сет-степеней и прямой степени конечного множества). Для N-множества $\mathfrak{X}$ его сет-степень*

$$\mathfrak{X}^{\otimes n} = \underbrace{\mathfrak{X} \otimes \ldots \otimes \mathfrak{X}}_{n},$$

упорядоченная сет-степень

$$\mathfrak{X}^{\times n} = \underbrace{\mathfrak{X} \times \ldots \times \mathfrak{X}}_{n}$$

и прямая степень n-го порядка, $n = 0, 1, \ldots, N$,

$$\mathfrak{X}^{\cdot n} = \underbrace{\mathfrak{X} \cdot \ldots \cdot \mathfrak{X}}_{n}$$

имеют следующие мощности:

$$\left|\mathfrak{X}^{\otimes n}\right| = C_N^n = \frac{N!}{n!(N-n)!},$$

$$\left|\mathfrak{X}^{\times n}\right| = A_N^n = \frac{N!}{(N-n)!},$$

$$|\mathfrak{X}^{\cdot n}| = N^n.$$

Доказательство комбинаторное.

Лемма 3 *(террасное представление произведений множеств). Пусть $\mathfrak{S} = \{\mathfrak{X}_1, \ldots, \mathfrak{X}_n\}$ — n-множество конечных множеств. Тогда*

$$\bigotimes_{\mathfrak{X}\in\mathfrak{S}} \mathfrak{X} = \bigcup_{\substack{\mathfrak{x}_1\in S_1\subseteq\mathfrak{S} \\ \mathfrak{x}_n\in\ddot{S}_n\subseteq\mathfrak{S}}} \left(\bigotimes_{i=1}^{n} \mathsf{ter}(S_i /\!\!/ \mathfrak{S})\right),$$

$$\bigtimes_{i=1}^{n} \mathfrak{X}_i = \bigcup_{\substack{\mathfrak{x}_1\in S_1\subseteq\mathfrak{S} \\ \mathfrak{x}_n\in\ddot{S}_n\subseteq\mathfrak{S}}} \left(\bigtimes_{i=1}^{n} \mathsf{ter}(S_i /\!\!/ \mathfrak{S})\right),$$

$$\prod_{i=1}^{n} \mathfrak{X}_i = \sum_{\substack{\mathfrak{x}_1\in S_1\subseteq\mathfrak{S} \\ \mathfrak{x}_n\in\ddot{S}_n\subseteq\mathfrak{S}}} \left(\prod_{i=1}^{n} \mathsf{ter}(S_i /\!\!/ \mathfrak{S})\right).$$

Доказательство.

$$\bigotimes_{\mathfrak{X}\in\mathfrak{S}} \mathfrak{X} = \bigotimes_{\mathfrak{X}\in\mathfrak{S}} \left(\sum_{\mathfrak{X}\in S\subseteq\mathfrak{S}} \mathsf{ter}(S /\!\!/ \mathfrak{S})\right) =$$

$$= \bigcup_{\substack{\mathfrak{x}_1\in S_1\subseteq\mathfrak{S} \\ \mathfrak{x}_n\in\ddot{S}_n\subseteq\mathfrak{S}}} \left(\bigotimes_{i=1}^{n} \mathsf{ter}(S_i /\!\!/ \mathfrak{S})\right),$$

$$\bigtimes_{i=1}^{n} \mathfrak{X} = \bigtimes_{i=1}^{n} \left(\sum_{\mathfrak{X}\in S\subseteq\mathfrak{S}} \mathsf{ter}(S /\!\!/ \mathfrak{S})\right) =$$

$$= \bigcup_{\substack{\mathfrak{x}_1\in S_1\subseteq\mathfrak{S} \\ \mathfrak{x}_n\in\ddot{S}_n\subseteq\mathfrak{S}}} \left(\bigtimes_{i=1}^{n} \mathsf{ter}(S_i /\!\!/ \mathfrak{S})\right),$$

$$\prod_{i=1}^{n} \mathfrak{X}_i = \prod_{i=1}^{n} \left(\sum_{\mathfrak{X}_i\in S\subseteq\mathfrak{S}} \mathsf{ter}(S /\!\!/ \mathfrak{S})\right) =$$

$$= \sum_{\substack{\mathfrak{x}_1\in S_1\subseteq\mathfrak{S} \\ \mathfrak{x}_n\in\ddot{S}_n\subseteq\mathfrak{S}}} \left(\prod_{i=1}^{n} \mathsf{ter}(S_i /\!\!/ \mathfrak{S})\right).$$

Лемма 4 *(мощность сет-произведения двух множеств). Для двух произвольных конечных множеств $\mathfrak{X}_1, \mathfrak{X}_2$ мощности их сет-произведения и упорядоченного сет-произведения равны:*

$$|\mathfrak{X}_1 \otimes \mathfrak{X}_2| = |\mathfrak{X}_1| \cdot |\mathfrak{X}_2| - \frac{1}{2}|\mathfrak{X}_1 \cap \mathfrak{X}_2|^2 - \frac{1}{2}|\mathfrak{X}_1 \cap \mathfrak{X}_2|,$$

$$|\mathfrak{X}_1 \times \mathfrak{X}_2| = |\mathfrak{X}_1| \cdot |\mathfrak{X}_2| - |\mathfrak{X}_1 \cap \mathfrak{X}_2|.$$

Доказательство. Рассмотрим "террасное" доказательство. Пусть

$$\mathfrak{S} = \{\mathfrak{X}_1, \mathfrak{X}_2\}$$

— дуплет произвольных конечных подмножеств некоторого базового множества $\mathcal{A}$: $\mathfrak{X}_1 \subset \mathcal{A}$, $\mathfrak{X}_2 \subset \mathcal{A}$. Тогда

$$\mathcal{A} = \sum_{S\subseteq\mathfrak{S}} \mathsf{ter}(S /\!\!/ \mathfrak{S}),$$

где

$$\mathsf{ter}(S /\!\!/ \mathfrak{S}) = \bigcap_{X\in S} X \bigcap_{X\in\mathfrak{S}-S} (\mathcal{A} - X)$$

— *S-террасное множество 1-го рода*[10], порожденное дуплетом $\mathfrak{S}$.

Заметим, что

$$\mathfrak{X}_1 = \sum_{\mathfrak{X}_1\in S\subseteq\mathfrak{S}} \mathsf{ter}(S /\!\!/ \mathfrak{S}) =$$

$$= \mathsf{ter}(\{\mathfrak{X}_1\} /\!\!/ \mathfrak{S}) + \mathsf{ter}(\{\mathfrak{X}_1, \mathfrak{X}_2\} /\!\!/ \mathfrak{S}),$$

$$\mathfrak{X}_2 = \sum_{\mathfrak{X}_2\in S\subseteq\mathfrak{S}} \mathsf{ter}(S /\!\!/ \mathfrak{S}) =$$

$$= \mathsf{ter}(\{\mathfrak{X}_2\} /\!\!/ \mathfrak{S}) + \mathsf{ter}(\{\mathfrak{X}_1, \mathfrak{X}_2\} /\!\!/ \mathfrak{S}).$$

[10] Если базовое множество $\mathcal{A}$ конечно, то все террасные множества 1-го рода конечны. Если же $\mathcal{A}$ бесконечно, то среди S-террасных множеств, порожденных совокупностью $\mathfrak{S}$ конечных множеств, только одно $\emptyset$-террасное множество $\mathsf{ter}(\emptyset /\!\!/ \mathfrak{S})$ бесконечно. Поэтому при $S \neq \emptyset$ всегда определена мощность $|\mathsf{ter}(S /\!\!/ \mathfrak{S})|$, равная числу элементов в S-террасном множестве $\mathsf{ter}(S /\!\!/ \mathfrak{S})$. Террасные множества 1-го рода, порожденные множеством $\mathfrak{S}$, не пересекаются между собой:

$$\mathsf{ter}(S /\!\!/ \mathfrak{S}) \cap \mathsf{ter}(S' /\!\!/ \mathfrak{S}) = \begin{cases} \mathsf{ter}(S /\!\!/ \mathfrak{S}), & S = S', \\ \emptyset, & \text{иначе.} \end{cases}$$

Часть 1. Сет-произведение. Имеем

$$\mathfrak{X}_1 \otimes \mathfrak{X}_2 =$$
$$= \left(\sum_{\mathfrak{X}_1 \in S \subseteq \mathfrak{S}} \mathsf{ter}(S/\!\!/\mathfrak{S})\right) \otimes \left(\sum_{\mathfrak{X}_2 \in S \subseteq \mathfrak{S}} \mathsf{ter}(S/\!\!/\mathfrak{S})\right) =$$
$$= \mathsf{ter}(\{\mathfrak{X}_1, \mathfrak{X}_2\}/\!\!/\mathfrak{S})^{\otimes 2} +$$
$$+ \mathsf{ter}(\{\mathfrak{X}_1\}/\!\!/\mathfrak{S}) \otimes \mathsf{ter}(\{\mathfrak{X}_2\}/\!\!/\mathfrak{S}) +$$
$$+ \mathsf{ter}(\{\mathfrak{X}_1\}/\!\!/\mathfrak{S}) \otimes \mathsf{ter}(\{\mathfrak{X}_1, \mathfrak{X}_2\}/\!\!/\mathfrak{S}) +$$
$$+ \mathsf{ter}(\{\mathfrak{X}_1, \mathfrak{X}_2\}/\!\!/\mathfrak{S}) \otimes \mathsf{ter}(\{\mathfrak{X}_2\}/\!\!/\mathfrak{S}).$$

Отсюда — аналогичное равенство для мощностей этих террасных множеств:

$$|\mathfrak{X}_1 \otimes \mathfrak{X}_2| = \left|\mathtt{ter}_{12}^{\otimes 2}\right| +$$
$$+ |\mathtt{ter}_1 \otimes \mathtt{ter}_2| + |\mathtt{ter}_1 \otimes \mathtt{ter}_{12}| + |\mathtt{ter}_2 \otimes \mathtt{ter}_{12}|.$$

В силу лемм 1 и 2 получаем то, что требуется:

$$|\mathfrak{X}_1 \otimes \mathfrak{X}_2| = C^2_{\mathtt{t}_{12}} +$$
$$+ \mathtt{t}_1 \cdot \mathtt{t}_2 + \mathtt{t}_1 \cdot \mathtt{t}_{12} + \mathtt{t}_2 \cdot \mathtt{t}_{12} =$$
$$= \mathtt{n}_1 \cdot \mathtt{n}_2 - \frac{1}{2}\mathtt{n}_{12}^2 - \frac{1}{2}\mathtt{n}_{12}.$$

Часть 2. Упорядоченное сет-произведение. Имеем

$$\mathfrak{X}_1 \times \mathfrak{X}_2 =$$
$$= \left(\sum_{\mathfrak{X}_1 \in S \subseteq \mathfrak{S}} \mathsf{ter}(S/\!\!/\mathfrak{S})\right) \times \left(\sum_{\mathfrak{X}_2 \in S \subseteq \mathfrak{S}} \mathsf{ter}(S/\!\!/\mathfrak{S})\right) =$$
$$= \mathsf{ter}(\{\mathfrak{X}_1, \mathfrak{X}_2\}/\!\!/\mathfrak{S})^{\times 2} +$$
$$+ \mathsf{ter}(\{\mathfrak{X}_1\}/\!\!/\mathfrak{S}) \times \mathsf{ter}(\{\mathfrak{X}_2\}/\!\!/\mathfrak{S}) +$$
$$+ \mathsf{ter}(\{\mathfrak{X}_1\}/\!\!/\mathfrak{S}) \times \mathsf{ter}(\{\mathfrak{X}_1, \mathfrak{X}_2\}/\!\!/\mathfrak{S}) +$$
$$+ \mathsf{ter}(\{\mathfrak{X}_1, \mathfrak{X}_2\}/\!\!/\mathfrak{S}) \otimes \mathsf{ter}(\{\mathfrak{X}_2\}/\!\!/\mathfrak{S}).$$

Отсюда — аналогичное равенство для мощностей этих террасных множеств:

$$|\mathfrak{X}_1 \otimes \mathfrak{X}_2| = \left|\mathtt{ter}_{12}^{\otimes 2}\right| +$$
$$+ |\mathtt{ter}_1 \otimes \mathtt{ter}_2| + |\mathtt{ter}_1 \otimes \mathtt{ter}_{12}| + |\mathtt{ter}_2 \otimes \mathtt{ter}_{12}|.$$

В силу лемм 1 и 2 получаем то, что требуется:

$$|\mathfrak{X}_1 \otimes \mathfrak{X}_2| = C^2_{\mathtt{t}_{12}} +$$
$$+ \mathtt{t}_1 \cdot \mathtt{t}_2 + \mathtt{t}_1 \cdot \mathtt{t}_{12} + \mathtt{t}_2 \cdot \mathtt{t}_{12} =$$
$$= \mathtt{n}_1 \cdot \mathtt{n}_2 - \frac{1}{2}\mathtt{n}_{12}^2 - \frac{1}{2}\mathtt{n}_{12}.$$

Следствие 1 ***(неравенство для мощностей сет-произведений пары конечных множеств).***

$$|\mathfrak{X}_1 \otimes \mathfrak{X}_2| \leqslant |\mathfrak{X}_1 \times \mathfrak{X}_2|.$$

Лемма 5 ***(мощность произведений трех множеств).*** *Для трех произвольных конечных множеств $\mathfrak{X}_1, \mathfrak{X}_2, \mathfrak{X}_3$ мощности их сет-произведения, упорядоченного сет-произведения и прямого произведения соответственно равны:*

$$|\mathfrak{X}_1 \otimes \mathfrak{X}_2 \otimes \mathfrak{X}_3| = C^3_{\mathtt{n}_{123}} +$$
$$+ C^2_{\mathtt{n}_{123}}(\mathtt{t}_{12} + \mathtt{t}_{13} + \mathtt{t}_{23} + \mathtt{t}_1 + \mathtt{t}_2 + \mathtt{t}_3) +$$
$$+ \mathtt{n}_1 C^2_{\mathtt{n}_{23}} + \mathtt{n}_2 C^2_{\mathtt{n}_{13}} + \mathtt{n}_3 C^2_{\mathtt{n}_{12}} +$$
$$+ \mathtt{t}_{123}\mathtt{t}_{12}(\mathtt{t}_{13} + \mathtt{t}_1) +$$
$$+ \mathtt{t}_{123}\mathtt{t}_{12}(\mathtt{t}_{12} + \mathtt{t}_{13} + \mathtt{t}_1) +$$
$$+ \mathtt{t}_{123}\mathtt{t}_2(\mathtt{t}_{12} + \mathtt{t}_{13} + \mathtt{t}_1) +$$
$$+ \mathtt{t}_{12}\mathtt{t}_{23}(\mathtt{t}_{13} + \mathtt{t}_1) +$$
$$+ \mathtt{t}_{23}\mathtt{t}_2\mathtt{n}_1 + \mathtt{t}_{23}\mathtt{t}_1\mathtt{n}_2 + \mathtt{t}_{12}\mathtt{t}_{23}\mathtt{t}_2 +$$
$$+ \mathtt{t}_{123}\mathtt{t}_3(\mathtt{t}_{12} + \mathtt{t}_{13} + \mathtt{t}_1) +$$
$$+ \mathtt{t}_{12}\mathtt{t}_3(\mathtt{t}_{13} + \mathtt{t}_1) +$$
$$+ \mathtt{t}_{23}\mathtt{t}_3\mathtt{n}_1 + \mathtt{t}_2\mathtt{t}_3\mathtt{n}_1. \tag{3.1}$$

$$|\mathfrak{X}_1 \times \mathfrak{X}_2 \times \mathfrak{X}_3| = A^3_{\mathtt{n}_{123}} +$$
$$+ A^2_{\mathtt{n}_{123}}(\mathtt{t}_{12} + \mathtt{t}_{13} + \mathtt{t}_{23} + \mathtt{t}_1 + \mathtt{t}_2 + \mathtt{t}_3) +$$
$$+ \mathtt{n}_1 A^2_{\mathtt{n}_{23}} + \mathtt{n}_2 A^2_{\mathtt{n}_{13}} + \mathtt{n}_3 A^2_{\mathtt{n}_{12}} +$$
$$+ \mathtt{t}_{123}\mathtt{t}_{12}(\mathtt{t}_{13} + \mathtt{t}_1) +$$
$$+ \mathtt{t}_{123}\mathtt{t}_{12}(\mathtt{t}_{12} + \mathtt{t}_{13} + \mathtt{t}_1) +$$
$$+ \mathtt{t}_{123}\mathtt{t}_2(\mathtt{t}_{12} + \mathtt{t}_{13} + \mathtt{t}_1) +$$
$$+ \mathtt{t}_{12}\mathtt{t}_{23}(\mathtt{t}_{13} + \mathtt{t}_1) +$$
$$+ \mathtt{t}_{23}\mathtt{t}_2\mathtt{n}_1 + \mathtt{t}_{23}\mathtt{t}_1\mathtt{n}_2 + \mathtt{t}_{12}\mathtt{t}_{23}\mathtt{t}_2 +$$
$$+ \mathtt{t}_{123}\mathtt{t}_3(\mathtt{t}_{12} + \mathtt{t}_{13} + \mathtt{t}_1) +$$
$$+ \mathtt{t}_{12}\mathtt{t}_3(\mathtt{t}_{13} + \mathtt{t}_1) +$$
$$+ \mathtt{t}_{23}\mathtt{t}_3\mathtt{n}_1 + \mathtt{t}_2\mathtt{t}_3\mathtt{n}, \tag{3.2}$$

$$|\mathfrak{X}_1 \otimes \mathfrak{X}_2 \otimes \mathfrak{X}_3| = \mathtt{n}_1\mathtt{n}_2\mathtt{n}_3. \tag{3.3}$$

Доказательство. Рассмотрим как и в лемме 4 “террасное” доказательство с элементами комбинаторики, которое отличается от доказательства леммы 4 для дуплета со́-бытий исключительно тем, что для триплета со́-бытий начинают появляться такие сет-произведения *различных* террасных множеств, которые состоят из одних и тех же 3-множеств, т.е., совпадают. Поэтому в помощь доказательству триплетной ситуации использован диаграммный перебор всех 64 сет-произведений террасных событий (см. рис. 7 и 8).

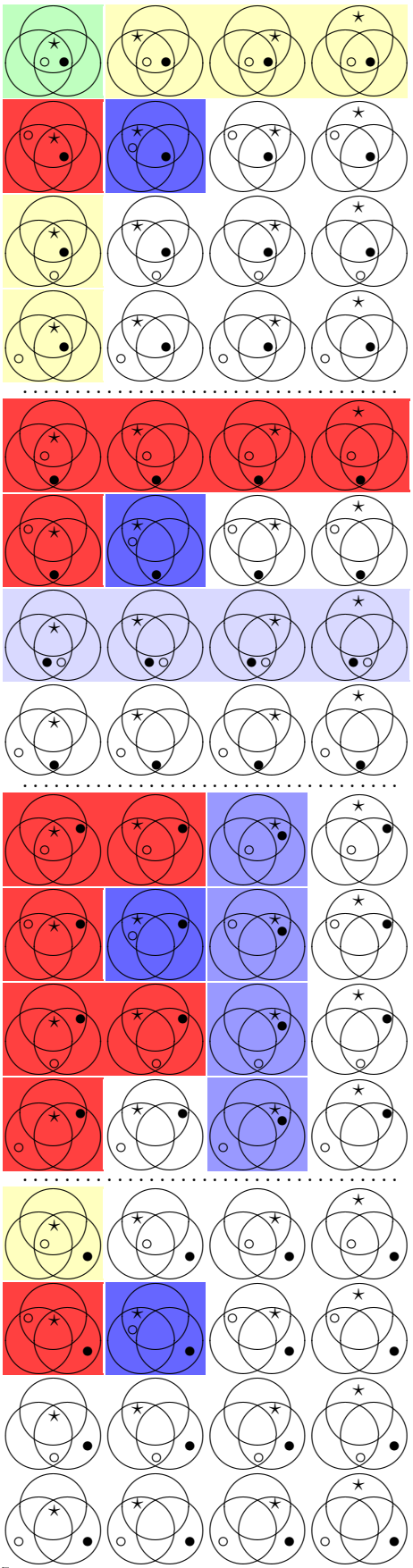


Рисунок 7: Диаграммное перечисление 64 слагаемых сет-произведения триплета $\mathfrak{S} = \{\mathfrak{X}_1, \mathfrak{X}_2, \mathfrak{X}_3\}$, составленного из 3-множеств $\{\star, \bullet, \circ\}$, содержащих по одному элементу из каждого множества триплета: $\star \in \mathfrak{X}_1$, $\bullet \in \mathfrak{X}_2$ и $\circ \in \mathfrak{X}_3$, выбранных из террасных множеств $\mathbf{ter}(S/\!/\mathfrak{S})$, $S \neq \emptyset$, порожденных $\mathfrak{S}$. 51 слагаемое из 64 не пересекаются между собой; 13 "дублей" выделены красным цветом.

$C^3_{t_{123}}$	$C^2_{t_{123}t_{12}}$	$C^2_{t_{123}t_{13}}$	$C^2_{t_{123}t_1}$
$C^2_{t_{123}t_{12}}$	$C^2_{t_{12}t_{123}}$	$t_{123}t_{12}t_{13}$	$t_{123}t_{12}t_1$
$C^2_{t_{123}t_{23}}$	$t_{123}t_{23}t_{12}$	$t_{123}t_{23}t_{13}$	$t_{123}t_{23}t_1$
$C^2_{t_{123}t_2}$	$t_{123}t_2t_{12}$	$t_{123}t_2t_{13}$	$t_{123}t_2t_1$
$C^2_{t_{123}t_{23}}$	$t_{123}t_{23}t_{12}$	$t_{123}t_{23}t_{13}$	$t_{123}t_{23}t_1$
$t_{12}t_{23}t_{123}$	$C^2_{t_{12}t_{23}}$	$t_{12}t_{23}t_{13}$	$t_{12}t_{23}t_1$
$C^2_{t_{12}t_{123}}$	$C^2_{t_{12}t_{12}}$	$C^2_{t_{12}t_{13}}$	$C^2_{t_{12}t_1}$
$t_1t_{23}t_{123}$	$t_1t_{23}t_{12}$	$t_1t_{23}t_{13}$	$t_1t_{23}t_1$
$C^2_{t_{123}t_{13}}$	$t_{13}t_{123}t_{12}$	$C^2_{t_{23}t_{123}}$	$t_{13}t_{123}t_1$
$t_{123}t_{13}t_{12}$	$C^2_{t_{12}t_{13}}$	$C^2_{t_{23}t_{12}}$	$t_{12}t_{23}$
$t_{13}t_{23}t_{123}$	$t_{13}t_{23}t_{12}$	$C^2_{t_{13}t_{23}}$	$t_{13}t_{23}t_1$
$t_{13}t_2t_{123}$	$t_{13}t_2t_{12}$	$C^2_{t_{13}t_2}$	$t_{13}t_2t_1$
$C^2_{t_{123}t_3}$	$t_{123}t_3t_{12}$	$t_{123}t_3t_{13}$	$t_{123}t_3t_1$
$t_{12}t_3t_{123}$	$C^2_{t_{12}t_3}$	$t_{12}t_3t_{13}$	$t_{12}t_3t_1$
$t_{23}t_3t_{123}$	$t_{23}t_3t_{12}$	$t_{23}t_3t_{13}$	$t_{23}t_3t_1$
$t_2t_3t_{123}$	$t_2t_3t_{12}$	$t_2t_3t_{13}$	$t_2t_3t_1$

Рисунок 8: Мощности 64 слагаемых сет-произведения триплета $\mathfrak{S} = \{\mathfrak{X}_1, \mathfrak{X}_2, \mathfrak{X}_3\}$, показанного на рис. 7, среди которых 13 "дублей" выделены красным цветом.

Пусть

$$\mathfrak{S} = \{\mathfrak{X}_1, \mathfrak{X}_2, \mathfrak{X}_3\}$$

— триплет произвольных конечных подмножеств некоторого базового множества $\mathcal{A}$: $\mathfrak{X}_1 \subset \mathcal{A}$, $\mathfrak{X}_2 \subset \mathcal{A}$, $\mathfrak{X}_3 \subset \mathcal{A}$. Тогда

$$\mathcal{A} = \sum_{S \subseteq \mathfrak{S}} \mathsf{ter}(S /\!\!/ \mathfrak{S}),$$

где

$$\mathsf{ter}(S /\!\!/ \mathfrak{S}) = \bigcap_{X \in S} X \bigcap_{X \in \mathfrak{S} - S} (\mathcal{A} - X)$$

— *S-террасное множество 1-го рода*[11], порожденное триплетом $\mathfrak{S}$.

Заметим, что

$$\mathfrak{X}_1 = \sum_{\mathfrak{X}_1 \in S \subseteq \mathfrak{S}} \mathsf{ter}(S /\!\!/ \mathfrak{S}) =$$

$$= \mathsf{ter}(\{\mathfrak{X}_1\} /\!\!/ \mathfrak{S}) + \mathsf{ter}(\{\mathfrak{X}_1, \mathfrak{X}_2\} /\!\!/ \mathfrak{S}) +$$

$$+ \mathsf{ter}(\{\mathfrak{X}_1, \mathfrak{X}_3\} /\!\!/ \mathfrak{S}) + \mathsf{ter}(\{\mathfrak{X}_1, \mathfrak{X}_2, \mathfrak{X}_3\} /\!\!/ \mathfrak{S}) =$$

$$= \mathsf{ter}_1 + \mathsf{ter}_{12} + \mathsf{ter}_{13} + \mathsf{ter}_{123},$$

$$\mathfrak{X}_2 = \sum_{\mathfrak{X}_2 \in S \subseteq \mathfrak{S}} \mathsf{ter}(S /\!\!/ \mathfrak{S}) =$$

$$= \mathsf{ter}(\{\mathfrak{X}_2\} /\!\!/ \mathfrak{S}) + \mathsf{ter}(\{\mathfrak{X}_1, \mathfrak{X}_2\} /\!\!/ \mathfrak{S}) +$$

$$+ \mathsf{ter}(\{\mathfrak{X}_2, \mathfrak{X}_3\} /\!\!/ \mathfrak{S}) + \mathsf{ter}(\{\mathfrak{X}_1, \mathfrak{X}_2, \mathfrak{X}_3\} /\!\!/ \mathfrak{S}) =$$

$$= \mathsf{ter}_2 + \mathsf{ter}_{12} + \mathsf{ter}_{23} + \mathsf{ter}_{123},$$

$$\mathfrak{X}_3 = \sum_{\mathfrak{X}_3 \in S \subseteq \mathfrak{S}} \mathsf{ter}(S /\!\!/ \mathfrak{S}) =$$

$$= \mathsf{ter}(\{\mathfrak{X}_3\} /\!\!/ \mathfrak{S}) + \mathsf{ter}(\{\mathfrak{X}_1, \mathfrak{X}_3\} /\!\!/ \mathfrak{S}) +$$

$$+ \mathsf{ter}(\{\mathfrak{X}_2, \mathfrak{X}_3\} /\!\!/ \mathfrak{S}) + \mathsf{ter}(\{\mathfrak{X}_1, \mathfrak{X}_2, \mathfrak{X}_3\} /\!\!/ \mathfrak{S}) =$$

$$= \mathsf{ter}_3 + \mathsf{ter}_{13} + \mathsf{ter}_{23} + \mathsf{ter}_{123}.$$

`1. Мощность сет-произведения.` Имеем

$$\mathfrak{X}_1 \otimes \mathfrak{X}_2 \otimes \mathfrak{X}_3 = \bigotimes_{\mathfrak{X} \in \mathfrak{S}} \mathfrak{X} = \bigotimes_{\mathfrak{X} \in \mathfrak{S}} \left(\sum_{\mathfrak{X} \in S \subseteq \mathfrak{S}} \mathsf{ter}(S /\!\!/ \mathfrak{S}) \right) =$$

$$= (\mathsf{ter}_{123})^{\otimes 3} +$$

$$+ (\mathsf{ter}_1 + \mathsf{ter}_2 + \mathsf{ter}_3 + \mathsf{ter}_{12} + \mathsf{ter}_{13} + \mathsf{ter}_{23}) \otimes (\mathsf{ter}_{123})^{\otimes 2} +$$

$$+ \mathfrak{X}_1 \otimes (\mathsf{ter}_{23})^{\otimes 2} + \mathfrak{X}_2 \otimes (\mathsf{ter}_{13})^{\otimes 2} + \mathfrak{X}_3 \otimes (\mathsf{ter}_{12})^{\otimes 2} +$$

$$+ \mathsf{ter}_{123} \otimes \mathsf{ter}_{12} \otimes (\mathsf{ter}_{13} + \mathsf{ter}_1) +$$

$$+ \mathsf{ter}_{123} \otimes \mathsf{ter}_{12} \otimes (\mathsf{ter}_{12} + \mathsf{ter}_{13} + \mathsf{ter}_1) +$$

$$+ \mathsf{ter}_{123} \otimes \mathsf{ter}_2 \otimes (\mathsf{ter}_{12} + \mathsf{ter}_{13} + \mathsf{ter}_1) +$$

$$+ \mathsf{ter}_{12} \otimes \mathsf{ter}_{23} \otimes (\mathsf{ter}_{13} + \mathsf{ter}_1) +$$

$$+ \mathsf{ter}_{23} \otimes \mathsf{ter}_2 \otimes \mathfrak{X}_1 +$$

$$+ \mathsf{ter}_{23} \otimes \mathsf{ter}_1 \otimes \mathfrak{X}_2 +$$

$$+ \mathsf{ter}_{12} \otimes \mathsf{ter}_{23} \otimes \mathsf{ter}_2 +$$

$$+ \mathsf{ter}_{123} \otimes \mathsf{ter}_3 \otimes (\mathsf{ter}_{12} + \mathsf{ter}_{13} + \mathsf{ter}_1) +$$

$$+ \mathsf{ter}_{12} \otimes \mathsf{ter}_3 \otimes (\mathsf{ter}_{13} + \mathsf{ter}_1) +$$

$$+ \mathsf{ter}_{23} \otimes \mathsf{ter}_3 \otimes \mathfrak{X}_1 +$$

$$+ \mathsf{ter}_2 \otimes \mathsf{ter}_3 \otimes \mathfrak{X}_1.$$

Отсюда следует утверждение леммы для мощности сет-произведения триплета со́-бытий.

`2. Мощность упорядоченного сет-произведения.` Доказательство отличается от только что проведенного только тем, что в силу леммы 3 мощность упорядоченной сет-степени n-го порядка множества $\mathfrak{X}$ равна не числу сочетаний: $C^n_{|\mathfrak{X}|}$, а числу размещений: $|\mathfrak{X}^{\times n}| = A^n_{|\mathfrak{X}|}$.

`3. Мощность прямого произведения.` Доказательство очевидно.

4 Перспективы

Определение со́-бытия как имени К-события, под которым понимается некоторое высказывание, характеризующее К-событие, подчеркивает очевидную связь теории со́-бытий с математической логикой. В следующих работах предполагается подробно исследовать свойства введенного в эвентологическую теорию понятия со́-бытия, его связь с понятием логического высказывания; и роль, которую играют в этой связи, новые эвентологические аксиомы со́-бытия и множества со́-бытий. Также предполагается провести сравнение теории со́-бытий и логики высказываний на языке террасного формализма.

Список литературы

[1] О. Ю. Воробьев. *Эвентология*. Сибирский федеральный университет, Красноярск, 2007, 435с., https://www.academia.edu/179393/Vorobyev_O.Yu._Eventology._-_Krasnoyarsk_Siberian_Federal_University._-_2007._-_435p.

[2] 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.

[3] A. Levy. *Basic Set Theory*. Springer-Verlag, 1979.

[4] O. Yu. Vorobyev. On elements of axiomatizing eventology. *Journal of Siberian Federal University. Mathematics & Physics*, 3(2):157–164, 2010.

[11] Если базовое множество $\mathcal{A}$ конечно, то все террасные множества 1-го рода конечны. Если же $\mathcal{A}$ бесконечно, то среди S-террасных множеств, порожденных совокупностью $\mathfrak{S}$ конечных множеств, только одно $\emptyset$-террасное множество $\mathsf{ter}(\emptyset /\!\!/ \mathfrak{S})$ бесконечно. Поэтому при $S \neq \emptyset$ всегда определена мощность $|\mathsf{ter}(S /\!\!/ \mathfrak{S})|$, равная числу элементов в S-террасном множестве $\mathsf{ter}(S /\!\!/ \mathfrak{S})$.

Эвентологическое обобщение границ Фреше для множества со́-бытий

Олег Юрьевич Воробьев
Институт математики и фундаментальной информатики
Сибирский федеральный университет
Красноярск
email: oleg.yu.vorobyev@gmail.com
url: sfu-kras.academia.edu/OlegVorobyev

Аннотация. *Предложен эвентологический взгляд на обобщение понятия границ Фреше для вероятности произвольной теоретико-множественной операции над конечным множеством со́-бытий.*
Ключевые слова. *Вероятность, многомерная статистика, событие, со́-бытие, эвентология, Фреше-граница, произвольная теоретико-множественная операция, слой, чердак, подвал.*

Морис Фреше [11, 1935] усовершенствовал понятие границ для вероятностей объединения и пересечения зависимых множеств событий, чтобы развить результаты, которые к тому времени уже были получены Булем [10, 1854] и Бонферрони [9, 1936]; позднее Хёффдинг [21, 22] и Фреше [14, 1951] определили верхние и нижние границы совместной функции распределения двух случайных величин при фиксированных маргинальных распределениях[1].

Математическая эвентология, как наука о событиях, с момента своего возникновения использовала понятия границ Фреше для вероятностей различных теоретико-множественных операций над множествами событий [2, 2007], которые получили развитие в работах [3, 2010], [4, 2013], [8, 2014]. В работе Новоселова [5, 2014] предложен функциональный подход к определению границ Фреше для вероятности произвольной теоретико-множественной операции[2] над множеством событий.

В этой работе представлен эвентологический взгляд на обобщение понятия границ Фреше для вероятности произвольной теоретико-множественной операции над конечным множеством со́-бытий, основанный на последних продвижениях в обосновании эвентологического метода [6, 2014].

© 2014 О.Ю.Воробьев
Олег Воробьев (ред.), Труды XIII ФАМЭМС'2014, Красноярск: СФУ

[1] См. в дополнение др. работы Фреше по этой теме: [12, 13, 15, 16, 17]; а также работы Рушендорфа [24], Хейльперина [18, 19, 20], Венециани [25] и Прекопы [23], в которых не совсем успешно, на мой взгляд, сделаны попытки определить границы вероятностей некоторых логических функций от событий, используя для этой цели методы линейного программирования.

[2] Террасное определение произвольной теоретико-множественной операции над конечным множеством см. в [1, 1993; стр. 8–14, 102].

1 Фреше-многогранник множества со́-бытий

Обозначим

$$\boldsymbol{p}^{[1]} = \boldsymbol{p}^{[1]}(\mathfrak{X}) = \{p(X/\!\!/\mathfrak{X}) : X \subseteq \mathfrak{X}\} \qquad (1.1)$$

— произвольное *частично упорядоченное множество*[3] из 2^N неотрицательных чисел $p(X/\!\!/\mathfrak{X}) \geqslant 0$, которые "занумерованы" подмножествами *неупорядоченного* N-множества $\mathfrak{X} = \{x_1, ..., x_N\}$ и сумма которых равна единице: $\sum_{X\subseteq\mathfrak{X}} p(X/\!\!/\mathfrak{X}) = 1$. Совокупность всех таких чу-множеств чисел обозначается

$$\mathcal{S}_{\mathfrak{X}}^{\left(2^N\right)} = \left\{ \boldsymbol{p}^{[1]} : p(X/\!\!/\mathfrak{X}) \geqslant 0, \sum_{X\subseteq\mathfrak{X}} p(X/\!\!/\mathfrak{X}) = 1 \right\} \qquad (1.2)$$

и называется 2^N-вершинным *сет-симплексом* [6], порожденным (*неупорядоченным*) N-множеством $\mathfrak{X}$.

N-множество

$$\check{p} = \check{p}_{\mathfrak{X}} = \{p_x = \mathbf{P}(x), x \in \mathfrak{X}\} \in [0,1]^{\otimes\mathfrak{X}},$$

которое служит элементом $\mathfrak{X}$-*степени (сет-степени)*[4] единичного отрезка, называется $\mathfrak{X}$-*вектором (сет-вектором)* чисел из $[0,1]$.

Определение 1 ***(Фреше-многогранник конечного множества).*** *Фреше-многогранником N-множества $\mathfrak{X}$* называется подмножество $\mathcal{S}_{\mathfrak{X}}(\check{p})$ сет-симплекса $\mathcal{S}_{\mathfrak{X}}^{\left(2^N\right)}$, определяемое $\mathfrak{X}$-вектором $\check{p}$ по формуле:

$$\mathcal{S}_{\mathfrak{X}}(\check{p}) = \left\{ \boldsymbol{p}^{[1]} \in \mathcal{S}_{\mathfrak{X}}^{\left(2^N\right)} : \sum_{x\in X} p(X) = p_x, x \in \mathfrak{X} \right\}. \qquad (1.3)$$

Пусть теперь $(\Omega, \mathcal{A}^{\mho}, \mathbf{P})$ — *всеобщее вероятностное пространство*, $(\Omega, \mathcal{A}, \mathbf{P})$ — некоторое *частное вероятностное пространство*, а $\mathfrak{X} \subseteq \mathcal{A}$ — (*неупорядоченное*) N-множество *со́-бытий*[5].

[3] Для термина *частично упорядоченное множество* можно использовать сокращение *чу-множество*, или *по-множество* по аналогии с английским сокращением *poset* для *partially ordered set*.

[4] Определение *сет-степени* и *сет-произведения* множеств см. в [6].

[5] Определения *всеобщего* и *частного вероятностного пространства*, а также понятия *со́-бытия* см. в [6].

Определение 2 ***(Фреше-граничные распределения множества со́-бытий)*** [3]. Вероятностные распределения 1-го рода

$$\dot{\breve{\boldsymbol{p}}}^{[1]}(\breve{p}) = \dot{\breve{\boldsymbol{p}}}^{[1]}(\breve{p};\mathfrak{X}) \in \mathcal{S}_{\mathfrak{X}}^{\left(2^N\right)} \tag{1.4}$$

террасной размерности (террасной мощности)[6] от 0 до N множества со́-бытий $\mathfrak{X}$ с фиксированным $\mathfrak{X}$-вектором вероятностей со́-бытий

$$\breve{p} = \{p_x, x \in \mathfrak{X}\} \in [0,1]^{\otimes \mathfrak{X}},$$

которые являются решениями следующей системы из $(N+1)$-го линейного уравнения:

$$\left\{ \begin{array}{rcl} \sum\limits_{X \subseteq \mathfrak{X}} p(X) & = & 1, \\ \sum\limits_{x \in X} p(X) & = & p_x, \quad x \in \mathfrak{X}, \end{array} \right. \tag{1.5}$$

называются *крайними Фреше-граничными распределениями* данного множества со́-бытий $\mathfrak{X}$. Все прочие решения из $\mathcal{S}_{\mathfrak{X}}^{\left(2^N\right)}$ системы (1.5), которые имеют террасные размерности от $N+1$ до 2^N-2 называются просто *Фреше-граничными распределениями* множества со́-бытий $\mathfrak{X}$ и обозначаются $\bar{\boldsymbol{p}}^{[1]}(\breve{p}) = \bar{\boldsymbol{p}}^{[1]}(\breve{p};\mathfrak{X}) \in \mathcal{S}_{\mathfrak{X}}^{\left(2^N\right)}$. Обозначим

$$\dot{\breve{\mathsf{L}}}^{[1]}(\breve{p};\mathfrak{X}) = \left\{ \sum_{X \subseteq \mathfrak{X}} \dot{\breve{p}}(\breve{p};X) = 1,\ 0 \leqslant \dot{\breve{p}}(\breve{p};X) \leqslant 1,\ X \subseteq \mathfrak{X} \right\}$$

ограничения на $\mathfrak{X}$-вектор $\breve{p}$, которые обеспечивают *легитимность (существование)*[7] крайнему Фреше-граничному распределению (1.4)

$$\dot{\breve{\boldsymbol{p}}}^{[1]}(\breve{p}) = \left\{\dot{\breve{p}}(\breve{p};X), X \subseteq \mathfrak{X}\right\},$$

а также

$$\dot{\breve{\boldsymbol{p}}}^{[1]}_1(\breve{p};\mathfrak{X}),\ \ldots,\ \dot{\breve{\boldsymbol{p}}}^{[1]}_M(\breve{p};\mathfrak{X}), \tag{1.6}$$

$$\dot{\breve{\mathsf{L}}}^{[1]}_1(\breve{p};\mathfrak{X}),\ \ldots,\ \dot{\breve{\mathsf{L}}}^{[1]}_M(\breve{p};\mathfrak{X}) \tag{1.7}$$

— все $M = M(\breve{p})$ крайних Фреше-граничных распределений м.с. $\mathfrak{X}$ с $\mathfrak{X}$-вектором $\breve{p} = \breve{p}_{\mathfrak{X}}$ и соответствующих M ограничений на $\breve{p}$, обеспечивающих легитимность (существование) этим распределениям, т.е. принадлежность сет-симплексу $\mathcal{S}_{\mathfrak{X}}^{\left(2^N\right)}$.

Определение 3 ***(множество крайних Фреше-граничных распределений множества со́-бытий)***. *Множество крайних Фреше-граничных распределений множества со́-бытий* $\mathfrak{X}$ определяется тремя эквивалентными способами:

$$\dot{\breve{\mathcal{S}}}_{\mathfrak{X}}(\breve{p}) = \left\{ \dot{\breve{\boldsymbol{p}}}^{[1]}(\breve{p}) \ :\ \dot{\breve{\boldsymbol{p}}}^{[1]}(\breve{p}) \in \mathcal{S}_{\mathfrak{X}}^{\left(2^N\right)} \right\} =$$

$$= \left\{ \dot{\breve{\boldsymbol{p}}}^{[1]}(\breve{p}) \ :\ \dot{\breve{\mathsf{L}}}^{[1]}(\breve{p};\mathfrak{X}) \right\} =$$

$$= \left\{ \dot{\breve{\boldsymbol{p}}}^{[1]}_1(\breve{p}),\ \ldots,\ \dot{\breve{\boldsymbol{p}}}^{[1]}_M(\breve{p}) \right\}, \tag{1.9}$$

где $M = M(\breve{p})$ — мощность этого множества, которая зависит от $\mathfrak{X}$-вектора $\breve{p} = \breve{p}_{\mathfrak{X}}$.

Теорема 1 ***(о Фреше-многограннике распределений множества со́-бытий)*** [3]. *Фреше-многогранник* $\mathcal{S}_{\mathfrak{X}}(\breve{p})$ *распределений* N*-множества со́-бытий* $\mathfrak{X}$ *с фиксированным* $\mathfrak{X}$*-вектором* $\breve{p}_{\mathfrak{X}}$ *вероятностей со́-бытий совпадает с выпуклой оболочкой всех крайних Фреше-граничных распределений* $\mathfrak{X}$*:*

$$\mathcal{S}_{\mathfrak{X}}(\breve{p}) = \mathsf{Conv}\left(\dot{\breve{\mathcal{S}}}_{\mathfrak{X}}(\breve{p})\right). \tag{1.8}$$

Доказательство (см. в [3]) основано на процедуре "зануления" всех избыточных[8] вероятностей, перебирающей не все, а лишь такие подозрительные граничные точки сет-симплекса, среди компонент которых не более, чем $N+1$ компонента[9] отлична от нуля. Такие граничные точки сет-симплекса $S_{\mathfrak{X}}^{\left(2^N\right)}$ принадлежат его границам размерности от нуля до N. Однако данный сет-симплекс, разумеется, имеет границы и бóльшей размерности — вплоть до размерности 2^N-2, на единицу меньшей размерности самого сет-симплекса: 2^N-1.

Теорема утверждает, что в сет-симплексе $\mathcal{S}_{\mathfrak{X}}^{\left(2^N\right)}$ *граничные точки* размерности бóльшей N, которые удовлетворяют указанным линейным ограничениям, принадлежат выпуклой оболочке *крайних граничных точек* размерности от 0 до N, которые удовлетворяют тем же ограничениям. Крайние Фреше-граничные распределения уже построены изложенным способом для дуплета и триплета со́-бытий и перечислены в [3, стр. 59, 61]: 4 крайних Фреше-граничных распределения — для дуплета и 70 — для триплета.

1.1 Пример крайних Фреше-граничных распределений для дуплета со́-бытий

Рассмотрим на примере[10] дуплета со́-бытий $\mathfrak{X} = \{x_1, x_2\}$ способ построения Фреше-граничных распределений — крайних точек Фреше-многогранника $\mathcal{S}_{\mathfrak{X}}(\breve{p})$, определяемого линейными ограничениями (1.5) на вероятности p_x со́-бытий $x \in \mathfrak{X}$, т.е. некоторого многогранного подмножества сет-симплекса $\mathcal{S}_{\mathfrak{X}}^{\left(2^N\right)}$ при $N = 2$.

[6] *Террасной размерностью (террасной мощностью)* распределения 1-го рода множества со́-бытий $\mathfrak{X}$ называется количество **без единицы** ненулевых вероятностей в чу-множестве вероятностей 1-го рода $\boldsymbol{p}^{[1]}(\mathfrak{X}) = \{p(X/\!/\mathfrak{X}) : X \subseteq \mathfrak{X}\}$.

[7] Вероятностное распределение м.с. называется *легитимным*, если оно существует, т.е. все его вероятности из $[0,1]$, а их сумма равна 1.

[8] В рассматриваемых системах линейных уравнений число уравнений избыточно в сравнении с числом неизвестных.

[9] $N = |\mathfrak{X}|$.

[10] Взят из [3].

Обозначим для краткости

$$p(1) = p(\{x_1\}), p(2) = p(\{x_2\}),$$
$$p(12) = p(\{x_1, x_2\})$$

— вероятности I-го рода дуплета со́-событий $\mathfrak{X} = \{x_1, x_2\}$. Тогда линейные ограничения глобальной нормировки и вероятностей со́-событий (1.4) запишутся в виде системы 3-х линейных уравнений с 4-мя неизвестными:

$$\begin{aligned} p(\emptyset) + p(12) + p(2) + p(1) &= 1, \\ p(12) + p(2) \qquad\quad &= p_2, \\ p(12) \qquad\quad + p(1) &= p_1. \end{aligned}$$

X	$\dot{\breve{p}}_1(X)$	$\dot{\breve{p}}_2(X)$	$\dot{\breve{p}}_3(X)$	$\dot{\breve{p}}_4(X)$
$\emptyset$	$1-S$	$1-p_1$	$1-p_2$	0
$\{x_1\}$	p_1	p_1-p_2	0	$1-p_2$
$\{x_2\}$	p_2	0	p_2-p_1	$1-p_1$
$\{x_1, x_2\}$	0	p_2	p_1	$S-1$
$\dot{\mathsf{L}}(\breve{p})$	$\dot{\mathsf{L}}_1(\breve{p}):$ $\{S \leqslant 1\}$	$\dot{\mathsf{L}}_2(\breve{p}):$ $\{p_2 \leqslant p_1\}$	$\dot{\mathsf{L}}_3(\breve{p}):$ $\{p_1 \leqslant p_2\}$	$\dot{\mathsf{L}}_4(\breve{p}):$ $\{1 \leqslant S\}$

Таблица 1: Четыре крайние Фреше-граничные распределения дуплета со́-событий $\mathfrak{X} = \{x_1, x_2\}$ при фиксированных вероятностях со́-событий, образующих $\mathfrak{X}$-вектор $\breve{p} = \{p_1, p_2\}$; а также соответствующие условия их легитимности (существования) $\dot{\mathsf{L}}_m(\breve{p}), m = 1, ..., 4$, где $S = p_1 + p_2$.

Способ построения крайних Фреше-граничных распределений, предложенный в [3], заключается в переборе всех $C_4^1 = 4$ вариантов "зануления" одного из четырех неизвестных и решения получающейся системы 3-х уравнений с 3-мя оставшимися неизвестными. Например, если положить равными нулю следующее неизвестное:

$$\dot{\breve{p}}(12) = 0,$$

то решением получившейся системы будет крайнее Фреше-граничное распределение

$$\dot{\breve{p}}(\emptyset) = 1 - p_1 - p_2,\ \dot{\breve{p}}(1) = p_1,\ \dot{\breve{p}}(2) = p_2,$$

расположенное в табл. 1 под номером 1:

$$\dot{\boldsymbol{\breve{p}}}_1^{[1]} = \dot{\boldsymbol{\breve{p}}}_1^{[1]}(\breve{p}) = \{1 - p_1 - p_2,\ p_1,\ p_2,\ 0\}$$

с ограничениями

$$\dot{\mathsf{L}}_1(\breve{p}) = \{p_1 + p_2 \leqslant 1\}.$$

В табл. 1 указаны все 4 крайние Фреше-граничные распределения дуплета со́-событий и соответствующие ограничения на вероятности, при которых эти распределения легитимны (существуют).

Из 4-х крайних Фреше-граничных распределений $\dot{\boldsymbol{\breve{p}}}_1^{[1]}, \dot{\boldsymbol{\breve{p}}}_2^{[1]}, \dot{\boldsymbol{\breve{p}}}_3^{[1]}, \dot{\boldsymbol{\breve{p}}}_4^{[1]}$ дуплета со́-событий $\mathfrak{X} = \{x_1, x_2\}$, приведенных в табл. 1, в зависимости от комбинации выполнения ограничений (см. рис. 1) образуются следующие 4 пары распределений, каждая из которых

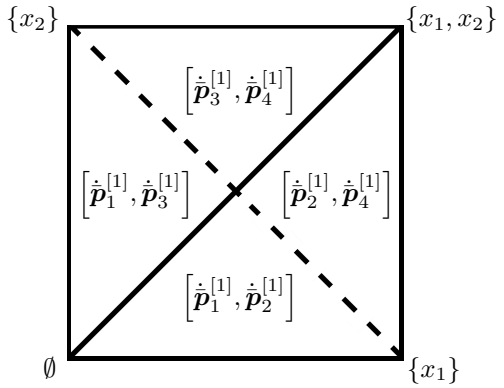


Рисунок 1: Проекция сет-симплекса дуплета со́-событий $\mathcal{S}^{(2^2)}_{\{x_1,x_2\}}$ на единичный квадрат $[0,1]^2$, разделенный ограничениями на 4 сектора, в точку $\breve{p} = \{p_1, p_2\}$ каждого из которых проектируется указанные на рисунке Фреше-отрезки распределений.

определяет соответствующий Фреше-многогранник $\mathcal{S}_{\mathfrak{X}}(\breve{p})$, который в случае дуплета $\mathfrak{X}$ представляет собой Фреше-отрезок распределений с фиксированными вероятностями со́-событий, составляющими $\mathfrak{X}$-вектор $\breve{p} = \{p_1, p_2\}$:

$$\mathcal{S}_{\mathfrak{X}}(\breve{p}) = \mathsf{Conv}\left\{\dot{\boldsymbol{\breve{p}}}^{[1]}, \dot{\breve{\boldsymbol{q}}}^{[1]}\right\} =$$

$$= \begin{cases} \left[\dot{\boldsymbol{\breve{p}}}_1^{[1]}, \dot{\boldsymbol{\breve{p}}}_2^{[1]}\right], & \text{при } \dot{\mathsf{L}}_1(\breve{p}) \wedge \dot{\mathsf{L}}_2(\breve{p}) = \{S \leqslant 1, p_2 \leqslant p_1\}, \\ \left[\dot{\boldsymbol{\breve{p}}}_1^{[1]}, \dot{\boldsymbol{\breve{p}}}_3^{[1]}\right], & \text{при } \dot{\mathsf{L}}_1(\breve{p}) \wedge \dot{\mathsf{L}}_3(\breve{p}) = \{S \leqslant 1, p_1 \leqslant p_2\}, \\ \left[\dot{\boldsymbol{\breve{p}}}_2^{[1]}, \dot{\boldsymbol{\breve{p}}}_4^{[1]}\right], & \text{при } \dot{\mathsf{L}}_2(\breve{p}) \wedge \dot{\mathsf{L}}_4(\breve{p}) = \{1 \leqslant S, p_2 \leqslant p_1\}, \\ \left[\dot{\boldsymbol{\breve{p}}}_3^{[1]}, \dot{\boldsymbol{\breve{p}}}_4^{[1]}\right], & \text{при } \dot{\mathsf{L}}_3(\breve{p}) \wedge \dot{\mathsf{L}}_4(\breve{p}) = \{1 \leqslant S, p_1 \leqslant p_2\}. \end{cases}$$

2 Фреше-границы для произвольной операции над множеством со́-событий

Итак, пусть $(\Omega, \mathcal{A}^{\mho}, \mathbf{P})$ — всеобщее вероятностное пространство, $(\Omega, \mathcal{A}, \mathbf{P})$ — некоторое частное вероятностное пространство, а $\mathfrak{A} \subset \mathcal{A}$ — произвольное N-*множество со́-бытий (м.с.)* $a \in \mathfrak{A}$, происходящих с вероятностями

$$p_a = \mathbf{P}(a),$$

которые образуют $\mathfrak{X}$-вектор

$$\breve{p}_{\mathfrak{A}} = \{p_a, a \in \mathfrak{A}\} \in [0,1]^{\otimes \mathfrak{A}}.$$

Тогда для любого со́-бытия $a \in \mathfrak{A}$ определено м.с. $\mathbb{X}_a(\breve{p}_{\mathfrak{A}}) \subset \mathcal{A}$, такое, что каждое со́-бытие $x_a \in \mathbb{X}_a(\breve{p}_{\mathfrak{A}})$ происходит с той же вероятностью, что и соответствующее ему со́-бытие $a \in \mathfrak{A}$, т.е.

$$\mathbf{P}(x_a) = p_a.$$

Для N-м.с. $\mathfrak{A} \subset \mathcal{A}$ определено *сет-произведение* [6] множеств со́-бытий

$$\mathbb{X}^{\times}(\breve{p}_{\mathfrak{A}}) = \bigotimes_{a \in \mathfrak{A}} \mathbb{X}_a(\breve{p}_{\mathfrak{A}}) \subset 2^{\mathcal{A}}, \qquad (2.1)$$

элементами которого служат N-множества со́-бытий

$$\mathfrak{X} = \{x_a, a \in \mathfrak{A}\} \in \mathbb{X}^{\times}(\breve{p}_{\mathfrak{A}}),$$

составленные из со́-бытий, которые происходят с теми же вероятностями, что и соответствующие им со́-бытия $a \in \mathfrak{A}$. Разумеется,

$$\mathfrak{A} \in \mathbb{X}^{\times}(\breve{p}_{\mathfrak{A}})$$

по определению.

Теорема 2 ***(Фреше-многогранник как образ сет-произведения множеств со́-бытий).*** *Пусть для каждого N-множества со́-бытий $\mathfrak{X} \in \mathbb{X}^{\times}(\breve{p})$*

$$\boldsymbol{p}^{[1]}(\mathfrak{X}) = \{p(X/\!\!/\mathfrak{X}), X \subseteq \mathfrak{X}\}$$

— его вероятностное распределение 1-го рода, а

$$\boldsymbol{p}^{[1]}\left(\mathbb{X}^{\times}(\breve{p})\right) = \{\boldsymbol{p}^{[1]}(\mathfrak{X}), \mathfrak{X} \in \mathbb{X}^{\times}(\breve{p})\}$$

— образ сет-произведения при отображении

$$\boldsymbol{p}^{[1]} : \mathbb{X}^{\times}(\breve{p}) \to \mathcal{S}_{\mathfrak{X}}^{\left(2^N\right)}$$

Тогда

$$\boldsymbol{p}^{[1]}\left(\mathbb{X}^{\times}(\breve{p})\right) = \mathcal{S}_{\mathfrak{X}}(\breve{p}) \tag{2.2}$$

— образ сет-произведения $\mathbb{X}^{\times}(\breve{p})$ при отображении $\boldsymbol{p}^{[1]}$ совпадает с Фреше-многогранником $\mathcal{S}_{\mathfrak{X}}(\breve{p})$ множества со́-бытий $\mathfrak{X}$ при любом $\mathfrak{X} \in \mathbb{X}^{\times}(\breve{p})$.

Доказательство следует из *аксиомы множества со́-бытий* (см. [6]).

Определение 4 ***(теоретико-множественная $\mathcal{O}$-операция над N-множеством со́-бытий)*** [1]. Для каждого подмножества $\mathcal{O} \subseteq 2^{\mathfrak{X}}$ *N-арная (N-валентная) теоретико-множественная операция над N-множеством со́-бытий $\mathfrak{X}$ ($\mathcal{O}$-операция)* определяется как

$$\underset{x\in\mathfrak{X}}{\mathcal{O}} x = \sum_{X\in\mathcal{O}} \mathsf{ter}(X/\!\!/\mathfrak{X}) \tag{2.3}$$

— сумма террасных со́-бытий 1-го рода

$$\mathsf{ter}(X/\!\!/\mathfrak{X}) = \bigcap_{x\in X} x \bigcap_{x\in\mathfrak{X}-X} x^c,$$

"занумерованных" подмножествами $X \in \mathcal{O}$. Подмножество $\mathcal{O} \subseteq 2^{\mathfrak{X}}$ называется *(террасным) генератором*[11] $\mathcal{O}$-операции, а мощность этого подмножества $0 \leqslant |\mathcal{O}| \leqslant 2^N$ — *террасной размерностью (террасной мощностью)* $\mathcal{O}$-операции.

Теорема 3 ***(вероятность $\mathcal{O}$-операции над N-множеством со́-бытий)***

$$\mathbf{P}\left(\underset{x\in\mathfrak{X}}{\mathcal{O}} x\right) = \sum_{X\in\mathcal{O}} p^{[1]}(X/\!\!/\mathfrak{X}). \tag{2.4}$$

[11] Термин *генератор* для подмножеств $\mathcal{O} \subseteq 2^{\mathfrak{X}}$, порождающих произвольную теоретико-множественную операцию над $\mathfrak{X}$, предложен в [5].

Доказательство очевидно.

Теорема 4 ***(произвольная теоретико-множественная операция над N-множеством со́-бытий)*** [1]. Для любой N-арной теоретико-множественной операции $A(\mathfrak{X}) \in \mathcal{A}$ над N-множеством со́-бытий $\mathfrak{X}$ найдется такой террасный генератор $\mathcal{O} \subseteq 2^{\mathfrak{X}}$, что

$$A(\mathfrak{X}) = \underset{x\in\mathfrak{X}}{\mathcal{O}} x.$$

Доказательство см. в [1].

Определение 5 ***(Фреше-границы вероятности $\mathcal{O}$-операции над N-множеством со́-бытий).*** *Нижней и верхней Фреше-границей вероятностей*

$$P_{\mathcal{O}_{\mathfrak{X}}}(\breve{p}) = \mathbf{P}\left(\underset{x\in\mathfrak{X}}{\mathcal{O}} x\right)$$

теоретико-множественной $\mathcal{O}$-операции над N-множествами со́-бытий $\mathfrak{X} = \{x_a, a \in \mathfrak{A}\} \in \mathbb{X}^{\times}(\breve{p})$ называются соответственно такие две функции

$$F^{-}_{\mathcal{O}_{\mathfrak{X}}} : [0,1]^{\times\mathfrak{X}} \to [0,1],$$

$$F^{+}_{\mathcal{O}_{\mathfrak{X}}} : [0,1]^{\times\mathfrak{X}} \to [0,1],$$

что

$$F^{-}_{\mathcal{O}_{\mathfrak{X}}}(\breve{p}) \leqslant P_{\mathcal{O}_{\mathfrak{X}}}(\breve{p}) \leqslant F^{+}_{\mathcal{O}_{\mathfrak{X}}}(\breve{p}) \tag{2.5}$$

для любого N-м.с. $\mathfrak{X} \in \mathbb{X}^{\times}(\breve{p})$. Иначе говоря,

$$F^{-}_{\mathcal{O}_{\mathfrak{X}}}(\breve{p}) = \min_{\mathfrak{X}\in\mathbb{X}^{\times}(\breve{p})} P_{\mathcal{O}_{\mathfrak{X}}}(\breve{p}), \tag{2.6}$$

$$F^{+}_{\mathcal{O}_{\mathfrak{X}}}(\breve{p}) = \max_{\mathfrak{X}\in\mathbb{X}^{\times}(\breve{p})} P_{\mathcal{O}_{\mathfrak{X}}}(\breve{p}), \tag{2.7}$$

где $\breve{p} = \breve{p}_{\mathfrak{A}}$.

3 Значения Фреше-границ для вероятностей $\mathcal{O}$-операции как решения задач линейного программирования

Из (2.4) получаем

$$P_{\mathcal{O}_{\mathfrak{X}}}(\breve{p}) = \sum_{X\subseteq\mathfrak{X}} \mathbf{1}_{\mathcal{O}}(X) p^{[1]}(X),$$

где

$$\mathbf{1}_{\mathcal{O}}(X) = \begin{cases} 1, & X \in \mathcal{O}, \\ 0, & \text{иначе}, \end{cases}$$

— индикатор террасного генератора $\mathcal{O} \subseteq 2^{\mathfrak{X}}$.

Из теоремы 2 следует, что

$$\min_{\mathfrak{X}\in\mathbb{X}^{\times}(\breve{p})} \left(\max_{\mathfrak{X}\in\mathbb{X}^{\times}(\breve{p})}\right) P_{\mathcal{O}_{\mathfrak{X}}}(\breve{p}) =$$

$$\min_{\boldsymbol{p}^{[1]}\in\mathcal{S}_{\mathfrak{X}}(\breve{p})} \left(\max_{\boldsymbol{p}^{[1]}\in\mathcal{S}_{\mathfrak{X}}(\breve{p})}\right) \sum_{X\subseteq\mathfrak{X}} \mathbf{1}_{\mathcal{O}}(X) p^{[1]}(X). \tag{3.1}$$

Так что, нижняя и верхняя границы Фреше $F^{-}_{\mathcal{O}}(\breve{p}), F^{+}_{\mathcal{O}}(\breve{p})$ для вероятности произвольной теоретико-множественной операции над множеством со́-событий $\mathfrak{X}$, порожденной террасным генератором $\mathcal{O} \subseteq 2^{\mathfrak{X}}$, являются решениями двух следующих задач линейного программирования

$$\min_{\boldsymbol{p}^{[1]} \in \mathcal{S}_{\mathfrak{X}}(\breve{p})} \left(\max_{\boldsymbol{p}^{[1]} \in \mathcal{S}_{\mathfrak{X}}(\breve{p})} \right) \sum_{X \subseteq \mathfrak{X}} \mathbf{1}_{\mathcal{O}}(X) p^{[1]}(X), \tag{3.2}$$

где экстремум вероятности ищется среди вероятностных распределений 1-го рода из Фреше-многогранника $\mathcal{S}_{\mathfrak{X}}(\breve{p})$.

Обе эти задачи линейного программирования (3.2) эквивалентным образом записываются в привычном виде:

$$\begin{cases} \min\limits_{\boldsymbol{p}^{[1]}} \left(\max\limits_{\boldsymbol{p}^{[1]}} \right) \sum\limits_{X \subseteq \mathfrak{X}} \mathbf{1}_{\mathcal{O}}(X) p^{[1]}(X), \\ \qquad\qquad \text{при условии} \\ \sum\limits_{X \subseteq \mathfrak{X}} p^{[1]}(X) = 1, \\ \sum\limits_{x \in X} p^{[1]}(X) = p_x, \quad x \in \mathfrak{X}, \\ 0 \leq p^{[1]}(X) \leq 1, \quad X \subseteq \mathfrak{X}. \end{cases} \tag{3.3}$$

Нет нужды сомневаться в том, что обе эти задачи линейного программирования относятся к классу *NP*-трудных, так как ищут экстремум линейного функционала, вообще говоря, в $(2^N - N - 1)$-мерном Фреше-многограннике. Для нахождения решения таких задач, т.е. для отыскания Фреше-границ, достаточно *перебрать* M_N крайних Фреше-граничных распределений $\dot{\boldsymbol{p}}^{[1]}_m$ из Фреше-многогранника $\mathcal{S}_{\mathfrak{X}}(\breve{p})$, чтобы выбрать из них те, для которых выполнены ограничения $\dot{\mathsf{L}}_m(\breve{p})$ и на которых вероятность $\mathcal{O}$-операции достигает экстремума:

$$F^{-}_{\mathcal{O}_{\mathfrak{X}}}(\breve{p}) = \min_{\substack{m=1,\ldots,M_N \\ \dot{\mathsf{L}}(\breve{p})}} \sum_{X \subseteq \mathfrak{X}} \mathbf{1}_{\mathcal{O}}(X) \dot{p}^{[1]}_m(X), \tag{3.4}$$

$$F^{+}_{\mathcal{O}_{\mathfrak{X}}}(\breve{p}) = \max_{\substack{m=1,\ldots,M_N \\ \dot{\mathsf{L}}(\breve{p})}} \sum_{X \subseteq \mathfrak{X}} \mathbf{1}_{\mathcal{O}}(X) \dot{p}^{[1]}_m(X). \tag{3.5}$$

В частности, для дуплета со́-событий $\mathfrak{X} = \{x, y\}$ и любого террасного генератора $\mathcal{O} \subseteq 2^{\{x,y\}}$ достаточно *перебрать* $M_2 = 4$ крайних Фреше-граничных распределения из табл. 2, а для триплета со́-событий $\mathfrak{X} = \{x, y, z\}$ и любого террасного генератора $\mathcal{O} \subseteq 2^{\{x,y,z\}}$ достаточно *перебрать* $M_3 = 79$ крайних Фреше-граничных распределения из табл. 3 [3].

3.1 Пример: Фреше-границы для вероятности симметрической разности дуплета со́-событий

Симметрическая разность дуплета со́-событий $\mathfrak{X} = \{x, y\}$ — это частный случай $\mathcal{O}$-операции, когда $\mathcal{O} = \{\{x\}, \{y\}\} \subset 2^{\{x,y\}}$:

$$x \Delta y = \sum_{X \in \mathcal{O}} \mathsf{ter}(X /\!\!/ \{x, y\}) =$$

$$= \mathsf{ter}(\{x\} /\!\!/ \{x, y\}) + \mathsf{ter}(\{y\} /\!\!/ \{x, y\}).$$

X	$\dot{\breve{p}}_1(X)$	$\dot{\breve{p}}_2(X)$	$\dot{\breve{p}}_3(X)$	$\dot{\breve{p}}_4(X)$
$\emptyset$	$1 - S$	$1 - p_x$	$1 - p_y$	0
$\{x\}$	p_x	$p_x - p_y$	0	$1 - p_y$
$\{y\}$	p_y	0	$p_y - p_x$	$1 - p_x$
$\{x, y\}$	0	p_y	p_x	$S - 1$
$\dot{\mathsf{L}}(\breve{p})$	$\dot{\mathsf{L}}_1(\breve{p}) :$ $\{S \leqslant 1\}$	$\dot{\mathsf{L}}_2(\breve{p}) :$ $\{p_y \leqslant p_x\}$	$\dot{\mathsf{L}}_3(\breve{p}) :$ $\{p_x \leqslant p_y\}$	$\dot{\mathsf{L}}_4(\breve{p}) :$ $\{1 \leqslant S\}$

Таблица 2: Четыре крайние Фреше-граничные распределения дуплета со́-событий $\mathfrak{X} = \{x, y\}$ при фиксированных вероятностях со́-событий, образующих $\mathfrak{X}$-вектор $\breve{p} = \{p_x, p_y\}$; а также соответствующие условия их легитимности (существования) $\dot{\mathsf{L}}_m(\breve{p}), m = 1, ..., 4$, где $S = p_x + p_y$.

Перебирая в табл. 2 крайние Фреше-граничные распределения и условия их легитимности (существования) при фиксированном сет-векторе $\breve{p}$, получаем Фреше-границы для вероятности симметрической разности дуплета со́-событий как следующие решения задач линейного программирования:

$$F^{-}_{\{\{x\},\{y\}\}}(\breve{p}) =$$

$$= \begin{cases} \min\{p_x + p_y, p_x - p_y\}, & \dot{\mathsf{L}}_1(\breve{p}) \wedge \dot{\mathsf{L}}_2(\breve{p}), \\ \min\{p_x + p_y, p_y - p_x\}, & \dot{\mathsf{L}}_1(\breve{p}) \wedge \dot{\mathsf{L}}_3(\breve{p}), \\ \min\{2 - p_x - p_y, p_x - p_y\}, & \dot{\mathsf{L}}_2(\breve{p}) \wedge \dot{\mathsf{L}}_4(\breve{p}), \\ \min\{2 - p_x - p_y, p_y - p_x\}, & \dot{\mathsf{L}}_3(\breve{p}) \wedge \dot{\mathsf{L}}_4(\breve{p}) \end{cases} =$$

$$= \begin{cases} \min\{p_x + p_y, |p_x - p_y|\}, & \dot{\mathsf{L}}_1(\breve{p}), \\ \min\{2 - p_x - p_y, |p_x - p_y|\}, & \dot{\mathsf{L}}_4(\breve{p}) \end{cases} =$$

$$= |p_x - p_y|, \tag{3.6}$$

$$F^{+}_{\{\{x\},\{y\}\}}(\breve{p}) =$$

$$= \begin{cases} \max\{p_x + p_y, p_x - p_y\}, & \dot{\mathsf{L}}_1(\breve{p}) \wedge \dot{\mathsf{L}}_2(\breve{p}), \\ \max\{p_x + p_y, p_y - p_x\}, & \dot{\mathsf{L}}_1(\breve{p}) \wedge \dot{\mathsf{L}}_3(\breve{p}), \\ \max\{2 - p_x - p_y, p_x - p_y\}, & \dot{\mathsf{L}}_2(\breve{p}) \wedge \dot{\mathsf{L}}_4(\breve{p}), \\ \max\{2 - p_x - p_y, p_y - p_x\}, & \dot{\mathsf{L}}_3(\breve{p}) \wedge \dot{\mathsf{L}}_4(\breve{p}) \end{cases} =$$

$$= \begin{cases} \max\{p_x + p_y, |p_x - p_y|\}, & \dot{\mathsf{L}}_1(\breve{p}), \\ \max\{2 - p_x - p_y, |p_x - p_y|\}, & \dot{\mathsf{L}}_4(\breve{p}) \end{cases} =$$

$$= \begin{cases} p_x + p_y, & \dot{\mathsf{L}}_1(\breve{p}), \\ 2 - p_x - p_y, & \dot{\mathsf{L}}_4(\breve{p}) \end{cases} =$$

$$= \min\{p_x + p_y, 2 - p_x - p_y\} =$$

$$= 1 - |1 - p_x - p_y|, \tag{3.7}$$

так как

$$p_x + p_y \geq |p_x - p_y|, \quad 2 - p_x - p_y \geq |p_x - p_y|.$$

Таблица 3: 79 крайних Фреше-граничных распределений триплета событий $\mathfrak{X} = \{x_1, x_2, x_3\}$ [3].

п/п	**1**	**2^2 71^2**	**3**	**4**	**5**	**6^2 72^2**	**7**
$\emptyset$	$1 - S$	$1 - S_{23}$, 0	$1 - S_{13}$	$1 - S_{12}$	$1 - S_{23}$	$1 - S_{13}$, 0	$1 - S_{12}$
3	p_3	p_3, $1 - p_2$	$p_3 - p_2$	0	$p_3 - p_1$	p_3, $1 - p_1$	0
2	p_2	p_2, $1 - p_3$	0	$p_2 - p_3$	p_2	0, 0	p_2
1	p_1	0, 0	p_1	p_1	0	p_1, $1 - p_3$	$p_1 - p_3$
23	0	0, $S_{23} - 1$	p_2	p_3	0	0, 0	0
13	0	0, 0	0	0	p_1	0, $S_{13} - 1$	p_3
12	0	0, 0	0	0	0	0, 0	0
123	0	0, 0	0	0	0	0, 0	0
$\dot{\mathbf{L}}(\breve{p})$	$S \leqslant 1$	$p_1 = 0$, $p_1 = 0$ $S_{23} \leqslant 1$, $1 \leqslant S_{23}$	$p_2 \leqslant p_3$ $S_{13} \leqslant 1$	$p_3 \leqslant p_2$ $S_{12} \leqslant 1$	$p_1 \leqslant p_3$ $S_{23} \leqslant 1$	$p_2 = 0$, $p_2 = 0$ $S_{13} \leqslant 1$, $1 \leqslant S_{13}$	$p_3 \leqslant p_1$ $S_{12} \leqslant 1$
п/п	**8**	**9**	**10**	**11**	**12**	**13^2 73^2**	**14**
$\emptyset$	$1 - p_3$	$1 - S_{12}$	$1 - S_{12}$	$1 - S_{23}$	$1 - S_{13}$	$1 - S_{12}$, 0	$1 - S_{13}$
3	$p_3 - S_{12}$	0	0	p_3	p_3	0, 0	$S_{13} - p_2$
2	0	$S_{12} - p_3$	0	$p_2 - p_1$	0	p_2, $1 - p_1$	0
1	0	0	$S_{12} - p_3$	0	$p_1 - p_2$	p_1, $1 - p_2$	0
23	p_2	$p_3 - p_1$	p_2	0	0	0, 0	$p_2 - p_1$
13	p_1	p_1	$p_3 - p_2$	0	0	0, 0	0
12	0	0	0	p_1	p_2	0, $S_{12} - 1$	p_1
123	0	0	0	0	0	0, 0	0
$\dot{\mathbf{L}}(\breve{p})$	$S_{12} \leqslant p_3$	$p_1 \leqslant p_3 \leqslant S_{12} \leqslant 1$	$p_2 \leqslant p_3 \leqslant S_{12} \leqslant 1$	$S_{23} \leqslant 1$ $p_1 \leqslant p_2$	$S_{13} \leqslant 1$ $p_2 \leqslant p_1$	$p_3 = 0$, $p_3 = 0$ $S_{12} \leqslant 1$, $1 \leqslant S_{12}$	$p_1 \leqslant p_2 \leqslant S_{13} \leqslant 1$
п/п	**15**	**16**	**17**	**18**	**19**	**20**	**21**
$\emptyset$	$1 - p_2$	$1 - S_{13}$	$1 - S_{23}$	$1 - S_{23}$	$1 - p_1$	$S/2$	$1 + p_1 - S_{23}$
3	0	0	$S_{23} - p_1$	0	0	0	$p_3 - p_1$
2	$p_2 - S_{13}$	0	0	$S_{23} - p_1$	0	0	$p_2 - p_1$
1	0	$S_{13} - p_2$	0	0	$p_1 - S_{23}$	0	0
23	p_3	p_3	0	0	0	$(S_{23} - p_1)/2$	0
13	0	0	$p_1 - p_2$	p_3	p_3	$(S_{13} - p_2)/2$	0
12	p_1	$p_2 - p_3$	p_2	$p_1 - p_3$	p_2	$(S_{12} - p_3)/2$	0
123	0	0	0	0	0	0	p_1
$\dot{\mathbf{L}}(\breve{p})$	$S_{13} \leqslant p_2$	$p_3 \leqslant p_2 \leqslant S_{13} \leqslant 1$	$p_2 \leqslant p_1 \leqslant S_{23} \leqslant 1$	$p_3 \leqslant p_1 \leqslant S_{23} \leqslant 1$	$S_{23} \leqslant p_1$	$p_1 \leqslant S_{23}$ $p_2 \leqslant S_{13}$ $p_3 \leqslant S_{12}$	$S_{23} \leqslant 1 + p_1$ $p_1 \leqslant p_3, p_1 \leqslant p_2$
п/п	**22**	**23**	**24**	**25**	**26^2 74^2**	**27**	**28^2 75^2**
$\emptyset$	$1 + p_2 - S_{13}$	$1 + p_3 - S_{12}$	$1 - p_3$	$1 - p_2$	$1 - S_{12}$, 0	$1 - p_3$	0, $1 - S_{23}$
3	$p_3 - p_2$	0	$p_3 - p_2$	0	0, 0	$p_3 - p_1$	0, 0
2	0	$p_2 - p_3$	0	$p_2 - p_3$	0, 0	0	$1 - p_3$, p_2
1	$p_1 - p_2$	$p_1 - p_3$	0	0	p_1, $1 - p_2$	0	0, 0
23	0	0	$p_2 - p_1$	$p_3 - p_1$	p_2, $1 - p_1$	0	0, 0
13	0	0	0	0	0, 0	$p_1 - p_2$	$1 - p_2$, p_3
12	0	0	0	0	0, 0	0	0, 0
123	p_2	p_3	p_1	p_1	0, $S_{12} - 1$	p_2	$S_{23} - 1$, 0
$\dot{\mathbf{L}}(\breve{p})$	$p_2 \leqslant p_1, p_2 \leqslant p_3$ $S_{13} \leqslant 1 + p_2$	$p_3 \leqslant p_1, p_3 \leqslant p_2$ $S_{12} \leqslant 1 + p_3$	$p_1 \leqslant p_2 \leqslant p_3$	$p_1 \leqslant p_3 \leqslant p_2$	$p_2 = p_3, p_2 = p_3$ $S_{12} \leqslant 1$, $1 \leqslant S_{12}$	$p_2 \leqslant p_1 \leqslant p_3$	$p_1 = p_3, p_1 = p_3$ $1 \leqslant S_{23}$, $1 \leqslant S_{23}$
п/п	**29**	**30**	**31^2 76^2**	**32**	**33**	**34**	**35**
$\emptyset$	$1 - p_1$	$1 - p_3$	$1 - S_{23}$, 0	$1 - p_2$	$1 - p_1$	$1 - p_2$	$1 - p_1$
3	0	0	p_3, $1 - p_2$	0	0	0	0
2	0	0	0, 0	$p_2 - p_1$	0	0	0
1	$p_1 - p_3$	0	0, 0	0	$p_1 - p_2$	0	0
23	0	$p_3 - p_1$	0, 0	0	0	$p_2 - p_1$	0
13	$p_3 - p_2$	$p_3 - p_2$	0, 0	0	0	0	$p_1 - p_2$
12	0	0	p_2, $1 - p_3$	$p_1 - p_3$	$p_2 - p_3$	$p_2 - p_3$	$p_1 - p_3$
123	p_2	$S_{12} - p_3$	0, $S_{23} - 1$	p_3	p_3	$S_{13} - p_2$	$S_{23} - p_1$
$\dot{\mathbf{L}}(\breve{p})$	$p_2 \leqslant p_3 \leqslant p_1$	$p_1 \leqslant p_3, p_2 \leqslant p_3$ $p_3 \leqslant S_{12}$	$p_1 = p_2, p_1 = p_2$ $S_{23} \leqslant 1$, $1 \leqslant S_{23}$	$p_3 \leqslant p_1 \leqslant p_2$	$p_3 \leqslant p_2 \leqslant p_1$	$p_3 \leqslant p_2, p_1 \leqslant p_2$ $p_2 \leqslant S_{13}$	$p_2 \leqslant p_1, p_3 \leqslant p_1$ $p_1 \leqslant S_{23}$
п/п	**36**	**37**	**38**	**39**	**40^2 77^2**	**41**	**42**
$\emptyset$	0	0	0	0	0, 0	0	0
3	$1 - S_{12}$	$1 - S_{12}$	$1 - S_{12}$	$1 - S_{12}$	0, 0	p_3	$1 - p_2$
2	$1 - S_{13}$	p_2	$1 - p_3$	0	0, $1 - p_3$	$1 - S_{13}$	$1 - S_{13}$
1	p_1	$1 - S_{23}$	0	$1 - p_3$	$1 - p_3$, 0	$1 - S_{23}$	0
23	$S - 1$	0	$S_{23} - 1$	p_2	p_2, $p_3 - p_1$	0	$S_{23} - 1$
13	0	$S - 1$	p_1	$S_{13} - 1$	$p_3 - p_2$, p_1	0	0
12	0	0	0	0	0, 0	$S - 1$	p_1
123	0	0	0	0	0, 0	0	0
$\dot{\mathbf{L}}(\breve{p})$	$1 \leqslant S \leqslant 2$ $S_{12} \leqslant 1, S_{13} \leqslant 1$	$1 \leqslant S \leqslant 2$ $S_{12} \leqslant 1, S_{23} \leqslant 1$	$1 \leqslant S_{23}$ $S_{12} \leqslant 1$	$S_{12} \leqslant 1, 1 \leqslant S_{13}$	$p_2 \leqslant p_3$, $p_1 \leqslant p_3$ $p_1 + p_2 = 1$	$1 \leqslant S \leqslant 2$ $S_{13} \leqslant 1$, $S_{23} \leqslant 1$	$S_{13} \leqslant 1, 1 \leqslant S_{23}$
п/п	**43^2 78^2**	**44**	**45^2 79^2**	**46**	**47**	**48**	**49**
$\emptyset$	0, 0	0	0, 0	0	0	0	0
3	$1 - p_2$, 0	0	$1 - p_1$, 0	$1 - p_1$	0	$2 - S$	0
2	0, 0	$1 - S_{13}$	0, $1 - p_1$	0	$1 - p_1$	0	$2 - S$
1	0, $1 - p_2$	$1 - p_2$	0, 0	$1 - S_{23}$	$1 - S_{23}$	0	0
23	$p_2 - p_1$, p_3	p_3	0, 0	0	0	$S_{23} - 1$	$S_{23} - 1$
13	0, 0	0	$p_1 - p_2$, p_3	$S_{13} - 1$	p_3	$S_{13} - 1$	$1 - p_2$
12	p_1, $p_2 - p_3$	$S_{12} - 1$	p_2, $p_1 - p_3$	p_2	$S_{12} - 1$	$1 - p_3$	$S_{12} - 1$
123	0, 0	0	0, 0	0	0	0	0
$\dot{\mathbf{L}}(\breve{p})$	$p_1 \leqslant p_2, p_3 \leqslant p_2$ $p_1 + p_3 = 1$	$1 \leqslant S_{12}$ $S_{13} \leqslant 1$	$p_2 \leqslant p_1$, $p_3 \leqslant p_1$ $p_2 + p_3 = 1$	$1 \leqslant S_{13}$ $S_{23} \leqslant 1$ $1 \leqslant S \leqslant 2$	$1 \leqslant S_{12}$ $S_{23} \leqslant 1$ $1 \leqslant S \leqslant 2$	$1 \leqslant S_{13}$ $1 \leqslant S_{23}$ $1 \leqslant S \leqslant 2$	$1 \leqslant S_{12}$ $1 \leqslant S_{23}$ $1 \leqslant S \leqslant 2$
п/п	**50**	**51**	**52**	**53**	**54**	**55**	**56**
$\emptyset$	0	0	0	0	0	0	0
3	0	$(p_3 - S_{12})/2 + 1/2$	$1 - p_2$	$p_3 - p_2$	0	$p_3 - p_1$	$1 - p_1$
2	0	$(p_2 - S_{13})/2 + 1/2$	$1 - p_3$	0	$p_2 - p_3$	$1 - p_3$	0
1	$2 - S$	$(p_1 - S_{23})/2 + 1/2$	0	$1 - p_3$	$1 - p_2$	0	$1 - p_3$
23	$1 - p_1$	0	$S_{23} - p_1 - 1$	$1 + p_2 - S_{13}$	$1 + p_3 - S_{12}$	0	0
13	$S_{13} - 1$	0	0	0	0	$1 + p_1 - S_{23}$	$S_{13} - p_2 - 1$
12	$S_{12} - 1$	0	0	0	0	0	0
123	0	$(S - 1)/2$	p_1	$S_{13} - 1$	$S_{12} - 1$	$S_{23} - 1$	p_2
$\dot{\mathbf{L}}(\breve{p})$	$1 \leqslant S_{13}$ $1 \leqslant S_{12}$ $1 \leqslant S \leqslant 2$	$S_{12} \leqslant p_3$ $S_{13} \leqslant p_2$ $S_{23} \leqslant p_1$ $1 \leqslant S$	$1 + p_1 \leqslant S_{23}$	$1 \leqslant S_{13} \leqslant 1 + p_2$ $p_2 \leqslant p_3$	$1 \leqslant S_{12} \leqslant 1 + p_3$ $p_3 \leqslant p_2$	$1 \leqslant S_{23} \leqslant 1 + p_1$ $p_1 \leqslant p_3$	$1 + p_2 \leqslant S_{13}$
п/п	**57**	**58^2**	**59**	**60**	**61**	**62**	**63**
$\emptyset$	0	0	0	0	0	0	0
3	0	0	0	0	$1 - p_2$	$1 - p_1$	0
2	$1 - p_1$	0	$1 - p_3$	0	$p_2 - p_1$	0	$1 - p_1$
1	$p_1 - p_3$	0	0	$1 - p_3$	0	$p_1 - p_2$	$1 - p_2$
23	0	$1 - p_1$	$p_3 - p_1$	$1 - p_1$	0	0	0
13	$1 + p_3 - S_{12}$	$1 - p_2$	$1 - p_2$	$p_3 - p_2$	0	0	0
12	0	0	0	0	$1 + p_1 - S_{23}$	$1 + p_2 - S_{13}$	$S_{12} - p_3 - 1$
123	$S_{12} - 1$	$S_{12} - 1$	$S_{12} - 1$	$S_{12} - 1$	$S_{23} - 1$	$S_{13} - 1$	p_3
$\dot{\mathbf{L}}(\breve{p})$	$1 \leqslant S_{12} \leqslant 1 + p_3$ $p_3 \leqslant p_1$	$p_3 = 1$ $1 \leqslant S_{12}$	$1 \leqslant S_{12}$ $p_1 \leqslant p_3$	$1 \leqslant S_{12}$ $p_2 \leqslant p_3$	$1 \leqslant S_{23} \leqslant 1 + p_1$ $p_1 \leqslant p_2$	$1 \leqslant S_{13} \leqslant 1 + p_2$ $p_2 \leqslant p_1$	$1 + p_3 \leqslant S_{12}$
п/п	**64**	**65^2**	**66**	**67**	**68**	**69^2**	**70**
$\emptyset$	0	0	0	0	0	0	0
3	$1 - p_2$	0	0	$1 - p_1$	0	0	0
2	0	0	0	0	$1 - p_1$	0	0
1	0	0	$1 - p_2$	0	0	0	0
23	$p_2 - p_1$	$1 - p_1$	$1 - p_1$	0	0	0	$1 - p_1$
13	0	0	0	$p_1 - p_2$	$1 - p_2$	$1 - p_2$	$1 - p_2$
12	$1 - p_3$	$1 - p_3$	$p_2 - p_3$	$1 - p_3$	$p_1 - p_3$	$1 - p_3$	$1 - p_3$
123	$S_{13} - 1$	$S_{13} - 1$	$S_{13} - 1$	$S_{23} - 1$	$S_{23} - 1$	$S_{23} - 1$	$S - 2$
$\dot{\mathbf{L}}(\breve{p})$	$1 \leqslant S_{13}$ $p_1 \leqslant p_2$	$p_2 = 1$ $1 \leqslant S_{13}$	$p_3 \leqslant p_2$ $1 \leqslant S_{13}$	$p_2 \leqslant p_1$ $1 \leqslant S_{23}$	$p_3 \leqslant p_1$ $1 \leqslant S_{23}$	$p_1 = 1$ $1 \leqslant S_{23}$	$2 \leqslant S$

4 Фреше-границы: некоторые обозначения и определения впрок

Пусть $p_x = \mathbf{P}(x), x \in \mathfrak{X}$ — вероятности со́-бытий из N-м.с. $\mathfrak{X}$. Для каждого $n\,(1 \leqslant n \leqslant N)$ обозначим

$$p^{(n)} = p^{(n)}_{x_n} = \mathbf{P}(x_n)$$

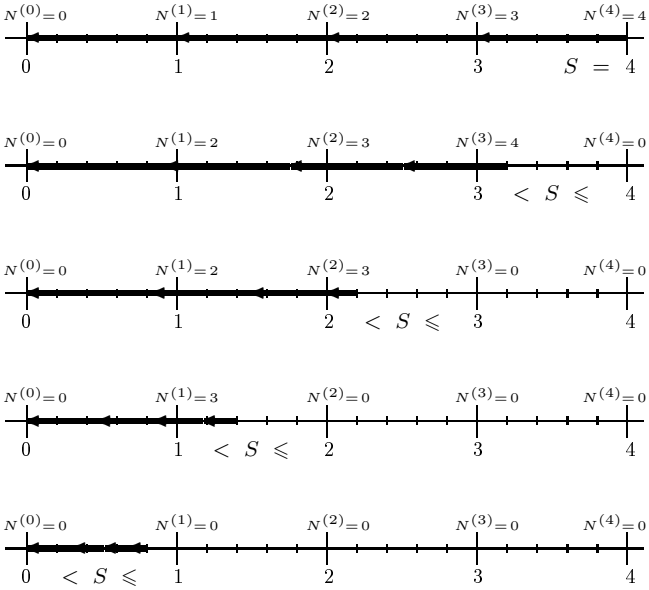
вероятность, которая принимает n-е значение в убывающем ряду значений, принимаемых вероятностями $p_x = \mathbf{P}(x), x \in \mathfrak{X}$. Итак, например, мы имеем

$$p^{(1)} \geqslant p^{(2)} \geqslant \ \ldots \ \geqslant p^{(N)}, \tag{4.1}$$

т.е.,

$$p^{(1)} = \max_{x\in\mathfrak{X}} p_x, \ \ldots, \ p^{(N)} = \min_{x\in\mathfrak{X}} p_x.$$

Упорядоченное множество вероятностей (4.1) называется *(убывающим) вариационным рядом* вероятностей со́-бытий из $\mathfrak{X}$.



$p^{(1)}$	$p^{(2)}$	$p^{(3)}$	$p^{(4)}$	S
1	1	1	1	4
0.92	0.83	0.75	0.70	3.2
0.82	0.65	0.52	0.20	2.2
0.48	0.36	0.33	0.23	1.4
0.30	0.20	0.17	0.13	0.8

Рисунок 2: Иллюстрация обозначений и определений, которые могут пригодиться для вывода Фреше-границ вероятностей "мощностных" теоретико-множественных операций над квартетом событий $\mathfrak{X} = \{x_1, x_2, x_3, x_4\}$ с различной суммой S вероятностей событий; внизу таблица соответствующих вариационных рядов вероятностей событий из $\mathfrak{X}$.

Теперь $\mathfrak{X}$ можно записать в виде упорядоченного множества со́-бытий

$$\mathfrak{X} = \{x_1, ..., x_N\}.$$

Для каждого $n\,(1 \leqslant n \leqslant N)$ обозначим

$$X_n = \{x_1, ..., x_n\} \subseteq \mathfrak{X}$$

n-подмножество первых n со́-бытий из $\mathfrak{X}$, обозначим также

$$X^c_{N-n} = \{x_{N-n+1}, ..., x_N\} \subseteq \mathfrak{X}$$

n-подмножество последних n со́-бытий из $\mathfrak{X}$.

Далее для каждого $n\,(1 \leqslant n \leqslant N)$ обозначим

$$S^{(n)} = \sum_{i=1}^{n} p^{(i)} = \sum_{x\in X_n} p_x$$

сумму первых n вероятностей из вариационного ряда (4.1); а

$$S = S^{(N)} = \sum_{i=1}^{N} p^{(i)} = \sum_{x\in\mathfrak{X}} p_x$$

сумму всех вероятностей со́-бытий из $\mathfrak{X}$; для $n = 0$ положим $S^{(0)} = 0$.

Наконец, пусть $N^{(n)}$ определяется формулой (см. рис. 2)

$$N^{(n)} = \begin{cases} 0, & n = 0, \\ m, & S^{(m-1)} < n \leqslant S^{(m)}, \\ & m = 1, ..., N, \\ 0, & S^{(N)} < n \leqslant N, \end{cases}$$

и пусть

$$p^{*(N^{(n)})} = S^{(N^{(n)})} - n$$

для $0 < n \leqslant S$.

4.1 Некоторые "мощностные" теоретико-множественные операции над множеством со́-бытий

Теоретико-множественная операция $\underset{x\in\mathfrak{X}}{\mathcal{O}} x$ над N-множеством со́-бытий $\mathfrak{X}$ называется "мощностной", если ее генератор $\mathcal{O}$ определяется равенствами или неравенствами для мощностей подмножеств $\mathfrak{X}$. Рассмотрим определения некоторых из этих операций.

• *n-слой* м.с. $\mathfrak{X}$ — это теоретико-множественная операция над со́-бытиями из $\mathfrak{X}$, которая определяется генератором $\mathcal{O} = C^n_{\mathfrak{X}} = \{X \subseteq \mathfrak{X} : |X| = n\} \subset 2^{\mathfrak{X}}$ как событие

$$\underset{x\in\mathfrak{X}}{\mathcal{O}} x = \underset{x\in\mathfrak{X}}{(C^2_{\mathfrak{X}})} x = \sum_{X\in C^n_{\mathfrak{X}}} \mathsf{ter}(X /\!\!/ \mathfrak{X}). \tag{4.2}$$

Обозначим

$$\pi_n = \pi_n(\mathfrak{X}) = \mathbf{P}\left(\underset{x\in\mathfrak{X}}{(C^2_{\mathfrak{X}})} x\right) = \sum_{X\in C^2_{\mathfrak{X}}} p(X /\!\!/ \mathfrak{X})$$

— вероятность n-слоя м.с. $\mathfrak{X}$.

• *n-чердак* множества со́-бытий из $\mathfrak{X}$ — это теоретико-множественная операция над $\mathfrak{X}$, которая определяется как *объединение всех n-пересечений* со́-бытий из

$\mathfrak{X}$ для $n = 0, 1, \ldots, N$. В террасных обозначениях *n-чердак* $\mathfrak{X}$ определяется формулой

$$\overset{n}{\underset{x\in\mathfrak{X}}{\odot}} x = \bigcup_{\substack{|X|=n\\ X\subseteq\mathfrak{X}}} \mathsf{ter}_{X/\!/\mathfrak{X}}, \tag{4.3}$$

где для $X \subseteq \mathfrak{X}$

$$\mathsf{ter}_{X/\!/\mathfrak{X}} = \bigcap_{x\in X} x \tag{4.4}$$

— *X-террасные со́-бытия 2-го рода*. Вероятности n-чердаков, т.е. *вероятности объединений n-пересечений со́-бытий*, имеют специальные обозначения:

$${}^{n}u_{\mathfrak{X}} = \mathbf{P}\left(\overset{n}{\underset{x\in\mathfrak{X}}{\odot}} x\right) = \sum_{\substack{X\subseteq\mathfrak{X}\\ |X|\geqslant n}} p(X/\!/\mathfrak{X}). \tag{4.5}$$

• *n-подвал* множества со́-бытий $\mathfrak{X}$ — это теоретико-множественная операций над $\mathfrak{X}$, которая определяется как *пересечение всех n-объединений* со́-бытий из $\mathfrak{X}$ для $n = 0, 1, \ldots, N$. В террасных обозначениях *n-подвал* $\mathfrak{X}$ определяется формулой

$$\overset{n}{\underset{x\in\mathfrak{X}}{\boxdot}} x = \bigcap_{\substack{|X|=n\\ X\subseteq\mathfrak{X}}} \mathsf{Ter}_{X/\!/\mathfrak{X}}. \tag{4.6}$$

где для $X \subseteq \mathfrak{X}$

$$\mathsf{Ter}_{X/\!/\mathfrak{X}} = \bigcup_{x\in X} x \tag{4.7}$$

— *X-террасное со́-бытие 4-го рода*. Вероятности n-подвалов, т.е. *вероятности пересечений n-объединений со́-бытий* имеют специальные обозначения:

$${}^{n}p_{\mathfrak{X}} = \mathbf{P}\left(\overset{n}{\underset{x\in\mathfrak{X}}{\boxdot}} x\right) = \sum_{\substack{X\subseteq\mathfrak{X}\\ |X|\geqslant N+1-n}} p(X/\!/\mathfrak{X}). \tag{4.8}$$

Таким образом произвольная теоретико-множественная $\mathcal{O}$-операция становится *n-слоем*, *n-чердаком* или *n-подвалом*, когда [7]

$$\mathcal{O} = \{X \subseteq \mathfrak{X} : |X| = n\}, \tag{4.9}$$

$$\mathcal{O} = \{X \subseteq \mathfrak{X} : |X| \geqslant n\}, \tag{4.10}$$

$$\mathcal{O} = \{X \subseteq \mathfrak{X} : |X| \geqslant N + 1 - n\} \tag{4.11}$$

соответственно.

Например, 2-слой триплета со́-бытий $\mathfrak{X} = \{x, y, z\}$:

$$(C^2_{\mathfrak{X}})\,x = \sum_{\substack{X\subseteq\mathfrak{X}\\ |X|=2}} \mathsf{ter}(X/\!/\mathfrak{X}) =$$

$$= \mathsf{ter}(\{x, y\}/\!/\mathfrak{X}) + \mathsf{ter}(\{x, z\}/\!/\mathfrak{X}) + \mathsf{ter}(\{y, z\}/\!/\mathfrak{X}),$$

и 2-чердак триплета со́-бытий $\mathfrak{X} = \{x, y, z\}$:

$$\overset{2}{\underset{x\in\mathfrak{X}}{\odot}} x = \sum_{\substack{X\subseteq\mathfrak{X}\\ |X|\geqslant 2}} \mathsf{ter}(X/\!/\mathfrak{X}) =$$

$$= \mathsf{ter}(\{x, y\}/\!/\mathfrak{X}) + \mathsf{ter}(\{x, z\}/\!/\mathfrak{X}) + \mathsf{ter}(\{y, z\}/\!/\mathfrak{X}) +$$
$$+ \mathsf{ter}(\{x, y, z\}/\!/\mathfrak{X}).$$

5 Примеры: Фреше-границы для вероятностей слоев и чердаков триплета со́-бытий

Пусть $p_x = \mathbf{P}(x), x \in \mathfrak{X}$ — вероятности со́-бытий из 3-множества $\mathfrak{X}$, а

$$p^{(1)} \geqslant p^{(2)} \geqslant p^{(3)}, \tag{5.1}$$

— *(убывающий) вариационный ряд* вероятностей со́-бытий из $\mathfrak{X}$. Заметим, что

$$p^{(1)} = \max_{x\in\mathfrak{X}} p_x, \quad p^{(3)} = \min_{x\in\mathfrak{X}} p_x.$$

следует из (5.1). Обозначим $S = p^{(1)} + p^{(2)} + p^{(3)}$, $S_{12} = p^{(1)} + p^{(2)}$, $S_{13} = p^{(1)} + p^{(3)}$, $S_{23} = p^{(2)} + p^{(3)}$ суммы вероятностей со́-бытий из $\mathfrak{X}$, и заметим, что

$$S \geqslant S_{12} \geqslant S_{13} \geqslant S_{23}$$

также следует из (5.1).

Нижние и верхние Фреше-границы вероятностей слоев и чердаков $\mathfrak{X}$ определяются как следующие функции от вероятностей со́-бытий из триплета $\mathfrak{X}$.

0-слой:[12]

$$\pi_1^- = \max\{0, 1 - S\}, \quad \pi_0^+ = 1 - p^{(1)};$$

1-слой:

$$\pi_1^- = \begin{cases} 0, & p^{(1)} \leqslant S_{23}, \\ p^{(1)} - S_{23}, & p^{(1)} > S_{23} \end{cases} =$$

$$= \max\left\{0,\ p^{(1)} - S_{23}\right\}, \tag{5.2}$$

$$\pi_1^+ = \begin{cases} S, & 0 < S \leqslant 1, \\ 2 - S, & 1 < S \leqslant 2, \\ 2 - S_{12}, & 2 < S \leqslant 3, p^{(3)} < S_{12} - 1, \\ 1 - p^{(3)}, & 2 < S \leqslant 3, p^{(3)} \geqslant S_{12} - 1, \end{cases} \tag{5.3}$$

2-слой:

$$\pi_2^- = \begin{cases} 0, & S_{12} \leqslant 1 + p^{(3)}, \\ S_{12} - 1 - p^{(3)}, & S_{12} > 1 + p^{(3)} \end{cases} =$$

$$= \max\left\{0,\ S_{12} - 1 - p^{(3)}\right\}, \tag{5.4}$$

$$\pi_2^+ = \begin{cases} p^{(1)} - S_{23}, & S_{23} \leqslant p^{(1)}, \\ 2p^{(1)}, & p^{(1)} < S_{23} \leqslant 1, \\ 1 + p^{(1)} - S_{23}, & 1 < S_{23}; \end{cases} \tag{5.5}$$

[12] Поскольку $\pi_0(\mathfrak{X}) = \pi_3(\mathfrak{X}^{(c)}) = p_{\mathfrak{X}^{(c)}}$.

3-слой:[13]

$$\pi_3^-=\max\{0,\ S-2\},\quad \pi_3^+=p^{(3)};$$

1-чердак:[14]

$$(\overset{1}{u}_{\mathfrak{X}})^-=p^{(1)},\quad (\overset{1}{u}_{\mathfrak{X}})^+=\min\{1,S\};$$

2-чердак:

$$(\overset{2}{u}_{\mathfrak{X}})^-=\begin{cases}\max\{0,S-1\} & S_{12}\leqslant 1,\\ \max\{p^{(3)},S_{12}-1\}, & 1<S_{12},\end{cases}\tag{5.6}$$

$$(\overset{2}{u}_{\mathfrak{X}})^+=\begin{cases}S_{23}, & S_{23}\leqslant p^{(1)},\\ p^{(1)}, & p^{(1)}<S_{23}\end{cases}=\min\{p^{(1)},S_{23}\};\tag{5.7}$$

3-чердак:[15]

$$(\overset{3}{u}_{\mathfrak{X}})^-=\max\{0,S-2\},\quad (\overset{3}{u}_{\mathfrak{X}})^+=p^{(3)}.$$

6 Выводы

Построение Фреше-границ для вероятностей произвольных теоретико-множественных операций над N-множеством со́-бытий — NP-трудная задача. Метод линейного программирования позволяет, решая NP-трудную задачу, вычислять лишь значения Фреше-границ для конкретных значений вероятностей со́-бытий из данного N-множества со́-бытий. Хотя Фреше-границы для вероятностей "мощностных" теоретико-множественных операций над N-множеством со́-бытий могут быть найдены аналитически для произвольного N, до сих пор они остаются неизвестными.

Благодарности

Автор благодарит Аркадия Арсеньевича Новоселова за его функциональный вклад в развитие теории границ Фреше.

Список литературы

[1] О. Ю. Воробьев. *Сет-суммирование*. Наука, Новосибирск, 1993, 137с.

[2] О. Ю. Воробьев. *Эвентология*. Сибирский федеральный университет, Красноярск, 2007, 435с., https://www.academia.edu/179393/Vorobyev_O.Yu._Eventology._-_Krasnoyarsk_Siberian_Federal_University._-_2007._-_435p.

[3] 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.

[4] О. Ю. Воробьев. Эвентологические особенности выбора точки из Фреше-многогранника в симплексе распределений. *Труды XII Международ. ФАМЭБ конференции по финансово-актуарной математике и эвентологии безопасности*, Красноярск: НИИППБ, СФУ (под ред. Олега Воробьева):134–142, 2013.

[5] А. А. Новоселов. Обобщение понятия границ Фреше. В *Трудах XIII ФАМЭМС конф. по финансово-актуарной математике и эвентологии многомерной статистики*, Красноярск, СФУ (под ред. Олега Воробьева):154–158, 2014.

[6] О. Ю. Воробьев. К обоснованиям эвентологического метода. *Труды XIII ФАМЭМС конференции по финансово-актуарной математике и эвентологии многомерной статистики*, Красноярск: СФУ (под ред. Олега Воробьева):112–121, 2014.

[7] О. Ю. Воробьев. Сет-дистограммные измерители разобщенности событий. *Труды XIII ФАМЭМС конференции по финансово-актуарной математике и эвентологии многомерной статистики*, Красноярск: СФУ (под ред. Олега Воробьева):95–111, 2014.

[8] О. Ю. Воробьев. Вне привычных измерений статистических взаимосвязей: эвентологически пересмотренный подход к измерению общности, ассоциации и статистической зависимости множеств событий. В *Трудах XIII ФАМЭМС конф. по финансово-актуарной математике и эвентологии многомерной статистики*, Красноярск, СФУ (под ред. Олега Воробьева):44–94, 2014.

[9] C.E. Bonferroni. Teoria statistica delle classi e calcolo delle probabilitá. *Pablicazioni del R. Instituto Superiore di Scienze Economiche e Commerciali di Firenze*, 8:1–62, 1936.

[10] 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, 298-299.

[11] M. Fréchet. Généralisations du théorème des probabilités totales. *Fundamenta Mathematica*, 25:379–387, 1935.

[12] M. Fréchet. *Les Probabilitiés, Associées a un Systéme d'Evénments Compatibles et Dépendants*. Premiere Partie. Hermann & Cie, Paris, 1940.

[13] M. Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l'Institut Henri Poincaré*, 10(4):215–310, 1948.

[14] M. Fréchet. Sur les tableaux de corrélation dont les marges sont données. *Annales de l'Université de Lyon. Section A*, 9:53–77, 1951.

[15] M. Fréchet. Les tableaux de corrélation dont les marges et des bornes sont données. *Annales de l'Université de Lyon. Sciences*, 20:13–31, 1957.

[16] M. Fréchet. Sur la distance de deux lois de probabilité. *Publ. Inst. Stat. Univ. Paris*, 6:185–198, 1957.

[17] M. Fréchet. Sur les tableaux dont les marges et des bornes sont données. *Revue Inst. de Stat.*, 28(1/2):10–32, 1960.

[18] T. Hailperin. Best possible inequalities for the probability of a logical function of events. *American Mathematical Monthly*, 72:343–359, 1965.

[19] T. Hailperin. Probability logic. *Notre Dame Journal of Formal Logic*, 25(3):198–212, 1984.

[20] T. Hailperin. *Boole's Logic and Probability*. North-Holland, Amsterdam, 1986.

[21] W. Hoeffding. Masstabinvariante korrelationstheorie. *Schriften des Matematischen Instituts und des Instituts für Angewandte Matematik de Universität Berlin (Reprinted as: Scale-invariant correlation theory. In: Fisher, N. I. and Sen, P. K., editors, (1994). The Collected Works of Wassily Hoeffding, 57-107. Springer, New York.)*, 5:179–233, 1940.

[22] W. Hoeffding. Masstabinvariante korrelationsmasse für diskontinuierliche verteilungen. *Arkiv für matematischen Wirtschaften und Sozialforschung (Reprinted as: Scale-invariant correlation measures for discontinuous distributions. In: Fisher, N. I. and Sen, P. K., editors, (1994). The Collected Works of Wassily Hoeffding, 109-133. Springer, New York.)*, 7:49–70, 1941.

[23] A. Précopa. A method of disaggregation for bounding probabilities of Boolean functions of events. RRR 21-97, Rutger University, Rutger Center for Operations research, New Jersey, January 1998.

[24] L. Rüschendorf. Fréchet-bounds and their applications. In G. Dall'Aglio, S. Kotz, and G. Salinetti, editors, *Advances in Probability Distributions with Given Marginals, Mathematics and Its Applications*, volume 67, pages 151–187. Kluwer, Dordrecht, 1991.

[25] P. Veneziani. New Bonferroni-type inequalities. RRR 2-2002, Rutger University, Rutger Center for Operations research, New Jersey, January 2002.

[13] Поскольку $\pi_3(\mathfrak{X})=p_{\mathfrak{X}}$.

[14] Поскольку $\overset{1}{u}_{\mathfrak{X}}=u_{\mathfrak{X}}$.

[15] Поскольку $\overset{3}{u}_{\mathfrak{X}}=p_{\mathfrak{X}}$.

Оценка общности покупательских сегментов по чековой статистике

Елена Евгеньевна Голденок
Сибирский федеральный университет
Институт математики
Красноярский государственный
торгово-экономический институт
Красноярск
ellen_gold@rambler.ru

Аннотация. *Предлагается применить сопряженные таблицы для эвентологического анализа общности покупательских сегментов на товарных рынках.*

Ключевые слова. *Вероятность, множество событий, вероятностное распределение, сопряженные таблицы, категориальное множество.*

1 Введение

В этой работе предпринята попытка построения сопряженных таблиц, известных в математической статистике, на основе чековой сиатистики для эвентологического анализа рынка товаров и услуг, в частности для анализа взаимодействия покупательских сегментов. Вычислены некоторые показатели общности этих сегментов.

2 Сопряженные таблицы для чековой статистики

Рассмотрим задачу построения сопряженных таблиц на основе "чековой статистики". "Чековая статистика" — это статистика выбора подмножеств из базового множества посредством отнесения каждого элемента базового множества к одной из возможных категорий; в частности, — это статистика выбора подмножеств товаров на товарном рынке из базового множества товаров, которое называется ассортиментом товаров, и отнесение каждого товара из ассортимента к одной из возможных стоимостных категорий; иначе говоря, это статистика покупок товаров из предлагаемого ассортимента, которая фиксируется в "чеках", в каждом из которых указано выбранное подмножество товаров, стоимости каждого товара и общая стоимость.

На товарный рынок можно взглянуть как на процесс выбора покупателями своего "мнения" о предлагаемых товарах, которое покупатель выражает в стоимости товара — сумме, которую он потратил на покупку товара. Покупая то или иное подмножество товаров, покупатели выступают в качестве "экспертов", относя товары, зафиксированные в "чеке", к той или иной стоимостной категории. Категориями служат *интервалы относительных стоимостей*, к которым покупатель относит каждый товар, совершая покупку. Например, если для каждого товара $a \in \mathcal{A}$ определена минимальная $\alpha_{\min}(a)$ и максимальная $\alpha_{\max}(a)$ стоимость его возможной покупки, то разбиение интервала стоимостей $[\alpha_{\min}(a), \alpha_{\max}(a)]$ на $M-1$ подинтервалов

$$(\alpha_{\min}(a), \alpha_{\max}(a)] =$$
$$= (\alpha_{\min}(a), \alpha_2(a)] + (\alpha_2(a), \alpha_3(a)] + \ldots$$
$$\ldots + (\alpha_{M-1}(a), \alpha_{\max}(a)]$$

определяет $M-1$ стоимостную категорию покупки данного товара (от 2-й до M-й). Сдвиг интервала стоимостей на $\alpha_{\min}(a)$ и нормировка на его длину $\alpha_{\max}(a) - \alpha_{\min}(a)$ дает единичный *интервал относительных стоимостных категорий* любого товара

$$(0, 1] = (0, \alpha_2] + (\alpha_2, \alpha_3] + \ldots + (\alpha_{M-1}, 1],$$

состоящий из $M-1$ категории, где

$$\alpha_i = \frac{\alpha_{i-1}(a) - \alpha_i(a)}{\alpha_{\max}(a) - \alpha_{\min}(a)}$$

для $i = 2, \ldots, M$ при $\alpha_1 = 0, \alpha_M = 1$.

Для учета в этом интервале ситуации непокупки товара (покупки товара нулевой стоимости), осталось добавить еще одну 1-ю категорию в виде вырожденного подинтервала $[\alpha_0, \alpha_1] = [0, 0]$, чтобы получить единичный *интервал относительных стоимостей* в окончательном виде:

$$[0, 1] = [0, \alpha_1] + (\alpha_1, \alpha_2] + (\alpha_2, \alpha_3] + \ldots + (\alpha_{M-1}, 1].$$

Множество покупателей на рынке можно рассматривать как можество экспертов, чтобы упростить модель можно также разбить это множество на m покупательских сегментов самыми разными способами, например, по доходам, по величине стоимости покупок, по времени покупок и т.д. Первый способ, пожалуй, самый трудноосуществимый, поскольку по чекам это определить невозможно, зато можно предположить, что более дорогие покупки совершает сегмент покупателей с большим уровнем дохода, это

предположение позволяет свести сегментацию покупателей по уровню дохода к сегментации по стоимости покупок. Сегментация множества покупателей по времени также вполне осуществима, поскольку на чеках указывается время совершения покупки.

В итоге получаем общее правило наблюдения за выбором покупателей (принадлежащих m покупательским сегментам) **товара** в одну из M категорий:

"*при исходе товарного рынка* $\omega \in \Omega$ *покупатель, принадлежащий* j*-тому сегменту покупателей* ($j = 1, \ldots, m$)*, выбирает товар* $a_\omega \in \mathcal{A}$ *в* i*-тую категорию* ($i = 1, \ldots, M$)*, когда происходит событие*

$$x_i^j = \left\{ \omega \in \Omega : \frac{\alpha(a_\omega) - \alpha_{\min}(a_\omega)}{\alpha_{\max}(a_\omega) - \alpha_{\min}(a_\omega)} \in [\alpha_{i-1}, \alpha_i] \right\}." \quad (\star)$$

Так определенные события ($\star$) образуют **для каждого товара** *сопряженное м.с.* [1]

$$\mathbb{X} = \{x_i^j, i = 1, \ldots, M; j = 1, \ldots, m\} = \\ = \sum_{i=1}^{M} \sum_{j=1}^{m} \{x_i^j\} = \sum_{i=1}^{M} \mathfrak{X}_i = \sum_{j=1}^{m} \mathbf{X}^j,$$

где

$$\mathfrak{X}_i = \{x_i^1, \ldots, x_i^m\}$$

— *категориальные м.с.*, а

$$\mathbf{X}^j = \{x_1^j, \ldots, x_M^j\}$$

— *экспертные м.с.*.

Поэтому в качестве статистической модели для описания "чековой" статистики **по каждому товару** можно выбрать m-мерную сопряженную $(\underbrace{M \times \ldots \times M}_{m})$ таблицу [1], вероятностное распределение которой характеризуется сопряженным м.с. $\mathbb{X}$ в задаче об общности мнений m экспертов (сегменты покупателей) по M категориям (товарные категории по стоимости).

В [1] предложено несколько эвентологических показателей общности мнений экспертов в (m, M) задаче: *каппа-близость и каппа-Фреше-близость*, которые могут быть использованы для оценки общности поведения сегментов покупателей на товарных рынках.

Кроме этого, интересно сравнить для товарных рынков эвентологические оценки общности в покупательских сегментах с классическим показателем общности *каппа Коэна* [2], а также с простейшим показателем общности — *диагональной каппа*.

2.1 Определение понятия товара в табличном описании "чековой" статистики

Определение понятия товара в описании "чековой" статистики сопряженными таблицами можно давать с точек зрения, которые различаются уровнем общности. Например, если, руководствуясь самой общей точкой зрения, считать товаром всё, что продается в данном гипермаркете, то каждый чек фиксирует покупку этого **общего товара** на общую сумму чека, размером которой покупатель относит **общий товар** к той или иной стоимостной категории. Каждый чек рассматривается как результат одного эксперимента с **общим товаром**, и его статистической моделью служит одна сопряженная таблица и одно сопряженное м.с.

Однако можно, разбив предварительно товары на n групп, т.е. создав ассортимент из n **групповых товаров**, рассматривать каждый чек как результат проведения n экспериментов, в каждом из которых покупатель относит купленный **групповой товар** к той или иной стоимостной категории размером суммы, которую он потратил на данный **групповой товар**. Теперь каждый чек рассматривается как результат n экспериментов с **групповыми товарами**, а их статистическими моделями служат n сопряженных таблиц и, соответственно, n сопряженных м.с.

3 Пример сопряженной таблицы на основе чековой статистики кафе для оценки общности бедного и богатого клиентов

На основе чековой статистики кафе за 10 дней можно составить сопряженные таблицы для оценки общности покупательских сегментов посетителей кафе. Всего за этот период 1078 чеков. Весь ассортимент блюд будем считать общим товаром. Все чеки за один день расположим в порядке возрастания их стоимости, разобьем их на три группы, равные по количеству чеков и получим разбиение покупателей на три сегмента по стоимости заказов. Первый сегмент условно назовем "бедным", второй – "средним" и третий – "богатым" клиентами. В каждом из полученных сегментов есть чеки с минимальной и максимальной ценами. Каждую совокупность чеков разделим на пять частей равных по количеству чеков и получим 5 стоимостных интервалов стоимостей общего товара. Нормировка этих стоимостных интервалов позволяет получить 5 общих для всех трех покупательских сегментов стоимостных категорий общего товара. Находим среднее значение стоимости чека "бедного" клиента и пронормировав его, смотрим в какую категорию он попал, аналогичным образом выясняем какой категории заказ сделан "средним" и "богатым" клиентами. Таким образом, для каждого из 10 дней получаем множество, состоящее из трех событий, элементами этих множеств являются события – категории общего товара, которые заказали сегменты клиентов.

Чтобы получить оценки вероятностей этих событий

применим алгоритм построения сопряженных таблиц.

3.1 Построение сопряженных таблиц на основе "чековой" статистики

Метод равновероятного деления на покупательские и товарные сегменты для $(3,5)$**-задачи.**

1) Дневная статистика делится на 3 равновероятных покупательских сегмента (т.е. в каждом сегменте равное количество чеков) и вычисляются стоимостные границы 3-х покупательских сегментов.

2) Каждый покупательский сегмент делится на 5 равновероятных товарных сегментов (т.е. в каждом сегменте равное количество чеков) и вычисляются стоимостные границы 5-ти товарных сегментов для каждого покупательского сегмента.

3) Вычисляется средняя стоимость чека для каждого из 3-х покупательских сегментов.

4) Оценивается в какой из 5-ти товарных сегментов попала каждая из 3-х средних стоимостей.

Замечание. Оценки сегментов, куда попадает средняя стоимость, вообще говоря д.б. различны, так как медиана не обязана совпадать с математическим ожиданием.

Алгоритм построения.

$s = \ \backslash\ j =$	1	2	3	алгоритм
$s=1$	2	3	4	$n[2,3,4] := n[2,3,4]+1$
$s=2$	3	3	4	$n[3,3,4] := n[3,3,4]+1$
$s=3$	3	2	5	$n[3,2,5] := n[3,2,5]+1$
$s=4$	3	3	4	$n[3,3,4] := n[3,3,4]+1$
$s=5$	1	2	3	$n[1,2,3] := n[1,2,3]+1$
$s=6$	3	4	4	$n[3,4,4] := n[3,4,4]+1$
$s=7$	5	4	4	$n[5,4,4] := n[5,4,4]+1$
$s=8$	2	3	4	$n[2,3,4] := n[2,3,4]+1$
$s=9$	3	3	4	$n[3,3,4] := n[3,3,4]+1$
$s=10$	3	4	4	$n[3,4,4] := n[3,4,4]+1$

Таблица 1: Алгоритм построения сопряженной $(3,5)$-таблицы на основе "чековой" статистики по $n=10$ наблюдениям.

Общая формула алгоритма. Для каждого результата $(j=1,2,3)$:

$$i_1(s), i_2(s), i_3(s),$$

s-го наблюдения $(s=1,\dots,n)$ из таблицы 1 накапливаются клеточные статистики $(3,5)$-сопряженной таблицы по формуле

$$n[i_1(s), i_2(s), i_3(s)] := n[i_1(s), i_2(s), i_3(s)] + 1.$$

Затем полученные клеточные статистики нормируются на n, чтобы получить оценки клеточных вероятностей по формулам

$$p[i_1, i_2, i_3] := \frac{n[i_1, i_2, i_3]}{n}$$

для $i_1, i_2, i_3 = 1, \dots, 5$.

статистика	число набл.	вероятности
$n[2,3,4]=2$	10	$\pi[2,3,4]=2/10$
$n[3,3,4]=3$	10	$\pi[3,3,4]=3/10$
$n[3,2,5]=1$	10	$\pi[3,2,5]=1/10$
$n[1,2,3]=1$	10	$\pi[1,2,3]=1/10$
$n[3,4,4]=2$	10	$\pi[3,4,4]=2/10$
$n[5,4,4]=1$	10	$\pi[5,4,4]=1/10$

Таблица 2: Оценки клеточных вероятностей из наблюдений в табл. 1.

По данным таблицы оценки клеточных вероятностей построим три двумерные сопряженные таблицы 5×5 для трех пар сегментов: первого и второго, первого и третьего, второго и третьего и одну трехмерную сопряженную таблицу.

	x_1^2	x_2^2	x_3^2	x_4^2	x_5^2	Ω
x_1^1	–	1/10	–	–	–	1/10
x_2^1	–	–	2/10	–	–	2/10
x_3^1	–	1/10	3/10	2/10	–	6/10
x_4^1	–	–	–	–	–	0
x_5^1	–	–	–	1/10	–	1/10
Ω	0	2/10	5/10	3/10	0	1

Таблица 3: Двумерная сопряженная таблица 5×5 для 1 и 2 сегментов клиентов по данным таблицы 2.

	x_1^3	x_2^3	x_3^3	x_4^3	x_5^3	Ω
x_1^1	–	–	1/10	–	–	1/10
x_2^1	–	–	–	2/10	–	2/10
x_3^1	–	–	–	5/10	1/10	6/10
x_4^1	–	–	–	–	–	0
x_5^1	–	–	–	–	1/10	1/10
Ω	0	0	1/10	7/10	2/10	1

Таблица 4: Двумерная сопряженная таблица 5×5 для 1 и 3 сегментов клиентов по данным таблицы 2.

	x_1^3	x_2^3	x_3^3	x_4^3	x_5^3	Ω
x_1^2	–	–	–	–	–	0
x_2^2	–	–	1/10	–	1/10	2/10
x_3^2	–	–	–	5/10	–	5/10
x_4^2	–	–	–	3/10	–	3/10
x_5^2	–	–	–	–	–	0
Ω	0	0	1/10	8/10	1/10	1

Таблица 5: Двумерная сопряженная таблица 5×5 для 2 и 3 сегментов клиентов по данным таблицы 2.

Эти три двумерные сопряженные таблицы являются проекциями трехмерной сопряженной таблицы $5\times5\times5$ в которой все вероятности кроме вероятностей, указанных в таблице, равны нулю.

статистика	ненулевые вероятности
$n[2,3,4]=2$	$\pi^{123}_{234}=2/10$
$n[3,3,4]=3$	$\pi^{123}_{334}=3/10$
$n[3,2,5]=1$	$\pi^{123}_{325}=1/10$
$n[1,2,3]=1$	$\pi^{123}_{123}=1/10$
$n[3,4,4]=2$	$\pi^{123}_{344}=2/10$
$n[5,4,4]=1$	$\pi^{123}_{544}=1/10$

Таблица 6: Ненулевые вероятности из сопряженной таблицы $5\times 5\times 5$ для трех сегментов клиентов из табл. 2.

категория	сегмент 1	сегмент 2	сегмент 3
1	$\pi^1_1=1/10$	$\pi^2_1=0$	$\pi^3_1=0$
2	$\pi^1_2=2/10$	$\pi^2_2=2/10$	$\pi^3_2=0$
3	$\pi^1_3=6/10$	$\pi^2_3=5/10$	$\pi^3_3=1/10$
4	$\pi^1_4=0$	$\pi^2_4=3/10$	$\pi^3_4=8/10$
5	$\pi^1_5=1/10$	$\pi^2_5=0$	$\pi^3_5=1/10$
Ω	1	1	1

Таблица 7: Таблица вероятностей выбора категорий 5×5 1-ым, 2 и 3 сегментами клиентов, найденных по сопряженным таблицам.

3.2 Вероятности событий из категориальных множеств

Каждая из этих двумерных сопряженных таблиц дает распределение сопряженных множеств событий, которые обозначим соответственно: $\mathbb{X}^{12}$ – сопряженное множество событий для 1 и 2 сегментов клиентов; $\mathbb{X}^{13}$ – сопряженное множество событий для 1 и 3 сегментов клиентов; $\mathbb{X}^{23}$ – сопряженное множество событий для 2 и 3 сегментов клиентов; $\mathbb{X}^{123}$ – сопряженное множество событий для 1, 2 и 3 сегментов клиентов; Рассмотрим например подробно первое сопряженное множество

$$\mathbb{X}^{12}=\{x^j_i, i=1,\dots,5; j=1,2\}=$$
$$=\sum_{i=1}^{5}\sum_{j=1}^{2}\{x^j_i\}=\sum_{i=1}^{5}\mathfrak{X}^{12}_i=\sum_{j=1}^{2}\mathbf{X}^j,$$

где $\mathfrak{X}^{12}_i$ – категориальные множества событий:

$$\mathfrak{X}^{12}_1=\{x^1_1,x^2_1\};$$
$$\mathfrak{X}^{12}_2=\{x^1_2,x^2_2\};$$
$$\mathfrak{X}^{12}_3=\{x^1_3,x^2_3\};$$
$$\mathfrak{X}^{12}_4=\{x^1_4,x^2_4\};$$
$$\mathfrak{X}^{12}_5=\{x^1_5,x^2_5\};$$

экспертные м.с.:

$$\mathbf{X}^1=\{x^1_1,\dots,x^1_5\};$$
$$\mathbf{X}^2=\{x^2_1,\dots,x^2_5\}.$$

Категориальные множества событий сопряженных множеств $\mathbb{X}^{12}$, $\mathbb{X}^{13}$, $\mathbb{X}^{23}$ представляют собой дуплеты событий выбора категорий товара первым и вторым, первым и третьим, вторым и третьим сегментами клиентов соответственно. Распределения этих дуплетов содержат по четыре вероятности. А категориальные множество событий $\mathbb{X}^{123}$ представляют собой триплеты и их распределения содержат 8 вероятностей. Обозначим вероятность события выбора i-той категории товара j-тым клиентом следующим образом:

$$\mathbf{P}(x^j_i)=\pi^j_i$$

На основе сопряженных таблиц находим индивидуальные вероятности событий, входящих в категориальные множества и также вносим их в таблицу.

3.3 Некоторые эвентологические показатели общности покупательских сегментов

Найдем некоторые эвентологические показатели общности покупательских сегментов. **Диагональная каппа сопряженного м.с.** $\mathbb{X}$

$$\kappa_{\mathrm{diag}}(\mathbb{X})=\sum_{i=1}^{4}\pi[i,i,i].$$

Классическая каппа сопряженного м.с. $\mathbb{X}$

$$\kappa(\mathbb{X})=\frac{\kappa_{\mathrm{diag}}(\mathbb{X})-\sum_{i=1}^{4}\Pi[i]}{1-\sum_{i=1}^{4}\Pi[i]}.$$

Каппа-близость сопряженного м.с. $\mathbb{X}$

$$\varkappa_{\Delta}(\mathbb{X})=\frac{1}{4}\sum_{i=1}^{4}(1-\Delta[i]).$$

Каппа-Фреше-близость сопряженного м.с. $\mathbb{X}$

$$\varkappa_{\mathbb{\Delta}}(\mathbb{X})=\frac{1}{4}\sum_{i=1}^{4}(1-\mathbb{\Delta}[i]).$$

На основе данных таблицы для 1 и 2 сегментов найдена **диагональная каппа сопряженного м.с.** $\mathbb{X}^{12}$

$$\kappa_{\mathrm{diag}}(\mathbb{X}^{12})=\sum_{i=1}^{5}\pi[i,i,i]=3/10.$$

Вычислена **классическая каппа сопряженного м.с.** $\mathbb{X}^{12}$

$$\kappa(\mathbb{X}^{12})=\frac{\kappa_{\mathrm{diag}}(\mathbb{X}^{12})-\sum_{i=1}^{5}\Pi[i]}{1-\sum_{i=1}^{5}\Pi[i]}=0,247,$$

где

$$\Pi[i]=\pi_1[i]\pi_2[i].$$

Каппа-близость сопряженного м.с. $\mathbb{X}^{12}$

$$\varkappa_{\Delta}(\mathbb{X}^{12}) = \frac{1}{5}\sum_{i=1}^{5}(1-\Delta[i]) = 0,86,$$

где $\Delta[i]$ – **сет-расстояние категориального м.с.** $\mathfrak{X}_i^{12}$

$$\Delta[i] = \begin{cases} u[i] - \frac{1}{5}\Sigma[i], & \mathfrak{X}_i^{12} \neq \emptyset, \\ 0, & \mathfrak{X}_i^{12} = \emptyset, \end{cases}$$

где $i = 1,2,3,4,5$, а

$$\Sigma[i] = \pi_1[i] + \pi_2[i],$$

– **суммы вероятностей событий** из категориальных м.с. $\mathfrak{X}_i^{12}$ Вычислена **Каппа-Фреше-близость сопряженного м.с.** $\mathbb{X}^{12}$

$$\varkappa_{\mathbb{\Delta}}(\mathbb{X}^{12}) = \frac{1}{5}\sum_{i=1}^{5}(1-\mathbb{\Delta}[i]) = 0,65,$$

где

$$1-\mathbb{\Delta}[i] = \begin{cases} 1-\Delta[i], & \max[i] = \Sigma[i] \leq 1 \text{ или} \\ & \max[i] = 1 < \Sigma[i], \\ \dfrac{\Sigma[i]-u[i]}{\Sigma[i]-\max[i]}, & \max[i] < \Sigma[i] \leq 1, \\ \dfrac{1-u[i]}{1-\max[i]}, & \max[i] < 1 < \Sigma[i]. \end{cases}$$

где $i = 1,2,3,4,5$ – **Фреше-близость категориального м.с.** $\mathfrak{X}_i^{12}$.

Для каждой из сопряженных таблиц найдены 4 характеристики, значения которых сведены в таблицу.

сегменты	$\kappa_{\text{diag}}(\mathbb{X})$	$\kappa(\mathbb{X})$	$\varkappa_{\Delta}(\mathbb{X})$	$\varkappa_{\mathbb{\Delta}}(\mathbb{X})$
12	3/10	0,247	0,86	0,65
13	0	0,075	0,81	0,6
23	3/10	0,014	0,86	0,77
123	0	-0,031	0,747	0,64

Таблица 8: Таблица 5×5 характеристик близости сегментов клиентов 2.

3.4 Выводы

Диагональная каппа для “бедного” и “среднего”, а также для “среднего” и “богатого” клиентов 0,3, а для “бедного” и “богатого” и для всех клиентов вместе – равны нулю, что говорит об отсутствии или о слабой общности выбора категорий товара разными сегментами клиентов.

Значение классической каппы для “бедного” и “среднего” клиентов является наибольшим из всех, но не превышает 0,4, что говорит о “ясной” общности этих клиентов (fair). Отрицательное значение классической каппы для всех трех сегментов в совокупности показывает “недостаточную” общность их выбора (poor). Для “бедного” и “богатого”, а также для “среднего” и “богатого” значения классических капп меньше 0,2, что говорит об их “незначительной” общности (slight).

Значения каппа-близости и каппа-Фреше-близости для “бедного” и “среднего” и для “среднего” и “богатого” клиентов имеют наибольшее значения, что говорит об их общности. Новые эвентологические показатели каппа-близость и каппа-Фреше-близость приняли большие значения, чем диагональная и классическая каппа это объясняется тем, что они учитывают неполные совпадения мнений, которые расположены вне главной диагонали сопряженной таблицы.

Список литературы

[1] О. Ю. Воробьев. Вне привычных измерений статистических взаимосвязей: эвентологически пересмотренный подход к измерению общности, ассоциации и статистической зависимости множеств событий. В *Трудах XIII ФАМЭМС конф. по финансово-актуарной математике и эвентологии многомерной статистики*, Красноярск, СФУ (под ред. Олега Воробьева):44–93, 2014.

[2] J. Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46, 1960.

Полярное уравнение и ϕ_p-оптимальные планы эксперимента для полиномиальной регрессии

Юрий Дмитриевич Григорьев
Санкт-Петербургский государственный
электротехнический университет (ЛЭТИ)
Санкт-Петербург
yuri_grigoriev@mail.ru

Аннотация. *Рассматриваются критерии D-, E- и A-оптимальности как представители семейства ϕ_p- функций Пукельсхайма. Обсуждаются современные подходы к построению соответствующих ϕ_p- планов для полиномиальной регрессии.*

Ключевые слова. *Полиномиальная регрессия, полярное уравнение, ϕ_p- планы, функции Пукельсхайма, теория исключения, базис Гребнера.*

1 Введение

К настоящему моменту в теории оптимального планирования эксперимента разработаны самые различные способы построения оптимальных планов эксперимента (ПЭ). В данном сообщении кратко рассматриваются результаты, достигнутые в области построения D- и E-оптимальных ПЭ для полиномиальной регрессии

$$\eta(x,\theta)=\sum_{i=0}^{s}\theta_i x^{s-i},\quad x\in[-1,1],\quad s\geq 1, \qquad (1)$$

при этом основное внимание уделено кругу вопросов, связанных с построением A-оптимальных планов.

2 ϕ_p-оптимальность

Критериями оптимальности ПЭ являются так называемые *информационные функции*, т. е. функции, удовлетворяющие определенным свойствам.

Определение 1. [18]. *Пусть $\mathcal{M}$ — множество неотрицательно определенных матриц порядка m. Информационной функцией ϕ на $\mathcal{M}$ называется функция $\phi\colon \mathcal{M}\to[0,\infty)$, которая положительно однородна, супераддитивна, неотрицательна, не является константой и полунепрерывна сверху.*

Из определения ϕ следует, что она является строго изотонической (сохраняет частичный порядок) на $\mathcal{M}_0$ [18, теорема 6.13]). Самые известные критерии, используемые в теории ПЭ, принадлежат семейству информационных функций $\{\phi_p\colon p\in(-\infty,1]\}$. В частности, оно включает в себя функционалы, соответствующие классическим критериям D-, A- и E-оптимальности. Согласно p-параметризации семейства $\mathcal{M}$, перечисленным критериям соответствуют следующие ϕ_p-функции:

$$\begin{aligned} D\text{-оптимальность} &: \phi_0(M)=|M|^{1/m},\\ E\text{-оптимальность} &: \phi_{-\infty}(M)=\lambda_{\min}(M),\\ A\text{-оптимальность} &: \phi_{-1}(M)=(m^{-1}\mathrm{tr}M^{-1})^{-1}. \end{aligned}$$

Определение 2. *План ξ^* называется ϕ_p-оптимальным, или ϕ_p-планом, если*

$$\xi^*=\arg\max_{\xi\in\Xi}\phi_p(M(\xi)).$$

Пусть $\mathcal{M}_0\subset\mathcal{M}$ — множество невырожденных матриц, $\langle A,B\rangle=\mathrm{tr}A^TB$ — скалярное произведение в $R^{n\times k}$.

Определение 3. [18]. *Пусть $C,D\in R^{m\times m}$. Полярной функцией $\phi^\infty\colon\mathcal{M}\to[0,\infty)$ для функции $\phi\colon\mathcal{M}_0\to(0,\infty)$ называется функция*

$$\phi^\infty(D)=\inf_{C>0}\frac{\langle C,D\rangle}{\phi(C)},\quad D\geq 0.$$

а матрица $D=C^{p-1}/\mathrm{tr}C^p$ — корнем полярного уравнения

$$\phi_p(C)\phi_p^\infty(D)=\langle C,D\rangle=1,\quad p\in(-\infty,1]. \qquad (2)$$

Обозначим $F=(f_1,...,f_m)\in R^{n\times m}$, f_i — регрессионные функции, $U=(FF^T)^{-1}F$, и пусть C в (2) имеет вид $C:=M=F^T\mathrm{diag}\{w_1,...,w_m\}F$. Тогда опираясь на полярное уравнение, можно получить обобщенный вариант теоремы эквивалентности для ϕ_p-критериев. Именно, имеет место

Теорема 1. [18]. *Пусть $M=\in M_0(\Xi)$ — невырожденная информационная матрица. Тогда*

(1) *план $\xi\in\Xi$ ϕ_p-оптимален, если и только, если*

$$f^T(x)M^{p-1}(\xi)f(x)\leq\mathrm{tr}M^p(\xi),\quad x\in X; \qquad (3)$$

© 2014 Ю.Д. Григорьев

Олег Воробьев (ред.), Труды XIII ФАМЭМС'2014, Красноярск: СФУ

(2) *неравенство* (3) *обращается в равенство, если* $x \in S(\xi)$, *где* $S(\xi)$ — *спектр любого* ϕ_p-*оптимального плана* $\xi \in \Xi$.

(3) *веса* $w_i = \xi(x_i)$, $i = 1, ..., n$ *точек* $x_i \in S(\xi)$ *спектра* ϕ_p-*оптимального плана* ξ *имеют вид*

$$w_i = \frac{\sqrt{b_{ii}}}{\sum_{j=0}^{s}\sqrt{b_{jj}}}, \quad B = UM^{p+1}U^T. \qquad (4)$$

Для $\phi_{-\infty}$-функции, соответствующей критерию E-оптимальности, полярное уравнение в случае кратности $\lambda_{\min}(C)$, большей 1, имеет множество решений. Поэтому этот случай требует отдельного рассмотрения, но также вписывается в общую схему ϕ_p-оптимальности (см. [16]).

3 ϕ_0- и $\phi_{-\infty}$-планы

Критерий D-оптимальности в настоящее время полностью изучен, в том числе и как ϕ_0-функция. Именно, имеет место

Теорема 2. (Федоров [4], Хоэл [13]). *D-оптимальный план ξ^* для регрессии* (1) *имеет равные массы в $m = s + 1$ нулях полинома $(1 - x^2)L'_s(x)$, $s \geq 1$, где $L_s(x)$ — s-ый полином Лежандра.*

Полиномы Лежандра определяются рекуррентно

$$(s+1)L_{s+1} = (2s+1)xL_s(x) - sL_{s-1}(x),$$
$$L_1(x) = x, \quad L_2(x) = \frac{1}{2}(3x^2 - 1).$$

Пример 1. Пусть $s = 3$. Используя соотношение

$$(1 - x^2)L'_s(x) = -sxL_s(x) + sL_{s-1}(x), \quad s \geq 2,$$

находим $\xi^* = \left\{\xi(\pm 1) = \frac{1}{2}, \quad \xi\left(\pm\frac{1}{\sqrt{5}}\right) = \frac{1}{2}\right\}$.

Впервые E-оптимальные планы для регрессии (1) были рассмотрены Лаптевым [5]. В данной работе они построены численно для $s = 2, 3, 4$ с использованием условий оптимальности, вытекающих из теоремы 1. В аналитическом виде соответствующий план для $s = 2$ найден Кифером [15]. После появления теоремы эквивалентности Меласа [6] $\phi_{-\infty}$-планы для регрессии (1) построены аналитически для всех s. Таким образом, здесь ситуация является той же, что и для ϕ_0-планов с тем отличием, что $\phi_{-\infty}$-планы оказались неинвариантными к заданию области планирования $X = [a, b]$ даже в симметричном случае. Это объясняется тем, что для соответствующих интервалов $[a, b]$ функционал $\phi_{-\infty}(M)$ в точке $M(\xi^*)$ оказывается недифференцируемым в силу появления кратности $\lambda_{\min}(M)$, большей 1. Полное решение этой проблемы представлено в [16].

При построении E-оптимальных планов главную роль играют полиномы Чебышева 1-го рода

$$T_s(x) = \cos[s \cdot \arccos(x)] = \sum_{i=0}^{s}\gamma_i x^{s-i}, \quad x \in [-1, 1].$$

Конкретные выражения для $T_s(x)$, $s = 1, ..., 8$, представлены в табл. 1.

Таблица 1. Полиномы Чебышева $T_s(x)$.

s	$T_s(x)$
1	x
2	$2x^2 - 1$
3	$4x^3 - 3x$
4	$8x^4 - 8x^2 + 1$
5	$16x^5 - 20x^3 + 5x$
6	$32x^6 - 48x^4 + 18x^2 - 1$
7	$64x^7 - 112x^5 + 56x^3 - 7x$
8	$128x^8 - 256x^6 + 160x^4 - 32x^2 + 1$

Пусть $m = s + 1$. Введем обозначения:

$F = ((-1)^j x_j^i)_{i=0,j=1}^{m-1,m}$;
$e_j \in R^{m-}$ канонический базис R^m;
$\gamma = (\gamma_0, \gamma_1, ..., \gamma_s) \in R^m$ — коэффициенты $T_s(x)$;
$A_j = e_j F^{-1}\gamma, \quad j = 1, ..., m.$

Предполагаем, что все $x_j \in [-1, 1]$ попарно различны, т. е. матрица F невырождена.

Теорема 3. (Мелас [16, с. 94]) *E-оптимальный план ξ^* для регрессии* (1) *удовлетворяет следующим условиям:*

(1) $supp\ \xi^* = (x_1^*, ..., x_m^*)$, *где x_j^* — экстремальные точки полинома $T_s(x)$;*

(2) $p_j = \xi(x_j^*) = \frac{|A_j|}{\sum_{k=1}^{m}|A_k|}$, $j = 1, ..., m$ — *веса точек плана;*

(3) $\lambda_{\min}^* := \lambda_{\min}[M(\xi^*)] = (\sum_{j=0}^{s}\gamma_j^2)^{-1}$ — *минимальное собственное число матрицы $M(\xi^*)$ кратности 1.*

Аналитические выражения $\phi_{-\infty}$-планов для $s = 2, ..., 5$ и $x \in [-1, 1]$ представлены в Табл. 2.

Таблица 2. E-оптимальные планы: $\text{supp}\ \xi^* = (\pm 1, \pm x^*, \pm y^*, 0)$, $\alpha = \frac{20\sqrt{5}}{681}$.

s	Узлы		Меры $\xi(\pm x_i^*)$ узлов x_i^*			
	x^*	y^*	± 1	$\pm x^*$	$\pm y^*$	0
1	—	—	1	—	—	—
2	—	—	$\frac{2}{5}$	—	—	$\frac{3}{5}$
3	$\frac{1}{2}$	—	$\frac{19}{75}$	$\frac{56}{75}$	—	—
4	$\frac{1}{\sqrt{2}}$	—	$\frac{24}{129}$	$\frac{64}{129}$	—	$\frac{41}{129}$
5	$\frac{\sqrt{5}+1}{4}$	$\frac{\sqrt{5}-1}{4}$	$\frac{167}{1135}$	$\frac{484}{1135} - \alpha$	$\frac{484}{1135} + \alpha$	—

Таблица 3. Значения $\lambda_{\min}^* = \phi_{-\infty}(M(\xi^*))$.

s	1	2	3	4	5
$\lambda_{\min}^*$	1	$\frac{1}{5}$	$\frac{1}{25}$	$\frac{1}{129} = 0.0077$	$\frac{1}{681} = 0.0015$

4 A-оптимальные планы

Исторически первый численный подход к построению A-оптимальных планов предложен Федоровым в [4]. Метод Федорова применим для линейных функционалов на множестве дисперсионных матриц $D = M^{-1}$, к числу которых относится и A-критерий. Однако он не лишен недостатков, свойственных всем процедурам итерационного типа, так как не учитывает дополнительной информации как о виде модели, так и о известных в настоящее время свойствах ϕ_{-1}-планов (см., например [1]. Эти моменты учитывает подход, связанный с *теорией исключения*, см. [17], [20], [12], [3].

4.1 Метод исключения

Задачей, для решения которой разрабатывалась теория исключения, являлось отыскание всех решений системы алгебраических уравнений

$$f_i(x_1, ..., x_n) = 0, \quad i = 1, ..., n. \tag{5}$$

Левые стороны этих уравнений являются какими-либо заданными полиномами от переменных $x_1, ..., x_n$ с коэффициентами из числового поля K. Очевидно, к этой задаче сводится построение ϕ_{-1}-планов, если в качестве отправной точки рассматривать систему уравнений

$$\frac{\partial \mathrm{tr} M^{-1}(\xi)}{\partial x_i}, \quad i = 1, ..., n, \tag{6}$$

где в качестве x_i рассматриваются неизвестные координаты ϕ_{-1}-плана ξ, включая как точки спектра, так и веса соответствующих точек.

Эта задача, очевидно, тем проще, чем меньше мы имеем переменных. Если $n = 1$, то мы предполагаем решение задачи полностью известным. Поэтому основная идея теории исключения заключается в преобразовании исходной системы (5) в эквивалентную ей систему $g_i(x_1, ..., x_{n-1}) = 0$, $i = 1, ..., m$ (m<n), содержащую меньшее число неизвестных, по крайней мере, в одном из уравнений. Затем этот процесс повторяется с целью получения в новой системе хотя бы одного уравнения с $n - 2$ переменными. Продолжая этот процесс, мы стремимся придти к системе, в которой одно из уравнений содержит только одну переменную, скажем x_1.

Определив возможные решения этого уравнения от переменной x_1, мы возвращаемся к предыдущей системе, содержащей, как минимум, одно уравнение с двумя переменными x_1 и, скажем, x_2. Определив x_2, продолжаем указанный процесс, пока не придем к окончательному решению системы (5). Очевидно, этот процесс аналогичен *схеме Гаусса* для решения систем линейных уравнений, но значительно сложнее в реализации, так как на пути его осуществления встречается много технических деталей, требующих преодоления.

Самым известным методом перехода от системы уравнений (5) к системе с $n - 1$ и т. д. переменными является метод исключения Кронекера. В общем случае он приводит к системе уравнений вида

$$G(x_1) = 0, \quad x_i - H_{i-1}(x_1) = 0, \quad i = 2, ..., n, \tag{7}$$

где *элиминанта* $G(x_1)$ — полином, а $H_i(x_1)$ — *рациональные* функции от x_1. Техника, позволяющая осуществить переход от системы (5) к системе (11), основана на понятии *результанта*.

4.2 Метод Пукельсхайма

Другой современный подход к построению A-отимальных планов связан с теорией информационных функций. Критерию ϕ_{-1}-оптимальности соответствует строго изотоническая $\phi_{-1}(M)$-функция. По тем же причинам, что и для других критериев, ϕ_{-1}-планы являются насыщенными и симметричными[18, раздел 9.5]. Веса $w_i = \xi(x_i)$ точек ϕ_{-1}-планов определяются согласно (4). Поскольку матрица $B = UU^T = (F^T F)^{-1}$ свободна от весов, то это дает возможность раздельно оценивать сначала точки плана, а затем их веса. Значение ϕ_{-1}-функции при этом имеет вид $\phi_{-1}(M) = (s + 1)\mathrm{tr}^{-1} M^{-1}$ [18], т. е.

$$\mathrm{tr} M^{-1}(\xi) = \Big(\sum_{j=0}^{s} \sqrt{b_{jj}}\Big)^2. \tag{8}$$

Выражения (4) и (8) именно для A-критерия, а не как частный случай теории информационных ϕ_p-функций, было также получено в [14].

4.3 Метод базисов Гребнера

Пусть K — некоторое поле. Обозначим через $K[x_1, ..., x_n]$ кольцо полиномов от n переменных x_i с коэффициентами из K. Если $\mathcal{F}$ — произвольное множество полиномов, то идеал, порождаемый множеством $\mathcal{F}$ — это множество $\langle \mathcal{F} \rangle$, состоящее из всех полиномиальных линейных комбинаций

$$\langle \mathcal{F} \rangle = \{p_1 f_1 + ... + p_n f_n \colon f_1, ..., f_n \in \mathcal{F}\},$$

где $p_1, ..., p_n \in K[x_1, ..., x_n]$.

Согласно теореме Гильберта о базисе, любой идеал $I \subset K[x_1, ..., x_n]$ имеет вид $I = \langle \mathcal{F} \rangle$, т. е. генерируется некоторым конечным множеством полиномов, образующим *базис* идеала. Базисов у конкретного идеала существует бесконечно много, но из существание базиса вытекает одно важное свойство: если полиномы базиса рассматривать как левые части системы уравнений (5), то множесто решений этой системы совпадает с множеством решений для любого другого базиса данного идеала.

Отсюда вытекает отсновная идея метода — найти базис заданного идеала вида (7) с той разницей, что входящие в него функции $H_i(x_1)$ — не рациональные функции, а *полиномы* от x_1. Базис такого вида называется *базисом Гребнера*. Алгоритм его отыскания указан в 1965 году Бухбергером [11]. В настоящее время его описание можно найти в [19], [10], [9], [8], [2], [7]. Пакеты программ по вычислению базиса Гребнера включены в системы AXIOM, Maple, Mathematica, REDUCE, Cali и др.

Пример 2. Пусть задана система уравнений

$$\begin{aligned} 4x^2+xy^2-z+\frac{1}{4} &= 0,\\ 2x+y^2z+\frac{1}{2} &= 0,\\ x^2z-\frac{1}{2}x-y^2 &= 0. \end{aligned}$$

Применяя алгоритм Бухбергера, получаем следующий базис Гребнера:

$$\begin{aligned} G_1 &= 10,980z-6,272x^6-45,504x^5+216x^4\\ &+ 3,640x^3-36,846x^2-412x-2,857,\\ G_2 &= 112x^6-1,1376x^4+1,772x+27,45y^2\\ &- 65x^3+756x^2+2,\\ G_3 &= 32x^7+232x^6-34x^4-44x^3+x^2+30x+8. \end{aligned}$$

Отсюда получаем два ответа (x,y,z):

$$\Big(-\frac{1}{2},\frac{1}{\sqrt{2}},1\Big),\quad \Big(-\frac{1}{2},-\frac{1}{\sqrt{2}},1\Big).$$

5 ϕ_{-1}-планы: методы построения

Случай $s=2$, тривиален [4]. В настоящее время в литературе представлены A-оптимальные планы для $s=2,...,8$ в работе [14], и для $s=2,...,10$ — в [18]. В данной работе кратко прокомментируем технику получения ϕ_{-1}-планов для полиномиальных регрессий степеней $s=3,4$ с помощью описанных выше методов.

5.1 Степень $s=3$.

Рассмотрим три метода — исключения, информационных функций Пукельсхайма и базисов Гребнера.

Метод исключения. Зададим план ξ в виде

$$\xi=\{\xi(\pm1)=2y,\quad \xi(\pm x)=1-2y\},$$

т. е. имеем неизвестный узел $x>0$ и вес y крайней точки интервала $[-1,1]$. Система (6) эквивалентна системе уравнений $f(x,y)=0$, $g(x,y)=0$, где

$$\begin{aligned} f &= -3x^2y+2x^6y-x^4-x^6+y, \qquad (9)\\ g &= -4x^6y+4x^6y^2-4y^2+x^6+x^2-4x^2y. \end{aligned}$$

Так как переменных всего две, то алгоритм сводится к процедуре вычисления результанта $R(f,g)$. Рассматривая f и g как полиномы относительно y с коэффициентами, зависящими от x и полагая $R(f,g)=0$, приходим к уравнению $3x^4+4x^2-1=0$, из которого находим значение x^2. Подставляя необходимое из двух решений в уравнение $f=0$, находим y. Окончательные выражения для координат ϕ_{-1}-плана ξ имеют вид

$$x=\frac{1}{3}\sqrt{3\sqrt{7}-6},\quad y=\xi(x)=\frac{1}{9}(4-\sqrt{7}), \qquad (10)$$

при этом $\mathrm{tr}D(\xi)=19+7\sqrt{7}$.

Метод Пукельсхайма. Максимизация информационной функции $\phi_{-1}(M)=\frac{4}{\mathrm{tr}M^{-1}(\xi)}$ эквивалентна минимизации функции

$$\mathrm{tr}M^{-1}(\xi)=\frac{1}{(1-x^2)^2x^2}(x\sqrt{x^4+1}+\sqrt{x^2+1}),$$

где x — неизвестный узел плана. Решение уравнения $\frac{\partial M^{-1}(\xi)}{\partial x}=0$ приводит к уже известному уравнению $3x^4+4x^2-1=0$, что и дает требуемый результат (10).

Метод базисов Гребнера. Применяя к базису (9), алгоритм Бухбергера, находим базис Гребнера

$$(G_1,G_2)=(9y^3-8y^2+y,18y^2-7y+3x^2). \qquad (11)$$

Элиминанта G_1 содержит только одну переменную. Это дает возможность снова получить тот же результат (10).

5.2 Степень $s=4$.

План ξ имеет вид

$$\xi=\{\xi(\pm1)=2p,\quad \xi(\pm x)=2q,\quad \xi(0)=1-2(p+q)\},$$

т. е. содержит три переменных, подлежащих определению — узел x и веса p и q.

Метод Пукельсхайма. Согласно (8) приходим к функционалу

$$\begin{aligned} &\mathrm{tr}M^{-1}(\xi)=\frac{1}{(1-x^2)x^2}\\ &\times\Big(x^2\sqrt{x^4+1}+\sqrt{x^2+1}+(1-x^2)\sqrt{x^4+x^2+1}\Big). \end{aligned}$$

Отсюда находим x, а затем, согласно (4), и оптимальные значения весов p и q. Они представлены в сводной табл. 6.

Метод исключения. Уравнения

$$\frac{\partial\mathrm{tr}M^{-1}}{\partial x}=0,\quad \frac{\partial\mathrm{tr}M^{-1}}{\partial p}=0,\quad \frac{\partial\mathrm{tr}M^{-1}}{\partial q}=0$$

приводят к системе уравнений

$$\begin{aligned} a_0q^2+a_1q+a_2=0,\\ b_0q^2+b_1q+b_2=0,\\ c_0q^2+c_1q+c_2=0, \end{aligned}$$

коэффициенты которой зависят от x и p. Элементарными преобразованиями приводим эту систему к треугольному виду

$$\begin{aligned} a_0q^2 + a_1q + a_2 &= 0, \\ g_1q + g_2 &= 0, \\ \delta &= 0, \end{aligned} \tag{12}$$

где $\delta = g_1h_2 - g_2h_1$ и

$$\begin{aligned} g_1 &= a_0b_1 - a_1b_0, \quad h_1 = a_0c_1 - a_1c_0, \\ g_2 &= a_0b_2 - a_2b_0, \quad h_2 = a_0c_2 - a_2c_0. \end{aligned}$$

Отсюда находим

$$q = -g_2g_1^{-1} \tag{13}$$

— функция переменных x и p, значения которых пока неизвестны. Подставляя (13) в первое уравнение системы (12), приходим к системе двух уравнений с двумя неизвестными

$$\begin{aligned} a_0g_2^2 - a_1g_1g_2 + a_2g_1^2 &= 0, \\ g_1h_2 - g_2h_1 &= 0. \end{aligned}$$

Решаем эту систему методом результантов. Для этого записываем ее уравнения по степеням p с коэффициентами, зависящими только от x:

$$\begin{aligned} G: &\quad u_0p^2 + u_1p + u_2 = 0, \\ H: &\quad v_0p^3 + v_1p^2 + v_2p + v_3 = 0, \end{aligned}$$

Соответствующий результант $R(G,H)$, с точностью до тривиальных множителей, равен произведению $R(G,H) = R_1R_2$, где

$$\begin{aligned} R_1 &= 9x^{18} + 36x^{16} + 18x^{14} - 120x^{12} \\ &- 267x^{10} - 292x^8 - 224x^6 - 72x^4 + 52x^2 + 32, \end{aligned}$$

— искомый полином, корень $x^* \in (0,1)$ которого приводит к системе $G(x^*,p) = 0$, $H(x^*,p) = 0$, дающей решение p^*. Таким образом, выполняя обратный ход, окончательно получаем

$$x = 0.6768, \quad p = 0.1045, \quad q = 0.2504, \quad r = 0.2903.$$

и

$$\mathrm{tr}M^{-1}(\xi) = 188.69, \quad \phi_{-1)}(M) = \frac{5}{\mathrm{tr}M^{-1}(\xi)} = 0.0265.$$

Так как размерность задачи сравнительно небольшая, то выражения для весов плана p и q можно представить в аналитическом виде. Другими словами, общее решение исходной системы уравнений (12), как и утверждает теория исключения, можно записать в следующем каноническом виде

$$\begin{aligned} \mathcal{X}(x_1) &= R_1(x_1) = 0, \\ x_1 &= x^*\text{— корень уравнения } \mathcal{X}(x) = 0, \\ x_2 &:= p = \vartheta_2(z) = \frac{P_1(z)}{(1-z)P_2(z)}, \\ x_3 &:= q = \vartheta_2(z) = \frac{Q_1(z)}{2z(1-z)(1+z^2)Q_2(z)Q_3(z)}, \end{aligned} \tag{14}$$

где $\mathcal{X}(x_1)$ — элиминанта, $z := x_1^2$, а P_i, Q_j $(i = 1,2), (j = 1,2,3)$ — полиномы, коэффициенты которых представлены в табл. 4 и 5.

Таблица 4. Коэффициенты полиномов $P_i(z) = \sum_{k=0}^{d} a_kz^k$, $d_1 = \deg\ P_1(z) = 15$, $d2 = \deg\ P_2(z) = 14$.

	Коэффициенты a_k			
k	0	1	2	3
$P_1(z)$	16	−104	212	−114
$P_2(z)$	−64	416	−620	−496
k	4	5	6	7
$P_1(z)$	−70	92	−76	246
$P_2(z)$	364	1,394	2,285	−108
k	8	9	10	11
$P_1(z)$	−1	−594	−750	−729
$P_2(z)$	−3,603	−5,193	−4,329	−1,530
k	12	13	14	15
$P_1(z)$	−387	36	108	27
$P_2(z)$	459	513	108	

Таблица 5. Коэффициенты полиномов $Q_i(z) = \sum_{k=0}^{d} a_kz^k$, $d_1 = \deg\ Q_1(z) = 26$, $d_2 = \deg\ Q_2(z) = 20$, $d_3 = \deg\ Q_3(z) = 10$.

	Коэффициенты a_k			
k	0	1	2	3
$Q_1(z)$	-256	2,816	−10,784	13,616
$Q_2(z)$	−64	416	−620	−496
$Q_3(z)$	8	−40	30	64

(*Продолжение табл. 5*).

k	4	5	6	7
$Q_1(z)$	6,936	−9,276	−19,216	−51,546
$Q_2(z)$	364	1,394	2,285	−108
$Q_3(z)$	10	−39	−135	−132
k	8	9	10	11
$Q_1(z)$	34,232	143,774	139,114	−29,846
$Q_2(z)$	−3,603	−5,193	−4,329	−1,530
$Q_3(z)$	−18	27	9	
k	12	13	14	15
$Q_1(z)$	−357,214	−511,305	−257,974	278,546
$Q_2(z)$	459	513	108	
k	16	17	18	19
$Q_1(z)$	802,623	956,163	677,481	255,582
k	20	21	22	23
$Q_1(z)$	−45,288	−135,711	−79,812	−11,664
k	24	25	26	
$Q_1(z)$	8,343	4,131	567	

Очевидно, решение (14) не является базисом Гребнера, получение которых как раз и является продолжением теории исключения.

5.3 Итог: численные результаты

В табл. 6 и 7 представлены численные результаты, полученные при построении A-оптимальных планов для модели (1). Отметим, что в [14] вычисления доведены до $s = 8$, а в [18] — до $s = 10$. Аналитический результат для $s = 3$ в данных работах не приводится.

Таблица 6. A-оптимальные планы: численные результаты (указана суммарная мера симметричных точек).

s	Узлы спектра S		Меры $\xi(\pm x_i^*)$ узлов x_i^*			
	x^*	y^*	± 1	0	$\pm x^*$	$\pm y^*$
1	–	–	1.000	–	–	–
2	–	–	0.500	0.500	–	–
3	0.464	–	0.310	–	0.690	–
4	0.677	–	0.210	0.290	0.500	–
5	0.789	0.291	0.160	–	0.374	0.466
6	0.853	0.479	0.130	0.205	0.295	0.370

Таблица 7. Значения функционала $\phi_{-1}(M(\xi^*)) = (s+1)\mathrm{tr}^{-1}M^{-1}(\xi^*)$.

s	1	2	3	4	5	6
$\phi_{-1}(M)$	1.0	0.375	0.107	0.026	0.006	0.001

Табл. 6 отражает закономерности изменения ϕ_{-1}-планов в зависимости от степени s полиномиальной модели регрессии.

Список литературы

[1] Голикова, Т. И., Панченко, Л. А. Непрерывные A- и Q-оптимальных планы второго порядка на кубе. In В. В. Налимов, editor, *Регрессионные эксперименты (Планирование и анализ)*, pages 71–84. МГУ, Москва, 1977.

[2] Кокс Д., Литтл Дж., О"Ши Д. *Идеалы, многообразия и алгоритмы: Введение в вычислительные аспекты алгебраической геометрии и коммутативной алгебры.* Мир, М., 2000.

[3] Калинина Е. А., Утешев А. Ю. *Теория исключения: Учебное пособие.* НИИ химии СПбГУ, СПб., 2002.

[4] Федоров В. В. *Теория оптимального эксперимента.* Наука, М., 1971.

[5] Лаптев В. Н. *Некоторые вопросы синтеза оптимальных планов регрессионных экспериментов с помощью ЭВМ.* PhD thesis, НЭТИ, Новосибирск, 1974.

[6] Мелас В. Б. Одна теорема двойственности и E-оптимальность. *Зав. лаборатория*, 48:295–296, 1982.

[7] Аржанцев И. В. *Базисы Гребнера и системы алгебраических уравнений.* МЦНМО, М., 2003.

[8] Бухбергер, Б. Базисы Гребнера: Алгоритмический метод в теории полиномиальных идеалов. In Р. Лоос Б. Бухбергер, Дж. Коллинз, editor, *Компьютерная алгебра. Символьные и алгебраические вычисления*, pages 331–372. Мир, Москва, 1986.

[9] Adams W. W., Loustaunau P. *An introduction to Gröbner Bases.* RI: AMS, Providence, 1994.

[10] Becker T., Kredel H., Weispfenning V. *Gröbner Bases: A Computational Approach to Commutative Algebra.* Springer Verlag, New York, 1993.

[11] Buchberger, B. *An algorithm for finding a basis for the residue class ring of a zero-dimensional polynomial ideal (in German).* PhD thesis, University Innsbruck, Austria, 1965.

[12] Gröbner, W. Über die eliminationstheorie. *Monatshefte für Mathematik*, 54:71–78, 1950.

[13] Hoel, P. G. Efficiency problems in polynomial estimation. *Ann. Math. Statist.*, 29:1134–1145, 1958.

[14] Huda, S. and Mukerdjee, R. A-optimal design measures for polynomial regression. *Pakistan J. Statist.*, 7(1) A:47–52, 1991.

[15] Kiefer, J. General equivalence theory for optimum designs (approximate theory). *Ann. Statist.*, 5:849–879, 1974.

[16] Melas V. B. *Functional approach to optimal experimental design.* Springer Science + Business Media, Inc., 2006.

[17] Netto, E. *Vorlesungen über Algebra.* Teubner, Leipzig, Bd. l, 1896. Bd. 2, 1900.

[18] Pukelsheim, F. *Optimal Design of Experiments.* John Wiley & Sons, New York, 1993.

[19] Sturmfels, B. *Gröbner Bases and Convex Polytopes.* RI: AMS, Providence, 1995.

[20] Weber, H. *Lehrbuch der Algebra.* Vieweg, Braunschweig, Bd. l, 1898.

О предельном распределении последовательностей случайных величин со случайным индексом

Абдусуннат Абдурахмонович Джамирзаев
Национальный Университет Узбекистана
Узбекистан
djamirzaev@rambler.ru

Аннотация. *В работе обобщается один результат Ф.Ж.Анскомбе о предельном распределении последовательностей случайных величин со случайным индексом.*

Ключевые слова. *Случайные величины, функция распределения, сходимость, перемешивания.*

1 Введение

Пусть на вероятностном пространстве $\{\Omega, F, P\}$ заданы последовательности случайных величин (с.в.) $\{\eta_n\}$ и $\{\nu_n\}$, где $\{\nu_n\}$ — последовательность положительных целочисленных с.в.. Ф.Ж.Анскомбе [2] доказал следующую теорему.

Теорема 1. *(F.J.Anscombe, 1952). Пусть выполнены следующие условия:*

(A) существует последовательность целых чисел k_n такая, что при $n \to \infty$, $k_n \to \infty$ и $\frac{\nu_n}{k_n} \xrightarrow{p} 1$;

(B) существует функция распределения $F(x)$ такая, что $\lim\limits_{n\to\infty} P\{\eta_n < x\} = F(x)$, $\forall x \in C(F)$, где $C(F)$ — множества точек непрерывности $F(x)$;

(C) для любых $\varepsilon_1 > 0$, $\varepsilon_2 > 0$ существуют число $c = c(\varepsilon_1, \varepsilon_2) > 0$ и номер $n_0 = n_0(\varepsilon_1, \varepsilon_2)$ такие, что при $n > n_0$

$$P\left\{\max_{|k-n|<cn} |\eta_k - \eta_n| \geq \varepsilon_1\right\} \leq \varepsilon_2.$$

Тогда

$$\lim_{n\to\infty} P\{\eta_{\nu_n} < x\} = F(x), \quad \forall x \in C(F).$$

Десятки работ авторов, таких как А.Реньи, В.Рихтер, С.Гуящу, Ш.Чёрго, Й.Модьороди, Т.А.Азларов, А.А.Джамирзаев и других посвящены обобщению теоремы 1. Например, В.Рихтер [4] доказал теорему 1 на случай когда $\{\eta_n\}$ - обладает свойством перемешивания в смысле А.Реньи, т.е. для любого $A \in F$, $P(A) > 0$,

$$\lim_{n\to\infty} P\{\eta_n < x,\ A\} = P(A) \cdot F(x), \quad \forall x \in C(F)$$

и

$$\frac{\nu_n}{k_n} \xrightarrow{p} \nu_0 > 0,$$

где ν_0 — дискретная с.в. Й.Модьороди [3] обобщил результат В.Рихтера [4] на случай когда ν_0 — необязательно дискретная с.в.

2 Основной результат

Пусть существуют последовательность целых чисел k_n, такая что при $n \to \infty$ $k_n \to \infty$ и двумерная функция распределения $\Psi(x, y)$, что (B1)

$$\lim_{n\to\infty} P\left\{\eta_n < x,\ \frac{\nu_n}{k_n} < y\right\} = \Psi(x, y), \quad \forall (x, y) \in C(\Psi).$$

Положим маргинальные распределения

$$\Psi(+\infty, y) = G(y), \qquad \Psi(x, +\infty) = F(x).$$

В работе доказана следующая

Теорема 2. *Пусть $G(+0) = 0$ и выполнены условия (B2), (C). Тогда*

$$\lim_{n\to\infty} P\left\{\eta_{\nu_n} < x,\ \frac{\nu_n}{k_n} < y\right\} = \Psi(x, y), \quad \forall (x, y) \in C(\Psi).$$

Отметим что из теоремы 2 при $G(1+0) - G(1-0) = 1$ следует теорема 1, но из результатов [4] и [3] не следует теорема 1. Заметим также, что если $\{\eta_n\}$ обладает свойством перемешивания в смысле А.Реньи и при

$$n \to \infty, \quad \frac{\nu_n}{k_n} \xrightarrow{p} \nu_0 > 0,$$

то (см. [1]) следует, что имеет место условия (B1).

Список литературы

[1] Джамирзаев А.А. О свойстве перемешивания в смысле А.Реньи для статистики критерия однородности. *Статистические методы оценивания и проверки гипотез, Пермь, Пермский Гос. Унв*, 21:33–37, 2008.

[2] Anscombe F.J. Large-sampl theory of sequential estimation. *Proc. Cambridje Phil. Soc.*, 48, №4:600–607, 1952.

[3] Mogourodi J. Limit distributions for sequences of random variables with randomdom indices. *Trans. 4 th Prague Conf.Inform. Theory, Statist. Decis. Finet. Ramdom Procese, Prague*, pages 463–470, 1965.

[4] Richter W. Ubertragung von Grenzaussagen für Folgen von zufälligen Grőssen auf Folgen mit zufälligen Indizes. *Теория вероят. и ее прим.*, 10, вып 1:82–83, 1965.

© 2014 А.А. Джамирзаев

Олег Воробьев (ред.), Труды XIII ФАМЭМС'2014, Красноярск: СФУ

Изучение различных постановок задачи о рюкзаке и методов их решения

Мария Михайловна Додонова
Сибирский Федеральный Университет,
Институт математики и фундаментальной информатики
Красноярск
mariya.dodonova.92@mail.ru

Аннотация. *В данной работе изучаются различные варианты постановки задачи о рюкзаке. Рассматриваются и реализуются четыре наиболее популярных алгоритма решения данной задачи: полный перебор, жадный и генетический алгоритмы, а также метод ветвей и границ. Проводится классификация и сравнение этих методов по скорости решения и точности нахождения решения. Также находится решение практической задачи о рюкзаке.*

Ключевые слова. *Задача о рюкзаке, оптимизация, комбинаторика, полный перебор, жадный алгоритм, генетический алгоритмы, метод ветвей и границ.*

1 Введение

Практика порождает все новые и новые задачи оптимизации, причем их сложность растет. Реальные прикладные задачи дискретной оптимизации очень сложны. Современные методы оптимизации далеко не всегда справляются с решением реальных задач без помощи человека. Требуются новые математические модели и методы, которые позволяют учитывать наличие многих критериев и проводят глобальный поиск оптимума. Следует отдавать предпочтение таким методам, которыми проще управлять в процессе решения задачи.

Целью данной работы является изучение различных вариантов постановки задачи о рюкзаке, реализация основных методов решения данных задач, классификация и сравнение этих методов.

Задача о ранце (рюкзаке) — одна из NP-трудных задач комбинаторной оптимизации [1]. Свое название задача получила от оптимизационной задачи укладки как можно большего числа ценных вещей в рюкзак при условии, что общий объём (или вес) всех предметов, способных поместиться в рюкзак, ограничен. Задача о рюкзаке является актуальной и достаточно востребованной с точки зрения ее приложения в реальной жизни. Задача о загрузке (о рюкзаке) и её модификации часто возникают в экономике, прикладной математике, криптографии, генетике и логистике для нахождения оптимальной загрузки транспорта (самолета, поезда, трюма корабля) или склада [3].

В работе рассматриваются четыре наиболее популярных алгоритма решения данной задачи: полный перебор, жадный и генетический алгоритмы, а также метод ветвей и границ (как сокращение метода полного перебора). Все методы разделены на две группы. Первую группу образуют точные методы, к которым относятся полный перебор и метод ветвей и границ. Вторая группа — приближенные методы, она представлена жадным и генетическим алгоритмом.

2 Задача о рюкзаке

2.1 Классическая постановка задачи

Пусть имеется набор предметов, каждый из которых имеет два параметра — вес и ценность. Имеется рюкзак с некоторым заданным значением вместимости. Задача заключается в том, чтобы собрать рюкзак с максимальной ценностью предметов внутри, соблюдая при этом весовое ограничение рюкзака.

Математически постановка задачи формулируется следующим образом:

Пусть имеется n предметов. Для каждого i-го предмета задан его вес $w_i > 0$ и стоимость (ценность) $p_i > 0$, $i = 1, 2, ..., n$. Задано ограничение на максимальный вес рюкзака — W. Каждый x_i может принимать только одно из двух значений: $x_i = 1$, если i-й предмет упаковывают в рюкзак, или $x_i = 0$, в противном случае.

Требуется выбрать из заданного множества предметов набор с максимальной суммарной стоимостью

$$\sum_{i=1}^{n} p_i x_i \tag{1}$$

при одновременном соблюдении ограничения на суммарный вес найденного набора

$$\sum_{i=1}^{n} w_i x_i \leq W. \tag{2}$$

© 2014 М.М.Додонова

Олег Воробьев (ред.), Труды XIII ФАМЭМС'2014, Красноярск: СФУ

Классическая задача о ранце (КЗР) относится к числу широко известных задач дискретной оптимизации. Впервые КЗР была сформулирована Д. Данцигом и с тех пор активно исследуется. Ее популярность, прежде всего, вызвана большим количеством приложений. Актуальность исследования предопределена широкой распространенностью и важностью прикладных проблем, формулируемых в рамках многокритериальных задач ранцевого типа.

2.2 Различные постановки задачи о рюкзаке

Существует несколько разновидностей задачи о ранце, отличия между которыми заключаются в условиях, наложенных на рюкзак, предметы или их выбор:

1. *Рюкзак 0-1* или *Классическая задача* — каждый предмет можно брать только один раз;
2. *Ограниченный рюкзак* — каждый предмет можно брать определенное количество раз ;
3. *Неограниченный рюкзак* — каждый предмет можно брать сколько угодно раз;
4. *Непрерывный рюкзак* — возможно брать любую дробную часть от предмета;
5. *Рюкзак с мультивыбором* — есть несколько классов предметов, из которых можно брать одного представителя, причём некоторые вещи имеют больший приоритет, чем другие;
6. *Мультипликативный рюкзак* — есть несколько рюкзаков с различной вместимостью;
7. *Многомерный рюкзак* — есть более одного ограничения на рюкзак.

Рюкзак 0-1
Это самая распространенная разновидность рюкзака. Постановка приведена ранее.

Ограниченный рюкзак

Обобщение классической задачи, когда любой предмет может быть взят некоторое количество раз. Каждый предмет x_i может быть выбран ограниченное число раз. Необходимо максимизировать ценность рюкзака

$$\sum_{i=1}^{n} p_i x_i \tag{3}$$

чтобы при этом выполнялось условие

$$\sum_{i=1}^{n} w_i x_i \leq W. \tag{4}$$

где $x_i \in (0, 1, ..., b_i)$ для всех $i = 1, 2, \ldots, n$. Число b_i называют границей.

Неограниченный рюкзак
Каждый предмет x_i может быть выбран неограниченное число раз. Необходимо максимизировать ценность рюкзака (4), чтобы при этом выполнялось условие ограничения на его вес (1) и целое $x_i \geq 0$ для всех $i = 1, 2, \ldots, n$.

Непрерывный рюкзак
Сходна с предыдущей, только x_i является вещественным числом. Данная задача представляет собой задачу линейного программирования.

Рюкзак с мультивыбором
Имеется множество предметов $Q = \{q_1, q_2, \ldots, q_n\}$, которые разделены на m классов $S_i = q_{ij}$, $i = 1, 2, \ldots, m$. Задано количество предметов в каждом классе $n_i = |S_i|$, $\sum\limits_{i=1}^{m} = n$. То есть

$$\sum_{i=1}^{m} \sum_{j=1}^{n_i} q_{ij} = Q.$$

Имеется рюкзак объема $W \in N$. Для каждого j-го предмета из i-го рюкзака q_{ij} задана его стоимость $p_{ij} > 0$ и определен вес $w_{ij} > 0$. Требуется упаковать рюкзак так, чтобы в нём обязательно присутствовал предмет — представитель каждого класса, причем общая ценность упакованных предметов была наибольшей, а их общий вес не превосходил W.

Для каждого предмета q_{ij} введем двоичные переменные x_{ij} следующим образом:

- $x_{ij} = 1$, если предмет выбран для упаковки,
- $x_{ij} = 0$, в противном случае.

Требуется найти

$$\sum_{i=1}^{m} \sum_{j=1}^{n_i} p_{ij} x_{ij} \to max \tag{5}$$

при выполнении ограничения на вес

$$\sum_{i=1}^{m} \sum_{j=1}^{n_i} w_{ij} x_{ij} \leq W \tag{6}$$

и условия на наличие представителя каждого класса

$$\sum_{j=1}^{n_i} x_{ij} > 0, \ i = 1, 2, \ldots, m. \tag{7}$$

Мультипликативный рюкзак
Пусть имеются n предметов и m рюкзаков ($m \leq n$). У каждого рюкзака определена своя вместимость P_i, $i = 1, 2, \ldots, m$, $x_i \in 0, 1$. Требуется выполнить максимизацию

$$\sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} \tag{8}$$

так чтобы выполнялось условие

$$\sum_{i=1}^{n} p_j x_{ij} \leq P_i \tag{9}$$

для всех $i = 1, 2, \ldots, n$. Причем $\sum_{j=1}^{m} x_{ij} = 1$.

Многомерный рюкзак

Задача о рюкзаке является m-мерной задачей, если есть более одного ограничения на рюкзак (например, объем и вес). Тогда формулировка задачи (для неограниченного варианта) имеет следующий вид: необходимо максимизировать

$$\sum_{i=1}^{n} c_i x_i \tag{10}$$

так чтобы

$$\sum_{i=1}^{n} p_{ij} x_i \leq P_j, \tag{11}$$

для $j = 1, 2, \ldots, m$ и $x_i \geq 0$ для всех $i = 1, 2, \ldots, n$.

3 Классификация и сравнение методов решения задачи о рюкзаке

Существует ряд методов, позволяющих находить решение задачи о ранце [4], [2]. В данной работе рассматривались следующие методы (для всех семи приведенных выше постановок задачи):

1. Полный перебор,
2. Метод ветвей и границ,
3. Жадный алгоритм,
4. Генетический алгоритм

Задача о рюкзаке является NP-трудной, т.е. для нее не существует полиномиального алгоритма, решающего ее за разумное время. Если перебирать всевозможные подмножества данного набора из n предметов, то получится решение сложности не менее чем $O(2^n)$. В настоящее время неизвестен (и, скорее всего, вообще не существует) алгоритм решения этой задачи, сложность которого является многочленом от n.

В работе рассматривался практический пример задачи о рюкзаке с максимальной вместимостью рюкзака $P = 30$ и количеством предметов $n = 20$.
С помощью перечисленных методов были найдены решения всех семи вариантов задачи о рюкзаке, а также выполнено сравнение методов по скорости и точности нахождения решения. Проведенное сравнение показано в таблице 1.

При выборе алгоритма решения приходится выбирать между точными алгоритмами, которые не применимы для рюкзаков большой размерности, и приближенными, которые работают быстро, но не обеспечивают оптимального решения задачи. Выбор использования того или иного метода является спорным вопросом. Всё зависит от постановки задачи, а также от того, какие цели поставлены. Если требуется найти точное решение, то необходимо использовать точные методы. В ситуации небольшого набора входных данных (до 10–20 предметов), лучше применять методы перебора или ветвей и границ, в силу простоты реализации. Если же точность решения не играет большой роли, или входные данные таковы, что ни один из точных методов не работоспособен, то для решения задачи можно использовать лишь приближенные алгоритмы.

	Предмет	**Вес**	**Стоимость**
1	рыба	6	5
2	молоко	4	3
3	сок	3	1
4	хлеб	2	3
5	телефон	5	6
6	мясо	7	9
7	вода	4	8
8	бумага	2	6
9	яблоки	9	10
10	сыр	3	7
11	колбаса	5	9
12	огурцы	6	9
13	клей	2	5
14	вата	4	5
15	тетрадь	5	10
16	кефир	7	8
17	кофе	4	7
18	курица	9	10
19	масло	5	7
20	шоколад	3	6

Таблица 1: Предметы для практической задачи рюкзака

Выводы

Поскольку метод полного перебора очень трудоемкий, то размер рюкзака и количество предметов при его использовании должны быть ограничены. Следовательно, этот метод неприменим для неограниченного и непрерывного рюкзаков. При решении перебором задач мультипликативного и ограниченного рюкзаков необходимо уменьшить количество предметов до десяти, чтобы решение было найдено.

Очевидным недостатком алгоритма метода ветвей и границ при решении задач большой размерности является необходимость перебрать слишком большое количество вариантов перед тем, как будет найден оптимальный. Несомненно, метод ветвей повто-

Метод	Сложность	Плюсы	Минусы
Полный перебор	$O(N!)$	простота реализации; точное решение	входные данные невелики; большая временная сложность
Метод ветвей и границ		возможно значительное сокращение времени; простота реализации	работает как полный перебор
Жадный алгоритм	$O(N * Log(N))$	высокая скорость; может работать с большими значениями N; простота реализации	решение неточное
Генетический алгоритм		высокая скорость; работа с большими N; независимость от вида исходных данных	не гарантирует нахождение оптимального решения
Метод динамического программирования	$O(W * N)$	независимость от вида исходных данных; точное решение	большой объем вычислительной работы

Таблица 2: Сравнительный анализ методов

ряет перебор. Однако, благодаря отсечению «бесполезных» решений, может самостоятельно ограничить бесконечное пространство решений. Поэтому его можно применить для неограниченного рюкзака, но нельзя применить для непрерывного рюкзака.

Жадный алгоритм является однокритериальным алгоритмом, поэтому он неприменим для многомерного рюкзака. Однако следует отметить, что для задачи непрерывного рюкзака именно этот алгоритм находит наиболее оптимальное решение. Генетический алгоритм является одним из самых быстрых алгоритмов. Но ограничением этого алгоритма является то, что его хромосомы кодируются дискретно. Соответственно, его нельзя использовать для решения непрерывных и бесконечных задач. Недостатком генетического алгоритма также является то, что он не гарантирует нахождение оптимального решения в некоторых ситуациях (находит локальный экстремум вместо глобального). Происходит это вследствие того, что алгоритм может заканчиваться не только при достижении оптимального решения, но и следующих условиях:

- пройдено максимальное заданное число итераций;
- прошло максимальное время, заданное для выполнения алгоритма;
- при переходе к новому поколению не происходит существенных изменений.

Список литературы

[1] Х. Х. Пападимитриу и К. Стайглиц. *Комбинаторная оптимизация. Алгоритмы и сложность.* Москва, Мир, 1985.

[2] И. Х. Сигал и А.П. Иванова. *Введение в прикладное дискретное программирование: модели и вычислительные алгоритмы.* Москва, Физматлит, 2003.

[3] Н. Н. Писарук. *Исследование операций.* Москва, МГУ, 2014.

[4] Т. Кормен. *Алгоритмы: построение и анализ.* Москва, Вильямс, 2005.

Решение обратной задачи теории риска для меры возмущенной вероятности

Есин Роман Витальевич
Сибирский Федеральный Университет
Институт космических и информационных технологий
Красноярск
surgeon14@mail.ru

Аннотация. *В данной работе излагается метод решения обратной задачи теории риска для меры возмущенной вероятности, основанный на вычислении функционала возмущенной вероятности, неприятия риска и на сравнении решений индивидуума с решениями, принимаемыми с использованием различных типов возмущающих функций.*

Ключевые слова. *Теория риска, когерентные меры риска, неприятие риска, функционал возмущенной вероятности.*

1 Введение

При принятии экономических или управленческих решений обычно не удается спрогнозировать результат каждого индивидуального решения с удовлетворительной точностью. Поэтому приходится признавать неопределенность последствий принятых решений, и, как следствие, для принятия рациональных решений использовать соответствующую теорию. Если индивидуум хочет принять решение на основании собственного отношения к риску, то 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. Обратной задачей будем считать построение меры риска, которая строится по каким-либо известным характеристикам предпочтения этого человека. Решим данную задачу в рамках меры возмущенной вероятности.

2 Основные понятия и определения

Рассмотрим вероятностное пространство $(\Omega, \mathcal{A}, \mathbf{P})$, где Ω — множество элементарных исходов, $\mathcal{A}$ - σ-алгебра, заданная на Ω, а $\mathbf{P}$ — вероятностная мера, определенная на множествах из $\mathcal{A}$.

© 2014 Р.В. Есин

Олег Воробьев (ред.), Труды XIII ФАМЭМС'2014, Красноярск: СФУ

Риском X на $(\Omega, \mathcal{A})$ называется произвольное измеримое отображение из Ω 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 $\preccurlyeq$, если это отношение обладает свойствами:

- полноты: для произвольной пары $r_1,\ r_2 \in \mathcal{R}$ либо $r_1 \preccurlyeq r_2$, либо $r_2 \preccurlyeq r_1$, либо справедливы оба этих отношения;
- транзитивности: если $r_1 \preccurlyeq r_2$ и $r_2 \preccurlyeq r_3$, то $r_1 \preccurlyeq r_3$.

Элементы $r_1,\ r_2 \in \mathcal{R}$, для которых верно $r_1 \preccurlyeq r_2$, $r_1 \preccurlyeq r_2$, называются эквивалентными: $r_1 \sim r_2$. Пусть на заданы отношения (частичного) порядка $\leq$ и предпочтения $\preccurlyeq$. Предпочтение $\preccurlyeq$ называют согласованным с порядком $\leq$, если $r_1 \leq r_2 \Longrightarrow r_1 \preccurlyeq r_2$. Данное отношение предпочтения отражает индивидуальное отношение к риску некоторого индивидуума.

Мерой риска называется произвольный функционал $\rho : \mathcal{X} \to \mathbf{R}$. Говорят, что ρ представляет отношение предпочтения на $\mathcal{X}$, клгда

$$\rho(X) \leq \rho(Y), \text{ если } X \preccurlyeq Y, \text{ где } X, Y \in \mathcal{X}.$$

Индивидуум, характеризуемый отношением предпочтения $\preccurlyeq$, не приемлет риск, если для произвольного риска $\Delta : E\Delta = 0$ и для произвольного $a \in \mathbf{R}$ имеет место $a + \Delta \prec a$. Неприятие риска можно истолковать следующим образом: если к детерминированной величине a прибавить некоторую невырожденную случайную величину Δ с нулевым средним, то результат $a + \Delta$ менее предпочтителен для индивидуума, чем a. При регулярном отношении предпочтения это означает, что найдется $c > 0$ такое, что $a + \Delta \sim a - c$

3 Когерентные меры риска

В работе [6], было введено понятие когерентной меры риска. Следуя работе [7] изложим их в терминах

супермодулярных функций, который являются эквивалентным описанием когерентных мер риска, точнее, если $\rho : \mathcal{X} \to \mathbf{R}$ - когерентная мера риска, то f - супермодулярная функция, связанная с ρ 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 $f : \mathcal{X} \to \mathbf{R}$, обладающий свойствами монотонности, супераддитивности, положительной однородности и инвариантности относительно сдвигов, соответственно:

- $X \leq Y \Longrightarrow f(X) \leq f(Y)$
- $f(X+Y) \geq f(X) + f(Y)$
- $f(\lambda X) = \lambda f(X),\ \lambda > 0$
- $f(X + aI) = f(X) + a,\ a \in \mathbf{R}$

Здесь X, Y — произвольные случайные величины из $\mathcal{X}$, а $I = (1, 1, \ldots, 1)$ — 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 непринятия 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 Функционал возмущенной вероятности

Мера возмущенной вероятности первоначально предназначалась для вычисления страховой премии, но может быть использована и в более широком классе задач, включая портфельный анализ.

Обозначим $\mathcal{X}$ множество всех вещественных случайных величин, а $\mathcal{X}_+$ - множество неотрицательных случайных величин:

$$\mathcal{X}_+ = \{X \in \mathcal{X} \mid \mathbf{P}\{X \geq 0\} = 1\}$$

Пусть, далее, $F_X(x) = \mathbf{P}\{X \leq x\},\ x \in \mathbf{R}$ есть функция распределения случайной величины $X \in \mathcal{X}$, $S_X(x) = \mathbf{P}\{X > x\} = 1 - F_X(x),\ x \in \mathbf{R}$ — ее дополнительная функция распределения.

Пусть $g : [0,1] \to [0,1]$ — неубывающая функция, причем $g(0) = 0,\ g(1) = 1$. $g(x)$ называется возмущающей функцией.

В [8] была введена мера риска

$$\pi_g(X) = \pi(X) = \int_0^\infty g(S_X(x))\ dt,\ X \in \mathcal{X}_+, \quad (1)$$

а в [9] предложена модификация для распределений на всей вещественной оси $\mathbf{R}$

$$\pi_g(X) = \pi(X) = \int_{-\infty}^0 [g(S_X(x)) - 1]\ dt + \int_0^{+\infty} g(S_X(x))\ dt \quad (2)$$

Для $b \in \mathbf{R}$ обозначим W_b вырожденную случайную величину: $\mathbf{P}\{W_b = b\} = 1$ и рассмотрим некоторые свойства π.

Мера риска π, обладающий следующими свойствами:

- $\pi(W_b) = b, b \in \mathbf{R}$
- $\pi(X + b) = \pi(X) + b,\ b \in \mathbf{R},\ X \in \mathcal{X}$
- $\pi(aX) = a\pi(X),\ a > 0, X \in \mathcal{X}$

Пусть распределение случайной величины X сосредоточено в конечном числе точек:

$$\mathbf{P}\{X = x_k\} = p_k,\ k = 1, \ldots, n;$$

$$p_1 \geq 0, \ldots, p_n \geq 0,\ p_1 + \cdots + p_n = 1$$

Без ограничения общности можно считать, что точки x_k расположены в порядке возрастания:

$$x_1 < x_2 < \cdots < x_n$$

Тогда значение меры возмущенной вероятности для такого распределения может быть вычислено по формуле

$$\pi_g(X) = \sum_{s=1}^{n} g(\sum_{k=s}^{n} p_k) x_{s,s-1} \quad (3)$$

где $x_0 = 0,\ x_{s,s-1} = x_s - x_{s-1}$.

Введем количественную характеристику для измерения степени неприятия риска. Поскольку добавление случайной величины Δ к начальному капиталу a снижает предпочтительность результата, распределение $a + \Delta$ покидает класс эквивалентности, содержащий W_a. Однако, новый класс эквивалентности, в который попало распределение $a + h\Delta$, где $h = \|\Delta\|_\infty$, как и любой другой класс эквивалентности, содержит некоторое вырожденное распределение, например, W_{a-c}, где $c > 0$ можно истолковать, как плату, которую потребует индивидуум за добавление к своему капиталу случайной величины (риска) .

Если начальный капитал индивидуума равен детерминированной величине a, то какова норма замещения рискового довеска $h\Delta$. Получаем

$$\pi(a + h\Delta) = a + h\pi(\Delta),\ \pi(a - c) = a - c$$

поэтому условие принадлежности $a+h\Delta$ и $a-c$ к одному классу эквивалентности приводит к уравнению $a + h\pi(\Delta) = a - c$, откуда

$$c = -h\pi(\Delta) \tag{4}$$

Из (2.8) ясно, что индивидуум, предпочтения которого представляются функционалом возмущенной вероятности, не приемлет риск, если $\pi(\Delta) < 0$ для произвольной невырожденной случайной величины с нулевым средним. Как показано в [3], последнее условие в терминах возмущающей функции g имеет вид $g(x) \leq x,\ x \in [0,1]$.

5 Обратная задача теории риска для меры возмущенной вероятности

Пусть мы являемся финансовой компанией, которая занимается вложением денег в рисковые активы. К нам пришел инвестор, который хочет вложить деньги в инвестиционный проект в соответствии со своим отношением предпочтения. Так как мы работаем мерой возмущенной вероятности, нам необходимо решить несколько задач:

- Подобрать параметр функционала π_g, в виде функции g, причем должны выполняться условия $g(0) = 0,\ g(1) = 1$;
- В работе Вонга [8], были предложены несколько классов функции g со входным параметром r, но все g являются выпуклыми вверх функциями. Для того чтобы функционал возмущенной вероятности учитывал неприятие риска, функция g должна удовлетворять условию: $g(x) \leq x,\ x \in [0,1]$, то есть g должна быть вогнутой на промежутке $[0,1]$;
- Выразить параметр r функции g через неприятие риска;
- На основе отношения предпочтения инвестора и его уровня неприятия риска подобрать для него возмущающую функцию g и параметр функции r.

5.1 Классы функции g

Рассмотрим несколько классов возмущающей функции $g(x)$:

1. Класс квадратичных функций $Q(x)$.
2. Класс радикальных функций $R(x)$.
3. Класс экспоненциальных функций $E(x)$.
4. Класс логарифмических функций $L(x)$.
5. Класс степенных функций $S(x)$.

5.1.1 Класс квадратичных возмущающих функций

$$Q(x) = (1+r)x - rx^2. \tag{5}$$

Найдем интервалы вогнутости функции, для этого найдем её вторую производную и решим неравенство $Q''(x) \geq 0$. Получаем, что функция $Q(x)$ является вогнутой $\forall x$ при $r \leq 0$, но при $r \leq -1$, функция $Q(x)$ является убывающей, а по условию в [4] функции из классов g должны быть неубывающей. Значит для функции $Q(x)$ параметр r находится на отрезке $-1 \leq r \leq 0$.

Попробуем выразить параметр функции $Q(x)$ через меру возмущенной вероятности, для этого воспользуемся формулой для дискретного распределения, получаем:

$$r = \frac{\pi_Q(X) + \sum\limits_{s=1}^{n}\left(\sum\limits_{k=s}^{n} p_k x_{k,k-1}\right)}{\sum\limits_{s=2}^{n}\left(\left(\sum\limits_{k=s}^{n} p_k - \sum\limits_{k=s}^{n} p_k^2\right)x_{s,s-1}\right)}.$$

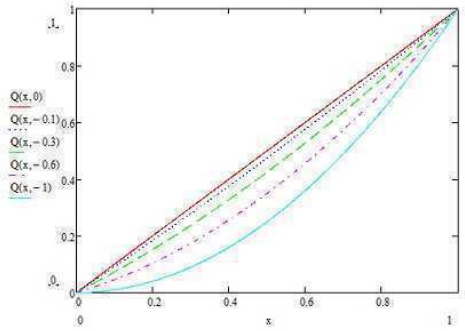


Рисунок 1: Вид квадратичной функции при разных параметрах r

5.1.2 Класс радикальных возмущающих функций

$$R(x) = \begin{cases} \frac{\sqrt{1+rx}-1}{\sqrt{1+r}-1}, & r \neq 0; \\ x, & r = 0. \end{cases} \tag{6}$$

Функция $R(x)$ является вогнутой $x \in [0,1]$ при $-1 < r < 0$.

Чтобы получить параметр r из этого класса функций $R(x)$, необходимо решить уравнение относительно r:

$$\pi_R(X) = x_{1,0} + x_{2,1}\frac{\sqrt{1+r(p_2+\cdots+p_n)}-1}{\sqrt{1+r}-1} + .$$

$$+\cdots+ x_{n,n-1}\frac{\sqrt{1+rp_n}-1}{\sqrt{1+r}-1}. \tag{7}$$

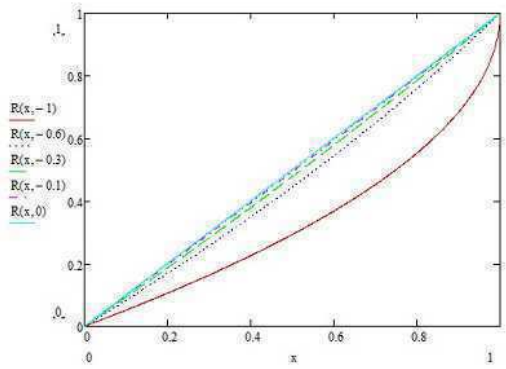


Рисунок 2: Вид радикальной функции при разных параметрах r

5.1.3 Класс экспоненциальных возмущающих функций

$$E(x) = \begin{cases} \frac{1-e^{-rx}}{1-e^{-r}}, & r \neq 0; \\ x, & r = 0. \end{cases} \tag{8}$$

Получаем, что функция $E(x)$ является вогнутой $\forall x$ при $r < 0$.

Чтобы получить параметр r из этого класса функций E, необходимо решить уравнение относительно r:

$$\pi_E(X) = x_{1,0} + x_{2,1}\frac{1-e^{-r(p_2+\cdots+p_n)}}{1-e^{-r}} + \\ + \cdots + x_{n,n-1}\frac{1-e^{-rp_n}}{1-e^{-r}} \tag{9}$$

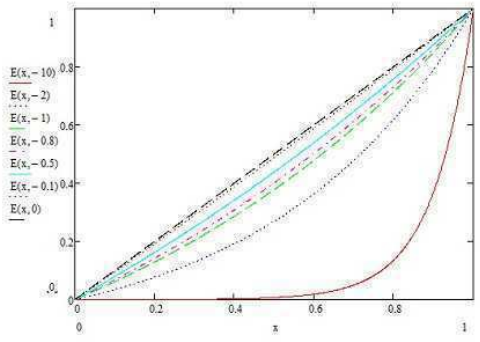


Рисунок 3: Вид экспоненциальной функции при разных параметрах r

5.1.4 Класс логарифмических возмущающих функций

$$L(x) = \begin{cases} \frac{\ln(1+rx)}{\ln(1+r)}, & r \neq 0; \\ x, & r = 0. \end{cases} \tag{10}$$

Получаем, что функция $L(x)$ является вогнутой $x \in [0,1]$ при $-1 < r < 0$.

Чтобы получить параметр r из этого класса функций L, необходимо решить уравнение относительно r:

$$\pi_L(X) = x_{1,0} + x_{2,1}\frac{\ln(1+r(p_2+\cdots+p_n))}{\ln(1+r)} + \\ + \cdots + x_{n,n-1}\frac{\ln(1+rp_n)}{\ln(1+r)}. \tag{11}$$

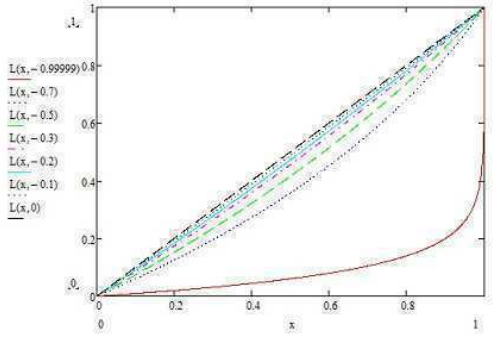


Рисунок 4: Вид логарифмической функции при разных параметрах r

5.1.5 Класс степенных возмущающих функций

$$S(x) = x^r \tag{12}$$

Функция $S(x)$ является вогнутой $\forall x$ при $r > 1$ Чтобы получить параметр r из этого класса функций S, необходимо решить уравнение относительно r:

$$\pi_S(X) = x_{1,0} + x_{2,1}(p_2+\cdots+p_n)^r + \cdots + x_{n,n-1}p_n^r. \tag{13}$$

Для $n = 2$ получим

$$r = \frac{\ln(\frac{\pi_S(X)-x_{1,0}}{x_{2,1}})}{\ln(p_2)}.$$

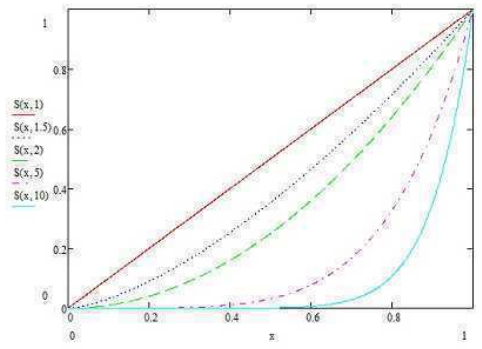


Рисунок 5: Вид степенной функции при разных параметрах r

5.2 Выбор класса функции g на основе отношения предпочтения

Для того чтобы определить уровень неприятия риска у инвестора и определить какой класс возмущающих функций наиболее точно описывает его отношение предпочтения проведем некоторый опрос.

Предложим инвестору некоторую лотерею, где риск соответствует доходу:

W	w_1	w_2	$\dots$	w_n
P	p_1	p_2	$\dots$	p_n

$p_1 + p_2 + \dots + p_n = 1$.

И спросим, какова минимальная оплата, за которую инвестор согласится сыграть в такую лотерею? Узнав, мы получим количественную характеристику неприятия риска у инвестора c. Значение c будет являться решением уравнения $\pi_g(a - c) = \pi_g(a + h\Delta)$ и будет зависеть от g и Δ. Обозначим это значение через $c_g(\Delta)$: его можно трактовать как цену, которую индивидуум запрашивает за "подмешивание"к его изначальному капиталу a случайного возмущения в "направлении"Δ. Далее представим

W	w_1	w_2	$\dots$	w_n
P	p_1	p_2	$\dots$	p_n

в виде некоторой детерминированной величины a и некоторой невырожденной случайной величины Δ с нулевым средним $E\Delta = 0$. Для этого решим СЛАУ методом Гаусса, лотереи будем подбирать так, чтобы СЛАУ имела решение, причем $a > 0$:

$$\begin{cases} w_1 = a + \Delta_1 \\ w_2 = a + \Delta_2 \\ \dots \\ w_n = a + \Delta_n \\ 0 = \Delta_1 p_1 + \Delta_2 p_2 + \dots + \Delta_n p_n. \end{cases} \tag{14}$$

В результате получим

Δ	Δ_1	Δ_2	$\dots$	Δ_n
P	p_1	p_2	$\dots$	p_n

со средним 0 и сможем представить $W = a + \Delta$. Отсюда найдем значение $h = \|\Delta\|_\infty = \max_i(|\Delta_i|)$. Условие принадлежности $a + h\Delta$ и $a - c$ к одному классу эквивалентности приводит нас к уравнению $a + h\pi(\Delta) = a - c$, откуда мы можем найти цену неприятия риска $c = -h\pi_g(\Delta)$, а так же $\pi_g(W) = a + \frac{-c}{h}$. Для того чтобы определить параметр r для класса квадратичных возмущающей функции g решим уравнение

$$r = \frac{\pi_Q(W) + \sum\limits_{s=1}^{n}\left(\sum\limits_{k=s}^{n} p_k w_{k,k-1}\right)}{\sum\limits_{s=2}^{n}\left(\left(\sum\limits_{k=s}^{n} p_k - \sum\limits_{k=s}^{n} p_k^2\right) w_{s,s-1}\right)},$$

чтобы найти параметр r для остальных классов нам необходимо численно решить уравнения (7), (9), (11), (13) относительно параметра r.

В итоге мы получим параметр r для каждого класса функций g. На последнем этапе нам необходимо определить какой класс возмущающих функций g наиболее точно описывает отношение предпочтения конкретного инвестора.

Предложим инвестору на выбор две лотереи, где риск соответствует доходу:

$Y1$	$y1_1$	$y1_2$	$\dots$	$y1_n$
P	p_1	p_2	$\dots$	p_n

$Z1$	$z1_1$	$z1_2$	$\dots$	$z1_n$
P	p_1	p_2	$\dots$	p_n

$p_1 + p_2 + \dots + p_n = 1$ и узнаем, участие в какой из них для него более предпочтительно. Для этой лотереи найдем значения:

$$d_1 = \pi_Q(Y1) - \pi_Q(Z1),$$
$$d_2 = \pi_R(Y1) - \pi_R(Z1),$$
$$d_3 = \pi_E(Y1) - \pi_E(Z1),$$
$$d_4 = \pi_L(Y1) - \pi_L(Z1),$$
$$d_5 = \pi_S(Y1) - \pi_S(Z1).$$

Мы будем выбирать $Y1$ и $Z1$ таким образом, чтобы некоторые d_i были больше 0, что означает, что для инвестора, предпочтения которого описаны соответствующими функциями: $Y1$ будет предпочтительнее $Z1$, а другие d_i были меньше 0, в противном случае. Таким образом, если инвестор выберет $Y1$, то далее мы будем выбирать функцию g среди тех, для которых $d_i > 0$, иначе будем выбирать среди тех функций, для которых $d_i < 0$. Далее предложим инвестору на выбор другие две лотереи, где риск является соответствующим доходом:

$Y2$	$y2_1$	$y2_2$	$\dots$	$y2_n$
P	p_1	p_2	$\dots$	p_n

$Z2$	$z2_1$	$z2_2$	$\dots$	$z2_n$
P	p_1	p_2	$\dots$	p_n

Поведя ту же самую процедуру, мы вновь на основе предпочтения инвестора определим какие функции наиболее точно описывают неприятие риска конкретного человека. Будем повторять этот алгоритм до тех пор, пока не определим одну единственную функцию, что в итоге даст нам необходимый класс возмущающих функций g. Зная класс возмущающих функций и параметр r, который мы получили на предыдущем шаге, мы можем предоставить инвестору конкретную функцию для вычисления меры возмущенной вероятности в соответствии с уровнем неприятия риска и отношением предпочтения конкретного индивида.

6 Итоги

В работе предложен метод решения обратной задачи теории риска, основанный на вычислении функционала, неприятия риска и на сравнении решений индивидуума с решениями, принимаемыми с использованием различных типов возмущающих функций. Изложенные методы для тестирования могут стать основой для создания процедуры подбора подходящего функционала возмущающей вероятности по индивидуальным предпочтениям. Это может быть более развернутое анкетирование, либо различные модели деловых игра, либо схема анализа ранее принятых инвестиционных решений.

Список литературы

[1] Новоселов А. А. *Математическое моделирование финансовых рисков. Теория измерения.* Наука, Новосибирск, 2001.

[2] Новоселов А. А. О неприяти риска и норме замещения риска доходностью. *ИВМ СО РАН*, 2001.

[3] Новоселов А. А. Непринятие риска: качественный подход и количественные оценки. *Автоматика и телемеханика*, 7:165–177, 2003.

[4] Кустицкая Т. А. Решение обратной задачи теории риска в рамках можеди обобщенных когерентных мер риска (для евклидовой нормы). *Тр. XVI Междун. ЭМ-конференции*, Красноярск, 2012.

[5] Acerbi C., Tasche D. Expected shortfall: a natural alternative to value-at-risk. *Economic notes*, 31:379–388, 2001.

[6] Artzner Ph., Delbaen F., Ebler J.-M., Heath D. Coherent measures of risk. *Mathmatical Finance*, 9:203–228, 1999.

[7] Delbaen F. Coherent risk measures on general probability spases. *Advance of Finance and Stochstics*, pages 10–38, 2002.

[8] Wang S. Premium calculation by transforming the layer premium density. *ASTIN Bulletin*, 1:71–92, 1996.

[9] Young V.R. Discussion of christofidesЎ conjecture regarding wangЎs premium principle. *ASTIN Bulletin*, 2:191Ц195, 1999.

Центральная предельная теорема для случайных сумм линейного процесса, порожденного последовательностью m-зависимых случайных

Талат Маруфович Зупаров
Национальный Университет Узбекистана
Узбекистан
tm_zuparov@rambler.ru

Ойбек Сафаров
Национальный Университет Узбекистана
Узбекистан

Аннотация. *В работе доказана центральная предельная теорема для случайных сумм линейного процесса, порожденного последовательностью m-зависимых случайных.*

Ключевые слова. *Стационарная последовательность случайных величин, линейный процесс, центральная предельная теорема, m-зависимые случайные величины, случайная сумма.*

1 Введение

Пусть $\{\xi_i, i \in Z\}$ — стационарная в узком смысле последовательность случайных величин (с.в.).Будем предполагать, что величины ξ_i имеют нулевые средние и конечные дисперсии. Пусть далее $\{a_{iN}, i \in Z, N \geq 1\}$ — последовательность чисел такая, что

$$\sum_{i=-\infty}^{\infty} a_{in}^2 < \infty.$$

Последовательность с.в. $\{X_{kn}, 1 \leq k \leq n, n \geq 1\}$ называется линейным процессом, имеющим коэффициенты $\{a_{iN}, i \in Z, N \geq 1\}$ и порожденными случайными величинами $\{\xi_i, i \in Z\}$, если ряд $\sum\limits_{i=-\infty}^{\infty} a_{in}\xi_{k-i}$ сходится с вероятностью 1 и

$$X_k = X_{kn} = \sum_{i=-\infty}^{\infty} a_{in}\xi_{k-i}.$$

Положим

$$S_n = \sum_{k=1}^{n} X_k, B_n^2 = DS_n, F_n(x) = P(S_n < xB_n),$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}} du,$$

$$\eta_n = \sum_{k=1}^{n} \xi_k, n = 1, 2, ...,$$

$$\Delta_x(F_n, \Phi) = |F_n(x) - \Phi(x)|,$$

$$\Delta(F_n, \Phi) = \sup_x \Delta_x(F_n, \Phi),$$

Асимптотическая нормальность S_n/B_n и оценки скорости сходимости к нулю величин $\Delta_x(F_n, \Phi)$ и $\Delta(F_n, \Phi)$ когда $\{\xi_i, i \in Z\}$ — слабозависимая последовательность с.в., изучена достаточно подробно (см.,например обзорную статью [2]).Можно перенести многие результаты, полученные в теории суммировании с.в. для линейных процессов, порожденных последовательностью m-зависимых с.в.

Последовательность с.в. $\{\xi_i, i \in Z\}$ называется m-зависимой, если любые два вектора вида $(\xi_{i_1}, \ldots, \xi_{i_k})$ и $(\xi_{i_k+j}, \xi_{j_1} \ldots, \xi_{j_k})$ независимы всякий раз. Когда $j > m$, где $-\infty \leq i_1 < i_2 < \cdots < i_k; i_k + j \leq j_1 < j_2 < \cdots < j_l \leq \infty$ k и произвольные целые положительные числа.

Центральная предельная теорема для последовательности независимых одинаково распределенных с.в., доказанной Леви, была обобщена к случаю m-зависимых с.в. в работах Hoefding и Robbins [8], Diananda [12], Orey [13], Bergstrom [7] и Зупаров и Чуянова [1]. В работе [4] рассмотрен случай, когда m может стремится к бесконечности.

В работе доказана центральная предельная теорема для случайных сумм линейного процесса, порожденного последовательностью m-зависимых с.в.

2 Основной результат

Классическая центральная предельная теорема для строго стационарной последовательности m-зависимых с.в. имеет следующий вид. Доказательство этого результата см.[8].

Теорема 1. *Пусть $\{\xi_i\}_{i\geq 1}$ строго стационарная последовательность m-зависимых случайных величин. Пусть $E\xi_i = a, 0 < D\xi_i = b^2 < \infty$. Тогда при $n \to \infty$*

$$\frac{\sqrt{n}}{\sigma}(\frac{\eta_n}{n} - a) \xrightarrow{d} N(0,1),$$

где $\sigma^2 = b^2 + 2\sum\limits_{i=2}^{m+1} Cov(\xi_1.\xi_i)$ и "$\xrightarrow{d} N(0,1)$" означает сходимость по распределению к стандартному нормальному распределению.

В работе [17], Y.Shang обобщил теорему 1 для сумм случайного числа m-зависимых с.в. Центральная предельная теорема для случайных сумм линейного

© 2014 Т.М.Зупаров и О.Сафаров

Олег Воробьев (ред.), Труды XIII ФАМЭМС'2014, Красноярск: СФУ

числа независимых с.в. были изучены в ранних работах Anscombe [6], Renyi [3], Blum, Hanson, Rosenblatt [4]. Основной результат из [4], независимо получена в работе Mogyorodi [10]. Более поздние работы относящиеся к случайным суммам можно найти в статьях [5, 9, 15, 14, 16].

В данной работе, используя результат работы Y.Shang(a) доказана центральная предельная теорема для случайных сумм линейного процесса, порожденного последовательностью m-зависимых с.в.

Теорема 2. *Пусть $\{\xi_i\}_{i\geq 1}$ строго стационарная последовательность m-зависимых случайных величин, с $E\xi_i = a, 0 < D\xi_i = b^2 < \infty$ и $\{X_k\}_{k\geq 1}$ — линейный процесс, порожденный последовательностью $\{\xi_i\}_{i\geq 1}$. Пусть, далее $\{N_n\}_{n\geq 1}$последовательность положительных целозначных с.в. такая,что*

$$\frac{N_n}{\nu} \xrightarrow{p} \nu,$$

при $n \to \infty$, где $\{\nu_n\}_{n\geq 1}$-произвольная последовательность, стремящаяся к $+\infty$ и ν положительная константа. Если выполнены условия:

(A1) Существуют $k_0 \geq 0$ и c>0 такие, что для любого $\lambda > 0$ и $n > k_0$

$$P\left(\max_{k_0<k_1\leq k_2\leq n} |\eta_{k_2} - \eta_{k_1} - (k_2 - k_1)a| \geq \lambda\right) \leq$$

$$\leq c\frac{D(\eta_n - \eta_{k_0})}{\lambda^2},$$

(A2) $Cov(\xi_1, \xi_2) \geq 0$ для i=2,3,...,m+1, тогда

$$\frac{\sqrt{N_n}}{\sigma^2}(\frac{S_{N_n}}{N_n} - a) \xrightarrow{d} N(0, 1),$$

при $n \to \infty$.

Список литературы

[1] Зупаров Т.М., Чуянов Х.Ч. Центральная предельная теорема для последовательностей m-зависимых случайных величин. *Труды научной конференыии «Проблемы современной математики» посвященной 20 летию независимости Республики Узбекистан, г. Карши*, pages 133–134, 2011.

[2] Зупаров Т.М. Предельные теоремы для линейных процессов. *Материалы респуб. Научно-практической конференции «Статистика и еч применения», Ташкент*, pages 112–123, 2012.

[3] Renyi A. On the asymptotic distribution of the sum of a random numer of independent random variables. *Acta. Math. Acad. Sci. Hung.*, 8:193–197, 1957.

[4] Blum I.K., Hanson D.I., Rosenblatt J.I. On the central limit theorem for the sum of a random number of independent random variables. *Probability Theory and Related Fields*, 1:389–393, 1963.

[5] Fakhre-Zakeri I., Farshidi J. Central limit theorem with random indices for stationary linear processes. *Stat.Prob.Let.*, 17:91–95, 1993.

[6] Anscombe F.J. Large-sample theory of sequential estimation. *Proc. Cambridge Philos.Soc.*, 49:.600–607, 1952.

[7] Bregstrom H. A comparison method for distribution functions of sums of independent and dependent random variables. *Theor. Probab.Appl.*, 15:430–457, 1970.

[8] Hoeffding W., Robbins H. The central limit theorem for dependent random variables. *Duke Math.J.*, 15:773–780, 1948.

[9] Hung T.L., Thanh T.T. Some results on asymptotic behaviors of random sums of independent identically distributed random variables. *Commun. Korean Math.Soc.*, 25:119–128, 2010.

[10] Mogyorodi J. A central limit theorem for the sum of a random number of independent random variables „p.409-424. *Publ.Hungar.Acad.Sci., Ser.A*, 7:409–424, 1962.

[11] Berk K.N. A central limit theorem for m-dependent random variables with unbounded m. *Ann.Probab.*, 1:352–354, 1973.

[12] Diananda P.H. The central limit theorem for m-dependent variables. *Proc.Cambridge Philos.Soc.*, 51:92–95, 1955.

[13] Orey S.A. Central limit theorems for m- dependent random variables. *Duke. Math.J.*, 25:543–546, 1958.

[14] Przystalski V. Asymptotics for products of a random number of partial sums. *Bull. Polish Acad.Sci. Math.*, 57:163–167, 2009.

[15] Korolev V.Y. The asymptotics of randomly indexed random sequences. indipendent indexes. *J.Math.Sci.*, 59:926–938, 1992.

[16] Shang Y. Central limit theorem for the sum of a random numer of dependent random variables. *Asian. J. Math.*, 4:168–173, 2011.

[17] Shang Y. *A central limit theorem for randomly indexed m-dependent random variables.* Filomat 26: 4 713-717. DOI 10, 2298/FIL. 12047135, 2012.

Численная идентификация старших коэффициентов в системе параболических уравнений

Клоков Андрей Владимирович
Сибирский федеральный университет
Институт математики и фундаментальной информатики
Красноярск
mrandreich@yandex.ru

Распопов Виталий Евгеньевич
Сибирский федеральный университет
Институт математики и фундаментальной информатики
Красноярск

Аннотация. *В данной работе рассматривается обратная задача для системы уравнений параболического типа. Обратные задачи могут быть связаны с интерпретацией и обработкой результатов реальных тепловых процессов. В действительности темп измерения температуры в удаленных внутренних точках может оказаться ниже темпа измерения температуры внешней поверхности. Таким образом, нарушается непрерывная зависимость результатов от входных температурных данных, поэтому обратные задачи как правило некорректны. Обратным коэффициентным задачам посвящено достаточно много работ(в частности смотри [1, 2, 3, 5]). Различные подходы к численному решению обратных задач обсуждаются, например, в работах [4, 5]. Обратная задача Коши, рассмотренная в [6] для одного уравнения в настоящей работе обобщена на систему двух уравнений.*

Ключевые слова. *Обратная задача, уравнение параболического типа, условия переопределения, численная идентификация, задача Коши*

Итак, рассмотрим задачу:

$$U_t = (k_1(x)\cdot U_x)_x + a_1(t,x)\cdot U_x + a_2(t,x)\cdot V_x + f_1(t,x) \qquad 0\le t\le 1, 0\le x\le 1, \tag{1}$$

$$V_t = (k_2(x)\cdot V_x)_x + b_1(t,x)\cdot U_x + b_2(t,x)\cdot V_x + f_2(t,x) \qquad 0\le t\le 1, 0\le x\le 1, \tag{2}$$

$$U(t,0) = \alpha_1(t), \tag{3}$$

$$U_x(t,0) = \beta_1(t), \tag{4}$$

$$k_1(0) = \gamma_1, \tag{5}$$

$$U(0,x) = r_1(x), \tag{6}$$

$$V(t,0) = \alpha_2(t), \tag{7}$$

$$V_x(t,0) = \beta_2(t), \tag{8}$$

$$k_2(0) = \gamma_2, \tag{9}$$

$$V(0,x) = r_2(x), \tag{10}$$

© 2014 А.В. Клоков, В.Е. Распопов

Олег Воробьев (ред.), Труды XIII ФАМЭМС'2014, Красноярск: СФУ

где (5),(2) — исходная система уравнений; (6),(4) — условия Коши для $U(t,x)$; (7),(8) — условия Коши для $V(t,x)$; (7),(6) и (9),(10) — трактуем как условия переопределения для $U(t,x)$ и $V(t,x)$ соответственно. Функции $\alpha_1(t), \beta_1(t), \gamma_1, r_1(x), \alpha_2(t), \beta_2(t), \gamma_2, r_2(x)$ — заданы. Требуется найти функции $U(t,x), V(t,x), k_1(x)$ и $k_2(x)$.

Поставленную задачу будем решать численно. Введем равномерную сетку на области $0\le t\le 1, 0\le x\le 1$ с шагами τ, h по t, x:

$$\omega_{h,\tau} = \{x_i = ih, i=\overline{0,M}, t_n = n\tau, n=\overline{0,N}, h=\frac{1}{M}, \tau=\frac{1}{N}\}.$$

Проведем следующие преобразования. В (5) и (2) аппроксимируем внешние производные по x разностными отношениями в точке (t_n, x_j). Имеем:

$$U_t(t_n,x_j) = \frac{k_{1,j+1}U^n_{x,j+1} - k_{1,j}U^n_{x,j}}{h} + a^n_{1,j}U^n_{x,j} + a^n_{2,j}V^n_{x,j} + f^n_{1,j},$$

$$V_t(t_n,x_j) = \frac{k_{2,j+1}V^n_{x,j+1} - k_{2,j}V^n_{x,j}}{h} + b^n_{1,j}U^n_{x,j} + b^n_{2,j}V^n_{x,j} + f^n_{2,j}.$$

Отсюда находим:

$$k_{1,j+1}U^n_{x,j+1} = k_{1,j}U^n_{x,j} + h\cdot\left[U^n_{t,j} - a^n_{1,j}U^n_{x,j} - a^n_{2,j}V^n_{x,j} - f^n_{1,j}\right], \tag{11}$$

$$k_{2,j+1}V^n_{x,j+1} = k_{2,j}V^n_{x,j} + h\cdot\left[V^n_{t,j} - b^n_{1,j}U^n_{x,j} - b^n_{2,j}V^n_{x,j} - f^n_{2,j}\right]. \tag{12}$$

Далее, вводя обозначения $\frac{dk_1(x)}{dx} = P_1(x)$, $\frac{dk_2(x)}{dx} = P_2(x)$, аппроксимируем производные по x, получаем:

$$\frac{k_{1,j+1}-k_{1,j}}{h} = P_{1,j}, \frac{k_{2,j+1}-k_{2,j}}{h} = P_{2,j},$$

откуда

$$k_{1,j+1} = k_{1,j} + hP_{1,j}, k_{2,j+1} = k_{2,j} + hP_{2,j}. \tag{13}$$

Запишем тождества:

$$\frac{\partial^2 U}{\partial x\partial t} = \frac{\partial^2 U}{\partial t\partial x}, \frac{\partial^2 V}{\partial x\partial t} = \frac{\partial^2 V}{\partial t\partial x}.$$

Аппроксимируя соответствующие производные, имеем:

$$\frac{\partial^2 U}{\partial t \partial x} = \frac{U^n_{t,j+1} - U^n_{t,j}}{h}; \frac{\partial^2 U}{\partial x \partial t} = \frac{U^{n+1}_{x,j} - U^{n-1}_{x,j}}{2\tau},$$

$$\frac{\partial^2 V}{\partial t \partial x} = \frac{V^n_{t,j+1} - V^n_{t,j}}{h}; \frac{\partial^2 V}{\partial x \partial t} = \frac{V^{n+1}_{x,j} - V^{n-1}_{x,j}}{2\tau}.$$

Отсюда находим:

$$U^n_{t,j+1} = U^n_{t,j} + h \cdot \frac{U^{n+1}_{x,j} - U^{n-1}_{x,j}}{2\tau}, \quad (14)$$

$$V^n_{t,j+1} = V^n_{t,j} + h \cdot \frac{V^{n+1}_{x,j} - V^{n-1}_{x,j}}{2\tau}. \quad (15)$$

Далее согласно (11),(12), имеем:

$$U^n_{x,j+1} = \frac{k_{1,j}U^n_{x,j} + h \cdot \left[U^n_{t,j} - a^n_{1,j}U^n_{x,j} - a^n_{2,j}V^n_{x,j} - f^n_{1,j}\right]}{k_{1,j+1}},$$

$$V^n_{x,j+1} = \frac{k_{2,j}V^n_{x,j} + h \cdot \left[V^n_{t,j} - b^n_{1,j}U^n_{x,j} - b^n_{2,j}V^n_{x,j} - f^n_{2,j}\right]}{k_{2,j+1}}. \quad (16)$$

Из исходных уравнений:

$$U_t = k_1 U_{xx} + k_1^{'} U_x + a_1 U_x + a_2 V_x + f_1,$$

$$V_t = k_2 V_{xx} + k_2^{'} U_x + b_1 U_x + b_2 V_x + f_2,$$

откуда при $t = 0$ имеем:

$$k^{'}_{1,j+1} = \frac{1}{U^0_{x,j+1}} \cdot \left[U^0_{t,j+1} - k_{1,j+1}U^0_{xx,j+1} - a^0_{1,j+1}U^0_{x,j+1} - a^0_{2,j+1}V^0_{x,j+1} - f^0_{1,j+1}\right], \quad (17)$$

$$k^{'}_{2,j+1} = \frac{1}{V^0_{x,j+1}} \cdot \left[V^0_{t,j+1} - k_{2,j+1}V^0_{xx,j+1} - b^0_{1,j+1}U^0_{x,j+1} - b^0_{2,j+1}V^0_{x,j+1} - f^0_{2,j+1}\right]. \quad (18)$$

Вводим обозначения:

$$Q^n_{1,j} = k_1(jh) \cdot U_x(jh, n\tau),$$

$$Q^n_{2,j} = k_2(jh) \cdot V_x(jh, n\tau),$$

$$R^n_{1,j} = U_x(jh, n\tau),$$

$$R^n_{2,j} = V_x(jh, n\tau),$$

$$W^n_{1,j} = U_t(jh, n\tau),$$

$$W^n_{2,j} = V_t(jh, n\tau),$$

$$A_{1,h} = k_1(jh),$$

$$A_{2,h} = k_2(jh).$$

Согласно предыдущим выкладкам, получаем расчетные формулы:

$$Q^n_{1,j+1} = Q^n_{1,j} + h \cdot \left[W^n_{1,j} - a^n_{1,j}R^n_{1,j} - a^n_{2,j}R^n_{2,j} - f^n_{1,j}\right],$$

$$Q^n_{2,j+1} = Q^n_{2,j} + h \cdot \left[W^n_{2,j} - b^n_{1,j}R^n_{1,j} - b^n_{2,j}R^n_{2,j} - f^n_{2,j}\right],$$

$$P_{1,j} = \frac{1}{R^0_{1,j}} \cdot \left(W^0_{1,j} - A_{1,j}U^0_{xx,j} - a^0_{1,j}R^0_{1,j} - a^0_{2,j}R^0_{2,j} - f^0_{1,j}\right),$$

$$P_{2,j} = \frac{1}{R^0_{2,j}} \cdot \left(W^0_{2,j} - A_{2,j}V^0_{xx,j} - b^0_{1,j}R^0_{1,j} - b^0_{2,j}R^0_{2,j} - f^0_{2,j}\right),$$

$$A_{1,j+1} = A_{1,j} + hP_{1,h},$$

$$A_{2,j+1} = A_{2,j} + hP_{2,h},$$

$$W^n_{1,j+1} = W^n_{1,j} + h \cdot \frac{R^{n+1}_{1,j} - R^{n-1}_{1,j}}{2\tau},$$

$$W^n_{2,j+1} = W^n_{2,j} + h \cdot \frac{R^{n+1}_{2,j} - R^{n-1}_{2,j}}{2\tau},$$

$$R^n_{1,j+1} = \frac{Q^n_{1,j+1}}{A^n_{1,j+1}},$$

$$R^n_{2,j+1} = \frac{Q^n_{2,j+1}}{A^n_{2,j+1}},$$

$$j = j + 1.$$

Для того, чтобы начать счет, необходимо знать все неизвестные при $j = 0$. из условий (6)-(10) имеем:

$$W^0_{1,0} = \left.\frac{\partial U}{\partial t}\right|_{t=0,x=0} = \alpha_1^{'}(0),$$

$$W^0_{2,0} = \left.\frac{\partial V}{\partial t}\right|_{t=0,x=0} = \alpha_2^{'}(0),$$

$$A_{1,0} = k_1(0) = \gamma_1,$$

$$A_{2,0} = k_2(0) = \gamma_2,$$

$$U^0_{xx,0} = r_1^{''}(0),$$

$$V^0_{xx,0} = r_2^{''}(0),$$

$$R^0_{1,0} = \left.\frac{\partial U}{\partial x}\right|_{t=0,x=0} = \beta_1(0),$$

$$R^0_{2,0} = \left.\frac{\partial V}{\partial x}\right|_{t=0,x=0} = \beta_2(0).$$

Согласно данного алгоритма, все производные $\frac{\partial U}{\partial t}, \frac{\partial U}{\partial x}$ и $\frac{\partial V}{\partial t}, \frac{\partial V}{\partial x}$ найдены численно.

Далее $U(t,x)$ и $V(t,x)$ находим в узлах сетки по формулам:

$$y^n_{j+1} = hR^n_{1,j} + y^n_j,$$

$$p^n_{j+1} = hR^n_{2,j} + p^n_j$$

Предложенный метод решения был апробирован на ряде тестов. В частности для теста:

$$
\begin{aligned}
&U(t,x) = \left(t^2+t+1\right)\left(x^2+5x+7\right),\\
&V(t,x) = \left(t^2-t+1\right)\left(x^2+3x+10\right),\\
&k_1(x) = 1+x,\\
&k_2(x) = 2+x,\\
&a_2(t,x) = e^t x^2; a_2 = e^t(2x),\\
&b_2(t,x) = e^t(x+1); b_2 = e^t(x^2+1),
\end{aligned}
$$

таблица максимальных по сетке погрешностей имеет вид:

Таблица максимальных погрешностей

Шаг		Относительная погрешность			
τ	h	$U(t,x)$	$V(t,x)$	$k_1(x)$	$k_2(x)$
1/64	1/64	1.2%	1.5%	0.68%	0.76%
1/128	1/128	0.5%	0.82%	0.21%	0.39%
1/256	1/256	0.3%	0.5%	0.1%	0.19%

На всех рассмотренных тестах, при уменьшении шагов сетки численное решение сходится к точному.

Список литературы

[1] Е.В. Кучунова, В.Е. Распопов. Численная идентификация коэффициентов параболических уравнений. *Вестник КрасГУ. Серия Физ.-мат. науки*, 5/2:7–14, 2004.

[2] Ю.В. Мандрик, В.Е. Распопов. Численная идентификация коэффициентов одномерного параболического уравнения. *Вестник КрасГУ*, 1, 2006.

[3] Жак Т.Ю., Распопов В.Е. Численая идентификация свободного члена специального вида в парабалическом уравнении. *Международная конференция Алгебра и ее приложения: Тезисы докладов. Красноярск*, page 186, 2007.

[4] А.А. Самарский, П.Н. Вабищевич. *Численные методы решения обратных задач математической физики.* Изданное кем-то, Москва, 2009.

[5] С.И. Кабанихин. *Обратные и некорректные задачи.* Сибирское научное издательство, Новосибирск, 2009.

[6] S.Zhan, D.A.Murio. *Identification of Parameters in One-Dimensional IHCP.* Computers Math. Applic. Vol. 35, No 3, pp. 1-16, USA, 1998.

О различных подходах к построению интегральных показателей

Крупкина Татьяна Валерьевна
Институт математики и фундаментальной информатики
Сибирский федеральный университет
Красноярск
tatiana.krupkina@gmail.com

Мажаров Владимир Федорович
ФГБУ НИИ комплексных проблем
гигиены и профзаболеваний СО РАМН
Новокузнецк
majarov@inbox.ru

Горный Борис Эмануилович
ФГБУ НИИ комплексных проблем
гигиены и профзаболеваний СО РАМН
Новокузнецк
bg1960@yandex.ru

Бабенышев Сергей Валерьевич
Сибирская пожарно-спасательная академия — филиал
Санкт-Петербургского университета ГПС МЧС России
Железногорск
sergei.babyonyshev@gmail.com

Аннотация. *Задачи по установлению зависимостей различных результативных показателей общественного здоровья от набора признаков, характеризующих среду, разделены на три типа. Определены методы построения интегральных показателей среды для каждого типа. Результаты получены на основе ряда работ, в которых среда характеризовалась социально-экономическими условиями.*

Ключевые слова. *Социально-экономические показатели, интегральные показатели, факторный анализ, главные компоненты.*

1 Введение

В течение ряда лет (1997–2013) мы проводили системные исследования по изучению влияния социально-экономических условий на здоровье населения. В рамках этой работы были, в частности, проведены типизации городских муниципальных образований Красноярского края по факторам окружающей среды, детерминирующим общественное здоровье и являющихся его фоном, в том числе, экологическим и климатическим (2002–2003); социально-гигиенические типизации сельских муниципальных образований Красноярского края с учетом особенностей геохимических ландшафтов и экологического неблагополучия (2002–2003); типизации территорий по индексу здоровья, социально-экономическому, климатическому, экологическому, интегральному риску (2004); типизации территорий по динамическим рядам показателей жизненного потенциала (2008); типизации территорий Красноярского края по потерям жизненного потенциала, показателям нереализованной фертильности и социально-экономическим показателям (2008–2009); типизации районов и городов Красноярского края по интегральным рискам социально-экономических условий на уровень потерь жизненного потенциала (2011); типизации территорий Красноярского края по демографическим и социально-экономическим показателям (2011). В данной статье мы попытаемся обобщить особенности применяемых нами методик.

© 2014 Т.В. Крупкина, В.Ф. Мажаров, Б.Э. Горный и С.В. Бабенышев

Олег Воробьев (ред.), Труды XIII ФАМЭМС'2014, Красноярск: СФУ

2 Выбор результативных показателей общественного здоровья

В работах по исследованию избранного аспекта общественного здоровья мы использовали общепринятые медико-демографические показатели. При анализе данных мониторинга демографической ситуации в Красноярском крае результативными показателями являлись естественный и миграционный приросты, численность населения, рождаемость, структурированная смертность. При изучении состояния здоровья населения результативные показатели — это структурированная заболеваемость и индекс здоровья [3]. Методика вычисления индекса Е предполагает переход от значения показателя x_i к его индивидуальной оценке — числу r_i согласно нормативной таблице, построенной для каждого показателя по среднему значению a и среднеквадратическому отклонению s. Все значения показателя разбиваются на интервалы с длиной, равной половине среднеквадратического отклонения, и если сравниваемые значения показателя принадлежат одному интервалу ряда распределения, то всем значениям, попавшим в i-й интервал, приписывается одно и то же число $r_i = r_I$. Числа r_i имеют явное выражение, связанное с энтропией:

$$r_i = \begin{cases} -p_i \log p_i, & i \le 7; \\ 1 + p_i \log p_i, & i > 7. \end{cases} \qquad (1)$$

В (1) p_i означает вероятность попадания в в i-й интервал, вычисленную исходя из предположении о распределении показателя. Чаще других используются нормальная и равномерная модели. Затем i-му показателю в m-ом районе X_{im} присваивается оценка r_{im}. Энтропийный индекс Е m-ого района рассчиты-

вается по формуле (2):

$$E_m = 100 - \frac{100 \cdot \sum_{i=1}^{k} r_{im}}{\max\{r_{im}\}k}. \qquad (2)$$

При более детальном исследовании процесса воспроизводства населения использовались стандартизованный по возрасту показатель нереализованной рождаемости, показатель общих потерь жизненного потенциала PYLL, показатели PYLL от отдельных причин. При выборе причин приоритет отдавался тем заболеваниям, которые влекли за собой максимальные потери, причем эти потери могли быть в некоторой степени устранимыми за счет улучшения социально-экономических условий, оздоровления экологической обстановки, изменения образа жизни людей в направлении уменьшения распространенности вредных и опасных для здоровья обычаев и привычек, развития современной медицины и здравоохранения. При исследовании репродуктивного здоровья женщин результативными показателями являлись рождаемость, смертность, заболеваемость женщин болезнями репродуктивной системы. В некоторых случаях имелось несколько вариантов одного и того же результативного показателя, как например, показателя потерь жизненного потенциала (удельный жизненный потенциал, число недожитых лет, ожидаемая продолжительность жизни, общие потери жизненного потенциала и др.). Если результативный показатель был нужен только для описания процесса воспроизводства, то предварительно проводили кластерный анализ и выбирали показатель, лучше разделяющий районы. Если же показатель был нужен для исследования влияния на него факторов окружающей среды, то показатель был тем лучше, чем сильнее различались его значения для разных групп, внутри которых содержались районы со схожими характеристиками комплексного риска окружающей природной и социальной среды для здоровья. Таким образом, принять окончательное решение можно было только после проведения типизации территорий по показателям окружающей среды (например, с помощью кластерного или дискриминантного анализов). Иногда оказывалось, что наилучшие результативные показатели были разными для городских и сельских местностей, для мужчин и женщин. Например, у мужчин Красноярского края основными причинами потерь жизненного потенциала являлись острая ИБС, заболевания, вызванные злоупотреблением алкоголем и отравлениями алкоголем и наркотиками, самоубийства, транспортные несчастные случаи, цереброваскулярные болезни, а у женщин на первом месте стояли заболевания, вызванные злоупотреблением алкоголем и отравлениями алкоголем и наркотиками, далее острая ИБС, цереброваскулярные болезни, транспортные несчастные случаи, злокачественные новообразования органов пищеварения. У выбранных результативных показателей исследовались их статистические свойства. Как правило, они были нормально распределены и (при наличии нескольких показателей) зависимы.

3 Исследование исходных показателей среды

Отдельные исследуемые массивы представляли собой значения 95 показателей (55 медико-демографических, 9 климатических, 9 экологических, 22 социально-экономических) за 1997 год для всей совокупности АТЕ Красноярского края, для 90 показателей за 2000 и 2001 годы; отдельно за те же годы для совокупности сельских АТЕ Красноярского края; 93 показателя, характеризующих состояние здоровья сельского населения, социальную среду, экономику и экологию за 2002 год; 38 социально-экономических и демографических показателей по городам (14 АТЕ) и районам (44 АТЕ) Красноярского края за 2007 год; 13 демографических показателей, являвшимися целевыми индикаторами системы здравоохранения и 49 социально-экономических показателей (предположительно демографически значимых) за 2008 год; 50 социально-экономических и демографических показателей по всем АТЕ Красноярского края за 2011 год; сводные демографические показатели за 1995–2005 годы; сводные демографические и социально-экономические показатели за 2005–2012 годы. Исходные показатели мы получали из официальных источников: климатические и экологические показатели данным по мониторинга за загрязнением атмосферного воздуха – из службы Гидрометцентра, геолого-геофизические из Геологоуправления, прочите – из статистических сборников федеральной службы государственной статистики по Красноярскому краю и социально-экономических паспортов муниципальных образований городов и районов. Эти документы за разные годы включают разные наборы показателей, поэтому база данных не была стабильной по составу. Сначала мы делали попытки отобрать показатели, обладающие специфической чувствительностью, измеряющие не просто социально-экономические условия, а связанный с ними риск для здоровья. Эксперты составляли список предположительно демографически значимых показателей и проводился корреляционный анализ, который на первой стадии давал, как правило, обнадеживающие результаты. Так, по данным за 2008 год среди коэффициентов корреляции между отобранными 49 социально-экономическими показателями и 13 демографическими показателями, являвшимися целевыми индикаторами системы здравоохранения, было более 200 достоверных ($p < 0,05$). Для всех достоверных коэффициентов корреляции были рассмотрены диаграммы рассеяния и коэффициенты частной корреляции, измеряющие силу линейной

зависимости двух величин, очищенную от влияния других величин. Количество установленных линейных зависимостей сильно сократилось. Например, корреляция демографического показателя D_1 (естественный прирост) с 17 социально-экономическими показателями подтвердилась только в трех случаях: с показателями E_{15} (общая площадь земель муниципального образования на 1000 чел.), E_{46} (бюджетные расходы на образование на 1000 чел.) и E_{49} (бюджетные расходы на здравоохранение на 1000 чел.). Младенческая смертность D_{10} линейно зависит от E_{14} (динамика индекса производства сельскохозяйственной продукции), прочие наблюдаемые связи обусловлены влиянием других показателей или выбросами. У остальных демографических показателей подтверждена зависимость только с E_{20} (фельдшерско-акушерские пункты) и E_{31} (текущие затраты на охрану окружающей среды).

Для более специфических результативных показателей, которые используются, например, при исследовании процесса воспроизводства населения, значимых парных коэффициентов кореляции на порядок меньше, а после устранения выбросов и влияния других величин оказалось, что линейной зависимости между показателем общих потерь жизненного потенциала PYLL и каким-либо социально-экономическим показателем не наблюдается. Из рисунка 1 видно, что при исключении трех выбросов связь отсутствует. Выбросы чаще всего соответствовали северным территориям, а также Новоселовскому, Идринскому, Большеулуйскому районам. Значимые параметры множественной линейной регрессии тоже не были обнаружены. Исследование различных формы нелинейной регрессии показало, что самый высокий коэффициент множественной корреляции дала кусочно-линейная форма зависимости, при которой зависимость между предикторами и переменной отклика различалась в разных областях значений независимых переменных. Кусочно-линейная зависимость имела одну точку разрыва и характеризовалась достаточно большим множественным коэффициентом детерминации R^2.

В целом массив социально-экономических показателей можно охарактеризовать как массив, состоящий из достаточно большого числа сильно взаимосвязанных признаков, которые частично дублируют друг друга, то есть массивы показателей среды во всех исследованиях имели одну и ту же природу. Однако структура связи результативных признаков с показателями среды была различной в разных задачах, это было тем более заметно, что с одним и тем же массивом показателей среды мы решали разные задачи, используя различные результативные признаки. Можно выделить три типа задач: 1) существуют индикаторы (отдельные показатели среды, с которыми сильно коррелируют результативные показатели); 2) индикаторов нет, но существует множество достоверных коэффициентов корреляции, значимость которых сохраняется и после удаления выбросов; 3) не удается обнаружить линейных зависимостей между результативными признаками и показателями среды, но регрессионный анализ показывает наличие некоторых нелинейных зависимостей.

4 Построение факторных (интегральных) показателей среды

Интегральные показатели среды выбираются из набора {принятые в предметной области изучения среды интегральные показатели, унифицированные квантильные индексы, главные компоненты массива первичных показателей} по критерию наибольшего влияния на избранный результативный показатель. Практически для всех массивов мы применяли все три способа и выбирали лучший. Оказалось, что если существует общепринятый интегральный показатель, с которым хорошо коррелирует результативный показатель, то целесообразно его и выбрать в качестве факторного показателя, поскольку другие способы дадут сходные результаты. Примером является использование показателя приведенного валового выброса для оценки экологического риска или показателя жесткости климата. В случае отсутствия такого интегрального показателя мы строили его сами аналогично (2). Например, так находили социально-экономический индекс. Во многих задачах этот индекс коррелирует с результативным показателем, в этих случаях его выбирали как интегральный показатель среды. Эти две ситуации соответствуют задаче первого типа. Заметим, что при наличии, например, показателя жесткости климата не требуется строить индекс жесткости климата, так как они взаимозаменяемы.

Задача второго типа отличается тем, что не удается подобрать или построить интегральный показатель, связанный с результативным, но существует множество слабых линейных связей между результативным показателем и показателями среды. Мы предполагаем, что в этом случае применима идеология факторного анализа: измерямые параметры являются косвенными характеристиками некоторого скрытого гипотетического фактора, действие которого проявляется через них опосредованно. Для поиска таких факторов мы использовали метод главных компонент. Мы неоднократно применяли также факторный анализ и рассматривали вращение главных факторов, используя разные алгоритмы, но наилучшими интерпретациями неизменно обладали первые главные компоненты. Наиболее информативные типообразующие факторы массива показателей состояния АТЕ являются устойчивыми. Как известно, разнообразие показателей определяется, главным образом, несколькими первыми обобщенными факторами массива признаков (главными компонента-

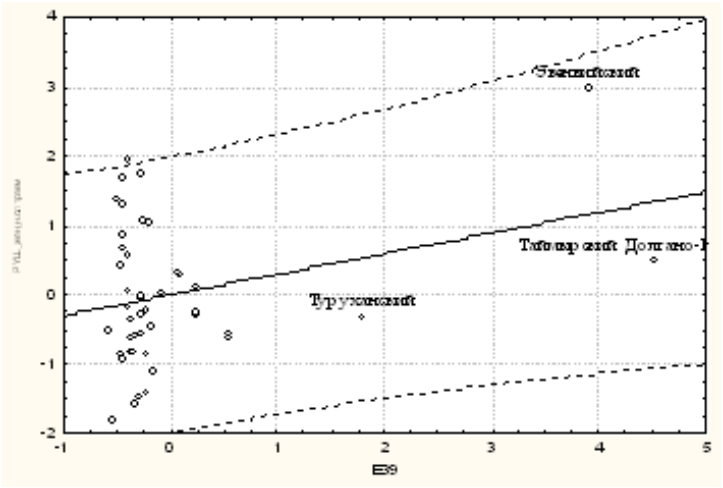


Рисунок 1: Диаграмма рассеяния показателей P_2 (PYLL мужской по новообразованиям) и E_9 (обрабатывающие производства) до исключения выбросов.

ми). За период с 1999 по 2013 год нами неоднократно проводился факторный анализ показателей среды, причем обновлялись не только данные, но изменялся и сам набор показателей. Помимо различных социально-экономических показателей в набор включались и другие, например, показатели уровня экологической загрязненности и дискомфортности климата. Независимо от этого обе первые главные компоненты всегда характеризовали социально-экономический профиль территории, причем одна из них описывала степень экономического развития, а также уровень социального развития и бюджетной обеспеченности. Например, по данным 1997 года максимальную нагрузку на первую компоненту имели следующие признаки: социально-экономический индекс, прибыль в промышленности, среднемесячные денежные доходы на душу населения, отношение зарплаты к прожиточному минимуму, объем промышленного производства, объем с/х производства, отношение зарплаты к прожиточному минимуму, валовой выброс, приведенный валовой выброс (фактор экономического развития), а вторая компонента зависела в основном от показателей: средний прожиточный минимум, отношение денежных доходов к стоимости набора основных продуктов питания, уровень безработицы, инвестиции в производственную сферу, инвестиции в непроизводственную сферу, расходы на образование, расходы на культуру, трансферты из федеральных и краевых бюджетов (фактор бюджетной обеспеченности). При добавлении в массив медико-демографические показателей в компоненту промышленного развития добавились показатели: общая заболеваемость взрослых и заболеваемость детей болезнями дыхательных путей. По данным 2007 года первая компонента определялась показателями: оборот розничной торговли, платные услуги, оказанные населению, строительство, среднесписочная численность работников организаций, инвестиции в основной капитал, фонд заработной платы, доходы местного бюджета, оборот общественного питания, что мы интерпретируем как фактор бюджетной обеспеченности и уровня дохода населения.

Наглядное представление многомерных данных методом проецирования на плоскость первых двух компонент показывает, что точки, соответствующие разным районам распадаются на несколько групп, однородных по социально-экономическим условиям. В задачах второго типа результативные показатели в этих группах, как правило, имеют неслучайные различия.

Наиболее сложной является задача третьего типа, когда не удается обнаружить корреляций между результативными признаками с одной стороны, и показателями среды, интегральными показателями, главными компонентами с другой стороны, но при этом гипотеза о наличии некоторой нелинейной связи не отвергается. С такими задачами мы сталкивались, например, при использовании в качестве результативных показателей показателя нереализованной рождаемости и потерь жизненного потенциала. В этом случае мы проводили факторный анализ массива показателей среды и затем корреляционно-регрессионный анализ зависимости результативных показателей от всего набора общих факторов. Выявились зависимости от некоторых полученных факторов, но факторы были малозначимыми (выделяли малую долю дисперсии) и не допускали убедительной интерпретации [2]. Поэтому дальше была разработана иерархическая модель влияния социально-экономических условий на уровень потерь жизненного потенциала [1]. Социально-экономические показатели были разбиты на 6 групп (доходов населения, социально-экономического развития, потребления, финансового состояния, инфраструктуры, качества жизни). По всем группам были найдены социально-экономические индексы. Эксперты определяли относительную важность этих факторов для

потерь жизненного потенциала, используя процедуру парных сравнений по шкале Саати. Путем обработки матриц парных сравнений был найден итоговый вектор приоритетов факторов; заметим, что он является нелинейной функцией исходных показателей. С помощью вектора приоритетов был рассчитан интегральный показатель социально-экономического риска для здоровья населения, который достоверно коррелировал с RYLL и имел при этом понятную структуру; так, самый большой вклад в него вносил фактор "Качество жизни".

Список литературы

[1] Т.В. Крупкина, В.Ф. Мажаров, Б.Э. Горный, А.М. Свиридов. Иерархическая модель зависимости потерь жизненного потенциала от социально-экономических факторов. *Труды X Международной конференции по финансово-актуарной математике и эвентоконвергенции технологий*, pages 188–192, 2011.

[2] Т.В. Крупкина, В.В. Мозолева, М.Г. Коломейцева. Устойчивость типизаций территорий. *Труды VIII Международной конференции по финансово-актуарной математике и смежным вопросам*, page 65, 2009.

[3] Т.В. Крупкина, С.В. Бабенышев, В.Ф. Мажаров. Интегральные показатели в синтезе оценок рисков потерь жизненного потенциала. *Труды XI Международной конференции по финансово-актуарной математике и эвентоконвергенции технологий*, pages 210–214, 2012.

Когерентные меры риска. Примеры, модификации, обобщения.

Кустицкая Татьяна Алексеевна
Институт космических и информационных технологий
Сибирский федеральный университет
Красноярск
m-tanika@yandex.ru

Аннотация. *В статье излагается обзор основных модификаций и обобщений когерентных мер риска, приводятся конкретные примеры. Проводится анализ описываемых классов функционалов меры риска с точки зрения представляемых возможностей для описания индивидуальных предпочтений.*

Ключевые слова. *Когерентные меры риска, выпуклые меры риска, обобщенные когерентные меры риска, CVaR, мера возмущенной вероятности.*

1 Введение

Когерентные меры риска — это функционалы, согласованные с «естественными» свойствами предпочтений участников рынка. В настоящее время они получают все большее распространение и подвергаются различным модификациям и обобщениям, позволяющим либо получить более удобный инструмент для решения тех или иных практических задач, либо более точно описать индивидуальные предпочтения.

В данной работе предлагается обзор основных примеров, модификаций и обобщений когерентных мер риска, существующих на сегодняшний день.

2 Представление предпочтения с помощью функционала меры риска

Рассмотрим вероятностное пространство $(\Omega, \mathcal{A}, P)$.

Определение 1. ***Риском*** *X на $(\Omega, \mathcal{A})$ называется произвольное измеримое отображение из Ω в $\mathbf{R}$.*

В теории финансовых рисков возможные значения риска интерпретируются как доход или убыток, получаемый неким субъектом. Мы будем рассматривать риск как величину возможного дохода.

Обозначим $\mathcal{X}$ совокупность всех рисков на $(\Omega, \mathcal{A})$. На множестве $\mathcal{X}$ можно задать отношение порядка.

Например, если каждый риск X отождествить с его функцией распределения F_X, то на $\mathcal{X}$ можно ввести порядок с помощью стохастического доминирования.

Определение 2. *Риск Y стохастически доминирует риск X ($X \leq_1 Y$), если*

$$F_X(x) \geq F_Y(x)\ \forall x \in \mathbf{R}$$

Одну из ключевых ролей в теории измерения риска играет отношение предпочтения, задаваемое на $\mathcal{X}$.

Определение 3. ***Отношением предпочтения*** *$\preceq$ на некотором множестве M называется произвольное полное транзитивное бинарное отношение, заданное на этом множестве.*

Будем считать, что отношение предпочтения $\preceq$, заданное на $\mathcal{X}$, отражает некоторое индивидуальное отношение к риску.

Учитывая то, что в большинстве случаев участники рынка стремятся достичь большей доходности, разумно потребовать, чтобы отношение предпочтения на $\mathcal{X}$ согласовывалось с порядком на $\mathcal{X}$.

Определение 4. *Пусть на $\mathcal{X}$ заданы отношение предпочтения $\preceq$ и отношение порядка $\leq$. Предпочтение $\preceq$ называется согласованным с порядком $\leq$, если из того, что $X \leq Y$ следует, что $X \preceq Y$.*

Отношение предпочтения на множестве рисков можно задать с помощью монотонных вещественных функционалов, определенных на $\mathcal{X}$.

Определение 5. ***Мерой риска*** *называется произвольный функционал $\rho : \mathcal{X} \to \mathbf{R}$. Говорят, что ρ представляет отношение предпочтения на $\mathcal{X}$ в одном из случаев:*

$$\rho(X) \leq \rho(Y),\ \text{если } X \preceq Y, \quad X, Y \in \mathcal{X} \tag{1}$$

$$\rho(X) \leq \rho(Y),\ \text{если } Y \preceq X, \quad X, Y \in \mathcal{X} \tag{2}$$

3 Когерентные меры риска

Рассмотрим аксиоматику классических когерентных мер риска. Будем работать с мерами ρ, представляющими отношение предпочтения как в (1).

Определение 6. ***Приемлемым*** *назовем такой риск, на который инвестор готов пойти при вложении нулевого капитала. Множество всех приемлемых рисков обозначим A ($A \subset \mathcal{X}$).*

Согласно [3], ***множество приемлемых рисков*** $A \subset \mathcal{X}$ удовлетворяет следующим аксиомам:

© 2014 Т.А. Кустицкая

Олег Воробьев (ред.), Труды XIII ФАМЭМС'2014, Красноярск: СФУ

Аксиома А1: $C_+ \subset A,\ \ C_+ = \{X \in \mathcal{X} :\ X \geq 0\}$
Аксиома А2: $A \bigcap C_- = \emptyset,\ C_- = \{X \in \mathcal{X} :\ X < 0\}$
Аксиома А3: A - выпуклый конус (если $X \in A,\ Y \in A$, то $\alpha_1 X + \alpha_2 Y \in A$ при $\alpha_1, \alpha_2 \geq 0$).

Определение 7. *Пусть A — множество приемлемых рисков, тогда* ***мерой риска, соответствующей множеству*** *A называется функционал ρ_A заданный на $\mathcal{X}$ следующим образом:*

$$\rho_A(X) = \sup\{b \in \mathbf{R} :\ X - bI \in A\},\ I = (1, \ldots, 1). \quad (3)$$

Функционал ρ_A обладает следующими свойствами:

1. **монотонность:**

$$\forall X, Y \in \mathcal{X},\ X \leq Y\ \ \rho_A(X) \leq \rho_A(Y); \quad (4)$$

2. **супераддитивность:**

$$\rho_A(X + Y) \geq \rho_A(X) + \rho_A(Y)\ \ \forall X, Y \in \mathcal{X}; \quad (5)$$

3. **положительная однородность:**

$$\rho_A(\lambda X) = \lambda \rho_A(X),\ \ \forall \lambda \geq 0,\ X \in \mathcal{X}; \quad (6)$$

4. **инвариантность относительно сдвига:**

$$\rho_A(X + aI) = \rho_A(X) + a,\ \ a \in \mathbf{R},\ X \in \mathcal{X}. \quad (7)$$

Определение 8. ***Когерентной мерой риска*** *называется любой функционал $\rho(X)$ на множестве $\mathcal{X}$, обладающий свойствами (4)-(7).*

Свойство монотонности гарантирует монотонность относительно естественных порядков на множестве рисков. Супераддитивность отражает стремление к диверсификации и наличие у задаваемого предпочтения неприятия риска. Положительная однородность отражает требоваие того, чтобы чтобы мера риска от n инструментов с одинаковыми характеристиками была в n раз больше меры риска, подсчитанной для одного инструмента. Свойство относительно сдвига указывает на линейность изменения функционала вдоль вектора, задающего безрисковый актив.

Определение 9. *Множество рисков A_ρ называется* ***множеством приемлемых рисков, соответствующим мере риска*** *ρ, если*

$$A_\rho = \{X \in \mathcal{X} :\ \rho(X) \geq 0\} \quad (8)$$

Граница множества приемлемых рисков ∂A_ρ определяется следующим образом:

$$\partial A_\rho = \{X \in \mathcal{X} :\ \rho(X) = 0\}$$

Также в [3] была доказана важная теорема о представлении когерентных мер риска в виде экстремального значения линейных функционалов.

Теорема 1. *Функционал ρ на $\mathcal{X}$ является когерентной мерой риска в том и только в том случае, когда найдется семейство вероятностных мер $\mathcal{P} = \mathcal{P}_\rho$ в S^n такое, что*

$$\rho(X) = \inf_{Q \in \mathcal{P}} E_Q X,\ \ X \in \mathcal{X}. \quad (9)$$

Содержательный смысл понятия когерентной меры риска относительно задачи портфельного анализа следующий:

1. *значение когерентной меры риска для приемлемых рисков - это наибольшее количество денежных средств, которое можно изъять из портфеля, чтобы риск оставался приемлемым для инвестора;*
2. *для неприемлемых рисков - это наименьшее количество денежных средств, которое необходимо добавить в портфель, чтобы риск полученного портфеля стал приемлемым.*

Анализируя риск с помощью когерентных мер, инвестор может регулировать риск портфеля с помощью вложения или изъятия денежных средств.

4 Примеры когерентных мер риска

4.1 Математическое ожидание

Простейшим представителем класса когерентных мер риска является математическое ожидание.

Рассмотрим вероятностную меру $Q = (q^1, \ldots, q^n) \in S^n$. Тогда ρ_Q является когерентной мерой риска, если задать этот функционал выражением

$$\rho_Q(X) = E_Q X = q^1 x^1 + \cdots + q^n x^n \quad (10)$$

В данном случае семейство вероятностных мер $\mathcal{P}$ состоит из одной точки $\mathcal{P} = \{Q\}$.

Множество приемлемых рисков, соответствующих математическому ожиданию есть полупространство $E_Q X \geq 0$. Получается, что эта мера риска - наименее консервативная среди всех когерентных мер.

4.2 Когерентные альтернативы мере VaR

В настоящее время наиболее широко используемой на практике мерой риска является VaR (Value at Risk), которая в является стандартом в измерении рыночных рисков.

Пусть риск X - случайный убыток. VaR представляет собой квантиль распределения уровня $\alpha \in (0, 1)$:

$$VaR_\alpha(X) = \inf\{x : F_X(x) \geq \alpha\} \quad (11)$$

$\alpha \in [0, 1]$ называется уровнем доверия ($\alpha \gg 0$).

В случае, когда X - случайный доход,

$$VaR_\beta(X) = -\inf\{x : F_X(x) \geq \beta\}, \quad \beta \ll 1. \qquad (12)$$

В таком случае, VaR - наибольший убыток, который не будет превышен с уровнем доверия $1-\beta$.

В работе [2] рассмотрены достоинства меры VaR, главным из которых можно считать возможность измерения рисков величиной потерь, соотнесенных с вероятностями их возникновения, что позволяет дать реальную оценку необходимого резервного капитала под риски портфеля.

Однако эта мера риска имеет ряд существенных недостатков:

- VaR не учитывает возможность больших потерь, которые могут произойти с вероятностью меньше α.
- Для большинства финансовых инструментов характерны значительные отклонения от законов нормального распределения по краям графика распределения (fat tails), вследствие чего расчетные значения VaR, как правило, несколько искажены.
- Для VaR не выполняется свойство субаддитивности (супераддитивности, если значение VaR - доход). В [6] приводятся примеры портфелей, для которых VaR самого портфеля превышает сумму VaR его составляющих, что противоречит здравому смыслу.

В качестве когерентной альтернативы VaR выступает мера риска CVaR [3], [5].

Пусть X - убыток, $\alpha \gg 0$ - уровень доверия.

$$CVaR_\alpha(X) = E(X : \ X \geq VaR_\alpha(X)) \qquad (13)$$

В [5] показано, что

$$CVaR_\alpha(X) = \frac{1}{1-\alpha}\int_\alpha^1 \overleftarrow{F}_X(p)dp, \qquad (14)$$

где $\overleftarrow{F}_X(p) = inf\{x : \ P(X \leq x) \geq p\}$ - функция, названная Ачерби обобщенным квантилем. Для абсолютно непрерывных распределений $\overleftarrow{F}_X(p) = F_X^{-1}(p)$.

Если случайная величина Y - доход, $\beta \ll 1$, то

$$CVaR_\beta(X) = -E(X : \ X \leq -VaR_\beta(X))$$

$$CVaR_\beta(X) = -\frac{1}{\beta}\int_0^\beta \overleftarrow{F}_X(p)dp$$

Мера $CVaR$ когерентна и более адекватно оценивает риск в случае «тяжелых хвостов», чем VaR. При этом она является более консервативной, поскольку величина CVaR всегда больше (если значение меры риска - доход - меньше), чем VaR того же уровня.

4.3 Мера возмущенной вероятности

В работе Ш. Вонга [11] предложено модифицировать дополнительную функцию распределения $S_X(t)$ по следующем правилу:

$$S_Y(t) = g(S_X(t)), \quad \text{где } g(t) : [0,1] \to [0,1],$$
$$g(0) = 0, \ g(1) = 1$$
$$\text{и } \ g(t) \ \text{ неубывающая функция.}$$

Полученная таким образом функция $S_Y(t)$ представляет собой дополнительную функцию распределения некоторой случайной величины Y.

Функционал возмущенной вероятности в работах [11] и [12] задается как математическое ожидание полученной случайной величины Y:

$$\pi(X) = \int_{-\infty}^0 [g(S_X(t)) - 1]dt + \int_0^\infty g(S_X(t))dt \qquad (15)$$

Для дискретной случайной величины X, принимающей значения $x^1, \ldots, x^n$ с вероятностями, равными соответственно $p^1, \ldots, p^n$:

$$\pi(X) = \sum_{k=1}^n g\left(\sum_{i=k}^n p^k\right)(x^k - x^{k-1}), \quad \text{где } x^0 = 0$$

Функционал $\pi(X)$ является положительно однородным и инвариантным относительно сдвигов, то есть для любых $a, b \in \mathbf{R}$ выполняется

$$\pi(aX + b) = a\pi(X) + b$$

4.4 Мера риска наихудшего случая

Мера риска наихудшего случая (worst case risk measure) определяется как

$$\rho_{\max}(X) = \inf_{\omega\in\Omega} X(\omega), \quad \forall X \in \mathcal{X}.$$

Значение $\rho_{\max}(X)$ - нижняя грань потенциальных доходов, измеренных на каждом сценарии или же верхняя грань потенциальных потерь. Соответствующий конус приемлемых рисков $A = C_+$.

Это наиболее консервативная мера риска в том смысле, что любая другая когерентная мера риска ρ на $\mathcal{X}$ удовлетворяет соотношениям

$$\rho(X) \geq \rho(\inf_{\omega\in\Omega} X(\omega)) = \rho_{\max}(X).$$

5 Некоторые модификации когерентных мер риска

5.1 Выпуклые меры риска

В работе [7] Г. Фелмером и А. Шидом предлагается обобщение класса когерентных мер риска - выпуклые

меры. Изложим информацию об этих мерах риска на языке супермодулярных функционалов, чтобы мера риска представляла предпочтения как в (2).

Согласно определению, мера риска $\rho : \mathcal{X} \to \mathbf{R}$ называется выпуклой, если она удовлетворяет условиям монотонности, инвариантности относительно сдвига и условию выпуклости в следующей формулировке:

$$\rho(\lambda X + (1-\lambda Y)) \geq \lambda\rho(X) + (1-\lambda)\rho(Y), \quad 0 \leq \lambda \leq 1.$$

Аксиома выпуклости придает точный смысл идее о том, что диверсификация не увеличивает риска.

Если функция ρ выпукла и нормирована, то

$$\rho(\lambda X) \leq \lambda\rho(X), \quad 0 \leq \lambda \leq 1,$$
$$\rho(\lambda X) \geq \lambda\rho(X), \quad \lambda \geq 1.$$

Если же, кроме того, выполняется условие положительной однородности:

$$\rho(\lambda X) = \lambda\rho(X), \quad \forall\lambda \geq 0,$$

то выпуклая мера ρ является когерентной.

Тем не менее, авторы не настаивают на выполнении условия положительной однородности в виду того, что во многих ситуациях риск может расти нелинейным образом при увеличении размера позиций.

Также, как и для когерентных мер риска имеет место представление выпуклых мер риска через математическое ожидание относительно некоторой вероятностной меры.

Обозначим через $M_{1,f} = M_{1,f}(\Omega, \mathcal{A})$ класс всех конечно-аддитивных функций множеств $Q : \mathcal{A} \to [0,1]$, нормированных единицей, то есть таких, что $Q(\Omega) = 1$. Через $E_Q[X]$ обозначим интеграл от X относительно $Q \in M_{1,f}$.

Пусть $\alpha : M_{1,f} \to \mathbf{R} \cup \{+\infty\}$ - такой функционал, что

$$\inf_{Q \in M_{1,f}} \alpha(Q) \in \mathbf{R}.$$

Для любой функции $Q \in M_{a,f}$ функционал

$$\rho(X) = \inf_{Q \in M_{1,f}} (E_Q[X] - \alpha(Q)) \tag{16}$$

является выпуклым, монотонным и инвариантным относительно сдвига на $\mathcal{X}$, и

$$\rho(0) = \sup_{Q \in M_{1,f}} \alpha(Q).$$

Функционал α называют ***штрафной функцией*** (penalty function) для ρ на $M_{1,f}$.

Теорема 2. *Любая выпуклая мера риска ρ на $\mathcal{X}$ имеет вид:*

$$\rho(X) = \max_{Q \in \mathcal{M}_{\infty,\{}} (E_Q[X] - \alpha_{\min}(Q)), \quad X \in \mathcal{X}, \tag{17}$$

где штрафная функция $\alpha_{\min}$ задается формулой

$$\alpha_{min}(Q) = \inf_{X \in A_\rho} E_Q[X], \quad Q \in M_{1,f}.$$

Более того, $\alpha_{\min}$ - это минимальная штрафная функция, которая представляет ρ, то есть любая штрафная функция α, для которой верно соотношение (16), удовлетворяет неравенству $\alpha(Q) \geq \alpha_{\min}(Q)$ для всех $Q \in M_{1,f}$.

5.2 $(\mathcal{M}, f)$-когерентные меры риска

В теории когерентных мер риска, множество приемлемых рисков A задается одновременно с вектором «безопасного актива» $I = (1, \ldots, 1)$ и соответствующая мера риска, которую мы можем ассоциировать с парой $(A,\ I)$ определяется как

$$\rho_{(A,I)}(X) = \sup\{b \in \mathbf{R} :\ X - bI \in A\}.$$

В работе [8] предлагается расширить перечень финансовых инструментов, используемых как инструменты хеджирования: ввести в рассмотрение подпространство или конус финансовых инструментов $\mathcal{M}$ и функционал стоимости $g \in \mathcal{X}^*$($\mathcal{X}^*$ - пространство, сопряженное к $\mathcal{X}$), значение которого $g(M)$ - есть цена инструмента $(-M)$, добавление которого в портфель к неприемлемому риску X, переводит получаемый портфель $X - M$ во множество приемлемых рисков.

Пусть $\mathcal{X}$ - линейное пространство. Множество $D \subseteq \mathcal{X}$, такое что $D + D \subseteq D$ и $\lambda D \subseteq D$, $\forall\lambda \in R_+$, называется ***конусом***. При этом конус, для которого $D \cap (\overline{D}) = \{0\}$ называется острым конусом.

Пусть на $\mathcal{X}$ задано отношение частичного порядка $\geq$. $C_+ = \{X \in \mathcal{X} :\ X \geq 0\}$ называется положительным конусом для частичного порядка на $\mathcal{X}$.

Множество приемлемых рисков A будем называть *когерентным*, если оно является конусом в пространстве рисков $\mathcal{X}$, таким что $C_+ \subseteq A$. Легко проверяется, что выполнение этого условия эквивалентно выполнению аксиом **A1**–**A3**. Подпространство $\mathcal{G}$ пространства $\mathcal{X}$ называется *конфинальным*, если для любого $X \in \mathcal{X}$ существует $Y \in \mathcal{G}$, что $Y \geq X$.

Функционал $g \in \mathcal{X}^*$ называется положительным функционалом на D, если $g(X) \geq 0$, $\forall X \in D$, и строго положительным функционалом на D, если $g(X) > 0$, $\forall C \in D \setminus D \cap (-D)$.

Пусть задано подпространство $\mathcal{M}$ пространства рисков $\mathcal{X}$, функционал $g \in \mathcal{X}^*$ и когерентное множество приемлемых рисков A : $\quad A \cap \mathcal{M} \neq \emptyset$.

Подобно тому, как определялась когерентная мера риска, связанная с парой (A, I) (классическая когерентная мера), определяется ***мера риска, связан-***

ная с тройкой $(A, \mathcal{M}, g)$:

$$\rho_{A,\mathcal{M},g}(X) = \sup\{g(M) \in \mathbf{R} : \ X - M \in A, \ M \in \mathcal{M}\} \tag{18}$$

Заметим, что классическую когерентную меру риска также можно определить через (18), если положить $M = [\{I\}]$, а за g взять любой линейный функционал, такой, что $g(I) = 1$.

Функционал g назван авторами *функционалом премии*. Подпространство $\mathcal{M}$ названо *подпространством хеджирования* – в нем для каждого риска X находится такой M, при вычитании которого получается безрисковая для инвестора позиция $X - M$.

Определение 10. ***$(\mathcal{M}, g)$-когерентная мера риска*** *(или когерентная мера риска, связанная с тройкой $(A, \mathcal{M}, g)$)- это функционал $\rho : \mathcal{X} \to \mathbf{R}$, удовлетворяющий свойствам супераддитивности, положительной однородности, монотонности и свойству $(\mathcal{M}, g)$-инвариантности относительно сдвига:*

$$\rho(X + M) = \rho(X) + g(M), \quad \forall X \in \mathcal{X}, \ M \in \mathcal{M};$$

Теорема 3. *Множество приемлемых рисков*

$$A_\rho = \{X \in \mathcal{X} : \quad \rho(X) \geq 0\}$$

*некоторой $(\mathcal{M}, g)$-когерентной меры риска является непустым множеством, удовлетворяющим аксиомам **A1–A3**, если существует некоторый ненулевой риск $E \in C_+ \cap \mathcal{M}$ и g - строго положительный функционал на C_+.*

Укажем, при каких условиях на подпространство хеджирования мера риска, связанная с тройкой $(A, \mathcal{M}, g)$ будет являться $(\mathcal{M}, g)$-когерентной.

Теорема 4. *Если A - когерентное множество приемлемых рисков, $\mathcal{M}$ - конфинальное подпространство $\mathcal{X}$ и g - строго положительный функционал на A, то $\rho_{A,M,g}$ – $(\mathcal{M}, g)$-когерентная мера риска.*

5.3 Модификация когерентных мер риска Джарроу и Пурнанандама

В работе Р. Джарроу и А. Пурнанандама [4] была предложена одна из модификаций когерентных мер риска, которую для краткости назовем ДП-мерой.

В этой модификации аксиоматика множеств приемлемых рисков совпадает с аксиоматикой в [3]. Соответствующая этому множеству мера риска вводится следующим образом:

$$j_A(X) = j_{A,\|\cdot\|}(X) = -\inf_{Y \in A} \|X - Y\| \tag{19}$$

То есть, значение меры риска - это наименьшее из расстояний от произвольного риска X до множества A (расстояние берется в смысле некоторой нормы в $\mathcal{X}$), взятое со знаком "минус".

Как следует из (19)

$$j_A(X) = 0 \ \forall X \in A \ \text{ и } j_A(X) < 0 \ \forall X \in A^c \tag{20}$$

Введенная таким образом мера обладает свойствами монотонности, супераддитивности и положительной однородности (4)–(6), а вместо свойства (7) появляется свойство ***кратчайшего пути***:
1) для любого $X \in A$ существует такой риск $X' \in A$, что X'- это элемент из множества A, находящий на наименьшем расстоянии от X;
2) для любого числа $-\infty < \lambda \leq -j_A(X)$ выполняется следующее соотношение:

$$j_A(X + \lambda u) = j_A(X) + \lambda, \quad \text{где } u = \frac{X' - X}{\|X' - X\|}. \tag{21}$$

Определение 11. ***ДП-мерой*** *называется любой функционал $j(X)$ на множестве $\mathcal{X}$, обладающий свойствами (4)–(6), (21)*

Соответствующее множество приемлемых рисков находится следующим образом:

$$A_j(X) = \{X \in \mathcal{X} : \ j(X) = 0\}. \tag{22}$$

Ключевое различие между когерентными и ДП-мерами заключается в свойстве (21). Свойство кратчайшего пути выражает линейность меры риска вдоль направления, задающего кратчайший путь до множества приемлемых рисков. Это отражает стремление инвестора регулировать риск портфеля добавлением рискованных активов.

Можно сказать, что *значение модуля ДП-меры для неприемлемых рисков - это наименьшая возможная цена такого рискованного актива, добавив который в портфель, инвестор получит портфель с приемлемым для него риском.*

Однако, ДП-мера не дает возможности сравнивать между собой приемлемые риски (это следует из (20)), что существенно ограничивает сферу ее возможного применения.

Кроме того, ДП-мера не является обобщением когерентных мер. Сравним свойство инвариантности относительно сдвига для когерентных мер и свойство кратчайшего пути для ДП-мер. Предположим, что вложение денежных средств является единственной возможностью регулировать риск портфеля. Тогда вектор u из (21) будет равен $I/\|I\|$. Для одного и того же множества A для когерентной меры ρ_A будет выполняться $\rho_A(X + \lambda I) = \rho_A(X) + \lambda \ \forall \lambda \in \mathbf{R}$, а для ДП-меры j_A свойство $j_A(X + \lambda I) = j_A(X) + \lambda$ выполняется только для $-\infty < \lambda \leq -j_A(X)/\|I\|$. То есть, свойство кратчайшего пути не вырождается в свойство инвариантности относительно сдвига.

5.4 Обобщенные когерентные меры риска

В работе А. Новоселова [10] предлагается модификация когерентных мер, названная обобщенными когерентными мерами риска.

Для модификации когерентных мер в работе Новоселова [1] используется прием, похожий на то, что используется в [4] - вычисляется кратчайшее расстояние, но не до множества приемлемых рисков, как в [4], а до границы этого множества. Это позволяет с помощью предложенных мер риска сравнивать между собой приемлемые риски и считать эти меры действительным обобщением когерентных мер.

Множество приемлемых рисков A удовлетворяет аксиомам **A1-A3**. По заданному A строится обобщенная когерентная мера риска f_A:

$$f_A(X) = f_{A,\|\cdot\|}(X) = \delta_A(X) \inf_{Y\in\partial A} \|X - Y\|, \quad \text{где } \delta_A(X) = \begin{cases} 1, & X \in A, \\ -1, & X \in A^c \end{cases}, \tag{23}$$

здесь ∂A - это граница множества A.

Определенный таким образом функционал $f_A(X)$ обладает свойствами монотонности, положительной однородности, супераддитивности и *свойству кратчайшего пути*:
а) $\forall X \in \mathcal{X}\ \exists\, X'(X) \in \partial A$
такой, что $\|X - X'(X)\| = \inf\limits_{Y\in\partial A} \|X - Y\|$;

б) $f_A(X + \lambda u(X)) = f_A(X) + \lambda, \quad -\infty < \lambda \le \lambda_A(X),$

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 $\|\cdot\|_\infty$

$$u(X) = I = (1, 1, \ldots, 1);$$

- для нормы $\|\cdot\|_1$

$$u(X) = e_i\ \forall i = 1, 2, \ldots, n,$$

где e_i — векторы из стандартного базиса пространства $\mathbf{R}^n$.

Для заданной обобщенной когерентной меры $f(X)$ можно можно найти соответствующее множество приемлемых рисков следующим образом:

$$A_f = \{X \in \mathcal{X} : \ f(X) \ge 0\}. \tag{24}$$

Граница множества приемлемых рисков, соответствующего мере f определяется как

$$\partial A_f = \{X \in \mathcal{X} \quad f(X) = 0\}.$$

Рассмотрим теорему о представлении обобщенной когерентной меры риска с помощью линейных функционалов.

Обозначим $\mathcal{X}^*$ пространство непрерывных линейных функционалов на $\mathcal{X}$, $\|\cdot\|_*$ — норму в этом пространстве, и рассмотрим в нем конус A^*, двойственный к A. Он состоит из всевозможных линейных непрерывных функционалов, принимающих на всех элементах из A только неотрицательные значения:

$$A^* = \{g \in \mathcal{X}^* : \ g(X) \ge 0,\ X \in A\}. \tag{25}$$

В этом двойственном конусе выделим подмножество функционалов, лежащих на единичной сфере :

$$A_1^* = \{g \in A^* : \ \|g\|_* = 1\}.$$

Теорема 5. *Пусть f_A — обобщенная когерентная мера риска, соответствующая множеству приемлемых рисков A. Тогда справедливо представление*

$$f_A(X) = \inf_{g\in A_1^*} g(X), \ \ X \in \mathcal{X}. \tag{26}$$

Верно и обратное: если $A_1^* \subseteq C_+^*$ — некоторое множество функционалов с единичной нормой, то (26) задает обобщенную когерентную меру риска.

Можно утверждать, что мера риска f является обобщением когерентных мер.

Теорема 6. *Пусть $|\Omega| = n$, а в $\mathcal{X} = \mathbf{R}^n$ задана норма $\|\cdot\|_\infty$. Тогда обобщенная когерентная мера риска представляет собой классическую когерентную меру риска.*

Благодаря введению нормы в $\mathcal{X}$ обобщенные когерентные меры риска являются более широким классом функционалов, чем когерентные меры, их можно считать более гибким инструментом для описания индивидуального отношения к риску.

В работе [9] в рамках модели обобщенных когерентных мер риска изучаются свойства множеств приемлемых рисков для предпочтений, обладающих неприятием риска и согласованных со стохастическим доминированием. Предлагается метод решения обратной задачи теории риска - нахождения обобщенной когерентной меры риска, согласованной с индивидуальными предпочтениями.

Список литературы

[1] Новоселов А.А. Обобщенные когерентные меры риска. *Тр. IV Всеросс. ФАМ-конференции, Красноярск, 2005*, 1:325–339, 2005.

[2] Филиппов К.В. Современные подходы к оценке рыночного риска инвестиционного портфеля ценных бумаг. *Горный информационно-аналитический бюллетень (научно-технический журнал)*, 2-12:363–374, 2009.

[3] Artzner Ph., Delbaen F., Eber J.-M., Heath D. Coherent measures of risk. *Mathematical Finance*, 9:203–228, 1999.

[4] Jarrow R.A. Purnanandam A.K. Generalized coherent risk measures: The firm's perspective. *Finance Research Letters*, 2:23–29, 2005.

[5] Acerbi C.,Tasche D. Expected shortfall: a natural alternative to value at risk. *Economic notes*, 31:379–388, 2001.

[6] Balasanov Y. Var is not appropriate measure for risk and economical capital. *Bank of America working report*, 1999.

[7] Follmer H.Schied A. Convex measures of risk and trading constraints. *Finance and Stochastics*, 6:429–447, 2002.

[8] Kounzakis C.E. Generalized coherent risk measures. *Applied Mathematical Sciences*, 3-49:2437–2451, 2009.

[9] Kustitskaya T. A. Representation of preferences by generalized coherent risk measures. *Журн. СФУ. Сер. Матем. и физ.*, 5:451–461, 2012.

[10] Novosyolov A. Generalized coherent risk measures in decision-making under risk. *Тр. Международной Научной Школы МА БР-2005 / ГОУ ВПО «СПбГУАП». СПб.*, pages 145–150, 2005.

[11] Wang S. Premium calculation by transforming the layer premium density. *ASTIN Bulletin*, 26-1:71–92, 1996.

[12] Young V.R. Discussion of christofide's conjecture regarding wang's premium principle. *ASTIN Bulletin*, 29-2:191–195, 1999.

Разработка вероятностного алгоритма поиска ассоциативных правил в статистике покупок

Кристина Игоревна Максимова
Сибирский Федеральный Университет
Институт математики
Красноярск
Sabaka_2007@mail.ru

Ирина Владимировна Баранова
Сибирский Федеральный Университет
Институт математики
Красноярск
irinabar@yandex.ru

Аннотация. *В работе рассматривается понятие интеллектуального анализа данных, его назначение и области его применения. Приводится подробное описание алгоритма поиска ассоциативных правил. Разрабатывается вероятностный аналог алгоритма поиска ассоциативных правил. Решается практическая задача поиска ассоциативных правил в статистике покупок продуктов.*

Ключевые слова. *Интеллектуальный анализ данных, ассоциативные правила, алгоритм Apriori, вероятностный алгоритм, закономерность, взаимосвязь*

1 Введение

Нахождение закономерностей между наборами статистических данных является одной из фундаментальных и актуальных задач современного статистического анализа данных. В последние годы одними из самых популярных методов обнаружения закономерностей в данных стали алгоритмы поиска ассоциативных правил. Эти алгоритмы являются наиболее востребованными методами интеллектуального анализа данных [2].

Интеллектуальный анализ данных (Data Mining) — мультидисциплинарная область, возникшая и развивающаяся на базе прикладной статистики, искусственного интеллекта, теории баз данных и др. 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 1. *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 [3].

К методам интеллектульного 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.

Как уже было сказано, наиболее востребованными методами интеллектуального анализа данных являются алгоритмы поиска ассоциативных правил [3]. Данные методы предназначены 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).

В работе приводится подробный алгоритм поиска ассоциативных правил. Предлагается вероятностный аналог алгоритма Apriori, а также проводится его модификация, которая позволяет улучшить конечные результаты поиска зависимостей в статистике. Приводится математическое обоснование разработанного алгоритма. Также в работе решается практический пример нахождения ассоциативных правил на основе реальной статистики покупок товаров.

2 Основные задачи и применение интеллектуального анализа данных

1. Классификация (Classification)

© 2014 К.И. Максимова, И.В. Баранова

Олег Воробьев (ред.), Труды XIII ФАМЭМС'2014, Красноярск: СФУ

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 следующие 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 генная 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 генной 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. Рассмотрим алгоритм работы метода ассоциативных правил в интеллектуальном анализе данных.

4.1 Постановка задачи поиска ассоциативных правил

Задача поиска ассоциативных правил представляет собой выявление статистических зависимостей между множествами объектов, связанных между собой. Под ассоциативным правилом понимается зависимость между двумя множествами объектов, при которой из наличия в наборе данных одного множества следует появление другого с некоторой заданной вероятностью.

Отличие задачи ассоциации от известных задач классификации и кластеризации данных заключается в том, что в данной задаче поиск закономерностей осуществляется не на основе свойств одного анализируемого объекта (события), а между несколькими событиями, которые происходят одновременно.

Приведем ряд необходимых определений, связанных с ассоциативными правилами, а также сформулируем постановку задачи поиска ассоциативных правил.

Пусть задано конечное множество объектов $Q = \{q_1, q_2, \dots q_n\}$ и множество событий $\mathfrak{X} = \{x_1, x_2, \dots x_n\}$. Здесь каждое событие x_i означает наличие в данных информации об объекте q_i.

Определение 2. *Транзакцией $T \subseteq Q$ называется множество событий, которые произошли одновременно $T = \{x \in \mathfrak{X}\}$, где $|T| \leq n$. (т.е. одновременно наблюдались соответствующие объекты из Q)*

Определение 3. *Пусть задана выборка наблюдений D (или, иначе, транзакционная база данных). Она представляет собой множество транзакций, т.е. множество наблюдаемых наборов событий, перечисленных в транзакциях:*

$$D = \{T_1, T_2, T_r, ..., T_m\}.$$

Теперь приведем понятие ассоциативного правила. К сожалению, проведенный обзор работ по методам поиска ассоциативных правил показал отсутствие математически корректного определения для ассоциативного правила, а также постановки задачи нахождения правил. В большинстве работ под ассоциативным правилом авторы понимают импликацию между наборами данных $I_B(X, Y) = X \Longrightarrow Y = \bar{X} \bigcup Y$, где $X \subseteq \mathfrak{X}$ и $Y \subseteq \mathfrak{X}$, что не корректно, поскольку классическая импликация $I_B(X, Y)$ справедлива для одиночных событий, а не их множеств.

Также довольно часто специалисты в области поиска ассоциативных правил и интеллектуального анализа данных ограничиваются словесной постановкой задачи: «если в транзакции встретился некоторый набор элементов X, то на основании этого можно сделать вывод о том, что другой набор элементов Y также же должен появиться в этой транзакции».

Поэтому устраним данный недостаток и введем более корректное определение ассоциативного правила.

Определение 4. *Под ассоциативным правилом будем понимать функцию*

$$f : [0, 1]X[0, 1] \to [0, 1],$$

, которая двум любым множествам событий $X \subseteq \mathfrak{X}$ и $Y \subseteq \mathfrak{X}$ ставит в однозначное соответствие вели-

чину

$$f(X,Y) = \frac{N(X \bigcup Y)}{m},$$

где

$$N(X \bigcup Y) = |T \supseteq (X \bigcup Y)|, T \in D,$$

а $m = |D|$.

Установление таких зависимостей дает нам возможность находить очень простые и интуитивно понятные правила.

Величину $\frac{N(X \bigcup Y)}{m}$ также принято называть поддержкой правила.

Определение 5. *Поддержка правила — количество транзакций, содержащих определенный набор данных.*

Сформулируем постановку задачи поиска ассоциативных правил: Требуется найти такие события $X \subseteq \mathfrak{X}$ и $Y \subseteq \mathfrak{X}$, для которых функция $f(X,Y) > h$, где h — минимально заданый уровень поддержки.

4.2 Классический алгоритм поиска ассоциативных правил Apriori

Приведем основные обозначения, связанные с основными классическими алгоритмами поиска ассоциативных правил.
$\mathbf{D}$ — множество транзакций $\mathbf{D} = \{\mathbf{T_1}, \mathbf{T_2}, \mathbf{T_r}, ..., \mathbf{T_m}\}$.
$\mathbf{L_k}$ — множество $\mathbf{k}$-элементных наборов, чья поддержка не меньше заданной пользователем.
$\mathbf{C_k}$ — множество сформированных элементов на $\mathbf{k}$-шаге.
$\mathbf{h}$ — минимально заданый уровень поддержки.

Классическим алгоритмом поиска ассоциативных правил является алгоритм Apriori, который имеет следующий вид:
Шаг 1. Присвоить $\mathbf{k = 1}$ и выполнить отбор $\mathbf{L_1}$.
Шаг 2. $\mathbf{k = k + 1}$.
Шаг 3. Если не удается $\mathbf{L_k}$, то завершить алгоритм, иначе выполнить следующий шаг.
Шаг 4. Создать множество $\mathbf{k}$-элементных наборов кандидатов из $\mathbf{L_{k-1}}$. Для этого необходимо объединить в $\mathbf{k}$-элементные кандидаты $(\mathbf{k-1})$-элементные частые наборы. Каждый кандидат будет формироваться путем добавления к $(\mathbf{k-1})$-элементному частому набору – p элемента из другого $(\mathbf{k-1})$-элементного частого набора – q. Причем добавляется последний элемент набора q, который по порядку выше, чем последний элемент набора p.
При этом все $(\mathbf{k-2})$ элемента обоих наборов одинаковы.
Шаг 5. $\forall \mathbf{T}$ из множества $\mathbf{D}$ выбрать кандидатов $\mathbf{C_t}$ из множества $\mathbf{C_k}$, присутствующих в транзакции $\mathbf{T}$. Для каждого набора из построенного множества $\mathbf{C_k}$ удалить набор, если хотя бы одно из его $(\mathbf{k-1})$ подмножеств не является часто встречающимся т.е. отсутствует во множестве $\mathbf{L_{k-1}}$.
Шаг 6. Выбрать только кандидатов $\mathbf{L_k}$ из множества $\mathbf{C_k}$, у которых значение поддержки не меньше заданной пользователем. Вернуться к шагу 2.
Результатом работы алгоритма является $\cup \mathbf{L_k} \forall \mathbf{k}$.

5 Вероятностный аналог алгоритма поиска ассоциативных правил Apriori

Приведем некоторые основные понятия теории вероятностей и математической статистики, необходимые для понимания алгоритма.

Определение 6. *Вероятностным пространством* называется тройка $(\Omega, \mathcal{F}, \mathbf{P})$, где Ω — пространство элементарных событий, $\mathcal{F}$ — алгебра событий и $\mathbf{P}$ — вероятность, определенная на элементах алгебры $\mathcal{F}$ — случайных событиях $x, y, \ldots \in \mathcal{F}$.

Определение 7. Конечное множество избранных событий $\mathfrak{X} \in \mathcal{F}$, выбранных из алгебры вероятностного пространства $(\Omega, \mathcal{F}, \mathbf{P})$ и состоящее из $N = |\mathfrak{X}|$ событий, называется *множеством случайных событий.*

Определение 8. *Случайное множество событий под* $\mathfrak{X}$, определяется как измеримое отображение

$$K : (\Omega, \mathcal{F}, \mathbf{P}) \to \left(2^{\mathfrak{X}}, 2^{2^{\mathfrak{X}}}\right),$$

где $2^{\mathfrak{X}}$ — множество всех подмножеств множества $\mathfrak{X}$.

Множество случайных событий $\mathfrak{X} \in \mathcal{F}$, выбранных из алгебры вероятностного пространства $(\Omega, \mathcal{F}, \mathbf{P})$ или случайное множество событий K под $\mathfrak{X}$ можно определить эквивалентным образом, задав на множестве $2^{\mathfrak{X}}$ всех подмножеств конечного множества $\mathfrak{X}$ одну из следующих функций множеств: эвентологическое распределение (Э-распределение) вероятностей пересечений событий эвентологическое распределение вероятностей пересечений событий

$$p(X) = \mathbf{P}\left(K = X\right) = \mathbf{P}\left(\bigcap_{x \in X} x \bigcap_{x \in X^c} x^c\right), X \subseteq \mathfrak{X},$$

либо Э-распределение вероятностей прямых пересечений событий

$$p_X = \mathbf{P}\left(X \subseteq K\right) = \mathbf{P}\left(\bigcap_{x \in X} x\right), X \subseteq \mathfrak{X},$$

либо Э-распределение вероятностей дополнительных пересечений событий

$$p^X = \mathbf{P}\left(K \subseteq X\right) = \mathbf{P}\left(\bigcap_{x \in X^c} x^c\right), X \subseteq \mathfrak{X},$$

либо Э-распределение вероятностей объединений событий

$$u(X) = 1 - \mathbf{P}\left(K = X^c\right) = \mathbf{P}\left(\bigcup_{x\in X} x \bigcup_{x\in X^c} x^c\right),\ X \subseteq \mathfrak{X},$$

либо Э-распределение прямых объединений событий

$$u_X = 1 - \mathbf{P}\left(K \subseteq X^c\right) = \mathbf{P}\left(\bigcup_{x\in X} x\right),\ X \subseteq \mathfrak{X},$$

либо Э-распределение вероятностей дополнительных объединений событий

$$u^X = 1 - \mathbf{P}\left(X^c \subseteq K\right) = \mathbf{P}\left(\bigcup_{x\in X^c} x^c\right),\ X \subseteq \mathfrak{X}.$$

Определение 9. *Сет–модой называется множество:* $\mathcal{E}K = \{x : p_x \geq h\}$.

Определение 10. *Слоями (n-слоями)*множества событий $\mathfrak{X}$ называются следующие множества подмно=-жеств событий:

$$C^n_{\mathfrak{X}} = \{X \subseteq \mathfrak{X} : |X| = n\},\ n = 0, \ldots, N.$$

Пусть задано конечное множество объектов $Q = \{q_1, q_2, \ldots, q_n\}$ и множество событий $\mathfrak{X} = \{x_1, x_2, \ldots x_n\}$. Здесь каждое событие x_i означает наличие в данных информации об объекте q_i.

Приведем обозначения, используемые в алгоритме:
$\mathbf{D}$ — множество транзакций (чеков покупок) $\mathbf{D} = \{D_1, D_2, ..., D_M\}$.
$\mathbf{L_k}$ — множество **k**-элементных наборов, чья вероятность не меньше заданной пользователем.
$\mathbf{C_k}$ — множество потенциально частых **k**-элементных наборов.
$\mathbf{p_{min}}$ — минимальное значение вероятности, заданное пользователем

Вероятность будем оценивать следующей формулой:

$$p(X) = \mathbf{P}\left(K = X\right) \approx \frac{N(X)}{m},$$

где $N(X) = |T_X|$ — количество чеков с данным множеством товаров , здесь $T_X = \{T_i \in D : X \subseteq T_i\}$, $m = |D|$. Алгоритм иеет вид:
Шаг 1. Присвоить $k = 1$ и выполнить отбор таких $X \subseteq \mathfrak{X}, p_x(X) \geq p_{min}$, т.е $L_1 = \{X \subseteq \mathfrak{X}, p_x(X) \geq p_{min}\}$. В качестве p_{min} принимаем значение $p_{min} \in [0, 1]$.
Шаг 2. Присвоить $k = k + 1$.
Шаг 3.Множество L_k формируется следующим образом: $\forall X \subseteq L_{k-1}$ и $\forall Y \subseteq L_{k-1}$, таких что $x_i \neq y_i$ получаем подмножество $X \bigcup Y$, добавляем $X \bigcup Y$ во множество L_k, если $X \bigcup Y$ удовлетворяет условию: $p_x(X \bigcup Y) \geq p_{min}$.
Шаг 4. Если L_k пусто, то завершить алгоритм, иначе перейти на шаг 2.
Результатом работы алгоритма являются множества L_{k-1}, $k = 2, 3...N$, которые не содержатся во множествах большей длины.

5.1 Вероятностный аналог алгоритма AprioriSome

Шаг 1. Присвоить $k = 1$ и выполнить отбор таких $X \subseteq \mathfrak{X}, p_x(X) \geq p_{min}$, т.е $L_1 = \{X \subseteq \mathfrak{X}, p_x(X) \geq p_{min}\}$. В качестве p_{min} принимаем значение $p_{min} \in [0, 1]$.
Шаг 2. Выбрать шаг, согласно рисунку 3.1. Присвоить $k = k + p$.
Шаг 3.Множество L_k формируется следующим образом: $\forall X \subseteq L_{k-p}$ и $\forall Y \subseteq L_{k-p}$, такого что $x_i \neq y_i$ получаем подмножество $X \bigcup Y$, добавляем $X \bigcup Y$ во множество L_k, если $X \bigcup Y$ удовлетворяет условию: $p_x(X \bigcup Y) \geq p_{min}$ Шаг 4. Если L_k пусто, то завершить алгоритм, иначе перейти на шаг 2.
Результатом работы алгоритма является $\cup L_{k-p}$,$k = 2, 3...N$, которые не содержатся во множествах большей длины.

5.2 Модифицированный вероятностный аналог алгоритма Apriori

Приведем алгоритм модифицированного вероятностного аналога алгоритма Apriori. Данный алгоритм отличается тем, что на шаге формирования множества L_k, дается возможность переопределения изначально заданного уровня p_{max} согласно условиям алгорима. Что в свою очередь позволяет не отбросить лишние случайные множества, а рассмотреть наиболее общую картину.
Шаг 1. Присвоить $k = 1$ и выполнить отбор таких $X \subseteq \mathfrak{X}$, для которых верно: $p_x(X) \geq p_{max}$, т.е формируем $L_1 = \{X \subseteq \mathfrak{X}, p_x(X) \geq p_{max}\}$. В качестве p_{max} принимаем значение $p_{max} \in [0, 1]$, чтобы $p_x(X) \geq p_{max}$ и мощность $|L_1| \leq |C_1|/2$.
Шаг 2. Присвоить $k = k + 1$.
Шаг 3. Пользователь задает значение $p_{max} \in [0, 1]$, такое, что $p_{max} \geq p_{min}$ и верно $|L_k| \leq |C_k|/2$, $L_k = \{X \subseteq \mathfrak{X}, p_x(X) \geq p_{min}\}$.
Шаг 4. Множество L_k формируется следующим образом: $\forall X \subseteq L_{k-1}$ и $\forall Y \subseteq \mathfrak{X}/X$, такого что $x_i \neq y_i$ получаем подмножество $X \bigcup Y$. Если вероятность $p_x(X \bigcup Y) \geq p_{max}$, то добавляем $X \bigcup Y$ во множество L_k.
Шаг 5. Если L_k пусто, то завершить алгоритм, иначе перейти на шаг 2.
Результатом работы алгоритма является $\cup L_{k-1}$, $k = 2, 3...N$ которые не содержатся во множествах большей длины.

5.3 Вероятностный аналог алгоритма Apriori по слоям

Данный алгоритм был разработан для нахождения частовстречающихся множеств случайных событий на слоях. В связи с этим, была переопределена оценка вероятности. Вероятность будем оценивать следующей формулой:

$$p(X) = \mathbf{P}\left(K = X\right) \approx \frac{N(X)}{M}$$

где M —количество множест элементов статистики, $N(X)$ —количество наборов X на данном слое.
Шаг 1. Присвоить $k = 1$ и выполнить отбор таких $X \subseteq \mathfrak{X}$, для которых верно: $p(X) \geq p_{min}, X \subseteq C^1_{\mathfrak{X}}$, т.е формируем $L_1 = \{X \subseteq \mathfrak{X}, p(X) \geq p_{min}, X \subseteq C^1_{\mathfrak{X}}\}$. В качестве p_{min} принимаем значение $p_{min} \in [0, 1]$, чтобы $p(X) \geq p_{min}$ и мощность $|L_1| \leq |C^1_{\mathfrak{X}}|/2$.
Шаг 2. Присвоить $k = k + 1$.
Шаг 3. Пользователь задает новое значение $p_{min} \in [0, 1]$, такое, что для $L_k = \{X \subseteq \mathfrak{X}, p(X) \geq p_{min}\}$ верно $|L_k| \leq |C^k_{\mathfrak{X}}|/2$.
Шаг 4. Множество L_k формируется следующим образом: $\forall X \subseteq L_{k-1}$ и $\forall Y \subseteq \mathfrak{X}/X$, такого что $x_i \neq y_i$ получаем подмножество $X \bigcup Y$. Если вероятность $p_x(X \bigcup Y) \geq p_{min}$, то добавляем $X \bigcup Y$ во множество L_k.
Шаг 5. Если L_k пусто, то завершить алгоритм, иначе перейти на шаг 2. Результатом работы алгоритма является множество L_{k-1}.

Данный алгоритм удобен при равномерном распределении случайного множества по слоям. При таком распределении алгоритм будет работать наилучшим образом.

6 Практический пример нахождения ассоциативных правил

Решим практическую задачу нахождения ассоциативных правил по описанному методу. Одним из продуктовых розничных магазинов г. Красноярска нам была предоставлена недельная статистика покупок продуктов за март 2012 года. Всего статистика содержала 1340 транзакций (чеков покупок). Всего в статистике содержалось 50 различных товаров: молоко, хлеб, сметана, творог, сыр, яйца, колбаса и т.д. На основе статистики было необходимо найти закономерности между покупками в наборе данных.

6.1 Вероятностнный аналог алгоритма Apriori

Для решения практической задачи нахождения ассоциативных правил в данной статистике были использованы вышеописанный классический метод поиска ассоциативных правил (алгоритм Apriori).

Согласно вероятностному аналогу алгоритма Apriori был определен уровень $p_{min} = 0,15$. На первом этапе были сформированы одноэлементные кандидаты. Наборы с уровнем p_{min} меньше установленного значения были отсечены. Результаты представлены в таблице 1.

Таблица 1: Одноэлементные кандидаты

Наименование	Уровень поддержки
Молоко	0,33
Хлеб пшеничный	0,63
Сыр твердый	0,29
Фрукты	0,24
Полуфабрикаты	0,2
Соусы	0,19
Сметана	0,16
Мясо куриное	0,18
Яйца	0,19
Слабоалкогольные напитки	0,18
Консервы	0,17

Далее находим множество двухэлементных кандидатов. При формировании двухэлементных кандидатов наибольшую поддержку $p_{min} = 0,15$ имеют наборы, представленные в таблице 2.

Таблица 2: Двухэлементные кандидаты

Наименование	Уровень поддержки
Сыр твердый, молоко	0,16
Хлеб, молоко	0,21
Фрукты, молоко	0,18
Сыр, хлеб	0,2
Хлеб, фрукты	0,17

На третьем этапе алгоритм останавливается, так как трехэлементные наборы имеют значения вероятностей меньше, чем p_{min} это можно увидеть из таблицы 3.

Таблица 3: Трехэлементные кандидаты

Наименование	Уровень поддержки
Молоко, сыр, хлеб	0,13
Молоко, сыр, фрукты	0,07
Молоко, хлеб, фрукты	0,12
Сыр, хлеб, фрукты	0,07

В результате исследования были обнаружены следующие ассоциативные правила:наибольшим числом

покупок стали пары товаров {Сыр твердый, молоко}, {Хлеб, молоко}, {Фрукты, молоко}, {Сыр, хлеб}, {Хлеб, фрукты}.

6.2 Решение модифицированным вероятностным аналогом алгоритма Aprirori

Рассмотрим те же статистические данные и решим задачу модифицированным вероятностным аналогом алгоритма Apriori на языке вероятностей.

Согласно алгоритму было определено значение вероятностей для данной задачи p_{max}, такое что бы оно удовлетворяло условию: $|L_1| = |C_1|/2$. На первом этапе было сформировано множество L_1, которое можно увидеть в таблице 26. Наборы с вероятностью меньше установленного значения (p_{max}=0,121) были отсечены. Стоит заметить, что значение веротяности на первом шаге алгоритма меньше значения вероятности в классическом алгоритме, это значит, что уже на первом шаге данный алгоритм рассматривает больше кандидатов в L_2, чем классический алгоритм Apriori.

Таблица 4: Множество L_1

Наименование	Значение вероятности
Молоко	0,334
Сметана	0,158
Сыр	0,288
Колбаса	0,13
Мясо куриное	0,177
Яйца	0,187
Полуфабрикаты	0,199
Хлеб	0,636
Конфеты	0,13
Слабоалкогольные напитки	0,181
Овощи	0,174
фрукты	0,242
Консервы	0,166
Чай	0,123
Соусы	0,188
Крупа	0,139
Йогурт	0,132

Далее находим множество двухэлементных кандидатов L_2. Для формировании L_2 согласно алгоритму выбираем новое значение p_{max}. Наборы с вероятностью меньше установленного значения (p_{max}=0,082) были отсечены. Результаты представлены в таблице 4. Далее, находим множество трехэлементных кандидатов L_3. Выбираем новое значение p_{max}. Наборы с вероятностью меньше установленного значения (p_{max}=0,046) были отсечены. Наборы, удовлетворяющие условиям алгоритма представлены в таблице 5.

Таблица 5: Множество L_2

Наименование	Значение вероятности
Молоко, сметана	0,094
Молоко,сыр	0,118
Молоко, хлеб	0,204
Молоко,фрукты	0,096
Сметана,сыр	0,093
Сметана, хлеб	0,11
Сыр, мясо куриное	0,099
Сыр, консервы	0,092
Сыр, крупа	0,083
Йогурт, хлеб	0,088
Сыр, фрукты	0,084
Хлеб, колбаса	0,102
Хлеб, мясо куриное	0,125
Хлеб, яйца	0,14
Хлеб, полуфабрикаты	0,12
Хлеб, фрукты	0,16
...	...

Таблица 6: Множество L_3

Наименование	Значение вероятности
Молоко, сметана, творог	0,054
Молоко, сметана, сыр	0,057
Молоко, сметана, соусы	0,048
Молоко, сыр, соусы	0,05
Молоко, сыр, колбаса	0,047
Молоко, сыр, хлеб	0,078
...	...

Формируем множество четырехэлементных кандидатов L_4. Выбираем новое значение p_{min}. Наборы с вероятностью меньше установленного значения ($p_{max} = 0,036 = p_{min} = 0,036$) были отсечены.Результаты представлены в таблице 6. Работа алгоритма на этом заканчивается, т. к. $p_{min} = p_{max}$ В результате работы данного алгоритма мы получили ассоциативные правила, с заданными границами значения вероятностей.

Таблица 7: Множество L_4

Наименование	Значение вероятности
Молоко, сметана, творог, сыр	0,038
Молоко, сметана, творог, колбаса	0,037
Молоко, сыр, хлеб, фрукты	0,05
Молоко, сыр, хлеб, мясо кур	0,04
Молоко, сметана, хлеб, фрукты	0,039
...	...

7 Сравнение классического и разработанных алгоритмов

В работе было проведено сравнение классического алгоритма Apriori и разработанных алгоритмов поиска ассциативных правил. Результаты проведенного анализа показали, что основными проблемами методов извлечения ассоциативных правил являются:

- значительная временная сложность поиска ассоциативных правил;
- необходимость нескольких проходов по исследуемым базам данных;
- необходимость обеспечения поиска интересных и нетривиальных правил.
- Большинство методов извлечения ассоциативных правил учитывают только уровень поддержки, при этом не анализируется интересность правил для исследуемой предметной области;
- проблемы выбора пороговых значений поддержки.

Несмотря на недостатки классического метода поиска ассоциативных правил Apriori (недостаточную обоснованность и сравнительно высокую вычислительную сложность), он очень популярен среди исследователей. Однако у алгоритма Apriori есть еще один существенный недостаток: он отбрасывает правила, у которых вероятность больших множеств меньше, чем у элементов их составляющих.

Разработанный модифицированный алгоритм Apriori на языке вероятностей решает эту проблему. С помощью разработанного метода ассоциативные правила получаются более объемные. Вычислительная сложность предложенного модифицированного метода меньше, чем у классического алгоритма Apriori.

Таблица 8: Сравнительная таблица работы алгоритмов

Название	Вычислительная сложность
Классический Apriori	$O(2^k), k <= n.$
Вероятностный аналог Apriori	$O(2^k), k <= n.$
Модифицированный вероятностный аналог	$O(2^{k/2}), k <= n.$
Вероятностный аналог алгоритма Apriori по слоям	$O(2^{k/2}), k <= n.$
Полный перебор	$O(2^n).$

Сравнительная характеристика алгоритмов приведена в таблице 8.

8 Итоги

В работе рассмотрены основные задачи и методы интеллектуального анализа данных, а также изучен метод ассоциативных правил в интеллектуальном анализе данных. Решен практический пример нахождения ассоциативных правил на основе статистики покупок товаров с помощью метода ассоциативных правил, а так же разработан вероятностный аналог алгоритма. Вероятностный аналог классического ассоциативного алгоритма позволяет пойти дальше и найти закономерности покупок товаров в этом наборе, т.е. получить правила такого вида: "Из покупки набора товаров A следует покупка набора товаров B".

Список литературы

[1] Барсегян А.А. , М.С. Куприянов, В.В. Степаненко, И.И. Холод. *Технологии анализа данных. Data Mining, Visual Mining, Text Mining, OLAP.* Спб., БХВ-Петербург, 2004.

[2] Афонин, А.Ю. и П.П. Макарычев. *Оперативный и интеллектуальный анализ данных.* Пенза, ПГУ, 2010.

[3] Чубукова И.А. *Data Mining.* Москва, Бином, 2006.

Проблемы сбалансированности знаковых графов

Игорь Владиславович Махонин
Сибирский Федеральный Университет
Красноярск
IMakhonin24ru@gmail.com

Виталий Анатольевич Демиденко
Сибирский Федеральный Университет
Красноярск
devit15@yandex.ru

Аннотация. *В работе сформулированы основные задачи исследования знаковых графов, предложен алгоритм проверки сбалансированности знаковых графов, приведены основные формулировки проблемы знакового баланса.*
Ключевые слова. *signed graph, balance.*

1 Введение

Неориентированные графы являются подходящей моделью для представления отношений между людьми. Люди изображаются вершинами графа, причем если между лицами a и b имеется симпатия или антипатия, то между вершинами a и b проводится ребро. Аналогичная модель возникает, если отношение "симпатия – антипатия" заменить отношением "общаться – избегать", "соглашаться – не соглашаться". Как правило, такие отношения симметричные. Их описывают с помощью знаковых графов [1].

Пусть задан неориентированный граф $G = (V, E)$, где V — конечное множество вершин графа, а E — множество ребер. Знаковым графом называется неориентированный граф, каждое ребро которого отмечено знаком "+" или "−".

Исследования знаковых графов сводятся к трем основным задачам:

- проверка сбалансированности знаковых графов;
- вычисление меры сбалансированности знаковых графов;
- установление знакового баланса — нахождение набора ребер графа (желательно минимального), изменение знаков которых обеспечивает сбалансированность заданного графа.

Американский математик Харари вместе с психологом Картрайт предложили следующее определение сбалансированности знакового графа: граф считается сбалансированным, если множество его вершин можно разбить на два подмножества таким образом, что любое ребро, соединяющее вершины из одного множества, имеет знак "+", а ребра, соединяющие вершины из разных подмножеств, знак "−". Причем второе подмножество может оказаться пустым.

© 2014 И.В.Махонин, В.А. Демиденко
Олег Воробьев (ред.), Труды XIII ФАМЭМС'2014, Красноярск: СФУ

2 Алгоритм проверки сбалансированности графа

Знаком цикла знакового графа называется произведение знаков всех его ребер. Если цикл имеет знак "+", то цикл называется положительным, в противном случае отрицательным.

Теорема 1. (Картрайт−Харари) [2]. *Следующие утверждения эквивалентны:*

1. $G = (V, E)$ *— сбалансированный граф;*
2. *любой простой цикл графа* G *положительный;*
3. *любые две простые* (u, v)*-цепи графа* G *имеют одинаковый знак* $(u, v \in V)$.

Данная теорема указывает эквивалентные определения свойства сбалансированности знакового графа и конструктивный способ проверки этого свойства. Например, такую проверку можно выполнить, используя алгоритм обхода вершин графа в глубину, за время $O(|V| + |E|)$

3 Меры сбалансированности графа

Вероятность того, что случайно выбранный знаковый граф сбалансирован, крайне мала. Поэтому интересен вопрос, насколько сбалансирован несбалансированный граф. Ответ на этот вопрос дает мера сбалансированности знакового графа. Возможны различные определения меры сбалансированности, однако в социологии и психологии ее определяют следующим образом [2]. Мера сбалансированности знакового графа G — доля простых положительных циклов к общему числу простых циклов этого графа:

$$b(G) = \frac{|C^+|}{|C|}.$$

Очевидно, что $0 \leqslant b(G) \leqslant 1$, причем $b(G) = 1$, если и только если все циклы этого графа положительны, т.е. граф сбалансирован.

Например, граф, изображенный на рис. 1, имеет меру сбалансированности $2/3$.

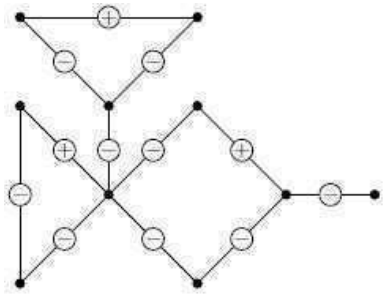


Рисунок 1: Несбалансированный граф с положительными и отрицательными циклами, имеет меру сбалансированности 2/3

Определение меры сбалансированности некорректно, если граф G не содержит ни одного цикла, однако по теореме Картрайт—Харари, такой граф является сбалансированным. Поэтому для него можно считать, что $b(G) = 1$.

4 Вычислительная сложность нахождения меры сбалансированности

При вычислении меры сбалансированности приходится последовательно выявлять все простые циклы этого графа и находить их знаки. Поэтому вычисление меры сбалансированности требует $O(n^2 \cdot 2^n)$ операций, где $n = |V|$. Например, для полного графа с n вершинами общее число простых циклов определяется формулой (эта ситуация отвечает худшему случаю):

$$|C| = \sum_{k=3}^{n} \frac{n!}{(n-k)! \cdot 2k} = O(2^n),$$

а число различных вариантов расстановки знаков ребер равно

$$2^{n(n-1)/2}.$$

5 Проблема знакового баланса

Задача приведения знакового графа к сбалансированному виду (ее называют проблемой баланса) связана с переменой знаков его ребер. Возможны три основные формулировки проблемы знакового баланса:

1. нахождение минимального (по вхождению) множества ребер графа, изменение знаков которых обеспечивает сбалансированность этого графа;
2. установление некоторого набора допустимых множеств ребер, каждое из которых гарантирует сбалансированность графа;
3. применение переборных процедур для перечисления всех допустимых вариантов сбалансированности знакового графа.

Приведенные выше формулировки проблемы знакового баланса приводят к трем комбинаторным NP-трудным задачам. Для них пока неизвестны алгоритмы нахождения точных решений за полиномиальное время. В первом случае речь идет о построении оптимального решения. Вторая формулировка приводит к некоторым вариантам допустимого решения. Третья формулировка — NP-трудная переборная задача. В данных формулировках условие сбалансированности графа можно заменить мерой сбалансированности. Однако это существенно увеличивает сложность вычислений.

В настоящий момент ведется разработка алгоритмов проблемы знакового баланса в оптимизационной формулировке с использованием матричных уравнений.

Список литературы

[1] Ф. С. Робертс. *Дискретные математические модели с приложениями к социальным и экологическим задачам.* М.: Наука, 1986.

[2] Ф. Харари. *Теория графов.* М.: Мир, 1973.

Регрессия как условное средневероятное событие

Яков Викторович Нартов
Институт математики
Сибирский федеральный университет
Красноярск
yanartov@yandex.ru

Аннотация. *В данной статье автор дает событийный взгляд на регрессию, основываясь на понятии условного средневероятного события.*

Ключевые слова. *Вероятность, регрессия, эвентология, средневероятное событие, условное средневероятное событие.*

1 Введение

1.1 Из истории регрессии

Понятие регрессии связано с именем Френсиса Гальтона[1], который внес неоценимый вклад в статистику помимо всего прочего и введением понятия корреляционной зависимости. В период с 1885 года по 1886 год Гальтоном было опубликовано несколько работ, касающихся наследственности роста человека, и самые основные положения этого исследования вошли в седьмую главу книги [7, 1889]. Гальтон рассматривает вопросы закономерностей наследования признаков, а именно он пытается установить общие статистические законы, которые лежат в основе этого явления. Гальтон собрал данные о росте 408 жителей Лондона и росте 928 их взрослых детей (Рисунок 1), затем постарался выяснить существует ли какая-нибудь связь между этими величинами [1]. Для этого он использовал следующий подход.

На левой стороне квадрата откладывается «усредненный» рост родителей в дюймах(для женщин вводился корректирующий коэффициент 1.08), а на верхней его стороне – средний рост их детей в дюймах. На полученной плоскости откладываются точки следующим образом: для значения «усредненного» роста родителей ищется среднее значение роста ребенка(Рисунок 2). Если бы рост родителей и рост детей в среднем совпадали друг с другом (т.е. регрессия равнялась бы единице), то отношение роста детей к росту родителей выразилось бы диагональю квадрата AB. Если бы, наоборот, дети вовсе не наследовали уклонения от средней роста родителей, то отношение между ростом тех и других выразилось бы вертикальной прямой EE. На рисунке точка M – точка среднего роста популяции(of the general Population). Гальтон обнаружил удивительную вещь. Дело в том, что в действительности дети высоких или низких родителей не наследуют их выдающийся рост, а стремятся к среднему росту популяции. Благодаря этому замечанию Гальтона и было введено слово «регрессия», показывая тем самым «регрессию к посредственности».

Работы Гальтона в последствии подвергались критике как специалистами в области генетики, так и статистами. В наше время данная работа [7] по большей части устарела и может быть интересна теперь главным образом для истории статистики и генетики. В последствии ученик Френсиса Гальтона – Карл Пирсон[2] – привел строгое описание корреляции и регрессии [8, 1969], ссылаясь при этом на Браве[3]. Но не смотря на это, работа Гальтона как нельзя лучше передает «дух» математического объекта, именуемым регрессия. Такой новаторский по временам Френсиса Гальтона способ мышления, который способен статистически смотреть на природу вещей, очень родственен современному, а сам Гальтон – один из «законодателей» регрессионного анализа.

1.2 Классическая регрессия

Определение 1. *Реграссией называется зависимость среднего значения какой-либо случайной величины Y от некоторой другой случайной величины X (или от вектора случайных величин). При каждом фиксированном значении x величины X величина Y имеет свое распределение вероятностей с математическим ожиданием, которое и является функцией от x*

$$\mathbf{E}(Y|x) = m(x)$$

и называется функцией теоретической регрессии(линией регрессии, кривой регрессии).

Так, например, если при каждом значении $X_i, i = 1, \ldots, N$ дискретной величины X наблюдается n_i зна-

© 2014 Я.В.Нартов

Олег Воробьев (ред.), Труды XIII ФАМЭМС'2014, Красноярск: СФУ

[1] Френсис Гальтон – (англ. Francis Galton; 16 февраля 1822 – 17 января 1911) английский исследователь, географ, антрополог и психолог; основатель дифференциальной психологии и психометрики

[2] Карл Пирсон – (англ. Karl (Carl) Pearson, 27 марта 1857 – 27 апреля 1936) английский математик, основатель математической статистики

[3] Auguste Bravais – (23 August 1811, Annonay, Ard?che Ц 30 March 1863) французский физик, известный своими работами в кристаллографии, концепцией решеток Браве и формулировкой законов Браве

TABLE 11 (R.F.F. Data).

NUMBER OF ADULT CHILDREN OF VARIOUS STATURES BORN OF 205 MID-PARENTS OF VARIOUS STATURES.

(All Female Heights have been multiplied by 1·08.)

Height of the mid-parents in inches.	Heights of the adult children.													Total number of		Medians or Values of M.
	Below	62.2	63·2	64·2	65·2	66·2	67·2	68·2	69·2	70·2	71·2	72·2	73·2	Above.	Adult children.	Mid-parents.	
Above 72·5...	...	...	...	...	...	...	...	...	...	...	...	1	3	...	4[1]	5[1]	
72·5...	...	...	...	...	...	...	...	1	2	1	2	7	2	4	19	6	72·2
71·5...	...	...	...	...	1	3	4	3	5	10	4	9	2	2	43	11	69·9
70·5...	1	...	1	...	1	1	3	12	18	14	7	4	3	3	68	22	69·5
69·5...	...	...	1	16	4	17	27	20	33	25	20	11	4	5	183	41	68·9
68·5...	1	...	7	11	16	25	31	34	48	21	18	4	3	...	219	49	68·2
67·5...	...	3	5	14	15	36	38	28	38	19	11	4	...	...	211	33	67·6
66·5...	...	3	3	5	2	17	17	14	13	4	...	...	...	...	78	20	67·2
65·5...	1	...	9	5	7	11	11	7	7	5	2	1	...	...	66	12	66·7
64·5...	1	1	4	4	1	5	5	...	2	...	...	...	...	..	23	5	65·8
Below	1	...	2	4	1	2	2	1	1	...	...	...	...	...	14	1	
Totals	5	7	32	59	48	117	138	120	167	99	64	41	17	14	928	205	
Medians	...	...	66·3	67·8	67·9	67·7	67·9	68·3	68·5	69·0	69·0	70·0					

Note.—In calculating the medians, the entries have been taken as referring to the middle of the squares in which they stand. The reason why the headings run 62·2, 63·2, &c., instead of 62·5, 63·5, &c., is that the observations are unequally distributed between 62 and 63, 63 and 64, &c., there being a strong bias in favour of integral inches. After careful consideration, I concluded that the headings, as adopted, best satisfied the conditions. This inequality was not apparent in the case of the mid-parents.

Рисунок 1: Фотография таблицы, содержащей данные об «усредненном» росте родителей и росте их детей, из книги [7].

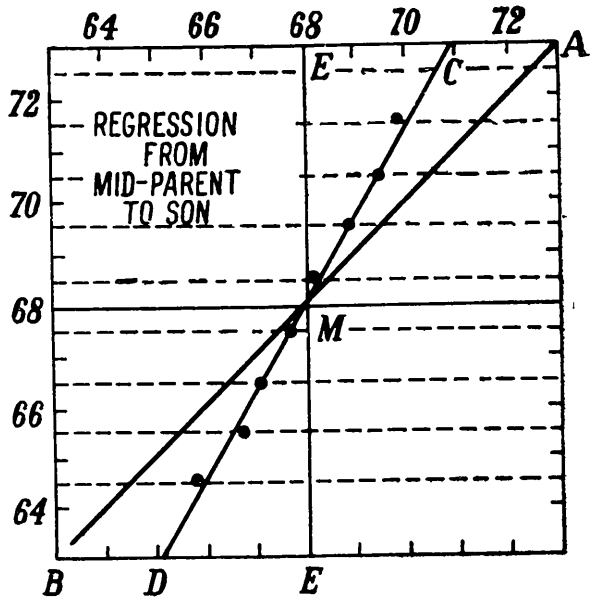


Рисунок 2: Фотография схемы регрессии из книги [7].

чений $y_{i1}, y_{i2}, \ldots, y_{in_i}$ дискретной случайной величины Y, то зависимость средних арифметических

$$\bar{y_i} = \frac{1}{n_i}\left(y_{i1} + y_{i2} + \cdots + y_{in_i}\right), i = 1, \ldots, N$$

этих значений от $X_i, i = 1, \ldots, N$ и есть регрессия.

2 Условное средневероятное событие

2.1 Основные положения

Определение 2. *Средневероятным событием для конечного множества однородных событий $\mathfrak{X} \subset \mathcal{F}$ называется такое всеобщее событие $\hat{x}_{\mathfrak{X}} \in \mathcal{F}$, удовлетворяющее включениям*

$$\sum_{|X|>m} \mathrm{ter}(X//\mathfrak{X}) \subseteq \hat{x}_{\mathfrak{X}} \subseteq \sum_{|X|\geq m} \mathrm{ter}(X//\mathfrak{X}), \quad (1)$$

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 $m \in \{0, 1, \ldots, |\mathfrak{M}|\}$ 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 [6] 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 $C^m_{\mathfrak{X}}$ [3] определяется как совокупность подмножеств событий

$$C^m_{\mathfrak{X}} = \{X \subseteq \mathfrak{X} : |X| = m\} \subseteq 2^{\mathfrak{X}}. \quad (4)$$

Множественный m-пласт $C^m_{\mathfrak{X}}$ [3] определяется как совокупность подмножеств событий

$$C^{m+}_{\mathfrak{X}} = \{X \subseteq \mathfrak{X} : |X| \geq m\} \subseteq 2^{\mathfrak{X}}. \quad (5)$$

Определим террасную операцию «m-слой» следующим образом:

$$\mathrm{terop}(C^m_{\mathfrak{X}}//2^{\mathfrak{X}}) = \sum_{X\in C^m_{\mathfrak{X}}} \mathrm{ter}(X//\mathfrak{X}) \subseteq \Omega \quad (6)$$

Аналогичным образом определим террасную операцию «m-пласт»:

$$\mathrm{terop}(C^{m+}_{\mathfrak{X}}//2^{\mathfrak{X}}) = \sum_{X\in C^{m+}_{\mathfrak{X}}} \mathrm{ter}(X//\mathfrak{X}) \subseteq \Omega \quad (7)$$

Определение 3. *Средневероятным событием для конечного однородного множества событий $\mathfrak{X} \subset \mathcal{F}$ называется любое такое всеобщее событие $\hat{x} \in \mathcal{F}$, которое содержится в террасном m-пласте, содержит террасный $m+1$-пласт:*

$$\mathrm{terop}(C^{(m+1)+}_{\mathfrak{X}}//2^{\mathfrak{X}}) \subseteq \hat{x}_{\mathfrak{X}} \subseteq \mathrm{terop}(C^{(m)+}_{\mathfrak{X}}//2^{\mathfrak{X}}), \quad (8)$$

и наступает со средней вероятностью

$$\mathbf{P}(\hat{x}) = \frac{1}{|\mathfrak{M}|}\sum_{\mu\in\mathfrak{M}} \mathbf{P}(x_\mu) \quad (9)$$

всякий раз, когда среди событий из $\mathfrak{X}$ наступает не менее, чем m событий, где $m \in \{1, 2, \ldots, |\mathfrak{X}|\}$ удовлетворяет неравенствам

$$\begin{gathered}\mathbf{P}(\mathrm{terop}(C^{(m+1)+}_{\mathfrak{X}}//2^{\mathfrak{X}})) \leq \\ \leq \mathbf{P}(\hat{x}) \leq \\ \leq \mathbf{P}(\mathrm{terop}(C^{m+}_{\mathfrak{X}}//2^{\mathfrak{X}}))\end{gathered} \quad (10)$$

Если не вдаваться в подробности, то минимальное число m наступающих событий в результате наступления средневероятного события выбирается из соображения, что вероятность средневероятного события должна быть не менее вероятности террасного $(m+1)$-пласта, но не более вероятности m-пласта.

Средневероятное событие обладает очень важным экстремальным свойством [5].

Определение 4. *Вероятностное расстояние события $\hat{x}_{\mathfrak{X}} \in \mathcal{F}$ до множества $\mathfrak{X} \subset \mathcal{F}$ определяется формулой:*

$$\begin{gathered}\rho(\hat{x}_{\mathfrak{X}}, \mathfrak{X}) = \sum_{x\in\mathfrak{X}} \mathbf{P}(x \triangle \hat{x}_{\mathfrak{X}}) = \\ = \sum_{x\in\mathfrak{X}} \mathbf{P}(\hat{x}_{\mathfrak{X}} \cap x^c) + \sum_{x\in\mathfrak{X}} \mathbf{P}((\hat{x}_{\mathfrak{X}})^c \cap x).\end{gathered} \quad (11)$$

Теорема 1. *Средневероятное событие $\hat{x}_{\mathfrak{X}}$ для множества событий $\mathfrak{X}$ минимизирует вероятностное расстояние до $\mathfrak{X}$:*

$$\rho(\hat{x}_{\mathfrak{X}}, \mathfrak{X}) = \min_{\alpha\in\mathcal{F}; \mathbf{P}(\alpha)=\hat{\mathbf{P}}_{\mathfrak{X}}} \rho(\alpha, \mathfrak{X}) \quad (12)$$

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 средневероятное событие не единственно [2] и все попытки визуализации средневероятного события – в сущности попытки визуализировать семейство событий, каждое из которых в праве претендовать на средневероятное.

2.2 О наименьшем числе наступления событий в средневероятном событии

Так как рассуждения в данном небольшом подразделе основаны на рассуждениях, описанных в работе [2], стоит заметить, что в [2] в формуле (23) автором допущена ошибка. В действительности формула (23) с учетом введенных в статье обозначений(хоть они и не совпадают с обозначениями, введенными в настоящей работе) должна иметь вид

$$P_{\mathfrak{M}}^{\{3\}} < P_{\mathfrak{M}}^{\{1\}} + 2P_{\mathfrak{M}}^{\{2\}} \leq 4P_{\mathfrak{M}}^{\{2\}} + P_{\mathfrak{M}}^{\{3\}}.$$

Введенные эквивалентные определения средневероятного события алгоритмизуемы, однако может быть полезно замечание следующего характера. Ключевым моментом в нахождении средневероятного события является определение наименьшего числа наступления событий(нахождение числа m в определении). Как уже ранее было сказано, основываясь на рассуждениях, описанных в [2], можно сделать вывод следующий.

Замечание Средневероятное событие для множества событий $\mathfrak{X}$ наступает всякий раз, когда среди $\mathfrak{X}$ наступает не менее, чем m событий, где m – решение системы неравенств:

$$\begin{aligned}&\sum_{k=m+1}^{|\mathfrak{X}|}(|\mathfrak{X}|-k)\mathbf{P}(\mathrm{terop}(\mathrm{C}_{\mathfrak{X}}^{\mathrm{k}})) < \\ &< \sum_{k=1}^{m} k\mathbf{P}(\mathrm{terop}(\mathrm{C}_{\mathfrak{X}}^{\mathrm{k}})) \leq \\ &\leq \sum_{k=m+1}^{|\mathfrak{X}|}(|\mathfrak{X}|-k)\mathbf{P}(\mathrm{terop}(\mathrm{C}_{\mathfrak{X}}^{\mathrm{k}})) + |\mathfrak{X}|\mathbf{P}(\mathrm{terop}(\mathrm{C}_{\mathfrak{X}}^{\mathrm{m}}))\end{aligned} \tag{13}$$

2.3 Определение средневероятного события через мощностной индикатор

Определение 5. *Случайная величина $\Upsilon_{\mathfrak{X}}$ равная мощности наступаемого множества $X \subseteq \mathfrak{X}$ называется мощностным индикатором множества $\mathfrak{X}$.*

Определение средневероятного события можно переписать через понятие мощностного индикатора

Определение 6. *Средневероятным событием для конечного однородного множества событий $\mathfrak{X} \subset \mathcal{F}$ называется любое такое всеобщее событие $\hat{x} \in \mathcal{F}$, которое содержится в террасном m-пласте, содержит террасный $m+1$-пласт:*

$$\mathrm{terop}(C_{\mathfrak{X}}^{(m+1)+}//2^{\mathfrak{X}}) \subseteq \hat{x}_{\mathfrak{X}} \subseteq \mathrm{terop}(C_{\mathfrak{X}}^{(m)+}//2^{\mathfrak{X}}), \tag{14}$$

наступает с вероятностью

$$\mathbf{P}(\hat{x}) = \frac{1}{|\mathfrak{X}|}\mathbf{E}\Upsilon_{\mathfrak{X}} \tag{15}$$

всякий раз, когда среди событий из $\mathfrak{X}$ наступает не менее, чем m событий, где $m \in \{1,2,\ldots,|\mathfrak{X}|\}$ удовлетворяет неравенствам

$$\mathbf{P}(\Upsilon_{\mathfrak{X}} \geq m+1) < \frac{1}{|\mathfrak{X}|}\mathbf{E}\Upsilon_{\mathfrak{X}} \leq \mathbf{P}(\Upsilon_{\mathfrak{X}} \geq m). \tag{16}$$

2.4 Понятие условного средневероятного события

Естественным дальнейшим развитием средневероятного события является условное средневероятное событие. Итак, рассмотрим 2 множества событий $\mathfrak{N}, \mathfrak{X} \subset \mathcal{F}$, которые имеют эвентологические распределения $\{p(N//\mathfrak{N}) = \mathbf{P}(\mathrm{ter}(N//\mathfrak{N}))|N \subseteq \mathfrak{N}\}$ и $\{p(X//\mathfrak{X}) = \mathbf{P}(\mathrm{ter}(X//\mathfrak{X}))|X \subseteq \mathfrak{X}\}$ соответственно. Задача стоит следующая: проанализировать поведение средневероятного события $\hat{x}_{\mathfrak{N}}$ для множества однородных событий $\mathfrak{N}$ при условии того, что $\mathfrak{X} \longrightarrow \mathrm{ter}(X//\mathfrak{X})$, то есть, при условии того, что множество $\mathfrak{X}$ случилось как $\mathrm{ter}(X//\mathfrak{X})$ [4].

Так как мы работаем в рамках всеобщего вероятностного пространства, то очевидно, что множество событий $\mathfrak{N}\bigcup\mathfrak{X} \subset \mathcal{F}$, составленное из объединений рассматриваемых множеств, имеет свое собственное эвентологическое распределение $\{p(N \cup X//\mathfrak{N} \cup \mathfrak{X})|N \subseteq \mathfrak{N}, X \subseteq \mathfrak{X}\}$.

Определение 7. *Средневероятным событием для конечного множества однородных событий $\mathfrak{N} \subset \mathcal{F}$ при условии, что $\mathfrak{X} \longrightarrow \mathrm{ter}(X//\mathfrak{X})$ называется такое всеобщее событие $\hat{\xi}_{\mathfrak{N}|\mathfrak{X}\longrightarrow\mathrm{ter}(X//\mathfrak{X})} \in \mathcal{F}$, для которого выполнены включения:*

$$\begin{aligned}&\sum_{|N|>m}\mathrm{ter}(N \cup X//\mathfrak{N} \cup \mathfrak{X}) \subseteq \\ &\subseteq \hat{\xi}_{\mathfrak{N}|\mathfrak{X}\longrightarrow\mathrm{ter}(X//\mathfrak{X})} \subseteq \\ &\subseteq \sum_{|N|\geq m}\mathrm{ter}(N \cup X//\mathfrak{N} \cup \mathfrak{X}),\end{aligned} \tag{17}$$

которое наступает с вероятностью

$$\mathbf{P}(\hat{\xi}_{\mathfrak{N}|\mathfrak{X}\longrightarrow\mathrm{ter}(X//\mathfrak{X})}) = \frac{1}{|\mathfrak{N}|}\sum_{\xi\in\mathfrak{N}}\mathbf{P}(\xi|\mathrm{ter}(X//\mathfrak{X})), \tag{18}$$

где $\mathbf{P}(\xi|\mathrm{ter}(X//\mathfrak{X}))$ – это условная вероятность события в классическом ее понимании. Событие $\hat{\xi}_{\mathfrak{N}|\mathfrak{X}\longrightarrow\mathrm{ter}(X//\mathfrak{X})}$ наступает всякий раз, когда среди событий из множества $\{\xi \cap \mathrm{ter}(X//\mathfrak{X})|\xi \in \mathfrak{N}\}$ наступает не менее, чем m событий, где $m \in \{0,1,\ldots,|\mathfrak{N}|\}$ и удовлетворяет неравенствам

$$\begin{aligned}&\sum_{|N|>m}p(N \cup X//\mathfrak{N} \cup \mathfrak{X}) \leq \\ &\leq \mathbf{P}(\hat{\xi}_{\mathfrak{N}|\mathfrak{X}\longrightarrow\mathrm{ter}(X//\mathfrak{X})}) \leq \\ &\leq \sum_{|N|\geq m}p(N \cup X//\mathfrak{N} \cup \mathfrak{X}).\end{aligned} \tag{19}$$

Как ранее уже было подмечено, в некоторых случаях очень удобно средневероятное событие формулировать на языке террасных m-слоев и террасных m-пластов.

Итак, определим множественный условный m-слой $C^m_{\mathfrak{N}|\mathfrak{X}\longrightarrow \mathrm{ter}(X//\mathfrak{X})} \subseteq 2^{\mathfrak{N}\cup\mathfrak{X}}$ и множественный условный m-пласт $C^m_{\mathfrak{N}|\mathfrak{X}\longrightarrow \mathrm{ter}(X//\mathfrak{X})} \subseteq 2^{\mathfrak{N}\cup\mathfrak{X}}$(оба при условии того, что множество $\mathfrak{X}$ случилось как $\mathrm{ter}(X//\mathfrak{X})$) как совокупность подмножеств событий

$$C^m_{\mathfrak{N}|\mathfrak{X}\longrightarrow \mathrm{ter}(X//\mathfrak{X})} = \{N \cup X : N \subseteq \mathfrak{N}, |N| = m\} \quad (20)$$

$$C^{m+}_{\mathfrak{N}|\mathfrak{X}\longrightarrow \mathrm{ter}(X//\mathfrak{X})} = \{N \cup X : N \subseteq \mathfrak{N}, |X| \geq m\} \quad (21)$$

Далее, определим террасные операции условного «m-слой»:

$$\begin{aligned} &\mathrm{terop}(C^m_{\mathfrak{N}|\mathfrak{X}\longrightarrow \mathrm{ter}(X//\mathfrak{X})}//2^{\mathfrak{N}\cup\mathfrak{X}}) = \\ &= \sum_{Y\in C^m_{\mathfrak{N}|\mathfrak{X}\longrightarrow \mathrm{ter}(X//\mathfrak{X})}} \mathrm{ter}(Y//\mathfrak{N}\cup\mathfrak{X}) \subseteq \Omega \end{aligned} \quad (22)$$

и условного «m-пласт»:

$$\begin{aligned} &\mathrm{terop}(C^{m+}_{\mathfrak{N}|\mathfrak{X}\longrightarrow \mathrm{ter}(X//\mathfrak{X})}//2^{\mathfrak{N}\cup\mathfrak{X}}) = \\ &= \sum_{Y\in C^{m+}_{\mathfrak{N}|\mathfrak{X}\longrightarrow \mathrm{ter}(X//\mathfrak{X})}} \mathrm{ter}(Y//\mathfrak{N}\cup\mathfrak{X}) \subseteq \Omega \end{aligned} \quad (23)$$

Определение 8. *Средневероятным событием для конечного однородного множества событий $\mathfrak{N} \subset \mathcal{F}$ при условии, что $\mathfrak{X} \longrightarrow \mathrm{ter}(X//\mathfrak{X})$ называется любое такое всеобщее событие $\hat{\xi}_{\mathfrak{N}|\mathfrak{X}\longrightarrow \mathrm{ter}(X//\mathfrak{X})} \in \mathcal{F}$, которое содержится в террасном условном m-пласте, содержит террасный условный $m+1$-пласт:*

$$\begin{aligned} &\mathrm{terop}(C^{(m+1)+}_{\mathfrak{N}|\mathfrak{X}\longrightarrow \mathrm{ter}(X//\mathfrak{X})}//2^{\mathfrak{N}\cup\mathfrak{X}}) \subseteq \\ &\hat{\xi}_{\mathfrak{N}|\mathfrak{X}\longrightarrow \mathrm{ter}(X//\mathfrak{X})} \subseteq \\ &\subseteq \mathrm{terop}(C^{(m)+}_{\mathfrak{N}|\mathfrak{X}\longrightarrow \mathrm{ter}(X//\mathfrak{X})}//2^{\mathfrak{N}\cup\mathfrak{X}}), \end{aligned} \quad (24)$$

и наступает со средней вероятностью

$$\mathbf{P}(\hat{\xi}_{\mathfrak{N}|\mathfrak{X}\longrightarrow \mathrm{ter}(X//\mathfrak{X})}) = \frac{1}{|\mathfrak{N}|}\sum_{\xi\in\mathfrak{N}} \mathbf{P}(\xi|\mathrm{ter}(X//\mathfrak{X})), \quad (25)$$

где $\mathbf{P}(\xi|\mathrm{ter}(X//\mathfrak{X}))$ – это условная вероятность события в классическом ее понимании. Событие $\hat{\xi}_{\mathfrak{N}|\mathfrak{X}\longrightarrow \mathrm{ter}(X//\mathfrak{X})}$ наступает всякий раз, когда среди событий из множества $\{\xi \cap \mathrm{ter}(X//\mathfrak{X})|\xi \in \mathfrak{N}\}$ наступает не менее, чем m событий, где $m \in \{0, 1, \ldots, |\mathfrak{N}|\}$ и удовлетворяет неравенствам

$$\begin{aligned} &\mathbf{P}(\mathrm{terop}(C^{(m+1)+}_{\mathfrak{N}|\mathfrak{X}\longrightarrow \mathrm{ter}(X//\mathfrak{X})}//2^{\mathfrak{N}\cup\mathfrak{X}})) \leq \\ &\leq \mathbf{P}(\{\xi \cap \mathrm{ter}(X//\mathfrak{X})|\xi \in \mathfrak{N}\}) \leq \\ &\leq \mathbf{P}(\mathrm{terop}(C^{(m+1)+}_{\mathfrak{N}|\mathfrak{X}\longrightarrow \mathrm{ter}(X//\mathfrak{X})}//2^{\mathfrak{N}\cup\mathfrak{X}})) \end{aligned} \quad (26)$$

3 Средне-событийная регрессия

Пусть множество $\mathfrak{N} \subset \mathcal{F}$ – это множество событий-регрессандов, а множество $\mathfrak{X} \subset \mathcal{F}$ – это множество событий-регрессоров.

Определение 9. *Функция теоретической среднесобытийной регрессии множества $\mathfrak{N} \subset \mathcal{F}$ принимает свои значения на $\mathrm{ter}(X//\mathfrak{X}), X \subseteq \mathfrak{X}$ равные условному средневероятному событию множества $\mathfrak{N}$ при условии наступления террасного регрессорного события:*

$$\mathcal{E}(\mathfrak{N}|\mathrm{ter}(X//\mathfrak{X})) = \varphi(\mathrm{ter}(X//\mathfrak{X})) \quad (27)$$

Иными словами, функция теоретической регрессии это отображение:

$$\varphi : 2^{\mathfrak{X}} \longrightarrow \Omega. \quad (28)$$

3.1 Пример

Рассмотрим элементарный пример. Будем наблюдать за покупками определенных продуктов в магазине покупателями в зависимости от их ежемесячного дохода. Данный пример полностью взят из головы автора, но уважаемый читатель не может не согласиться, что приведенные данные полностью соответствуют действительности.

Из общих соображений ясно, что благосостояние человека непременно влияет на его решение о покупке того или иного продукта в магазине. Поэтому будем считать, что множество событий-регрессоров $\mathfrak{Y} = \{\xi, \eta\}$, где

$\xi = \{$покупатель с ежемесячным доходом до 50 т.р.$\}$,
$\eta = \{$покупатель с ежемесячным доходом от 50 т.р.$\}$,

а множество событий-регрессантов $\mathfrak{X} = \{x, y, z\}$, где

$x = \{$покупателем приобретен хлеб$\}$,
$y = \{$покупателем приобретено молоко$\}$,
$z = \{$покупателем приобретена икра красная$\}$.

Предположим,то мы наблюдаем за первыми 110 покупателями и оказывается, что из них 30 имеют доход ежемесячный доход более 50 т.р., а 80 имеют ежемесячный доход ниже данного уровня. Основываясь на этих данных делаем вывод о том, что $\mathbf{P}(\xi) = \frac{8}{11}$, а $\mathbf{P}(\eta) = \frac{3}{11}$. При этом, $\xi \cap \eta = \emptyset$, поэтому $\mathrm{ter}(\{\xi\}//\mathfrak{Y}) = \xi$, а $\mathrm{ter}(\{\eta\}//\mathfrak{Y}) = \eta$, все остальные терраски пустые. Результатом наших наблюдений о характере покупок оказались данные, изображенные на рисунке 3, откуда можно сделать вывод об значениях условных вероятностей:

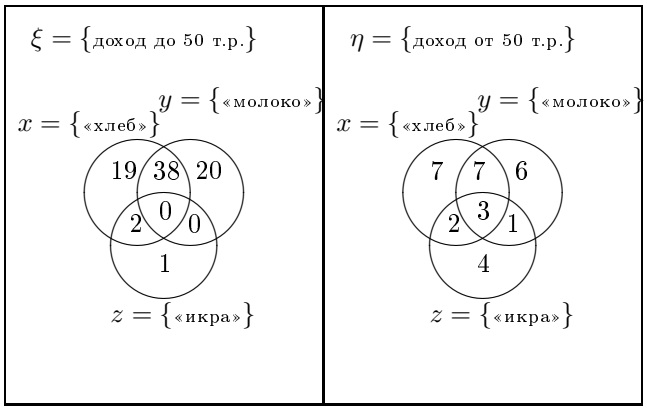


Рисунок 3: Данные о характере покупок людей с разным ежемесячным доходом.

$$p(\{x\}|\xi) = \frac{19}{80} \qquad p(\{x\}|\eta) = \frac{7}{30}$$

$$p(\{y\}|\xi) = \frac{20}{80} \qquad p(\{y\}|\eta) = \frac{6}{30}$$

$$p(\{z\}|\xi) = \frac{1}{80} \qquad p(\{z\}|\eta) = \frac{4}{30}$$

$$p(\{x,y\}|\xi) = \frac{38}{80} \qquad p(\{x,y\}|\eta) = \frac{7}{30}$$

$$p(\{x,z\}|\xi) = \frac{2}{80} \qquad p(\{x,z\}|\eta) = \frac{2}{30}$$

$$p(\{y,z\}|\xi) = 0 \qquad p(\{y,z\}|\eta) = \tfrac{1}{30}$$

$$p(\{x,y,z\}|\xi) = 0 \qquad p(\{x,y,z\}|\eta) = \frac{3}{30}$$

Не трудно вычислить вероятность условного средневероятного события:

$$\mathbf{P}(\mathcal{E}(\mathfrak{X}|\xi)) = \tfrac{120}{240} = 0,5;$$

$$\mathbf{P}(\mathcal{E}(\mathfrak{X}|\eta)) = \tfrac{46}{90} \approx 0,51$$

и определить, что

$$\mathcal{E}(\mathfrak{X}|\xi) = \mathrm{terop}(C^{2+}_{\mathfrak{X}|\mathfrak{Y}\longrightarrow\xi}//2^{\mathfrak{Y}\cup\mathfrak{X}}),$$

а также, что $\mathcal{E}(\mathfrak{X}|\eta)$ будет включать условный 2-пласт и включаться в условный 1-пласт:

$$\mathrm{terop}(C^{3+}_{\mathfrak{X}|\mathfrak{Y}\longrightarrow\eta}//2^{\mathfrak{Y}\cup\mathfrak{X}}) \subset$$

$$\subset \mathcal{E}(\mathfrak{X}|\eta) \subset$$

$$\subset \mathrm{terop}(C^{2+}_{\mathfrak{X}|\mathfrak{Y}\longrightarrow\eta}//2^{\mathfrak{Y}\cup\mathfrak{X}}).$$

Список литературы

[1] И.И Канаев, Л.Я. Бляхер, А.Т. Гриюръян, Я.Г. Дорфман, Б.М. Кедров. *Френсис Гальтон*, volume 139. НАУКА, 1972.

[2] Д.В.Семенова Н.А. Лукьянова, Я.В. Нартов. Визуализация средневероятного события через матричное представление террасок-событий. *Труды XVI Международной ЭМ конференции по эвентологической математике и смежным вопросам*, 205:145–158, 2012.

[3] О.Ю Воробьев. *Эвентология*. Красноярск: Сибирский федеральный университет, 2007.

[4] О.Ю. Воробьев. Об эвентологическом отношении «случается как» и «коллапсирует в». *Труды VII конференции по финансово-актуарной математике и смежным вопрсам*, 425:47–50, 2008.

[5] О.Ю. Воробьев. Средневероятное событие для множества событий. *Труды XI международной конференции по финансово-актуарной математике и эвентологии безопасности*, 425:139–147, 2012.

[6] 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.

[7] Fr. Galton. *Natural inheritance*, volume 259. НАУКА, 1889.

[8] K. Pearson. Mathematical contributions to the theory of evolution. iii. regression, heredity and panmixia. *Philosophical Transactions of the Royal Society of London. Series A*, pages 253–318, 1896.

Средне-феноменный портфельный анализ

Александр Сергеевич Нифонтов
Сибирский федеральный университет
Институт математики и фундаментальной информатики
Красноярск
nialeksander@gmail.com

Аннотация. *Рассматривается эвентологическая портфельная задача Марковица. Формулируется средне-феноменная постановка. Рассчитаны примеры решения портфельных задач в средне-феноменной постановке для двух, трех и четырех событийных портфелей.*

Ключевые слова. *Эвентология, теория вероятностей, событие, вероятность, портфельный анализ, портфельный анализ событий, средне-феноменная, Марковиц*

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 .

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:

$$\Upsilon_x = \mathbf{1}_x.$$

Тогда средняя доходность каждого портфельного события $x \in \mathfrak{X}$ является вероятностью этого события:

$$\mathbf{E}\Upsilon_x = \mathbf{E}\mathbf{1}_x = p_x,$$

дисперсия доходности портфельного события равна

$$\mathbf{D}\Upsilon_x = \mathbf{D}\mathbf{1}_x = p_x(1 - p_x) = \sigma_x^2,$$

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 $\mathbf{a}$:

$$\Upsilon_{\mathfrak{X}} = \sum_{x \in \mathfrak{X}} a_x \mathbf{1}_x.$$

Тогда средняя доходность портфеля событий определяется вероятностями его событий — это выпуклая комбинация вероятностей событий с коэффициентами, равными долям этих событий в портфеле:

$$\mathbf{E}\left(\Upsilon_{\mathfrak{X}}\right) = \sum_{x \in \mathfrak{X}} a_x p_x,$$

а дисперсия — это квадратичная форма от $\mathbf{a}$:

$$\mathbf{D}\left(\Upsilon_{\mathfrak{X}}\right) = \sum_{x \in \mathfrak{X}} \sum_{y \in \mathfrak{X}} a_x a_y \mathsf{Kov}_{xy}$$

с коэффициентами, равными $\mathsf{Kov}_{xy} = p_{xy} - p_x p_y$ — 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):*

$$\mathbf{E}\left(\Upsilon_{\mathfrak{X}}\right) = \langle \mathbf{a} \rangle,$$

$$\mathbf{D}\left(\Upsilon_{\mathfrak{X}}\right) \to \min_{\mathbf{a}}.$$

1.1 Примеры

Для построения графика задаются вероятности событий и все событийные ковариации. Генерируются случайные множества долей для активов и рассчитываются математическое ожидание и дисперсия по исходным данным. Полученные величины располагаются на плоскости: *по оси абсцисс - корень из дисперсии, а по оси ординат - математическое ожидание.* В результате получается график зависимости риска и доходности. Изображенная фигура, охватывающая все построенные точки, называется "пуля Марковица".

© 2014 А.С.Нифонтов

Олег Воробьев (ред.), Труды XIII ФАМЭМС'2014, Красноярск: СФУ

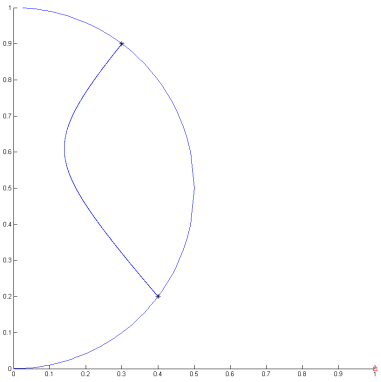


Рисунок 1: Пример пули Марковица для двух событий

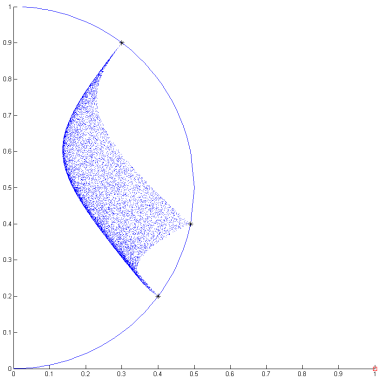


Рисунок 2: Пример пули Марковица для трех событий

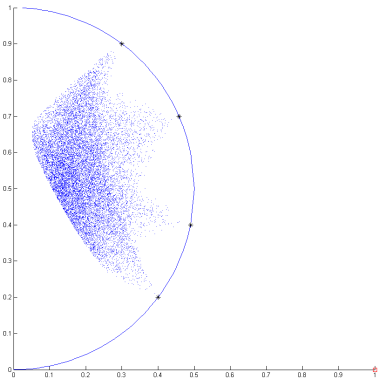


Рисунок 3: Пример пули Марковица для четырех событий

Нас интересует только эффективная граница пули Марковица - "Северо-западная"часть пули. Особенность эффективной границы пули в том, что на этой границе лежат точки с максимальной доходностью и в тоже время риск при данной доходности минимален. "Ноги"пули - точки лежащие на полуокружности радиуса 0.5 и с центром в $[0, 0.5]$. Если портфель состоит только из одного вида активов, то значение доходности и риска, как раз располагается в "ногах"пули.

2 Обобщенная постановка: доходность события как линейная функция его индикатора

В рассмотренной стандартной постановке присутствует недостаток: В реальных условиях сложно получить распределения в которых доходности и риск располагаются в одинаковых интервалах для всех портфельных активов. Для решения задач с распределениями, в которых доходность и/или риск портфельных активов не принадлежат одним интервалам, применяется масштабирование и приведение доходности и риска к одинаковым интервалам для всех активов. В данном подходе теряется связь результатов с реальностью. Для решения этой проблемы была сформулирована обобщенная постановка задачи Марковица.

Рассмотрим линейное обобщение прямой эвентологической портфельной задачи, 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:

$$\Upsilon_x = b_x \mathbf{1}_x + d_x.$$

Тогда средняя доходность каждого портфельного события $x \in \mathfrak{X}$ является линейной функцией вероятности события:

$$\mathbf{E}\Upsilon_x = b_x \mathbf{E}\mathbf{1}_x + d_x = b_x p_x + d_x,$$

дисперсия доходности портфельного события равна

$$\mathbf{D}\Upsilon_x = b_x^2 \mathbf{D}\mathbf{1}_x = b_x^2 p_x (1 - p_x) = b_x^2 \sigma_x^2,$$

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 $\mathbf{a}$:

$$\Upsilon_{\mathfrak{X}} = \sum_{x \in \mathfrak{X}} a_x (b_x \mathbf{1}_x + d_x).$$

Тогда средняя доходность портфеля событий определяется вероятностями его событий по формуле:

$$\mathbf{E}\left(\Upsilon_{\mathfrak{X}}\right) = \sum_{x \in \mathfrak{X}} a_x b_x p_x + \sum_{x \in \mathfrak{X}} a_x d_x,$$

а дисперсия — формулой:

$$\mathbf{D}\left(r_{\mathfrak{X}}\right) = \sum_{x\in\mathfrak{X}}\sum_{y\in\mathfrak{X}} a_x b_x a_y b_y \mathsf{Kov}_{xy},$$

где $\mathsf{Kov}_{xy} = p_{xy} - p_x p_y$ — 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):

$$\mathbf{E}\left(\Upsilon_{\mathfrak{X}}\right) = \langle \mathbf{a} \rangle,$$

$$\mathbf{D}\left(\Upsilon_{\mathfrak{X}}\right) \to \min_{\mathbf{a}}.$$

2.1 Примеры

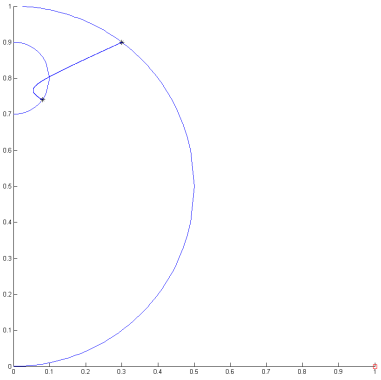


Рисунок 4: Пример пули Марковица для двух событий

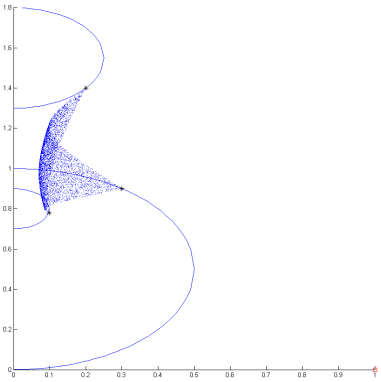


Рисунок 5: Пример пули Марковица для трех событий

В данной постановке каждая "нога"пули расположена на своей полуокружности радиуса $b_x/2$ и центр смещен в точку $[0, b_x/2 + d_x]$.
У пули также, как и в обычной постановке, определяется эффективная граница. Построенная пуля, получается растягиванием пули из стандартной постановки на полуокружности, характеризующие активы.

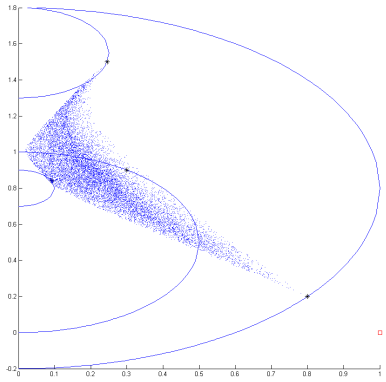


Рисунок 6: Пример пули Марковица для четырех событий

3 Средне-феноменная постановка

Каждый феномен это только один из 2^N всевозможных взглядов на рассматриваемый портфель. Используя средне-феноменную постановку, мы сможем оценивать портфель со свех точек зрения одновременно. В результате получим портфель, который одинаково эффективен во всех феноменах.

3.1 Средне-феноменая обобщенная задача Марковица

Рассмотрим средне-феноменную версию обобщенной постановки задачи Марковица. В задаче оценивается средняя доходность и средний риск по всевозможным сет-феноменам событий.

Пусть $\{\mathfrak{X}(\mathbf{p}), \mathbf{a}\}$ — портфель событий, где $\mathfrak{X}$ — множество портфельных событий с известным вероятностным распределением для X-феномена

$$\mathbf{p} = \{p(X) : X \subseteq \mathfrak{X}\},$$

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.

X-феномен стандартного портфельного множества событий определяется по формуле

$$\mathfrak{X}^{(c|X)} = X + (\mathfrak{X} - X)^{(c)}.$$

Мощностной индикатор портфельного множества событий - выпуклая комбинация линейных индикаторов событий из $\mathfrak{X}$ вычисляется по формуле:

$$\check{\Upsilon}_{\mathfrak{X}} = \sum_{x\in\mathfrak{X}} a_x (b_x 1_x + d_x).$$

Из свойств мощностного индикатора и сет-феномена следует, что

$$\check{\Upsilon}_{\mathfrak{X}^{(c|X)}} = \check{\Upsilon}_X + \check{\Upsilon}_{(\mathfrak{X}-X)^{(c)}} =$$

$$= \sum_{x \in X} a_x (b_x \mathbf{1}_x + d_x) + \sum_{x \in \mathfrak{X} - X} a_x (1 - (b_x \mathbf{1}_x + d_x)).$$

Средняя доходность портфеля событий определяется по формуле

$$\langle \mathbf{E} \check{\Upsilon}_{\mathfrak{X}^{(c|X)}} \rangle_{X \subseteq \mathfrak{X}} = |\mathfrak{X}|_{a(b+d)} - 2 \sum_{x \in \mathfrak{X}} a_x^2 b_x^2 \sigma_x^2,$$

где

$$|\mathfrak{X}|_{a(b+d)} = \sum_{x \in \mathfrak{X}} a_x (b_x + d_x)$$

$$\langle \mathbf{D} \check{\Upsilon}_{\mathfrak{X}^{(c|X)}} \rangle_{X \subseteq \mathfrak{X}} =$$

$$= \sum_{x \in \mathfrak{X}} a_x^2 b_x^2 \sigma_x^2 + 2 \sum_{\{x,y\} \subseteq \mathfrak{X}} a_x a_y b_x b_y \langle \mathrm{Kov}_{\{x,y\}^{(c|X)}} \rangle_{X \subseteq \{x,y\}}$$

$$\sigma_x^2 = p_x (1 - p_x)$$

и

$$\mathsf{Kov}_{xy} = p_{xy} - p_x p_y.$$

Постановка прямой средне-феноменной 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):

$$\langle \mathbf{E} \check{\Upsilon}_{\mathfrak{X}^{(c|X)}} \rangle_{X \subseteq \mathfrak{X}} = \langle \mathbf{a} \rangle,$$

$$\langle \mathbf{D} \check{\Upsilon}_{\mathfrak{X}^{(c|X)}} \rangle_{X \subseteq \mathfrak{X}} =\!\!\rightarrow \min_{\mathbf{a}}.$$

3.2 Примеры

Также как и в стандартной постановке, в средне-феноменной определяется эффективная граница. Назовем эту границу средне-феноменной эффективной границей пули Марковица. Распределения множества долей портфеля, которые формируют средне-феноменную эффективную границу, создают одинаково результативные портфели для всех феноменов. В средне-феноменной постановке "ноги"у пули лежат на перевернутых параболах центры, которых находится в точках $[0, b_x + d_x]$, а ветви растянуты в два раза вдоль оси абсцисс.

4 Сравнение результатов стандартной и средне-феноменной постановок

Для сравнения результатов решаем три задачи с разными распределениями. Для упрощения анализа сравнивать будем постановки с весами у всех событий равными $b_x = 1$ и $d_x = 0$.
Феномен, вероятности, которого задаются при формулировке задачи, будем называть *основным феноменом.*

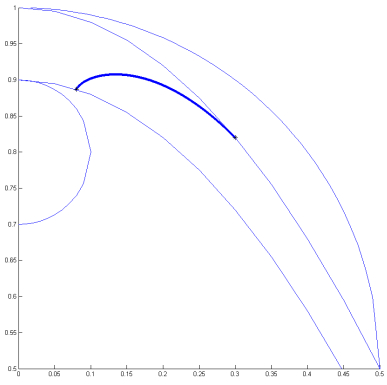


Рисунок 7: Пример средне-феноменной пули Марковица для двух событий

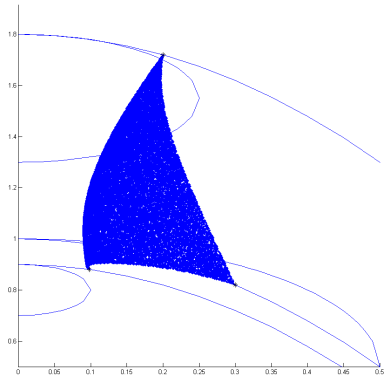


Рисунок 8: Пример средне-феноменной пули Марковица для трех событий

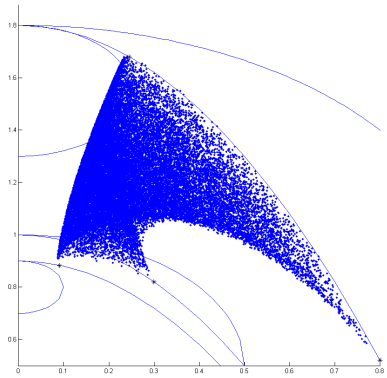


Рисунок 9: Пример средне-феноменной пули Марковица для четырех событий

4.1 Два события

Решим средне-феноменную задачу Марковица для двух событий с вероятностями $[0.9, 0.2]$. Построим график пули и сразу определим соотношения активов в двух точках: *минимальный риск* и *максимальная доходность*. Эти точки на графике отмечены зеленым и красным соответственно.

Увеличив область, отмеченную красным квадратом, увидим средне-феноменную эффективную границу Марковица.

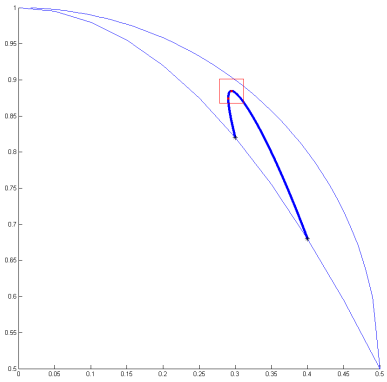


Рисунок 10: Средне-феноменная пуля Марковица для двух событий

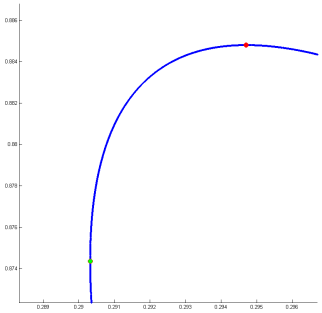


Рисунок 11: Средне-феноменная эффективная граница Марковица для двух событий.

С полученными значениями активов решаем стандартную постановку для основного феномена. Отрезок между точками, сформированными при решении средне-феноменной задачи, расположен на части эффективной границы. Это означает, что множество решений средне-феноменной постановки является подмножеством множества решений стандартной постановки для двух событий. Точки *красного* цвета обозначают значения построенные для портфеля со средне-феноменной максимальной доходностью, а *зеленые* - со средне-феноменным минимальным риском.

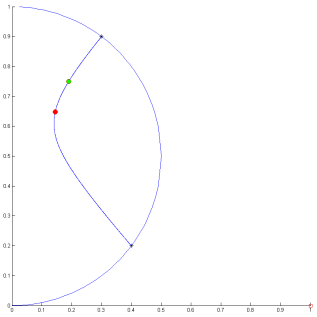


Рисунок 12: Стандартная пуля Марковица для двух событий.

И строим для всех феноменов пули, возможные при данном распределении, на одном графике и располагаем значения полученные при найденных активах.

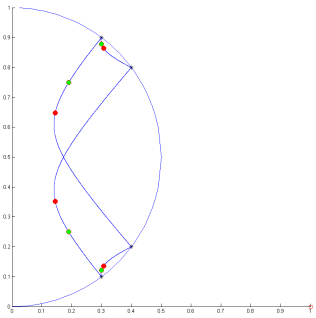


Рисунок 13: Стандартные пули Марковица всевозможных феноменов для двух событий.

4.2 Три события

Решим средне-феноменную задачу Марковица для трех событий с вероятностями $[0.9, 0.4, 0.2]$. Построим график пули и сразу определим соотношения активов в двух точках: *минимальный риск* и *максимальная доходность*. Эти точки на графике отмечены зеленым и красным соответственно.

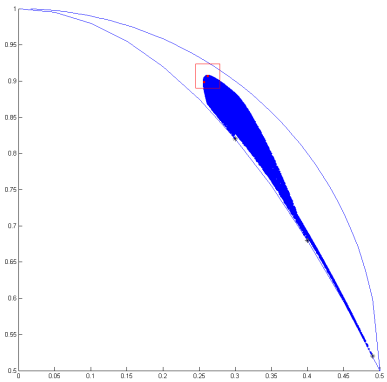


Рисунок 14: Средне-феноменная пуля Марковица для трех событий

Увеличив область, отмеченную красным квадратом, увидим средне-феноменную эффективную границу Марковица.

С полученными значениями активов решаем стандартную постановку для основного феномена.

И строим для всех феноменов пули, возможные при данном распределении, на одном графике и располагаем значения полученные при найденных активах. Значения полученные для средне-феноменного эффективного портфеля располагаются вблизи эффективной границы пули Марковица.

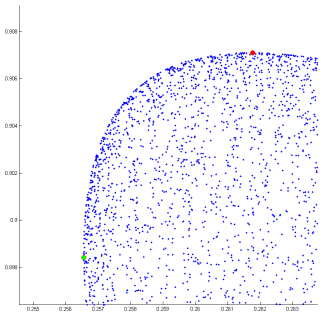


Рисунок 15: Средне-феноменная эффективная граница Марковица для трех событий.

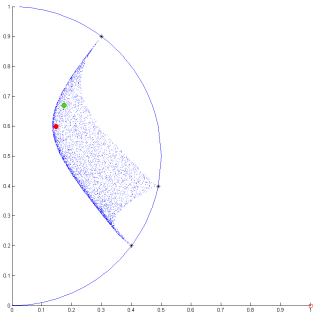


Рисунок 16: Стандартная пуля Марковица для трех событий.

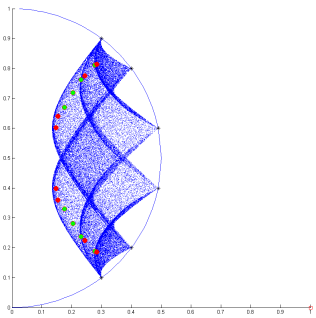


Рисунок 17: Стандартные пули Марковица всевозможных феноменов для трех событий.

4.3 Четыре события

Решим средне-феноменную задачу Марковица для четырех событий с вероятностями $[0.9, 0.7, 0.4, 0.2]$. Построим график пули и сразу определим соотношения активов в двух точках: *минимальный риск* и *максимальная доходность*. Эти точки на графике отмечены зеленым и красным соответственно.

Увеличив область, отмеченную красным квадратом, увидим средне-феноменную эффективную границу Марковица.

С полученными значениями активов решаем стандартную постановку для основного феномена.

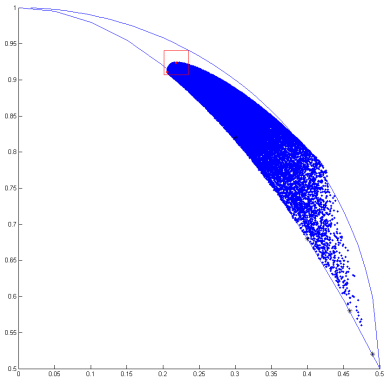


Рисунок 18: Средне-феноменная пуля Марковица для четырех событий

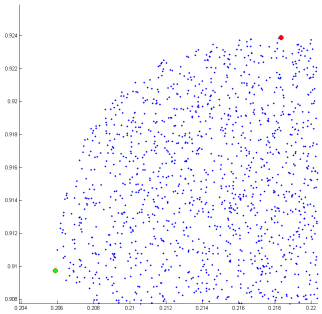


Рисунок 19: Средне-феноменная эффективная граница Марковица для четырех событий.

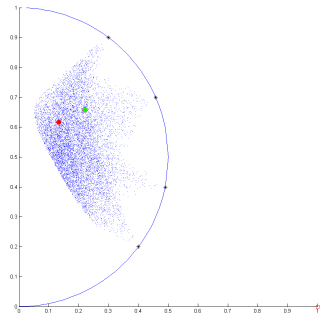


Рисунок 20: Стандартная пуля Марковица для четырех событий.

И строим для всех феноменов пули, возможные при данном распределении, на одном графике и располагаем значения полученные при найденных активах.

4.4 Вывод

Средне-феноменная постановка может применяться наравне со стандартной постановкой, но при увеличении количества активов в портфеле, результат от-

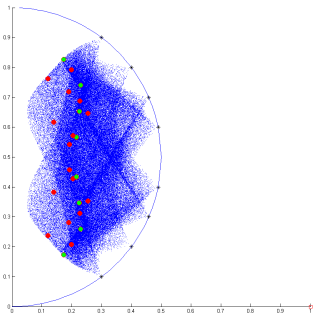


Рисунок 21: Стандартные пули Марковица всевозможных феноменов для четырех событий.

даляется от края эффективной границы стандартной пули к центру пули.
В свою очередь средне-феноменная постановка оценивает портфель со всех сторон, и в результате применения средне-феноменной теории получается портфель, который наиболее оптимален по всем феноменам.

5 Смысл средне-феноменной постановки

Игра на фондовых рынках - это сражение одного против всех, т.е. каждый инвестор стремится переиграть остальных инвесторов, как совокупность, а не каждого по отдельности. Каждый феномен отображает взгляд на активы одного из инвесторов. Путем перебора всех феноменов мы оценим в каких долях вкладывать активы в феномены, затем применяя полученное распределение активов, выберем феномен с максимальной доходностью. В результате получим портфель, который в среднем эффективнее других.

6 Дальнейшее развитие

Логичным развитием средне-феноменной портфельной теории будет являться применение методики на различных фондовых рынках. Так же, мне кажется, получится описать различные стратегии инвестирования путем применения средне-феноменной постановки.

Стресс-тестирование риск-моделей: некоторые подводные камни

Аркадий Арсеньевич Новоселов
Сибирский федеральный университет
Институт математики и фундаментальной информатики
Красноярск
arcady@novosyolov.com

Аннотация. *В работе описан метод стресс-тестирования риск-моделей и указана одна проблема, с которой можно столкнуться при построении стрессовых сценариев описанным методом. Предлагается способ решения указанной проблемы.*
Ключевые слова. *Риск, модель, условное распределение, стресс-тест, сценарий.*

1 Введение

Управление финансовым риском имеет давнюю историю, и на интуитивном уровне применялось, по-видимому, еще в древнем мире, во времена зарождения торговых отношений. В двадцатом веке эта деятельность стала приобретать формализованный характер, активно привлекая математические методы. Так возникла теория портфельного управления Марковица, сначала в виде статьи [6], а потом - и в виде систематического изложения в книге [7]. Далее это теория была развита в части включения в портфель безрисковых инструментов Тобиным [11], и доведена Шарпом до теории взаимодействия с рынком в целом в [10].

С повышением изменчивости поведения инструментов финансового рынка росла и потребность в специализированных инструментах управления риском. Так, в 1970-х годах широкое использование так называемые производные инструменты: опционы, фьючерсы, форварды, а также многие другие. Математическая теория таких инструментов была заложена Блэком, Шоулзом м Мертоном в работах [2, 8], и с тех пор стремительно развивалась, см. например последнее издание книги Халла [3].

В последние десятилетия финансовые кризисы приобрели более выраженную форму. Так в 1997-1998 годах произошли азиатский кризис и российский дефолт, в 2000-2001 годах наблюдалось лопание интернет-пузыря, в 2008-2009 году имел место мировой финансовый кризис. Обеспечение устойчивости финансовой системы фирмы, компании или хедж-фонда в таких условиях требует проведения специальных мероприятий. Одним из таких комплексов мероприятий стала методология стресс-тестирования, основные принципы которой были сформулированы в 1998 году в работе [5], и получили развитие далее в статьях [9, 1, 4] и многих других.

© 2014 А.А.Новоселов

Олег Воробьев (ред.), Труды XIII ФАМЭМС'2014, Красноярск: СФУ

2 Распределение факторов и разбиение

Рассмотрим методологию стресс-тестирования, предложенную в [5]. Обозначим $X = (X_1, \ldots, X_n)'$ вектор факторов, имеющих нормальное распределение с нулевым средним и ковариационной матрицей $C = \mathbf{E}(XX')$, а $w = (w_1, \ldots, w_n)'$ — вектор весов. Из этих объектов образуется портфель по правилу $U = w'X$. Вектор весов предполагается ненулевым, других ограничений не накладывается. Другими словами, в портфель могут входить как длинные (положительные веса), так и короткие (отрицательные веса) позиции. Доходность портфеля, очевидно, равна нулю, дисперсия и стандартное отклонение задаются выражениями $w'Cw$, $\sigma(X) = \sqrt{w'Cw}$.

Множество индексов $\{1, \ldots, n\}$ разбивается на два подмножества I, K таким образом, чтобы в множество индексов I попадали номера факторов, подвергаемых стрессу, а в множество K — номера факторов, которые изменяются в соответствии со связями, задаваемыми исходным распределением случайного вектора X. Для удобства записи блоков векторов и матриц будем считать, что при некотором m эти подмножества имеют вид[1] $I = \{1, \ldots, m\}$ и $K = \{m+1, \ldots, n\}$. Тогда вектор X разбивается на $X = (X_I', X_K')'$ с $X_I = (X_1, \ldots, X_m)'$ и $X_K = (X_{m+1}, \ldots, X_n)'$, а ковариационная матрица представляется в блочном виде

$$C = \begin{pmatrix} C_{II} & C_{IK} \\ C_{KI} & C_{KK} \end{pmatrix},$$

с очевидным составом блоков, причем $C_{IK} = C_{KI}'$.

3 Условное распределение

Зафиксируем значения компонент подвектора $X_I = a = (a_1, \ldots, a_m)$. Известно [5], что условное распределение оставшихся факторов X_K при этом условии

[1] Для реализации расчетного алгоритма такое упорядочение несущественно, факторы можно оставлять на своих местах.

является нормальным, со средним

$$\mu(a) = \mathbf{E}(X_K | X_I = a) = C_{KI} C_{II}^{-1} a \qquad (1)$$

и ковариационной матрицей

$$C(a) = C_{KK} - C_{KI} C_{II}^{-1} C_{IK}. \qquad (2)$$

Отметим, что корреляционная структура условного распределения не зависит от размера стресса a, а зависит лишь от способа разбиения факторов на фиксированные (возмущаемые) I и "свободные" K.

Результат стресса может быть записан для ожидаемых значений факторов в виде $\begin{pmatrix} a \\ \mu(a) \end{pmatrix}$, а также в виде доверительных интервалов уровня α, ширина которых равна нуля для факторов с номерами из I; для факторов с номерами из K стандартные отклонения τ_k, $k = 1, \ldots, n-m$ равны квадратным корням из диагональных элементов матрицы $C(a)$, а доверительные интервалы уровня доверия α имеют границы $\mu_k(a) \pm q_{(1+\alpha)/2}\tau_k$, $k = 1, \ldots, n-m$, где квантиль q заданного уровня определен в (**??**).

4 Интуитивное представление о реакции на стресс

Интуитивно ясно, что при положительной корреляции двух факторов положительный стресс одного из них вызовет положительную среднюю реакцию другого, а при отрицательной корреляции знак средней реакции будет противоположен знаку стресса. Эта связь проиллюстрирована на рисунках 1 и 2.

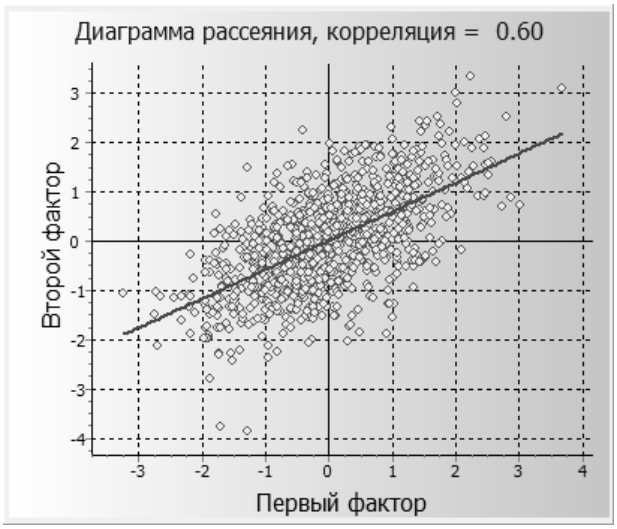


Рисунок 1: Диаграмма рассеяния двумерного нормального распределения при коэффициенте корреляции 0.6 и равных стандартных отклонениях; линия регрессии второго фактора на первый $x_2 = 0.6 \cdot x_1$.

На рисунке 1 показана диаграмма рассеяния двумерного нормального распределения с коэффициентом корреляции компонент (факторов) 0.6, а также линия регрессии второго фактора на первый, которая имеет в данном случае уравнение $x_2 = 0.6 \cdot x_1$. Шоку первого фактора величины a соответствует в среднем шок второго фактора величины $0.6 \cdot a$.

На рисунке 2 представлена аналогичная картина для случая отрицательной корреляции между факторами. Видно, что в данном случае шок первого фактора величины a вызывает среднюю реакцию второго фактора, по величине равную $-0.8 \cdot a$.

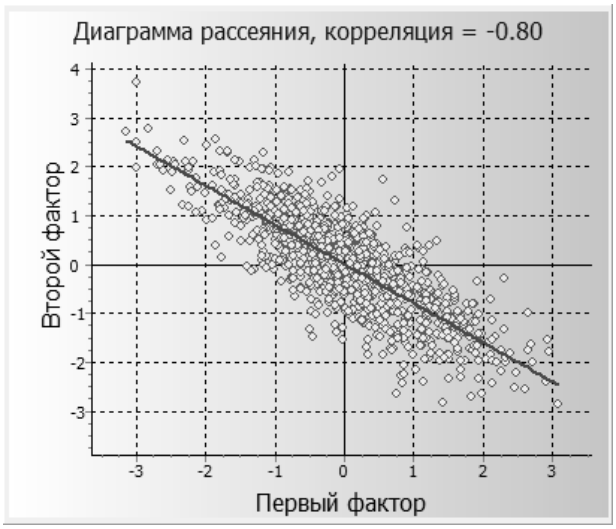


Рисунок 2: Диаграмма рассеяния двумерного нормального распределения при коэффициенте корреляции -0.8 и равных стандартных отклонениях; линия регрессии второго фактора на первый $x_2 = -0.8 \cdot x_1$.

5 Примеры контринтуитивного поведения

Рассмотрим трехфакторную модель ($n = 3$) с ковариационной матрицей

$$C = \begin{pmatrix} 170 & -50 & -6 \\ -50 & 25 & 5 \\ -6 & 5 & 1.5 \end{pmatrix}$$

и соответствующей корреляционной матрицей

$$R = \begin{pmatrix} 1 & -0.77 & -0.38 \\ -0.77 & 1 & 0.82 \\ -0.38 & 0.82 & 1 \end{pmatrix}$$

и матрицей стандартных отклонений

$$\Lambda = \begin{pmatrix} 13.0 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1.22 \end{pmatrix},$$

так что, в частности, $C = \Lambda K \Lambda$. Для удобства описания снабдим факторы фиктивными наименованиями, например, "энергетика", "медиа" и "золото".

Подвергнем стрессу фактор "золото", задав ему значение -2. Среднее значение реакции первых двух факторов, в соответствии с (1), приведено в табл. 1. Как и ожидалось, отрицательно коррелированный с "золотом" фактор "энергетика" реагирует в среднем в противоположную (положительную) сторону, а положительно коррелированный фактор "медиа" - в ту же (отрицательную) сторону.

Таблица 1: Пример стресс-теста с возмущением фактора "золото"

Фактор	Значение
Энергетика	8
Медиа	-6.67
Золото	-2

Теперь подвергнем стрессу -2 фактор "медиа". Результат показан в табл. 2. Здесь знаки реакции также соответствуют ожиданиям, основанным на знаке корреляции.

Таблица 2: Пример стресс-теста с возмущением фактора "медиа"

Фактор	Значение
Энергетика	4
Медиа	-2
Золото	-0.4

Наконец, подвергнем одновременно стрессу -2 два фактора, "медиа" и "золото". Среднее значение первого фактора, приведенное в табл., оказывается также отрицательным, несмотря на отрицательное значение корреляции этого фактора с каждым из стрессовых факторов.

Таблица 3: Пример стресс-теста с возмущением факторов "медиа"и "золото"

Фактор	Значение
Энергетика	-8.8
Медиа	-2
Золото	-2

Это неожиданное явление объясняется следующим образом. Средняя реакция на шок соответствует (в условиях симметричных распределений) наиболее вероятной ситуации (моде распределения). Значительные отклонения от такого наиболее вероятного поведения возможны, хотя и довольно маловероятны. Рассмотренный шок пары факторов, как видно из рисунка 3, лежит далеко от области наиболее вероятных значений этой пары факторов; поведение факторов в этой области плохо описывается средними характеристиками, что мы и увидели из примера. Наиболее вероятны значения этой пары факторов, в которых величина фактора "медиа" вчетверо превосходит величину фактора "золото" (см. матрицу стандартных отклонений факторов Λ). Сложные многофакторные стрессовые сценарии имеет смысл подбирать с учетом рассмотренного обстоятельства, или пользоваться специализированными методами, разрабатываемыми в рамках обратного стресс-тестирования [4].

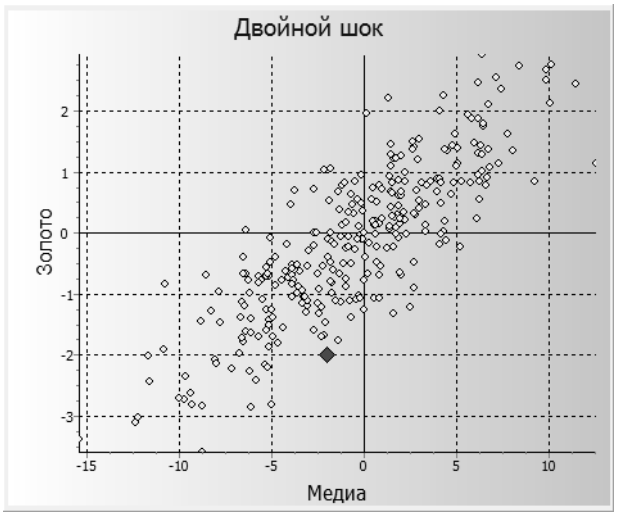


Рисунок 3: Диаграмма рассеяния распределения пары факторов "медиа-золото"; двойной шок (ромб).

Список литературы

[1] А.А. Новоселов. Моделирование роста корреляций для стресс-тестирования. *Риск-менеджмент в кредитной организации*, 3 (1):92–103, 2013.

[2] F. Black and M. Scholes. The pricing of options and corporate liabilities. *The Journal of Political Economy*, 81:637–654, 1973.

[3] J. Hull. *Options, Futures and Other Derivatives, 8th ed.* Pearson, 2011.

[4] Y. Kopeliovich, A. Novosyolov, D. Satchkov, and B. Schachter. Robust risk estimation and hedging: A reverse stress testing approach (submitted). *Journal of Derivatives*, 22:1–35, 2014.

[5] P. Kupiec. Stress testing in a value at risk framework. *The Journal of Derivatives*, 6:7–24, 1998.

[6] H. Markowitz. Portfolio selection. *The Journal of Finance*, 7:77–91, 1952.

[7] H. Markowitz. *Portfolio Selection: Efficient Diversification of Investments.* John Wiley & Sons, New York, 1959.

[8] R. Merton. On the pricing of corporate debt: The risk structure of interest rates. *The Journal of Finance*, 29:449–470, 1974.

[9] A. Novosyolov and D. Satchkov. Portfolio crash testing: making sense of extreme event exposures. *Journal of Risk Model Validation*, 4:53–67, 2010.

[10] W. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19:425–442, 1964.

[11] J. Tobin. Liquidity preference as behavior towards risk. *The Review of Economic Studies*, 25:65–86, 1958.

Обобщение понятия границ Фреше

Аркадий Арсеньевич Новоселов
Сибирский федеральный университет
Институт математики и фундаментальной информатики
Красноярск
arcady@novosyolov.com

Аннотация. *В работе формулируется задача вычисления границ Фреше для произвольной теоретико-множественной операции, вводится понятие двойственных операций, изучается связь границ Фреше для двойственных операций, вводится понятие (ко)валентности событий, вычислены границы Фреше для всех операций на дуплете событий.*

Ключевые слова. *Событие, функция распределения, граница Фреше, терраска, валентность, двойственность, генератор.*

1 Введение

Первые упоминания о неравенствах типа границ Фреше появились, по-видимому, в работах Буля [3], а строгие формулировки и доказательства были даны Фреше в работах [4, 5].

В [6] описано понятие копулы и использованы границы Фреше для значений копулы (вероятности пересечения конечного набора событий).

Аналогичное понятие можно ввести и для других теоретико-множественных операций над n множествами.

Известно [1], что над n множествами можно определить $2^{(2^n)}$ различных теоретико-множественных операций. В настоящей работе обсуждается проблемы вычисления верхней и нижней границ Фреше для каждой такой операции.

В [2] была высказана идея о распространении понятия границ Фреше на произвольную теоретико-множественную операцию (объединение произвольного набора терраск, получаемых из заданного набора событий). В настоящей работе описывается задача вычисления границ Фреше для произвольной теоретико-множественной операции над n событиями, вводится понятие двойственных операций и изучается связь границ Фреше для таких операций, дано исчерпывающее описание границ Фреше для операций на дуплете событий.

© 2014 А.А.Новоселов

Олег Воробьев (ред.), Труды XIII ФАМЭМС'2014, Красноярск: СФУ

2 Основные определения

Пусть $(\Omega, \mathcal{F}, \mathbf{P})$ — вероятностное пространство, зафиксируем целое положительное число n, обозначим $N = \{1, \ldots, n\}$ и рассмотрим вектор $u \in [0,1]^n$, компоненты которого будем интерпретировать, как вероятности. Выберем n событий $x_1, \ldots, x_n$ из σ-алгебры $\mathcal{F}$ таким образом, что $\mathbf{P}(x_i) = u_i$, $i \in N$, а в остальном произвольных. Какова максимальная и минимальная вероятность пересечения всех этих событий при сформулированных условиях? То есть, чему равны экстремальные значения

$$F_{\cap}^{-}(u) = \min \mathbf{P}\left(\bigcap_{i \in N} x_i\right), \tag{1}$$

$$F_{\cap}^{+}(u) = \max \mathbf{P}\left(\bigcap_{i \in N} x_i\right), \tag{2}$$

при условии

$$\mathbf{P}(x_i) = u_i,\ i \in N.$$

Ответ на этот вопрос дают классические границы Фреше. Когда вектор $(u_x, x \in \mathfrak{X})$ пробегает весь гиперкуб $[0,1]^n$, выражения (1), (2) задают на этом гиперкубе функции, называемые, соответственно, нижней и верхней границами Фреше для операции пересечения множеств.

Точные выражения границ Фреше для вероятности пересечения n событий тесно связаны с понятием копулы [6], поскольку значение совместной функции распределения случайного вектора $K = (K_1, \ldots, K_n)$ совпадает с вероятностью пересечения событий $\{K_i \le u_i\}$, $i \in N$:

$$F_K(u) = \mathbf{P}\left(\bigcap_{i \in N} \{K_i \le u_i\}\right),$$

и выражается через функцию копулы $C : [0,1]^n \to [0,1]$ и маргинальные функции распределения F_i, $i \in N$ посредством

$$F_K(u) = C\left(F_1(u_1), \ldots, F_n(u_n)\right).$$

Таким образом, функции (1), (2) являются верхними и нижними границами значений функций копулы C на гиперкубе $[0,1]^n$:

$$F_{\cap}^{-}(u) = \min_{u \in [0,1]^n} C(u),$$

$$F_{\cap}^{+}(u) = \max_{u\in[0,1]^n} C(u).$$

Границы для функции копулы известны и равны

$$F_{\cap}^{-}(u) = \max\left(0, \sum_{i\in N} u_i - (n-1)\right), \quad (3)$$

$$F_{\cap}^{+}(u) = \min(u_i,\, i \in N). \quad (4)$$

С помощью принципа дополнительности легко получить границы Фреше для операции объединения всех n событий. Действительно,

$$\bigcup_{i\in N} x_i = \left(\bigcap_{i\in N} x_i^c\right)^c,$$

так что

$$\begin{aligned}
&\mathbf{P}\left(\bigcup_{i\in N} x_i\right) = \left(1 - \mathbf{P}\left(\bigcap_{i\in N} x_i^c\right)\right)\\
=\;& 1 - \mathbf{P}\left(\bigcap_{i\in N} x_i^c\right) \geq 1 - \min_{i\in N}\left(1 - \mathbf{P}(x_i)\right)\\
=\;& \max_{i\in N} \mathbf{P}(x_i),
\end{aligned}$$

откуда нижняя граница для операции объединения всех событий записывается в виде функции

$$F_{\cup}^{-}(u) = \max_{i\in N} u_i. \quad (5)$$

С другой стороны,

$$\begin{aligned}
&\mathbf{P}\left(\bigcup_{i\in N} x_i\right) = 1 - \mathbf{P}\left(\bigcap_{i\in N} x_i^c\right)\\
\leq\;& 1 - \max\left(0, \sum_{i\in N}(1 - \mathbf{P}(x_i)) - (n-1)\right)\\
=\;& \min\left(1, \sum_{i\in N} \mathbf{P}(x_i)\right),
\end{aligned}$$

поэтому верхняя граница Фреше для объединения имеет вид

$$F_{\cup}^{+}(u) = \min\left(1, \sum_{i\in N} u_i\right). \quad (6)$$

3 Другие операции: подходы

3.1 Представление произвольной операции генератором

Множество из n событий разбивает Ω на 2^n террасок, каждой из которых может быть поставлено в соответствие множество наступивших событий. Обозначим множество номеров этих событий $I \subseteq N$, а соответствующую терраску

$$ter(I) = \left(\bigcap_{i\in I} x_i\right) \bigcap \left(\bigcap_{i\in N-I} x_i^c\right) \quad (7)$$

Так, например, терраске пересечения всех событий соответствует множество событий $\{x_1, \ldots, x_n\}$ или множество их индексов N. Другим террасскам соответствуют подмножества индексов $I \subseteq N$. А произвольная теоретико-множественная операция над n событиями представляется в виде объединения некоторого множества террасок, которому можно поставить в соответствие набор подмножеств индексов $A \subseteq 2^N$.

Рассмотрим эти понятия на примерах. Дуплет событий $\{x_1, x_2\}$ разбивает Ω на 4 терраски, каждая из которых соответствует одному их множеств событий $\emptyset, \{x_1\}, \{x_2\}, \{x_1, x_2\}$. Пересечение всех множеств представляется терраской $ter(\{x_1, x_2\})$, а объединение всех множеств — суммой террасок $ter(\{x_1\}) + ter(\{x_2\}) + ter(\{x_1, x_2\})$. Операции объединения соответствует набор террасок, задаваемых набором множеств индексов $A = \{\{1\}, \{2\}, \{1, 2\}\} \subseteq 2^N$. Будем называть множество террасок A, порождающее операцию, *генератором* этой теоретико-множественной операции.

Таким образом, для произвольного конечного множества событий произвольная теоретико-множественная операция над событиями из этого множества формально определяется своим генератором $A \subseteq 2^N$ посредством

$$\mathcal{O}_A = \sum_{I\in A} ter(I). \quad (8)$$

3.2 Границы Фреше для произвольной операции

Формализуем понятие границ Фреше для произвольной теоретико-множественной операции над событиями из конечного множества событий. Нижние и верхние границы Фреше для операции $\mathcal{O}_A$ будем обозначать, соответственно $F_A^-(u)$, $F_A^+(u)$, $u \in [0,1]^n$.

Зафиксируем число событий n, а также генератор A, представляющий операцию. Для каждой точки $u \in [0,1]^n$ и всевозможных наборов n событий из $\mathcal{F}$, удовлетворяющих условиям $\mathbf{P}(x_i) = u_i$, $i \in N$ вычислим наименьшее и наибольшее значение вероятности события $\mathbf{P}(\mathcal{O}_A(x_1, \ldots, x_n))$. Другими словами, нижняя граница Фреше является решением экстремальной задачи

$$F_A^-(u) = \min_{x_1,\ldots,x_n\in\mathcal{F}} \mathbf{P}(\mathcal{O}_A(x_1, \ldots, x_n)) \quad (9)$$

при условии

$$\mathbf{P}(x_i) = u_i,\, i \in N, \quad (10)$$

а верхняя граница Фреше - решением экстремальной задачи

$$F_A^+(u) = \max_{x_1,\ldots,x_n\in\mathcal{F}} \mathbf{P}(\mathcal{O}_A(x_1, \ldots, x_n)) \quad (11)$$

при том же условии (10).

Мощность генератора A будем называть *валентностью* операции $\mathcal{O}_A$: $v(\mathcal{O}_A) = |A|$; валентность может принимать значения от 0 до $2^n = 2^{|N|}$. Дополнение валентности до 2^n будем называть *ковалентностью* операции $\mathcal{O}_A$: $c(\mathcal{O}_A) = 2^{|N|} - |A|$.

Очевидно, $\mathcal{O}_\emptyset = \emptyset$ и $\mathcal{O}_{2^N} = \Omega$, поэтому

$$F^-_\emptyset(u) = F^+_\emptyset(u) = 0,\ F^-_{2^N}(u) = F^+_{2^N}(u) = 1, \qquad (12)$$

что исчерпывает вопрос о вычислении границ Фреше для операций нулевой валентности и полной валентности.

3.3 Двойственность

Элементарно выводятся следующие соотношения двойственности:

$$F^-_A(u) + F^+_{2^N-A}(u) = 1,\ F^+_A(u) + F^-_{2^N-A}(u) = 1. \quad (13)$$

Вообще, назовем операции, соответствующие взаимно дополнительным генераторам A и $2^N - A$, *двойственными*. Ввиду соотношений двойственности (13) вычисление границ Фреше достаточно произвести только для половины операций, например, для операций небольшой валентности (не превышающей 2^{n-1}).

4 Границы Фреше для операций малой валентности

4.1 Одновалентная операция

Одноэлементный генератор $A = \{I\}$, где $I \subseteq N$ — некоторое подмножество индексов, порождает операцию, значением которой является терраска (7), поэтому нижняя граница Фреше имеет вид

$$\begin{aligned} & F^-_{\{I\}}(u) \\ = \ & \max\left(0, \sum_{i\in I} u_i + \sum_{i\in N-I}(1-u_i) - (n-1)\right) \\ = \ & \max\left(0, \sum_{i\in I} u_i - \sum_{i\in N-I} u_i - |I| + 1)\right), \end{aligned} \qquad (14)$$

а верхняя граница Фреше —

$$F^+_{\{I\}}(u) = \min(\min_{i\in I} u_i, \min_{i\in N-I}(1-u_i)). \qquad (15)$$

Из соотношений двойственности (13) легко выводятся и границы Фреше для операций A ковалентности 1. В этом случае дополнение к A является одноэлементным, и границы получаются из (14), (15) в виде

$$\begin{aligned} & F^-_{2^N-\{I\}}(u) = 1 - F^+_{\{I\}}(u) \\ = \ & \max\left(\max_{i\in N}(1-u_i), \max_{i\in N-I} u_i\right), \end{aligned} \qquad (16)$$

$$\begin{aligned} & F^+_{2^N-\{I\}}(u) = 1 - F^-_{\{I\}}(u) \\ = \ & 1 - \max\left(0, \sum_{i\in I} u_i - \sum_{i\in N-I} u_i - |I| + 1\right) \\ = \ & \min\left(1, \sum_{i\in I}(1-u_i) + \sum_{i\in N-I} u_i\right) \end{aligned} \qquad (17)$$

4.2 Дуплет

Для дуплета событий $n = 2$, так что $\mathfrak{X} = \{x_1, x_2\}$. При этом валентность операции может изменяться от 0 до $2^2 = 4$. Ввиду соотношений двойственности достаточно вычислить границы Фреше для операций до валентности 2 включительно. Всего имеется одна нуль-валентная операция, $C^1_4 = 4$ одновалентных операций, $C^2_4 = 6$ двухвалентных операций, а также 4 трехвалентных операции и одна четырехвалентная.

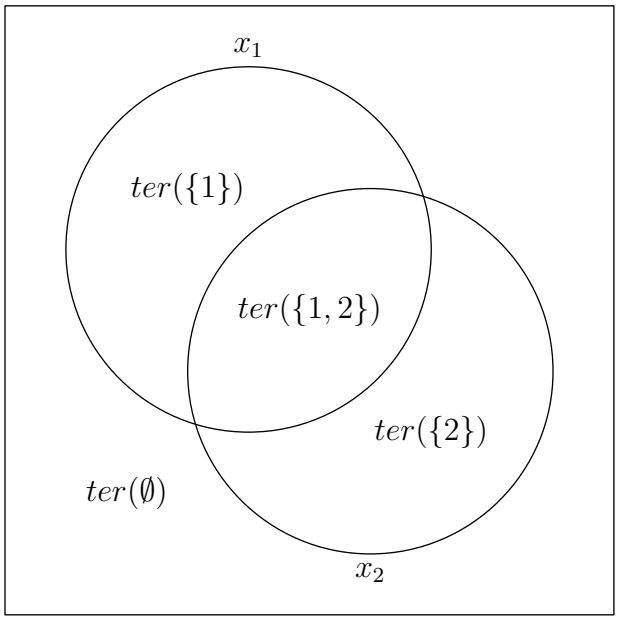


Рисунок 1: Диаграмма Венна для дуплета событий $\{x_1, x_2\}$.

Для нуль-валентной и двойственной к ней четырехвалентной операции границы Фреше указаны в (12). Для одновалентных операций эта работа была проделана в параграфе 4.1, а выводы для трехвалентных операций получаются из соотношений двойственности (13), см. (16), (17). Здесь сразу перейдем к двухвалентным операциям, $|A| = 2$.

Шесть двухвалентных операций представлены в табл. 1 в виде двух групп (основные и двойственные)

A	$\mathcal{O}_A$	$2^N - A$	$\mathcal{O}_{2^N-A}$
$\{\{1\}, N\}$	x_1	$\{\{2\}, \emptyset\}$	x_1^c
$\{\{2\}, N\}$	x_2	$\{\{1\}, \emptyset\}$	x_2^c
$\{\{1\}, \{2\}\}$	$x_1 \Delta x_2$	$\{\emptyset, N\}$	$ter(\emptyset) + ter(N)$

Таблица 1: Двухвалентные операции на дуплете и их результат

Из таблицы видно, что результат операции для генератора $A = \{\{1\}, N\}$ не зависит от взаимного расположения событий x_1, x_2, и равен x_1, поэтому верхние и нижние границы Фреше легко вычисляются:

$$F^-_{\{\{1\},N\}}(u) = F^+_{\{\{1\},N\}}(u) = u_1. \quad (18)$$

Аналогичный вывод справедлив для операции с генератором $\{\{2\}, N\}$: ее результат операции всегда равен x_2, а границы Фреше имеют вид

$$F^-_{\{\{2\},N\}}(u) = F^+_{\{\{2\},N\}}(u) = u_2. \quad (19)$$

Из соотношений двойственности (13) непосредственно выводим

$$F^-_{\{\{2\},\emptyset\}}(u) = F^+_{\{\{2\},\emptyset\}}(u) = 1 - u_1 \quad (20)$$

и

$$F^-_{\{\{1\},\emptyset\}}(u) = F^+_{\{\{1\},\emptyset\}}(u) = 1 - u_2. \quad (21)$$

Далее, для генератора $\{\{1\}, \{2\}\}$ значением операции является симметрическая разность событий, что позволяет легко вычислить границы Фреше. Действительно, вероятность симметрической разности событий служит метрикой, наименьшее значение которой достигается при вложенном положении событий; метрика при этом равна модулю разности вероятностей событий, так что

$$F^-_{\{\{1\},\{2\}\}}(u) = |u_1 - u_2|. \quad (22)$$

Наибольшее значение расстояния между событиями достигается при максимальном их разделении, когда они либо не пересекаются, либо заполняют все Ω, что дает выражение для верхней границы Фреше

$$\begin{aligned} F^+_{\{\{1\},\{2\}\}}(u) &= \min(u_1 + u_2, 2 - (u_1 + u_2)) \\ &= 1 - |u_1 + u_2 - 1|. \end{aligned} \quad (23)$$

Границы Фреше для генератора $\{\emptyset, \{N\}\}$ получаются из (22), (23) посредством соотношений двойственности (13):

$$\begin{aligned} F^-_{\{\emptyset,\{N\}\}}(u) &= |u_1 + u_2 - 1|, \\ F^+_{\{\emptyset,\{N\}\}}(u) &= 1 - |u_1 - u_2|. \end{aligned}$$

4.3 Триплет

Для операций на триплете событий максимальная валентность составляет 8, и вычисление придется проводить до операций валентности 4 включительно. Эту работу еще предстоит проделать в будущем.

5 Заключение

В работе поставлена задача вычисления границ Фреше для произвольной теоретико-множественной операции. Для решения этой задачи предложено представление произвольной теоретико-множественной операции генератором, введены понятия двойственной операции, валентности и ковалентности операции. Проведено полное вычисление границ Фреше для операций на дуплете событий.

Благодарности

Автор признателен Олегу Юрьевичу Воробьеву за идею обобщения понятия границы Фреше на произвольные теоретико-множественные операции над n множествами (событиями).

Список литературы

[1] О. Ю. Воробьев. *Эвентология.* Сибирский федеральный университет, Красноярск, 2007, 435с., https://www.academia.edu/179393/Vorobyev_O.Yu._Eventology._-_Krasnoyarsk_Siberian_Federal_University._-_2007._-_435p.

[2] О. Ю. Воробьев. Границы Фреше для произвольной теоретико-множественной операции. Устное сообщение, 2011.

[3] G. Boole. An investigation of the laws of thought, on which are founded the mathematical theories of logic and probability. 1854.

[4] M. Frechet. Generalisations du theoreme des probabilites totales. *Fundamenta Mathematica*, 25:379–387, 1935.

[5] M. Frechet. Sur les tableaux de correlation dont les marges sont donnees. *Annales de l'Universite de Lyon, Section A: Sciences mathematiques et astronomie*, 9:53–77, 1951.

[6] A. Sklar. Fonctions de repartition a n dimensions et leures marges. *Publications de l'Institut de Statistique de L'Universite de Paris*, 8:229–231, 1959.

Измеримость случайного множества событий

Аркадий Арсеньевич Новоселов
Сибирский федеральный университет
Институт математики и фундаментальной информатики
Красноярск
arcady@novosyolov.com

Аннотация. *В работе доказана безусловная измеримость случайного множества событий, вытекающая непосредственно из структуры этого случайного элемента безо всяких дополнительных условий и предположений.*
Ключевые слова. *Событие, множество событий, алгебра, измеримость, терраска.*

1 Введение

В работах О.Ю.Воробьева [1, 2, 3] изучался случайный элемент, называемый случайным множеством. В дальнейшем возникло эвентологическое направление [4], центральным объектом в котором стало специфическое случайное множество, а именно - случайное множество событий.

И абстрактное случайное множество, и случайное множество событий, будучи случайными элементами, обязаны обладать свойством измеримости. В настоящей работе показано, что случайное множество событий обладает свойством измеримости "по статусу", безо всяких дополнительных предположений.

Во втором разделе статьи вводятся основные понятия и определения, а в третьем разделе доказана основная теорема.

2 Случайные элементы и измеримость

Пусть $(\Omega, \mathcal{F}, \mathbf{P})$ — вероятностное пространство, U — произвольное множество, $\mathcal{A}$ — некоторая σ-алгебра его подмножеств (отметим, что в случае, когда U — конечное множество, можно ограничиться понятием алгебры подмножеств). Случайным элементом K на этом вероятностном пространстве со значениями в U называется отображение $K : \Omega \to U$, измеримое относительно пары σ-алгебр $(\mathcal{F}, \mathcal{A})$ в том смысле, что для всякого $A \in \mathcal{A}$ справедливо $K^{-1}(A) \in \mathcal{F}$.

Для дальнейшего отметим, что измеримость относительно σ-алгебры на множестве значений является *монотонной по включению*, то есть, если две σ-алгебры $\mathcal{A}_1, \mathcal{A}_2$ на U таковы, что $\mathcal{A}_1 \subseteq \mathcal{A}_2$, то измеримость K относительно пары σ-алгебр $(\mathcal{F}, \mathcal{A}_2)$ влечет измеримость относительно пары σ-алгебр $(\mathcal{F}, \mathcal{A}_1)$. В частности, самой сильной измеримостью является измеримость относительно пары σ-алгебр $(\mathcal{F}, 2^U)$.

Случайным конечным множеством называется случайный элемент, значения которого являются подмножествами некоторого конечного множества $\mathcal{X}$, то есть, элементы $2^{\mathcal{X}}$. Измеримость случайного конечного множества обычно требуется относительно полной алгебры подмножеств $2^{\mathcal{X}}$, которая естественно обозначается $2^{(2^{\mathcal{X}})}$.

Случайное конечное множество событий является частным случаем случайного конечного множества. Оно задается следующим образом. Фиксируется некоторый конечный набор событий $\mathcal{X} \subset \mathcal{F}$, и случайный элемент порождается соотношением

$$K(\omega) = \{x \in \mathcal{X} : \omega \in x\}. \tag{1}$$

Выражение (1) может быть истолковано, как "случайное множество наступивших событий", поскольку элементарному исходу эксперимента $\omega \in \Omega$ ставится в соответствие некоторое подмножество событий $X \in \mathcal{X}$, точнее, совокупность всех событий, в которые входит данный исход эксперимента ω, другими словами, все те (и только те) события, которые наступили в данном испытании (эксперименте).

3 Основная теорема

Нам придется рассматривать события в двух различных пространствах событий: в σ-алгебре $\mathcal{F}$ и в алгебре $2^{2^{\mathcal{X}}}$. Для удобства различения этих понятий будем последние называть *гиперсобытиями*.

Теорема 1. *Случайное множество событий K является измеримым отображением из Ω в $2^{\mathcal{X}}$ относительно полной алгебры гиперсобытий $2^{2^{\mathcal{X}}}$ (и, следовательно, относительно любой алгебры подмножеств $2^{\mathcal{X}}$, см. монотонность измеримости по включению).*

Доказательство. Требуется показать, что прообраз любого элемента алгебры гиперсобытий $2^{2^{\mathcal{X}}}$ является элементом $\mathcal{F}$. Поскольку алгебра $2^{2^{\mathcal{X}}}$ конечна, любой ее элемент A является конечным объединением моноплетов из $2^{2^{\mathcal{X}}}$,

$$A = \sum_{X \in \Lambda} X, \tag{2}$$

© 2014 А.А.Новоселов
Олег Воробьев (ред.), Труды XIII ФАМЭМС'2014, Красноярск: СФУ

где Λ — множество индексов, характеризующее гиперсобытие A. Например, гиперсобытие

$$A_0 = \{\{a\}, \{a, b\}\}$$

представимо в виде суммы двух моноплетов

$$\{\{a\}, \{a, b\}\} = \{\{a\}\} + \{\{a, b\}\};$$

здесь $a, b \in \mathcal{F}$ — события в исходном вероятностном пространстве. Прообраз гиперсобытия A по определению является объединением прообразов своих элементов

$$K^{-1}(A) = \sum_{X \in \Lambda} K^{-1}(\{X\}) = \sum_{X \in \Lambda} K^{-1}(X),$$

например, для гиперсобытия A_0 получаем

$$\begin{aligned} K^{-1}(A_0) &= K^{-1}(\{\{a\}\}) + K^{-1}(\{\{a, b\}\}) \\ &= K^{-1}(\{a\}) + K^{-1}(\{a, b\}). \end{aligned}$$

Осталось отметить, что каждое выражение вида $K^{-1}(X)$ для $X \subseteq \mathcal{X}$ представляет собой терраску $ter(X//\mathcal{X})$, задаваемую подмножеством событий X под $\mathcal{X}$, а объединение таких террасок с необходимостью является элементом $\mathcal{F}$, что и требовалось.

4 Заключение

В работе показано, что случайное множество событий является измеримым отображением из пространства элементарных исходов в множество своих значений, поэтому необходимость в высказывании предположения измеримости этого случайного элемента при его определении отпадает.

Благодарности

Автор признателен Олегу Юрьевичу Воробьеву за многочисленные обсуждения проблем случайных множеств событий, а также Людмиле Юрьевне Шангареевой за вопросы, стимулировавшие размышления над проблемой измеримости случайных множеств событий.

Список литературы

[1] О. Ю. Воробьев. *Вероятностное множественное моделирование.* Наука, Новосибирск, 1978, 160с.

[2] О. Ю. Воробьев. *Среднемерное моделирование.* Наука, Москва, 1984, 133с.

[3] О. Ю. Воробьев. *Сет-суммирование.* Наука, Новосибирск, 1993, 137с.

[4] О. Ю. Воробьев. *Эвентология.* Сибирский федеральный университет, Красноярск, 2007, 435с., https://www.academia.edu/179393/Vorobyev_O.Yu._Eventology._-_Krasnoyarsk_Siberian_Federal_University._-_2007._-_435p.

Скорость сходимости в последовательном оценивании интервалами фиксированной ширины асимптотической дисперсии ранговых оценок параметра сдвига

Гулноза Гафуровна Рахимова
Национальный Университет Узбекистана
Ташкент

Калмурза Сапарбаевич Сагидуллаев
Национальный Университет Узбекистана
Ташкент, khalmurza@mail.ru

Гафуржан Тоирович Турсунов
Высшая техническая школа пожарной безопасности МВД РУз
Ташкент

Аннотация. *В работе исследована скорость сходимости для последовательности моментов остановки и ширины доверительных интервалов.*
Ключевые слова. *Ранговая оценка, асимптотическая нормальность, момент остановки, интервалы фиксированной ширины .*

1 Введение

Рассмотрим последовательность независимых, одинаково распределенных случайных величин

$$\xi_1,\ \xi_2, \ldots,\ \xi_n, \ldots$$

с неизвестной функцией распределения $F(x)$, $x \in R_1$ и плотностью вероятности $f(x) = F'(x)$, $x \in R_1$ симметричной относительно параметра сдвига θ. Если $\theta_n = \theta_n(\xi_1, \xi_2, \ldots, \xi_n)$ ранговая оценка Ходжеса-Лемана [5], то известно (см. например [8]), что $\sqrt{n}\,(\theta_n - \theta)$ асимптотически нормально со средним 0 и дисперсией $\sigma^2 = \frac{1}{3A^2}$, где $A = \int\limits_{-\infty}^{\infty} f^2(x)dx$. В дальнейшем если не указаны пределы интегралов, то интегрирование осуществляется от $-\infty$ до $+\infty$. Статистические оценки дисперсии σ^2, а следовательно и A имеют большое значение в изучении асимптотической эффективности ранговых критериев и в других задачах непараметрического анализа. Статистические оценки A, основанные на оценке плотности вероятности $f(x)$ изучены авторами [8, 4, 6, 2]. Пусть

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(\xi_i < x)$$

эмпирическая функция распределения, где $I(A)$ — индикатор события ,

$$f_n(x) = \int\limits_{-\infty}^{+\infty} K_n(x,y)dF_n(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{a_n}K\left(\frac{x-\xi_i}{a_n}\right)$$

— ядерная оценка Розенблатта-Парзена для плотности вероятности. Здесь

$$K_n(x,y) = \frac{1}{a_n}K\left(\frac{x-y}{a_n}\right)$$

и $\{a_n > 0,\ n \geq 1\}$ — последовательность действительных чисел такая, что $\lim\limits_{n\to\infty} a_n = 0$ и $K(x)$, $x \in R_1$ — некоторая ограниченная плотность вероятности. Используя эмпирическую функцию распределения $F_n(x)$, и оценку плотности вероятности $f_n(x)$, построим оценку для

$$A = \int f^2(x)dx = \int f(x)dF(x)$$

следующим образом

$$A_n = \int f_n(x)dF_n(x) = \iint K_n(x,y)dF_n(x)dF_n(y).$$

Эта оценка изучена в работах [8, 4, 6] и получены условия асимптотической несмещенности, состоятельности и нормальности. Пусть плотности вероятности $f(x)$ и $K(x)$ удовлетворяют следующим условиям: (K): $\lim\limits_{|t|\to\infty} |t|K(t) = 0$, $\sup K(t) < \infty$, $\int tK(t)dt < \infty$, $\int t^2K(t)dt < \infty$, $\operatorname{var} K(t) < \infty$ (f): $\sup f(t) < \infty$, $\sup|f'(t)| < \infty$, $sup\,|f''(t)| < \infty$. В дальнейшем если не указана область, то sup вычисляется по всей прямой $R = (-\infty;\ \infty)$.

Теорема 1. *[8] Пусть $\sqrt{n}a_n \to \infty$ при $n \to \infty$ и выполнены условия (K) и (f). Тогда $\sqrt{n}\,(A_n - A)$ асимптотически нормальна со средним 0 и дисперсией*

$$\sigma^2 = 4\int f^3(x)dx - 4\left(\int f^2(x)dx\right)^2.$$

В работе [2] этот результат обобщен на выборки со случайным объемом N_n.

Теорема 2. *[2] Пусть выполнены условия теоремы 1 и $b(n), n > 1$ последовательность неслучайных чисел такая, что $b(n) = O(n)$ при $n \to \infty$ и (N):*

$$\frac{N_n}{b(n)} \xrightarrow{p} N$$

при $n \to \infty$, где $N > 0$ случайная величина. Тогда $\sqrt{N_n}\,(A_{N_n} - A)$ асимптотически нормальна со средним 0 и дисперсией σ^2.

Из этой теоремы видно, чтобы построить доверительный интервал фиксированной ширины для необходимо оценить дисперсию σ^2. В качестве оценки

© 2014 Г.Г.Рахимова, К.С.Сагидуллаев и Г.Т.Турсунов
Олег Воробьев (ред.), Труды XIII ФАМЭМС'2014, Красноярск: СФУ

дисперсии σ^2 возьмем следующую статистику $V_n = 4\left(T_n - A_n^2\right)$, где $T_n = \int f_n^2(x)dF_n(x)$. Пусть $0 < \gamma < 1$,

$$c = \Phi^{-1}\left(\frac{1+\gamma}{2}\right),$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{u^2}{2}} du.$$

В работе [2] изучены свойства момента остановки

$$N_\varepsilon = \min\left(n \geq 1 : \quad n \geq \frac{c^2}{\varepsilon^2} V_n\right),$$

$\varepsilon > 0$ и доверительного интервала фиксированной ширины

$$I_{\varepsilon,\gamma}\left(A_{N_\varepsilon}\right) = \left(A_{N_\varepsilon} - \varepsilon, \quad A_{N_\varepsilon} + \varepsilon\right).$$

Теорема 3. *[2] Если выполнены условия теоремы 1, тогда выполняется (N) для*

$$b_\varepsilon = \frac{c^2\sigma^2}{\varepsilon^2}$$

и $N = 1$, *а также*

$$\lim_{\varepsilon \to 0} P\left(A \in I_{\varepsilon,\gamma}\left(A_{N_\varepsilon}\right)\right) = \gamma. \qquad (1)$$

В данной работе изучается скорость сходимости в предельных соотношениях (N) и (1).

2 Основные результаты

Предположим, что $a_n = n^{-p}$, $p > 0$.

Теорема 4. *Пусть выполнены условия (K), (f) и* $0 < p < 2/3$. *Тогда*

$$P\left(\left|\frac{N_\varepsilon}{b_\varepsilon} - 1\right| \geq \varepsilon^{2p}\right) = O\left(\varepsilon^p\right) \quad \textit{при } \varepsilon \to 0.$$

Теорема 5. *Пусть выполнены условия (K), (f) и* $0 < p < 1/3$, *а также функции* $f^4(x)$ *и* $K^4(x)$ *интегрируемы. Тогда*

$$P\left(A \in I_{\varepsilon,\gamma}\left(A_{N_\varepsilon}\right)\right) = \gamma + O\left(\varepsilon^p\right) \quad \textit{при } \varepsilon \to 0.$$

Список литературы

[1] Боровских Ю.В. Королюк В.С. *Теория U-статистик.* Науково думка. Киев, 1989.

[2] Турсунов Г.Т. Об оценивании асимптотической дисперсии ранговых оценок параметра сдвига по выборке случайного объема. *Теория случайных процессов. «Науково думка». Киев*, 15:97–102, 1987.

[3] Csenki A. A theorem on the departure of randomly indexed u-statistics from normality with an application in fixed-width sequential interval estimation. *Sankhya, Indian J.Statist.*, 43, ser. A, 1:84–99, 1981.

[4] Schuster E.F. On the rate of convergence of an estimate of a functional of a probability density. *Scand.Actuarial J.*, 1:103–107, 1974.

[5] Lehman E. L. Hodges J.L. Estimates of location based on rank tests. *Ann. Math. Statist.*, 34:598–611, 1963.

[6] Ahmad I.A. On asymptotic properties of an estimate of a functional of a probability density. *Scand. Actuarial J.*, 3:176–181, 1976.

[7] Pfanzagl J. Michel R. The accuracy of the normal approximation for minimum contrast estimates. *Z. Wahr. verw.Geb.*, 18:73–84, 1971.

[8] Schweder T. Window estimation of the asymptotic variance of rank estimators of location. *Scand. J. Statist.*, 2:113–126, 1975.

Правила Чаргаффа и супер-симметрия в геномах

Михаил Георгиевич Садовский
Институт вычислительного моделирования СО РАН,
660036 Красноярск
msad@icm.krasn.ru

Ярослав Владимирович Гребнев
Сибирский федеральный университет, ИФБиБт
yaroslav.grebnev@yandex.ru

Аннотация. *В работе представлены предварительные результаты исследования супер-симметрии геномов различных организмов. Под супер-симметрией понимается равенство частот слова и его комплиментарного палиндрома (второе правило Чаргаффа), наблюдаемое в пределах одного стрэнда ДНК. Рассмотрены простейшие модели ДНК, порождающие такую супер-симметрию и её нарушения. Данная модель сравнена с реальными данными, полученными для митохондрий, хлоропластов, геномов дрожжей и некоторых других геномов высших животных.*

Ключевые слова. *ДНК, слово, частота, палиндром, комплиментарность, случайный процесс, Марковский процесс, упорядоченность.*

1 Введение

Первое правило Чаргаффа гласит, что в любой[1] последовательности ДНК число символов A равно числу символов T; аналогично, равны числа символов C и G. Это правило было первоначально сформулировано для двухцепочечной спирали ДНК, что во многом поспособствавало открытию самой двойной спирали. Позднее выяснилось, что данное правило также выполняется — с той точностью, которую нельзя не признать важной для биологии — и в пределах одного стрэнда двойной спирали.

Позже было сформулировано утверждение, получившее название второго правила Чаргаффа:

Биологическая «теорема» 1. *Рассмотрим пары слов (цепочек символов) длины q, которые являются комплиментарными палиндромами, т. е. читаются одинаково в противоположных направлениях с учётом замены символов в одном из слов по правилу комплиментарности, например,*

$$\mathsf{TTTACGACTAAACGCG} \Longleftrightarrow \mathsf{CGCGTTTAGTCGTAAA}\,.$$

Частоты таких слов также близки, либо равны с высокой точностью.

Данная «теорема» становится абсолютно верной и точной для двойной спирали ДНК: причина проста — двойная спираль по построению содержит абсолютно одинаковые количества (любых допустимых) слов и их комплиментарных пар. Тем удивительнее было открытие того, что такие пары слов имеют близкие частоты и для случая, когда слова подсчитываются только по одному стрэнду.

Упомянутая выше «Биологическая теорема № 1» для реальных генетических последовательностей выполняется не точно, но с некоторой погрешностью. Природа такой погрешности может быть двоякой: во-первых, это может быть математический артефакт — такое нарушение является следствием комбинаторных свойств рассматриваемой символьной последовательности; во-вторых, могут быть чисто биологические причины такого рода нарушений — возможно, процессы отбора привели к тому, что отобрались генетические тексты с ровно таким (комбинаторным) нарушением.

Основная цель нашей работы — оценить степень «биологичности» в нарушениях супер-симметрии, наблюдаемых для реальных последовательностей. Здесь возможны два подхода — **популяционный**: можно сравнивать реальные последовательности разных видов; **внутригеномный**: можно сравнивать отдельные фрагменты одной и той же последовательности. В рамках настоящей статьи мы будем следовать первому подходу.

Кроме того, возможен (и важен) ещё один подход к проблеме: моделирование происхождения такого свойства нуклеотидных последовательностей, как супер-симметричность. Одна из возможных моделей здесь — удвоение (и вообще, копирование) весьма больших фрагментов в последовательности с последующим разрушением точности копирования; заметим, что эта модель не является единственно возможной.

2 Супер-симметрия и её нарушения

Как уже было сказано выше, супер-симметрией в геномах (и вообще — в символьных последовательностях) мы будем считать точное выполнение второго правила Чаргаффа для частотного словаря, определённого по одному стрэнду (одной последовательности). Отклонения от этого точного выполнения второго правила мы будем считать нарушения-

© 2014 М.Г. Садовский и Я.В. Гребнев

Олег Воробьев (ред.), Труды XIII ФАМЭМС'2014, Красноярск: СФУ

[1] С точностью до ошибок, которыми можно смело пренебречь.

ми супер-симметрии. Для дальнейшего нам понадобится мера такого отклонения. Эту меру можно ввести многими различными способами (см. напр., [2]); в рамках настоящей работы мы будем пользоваться следующей величиной, определяющей меру нарушения супер-симметрии:

$$\mu(q) = \sqrt{\frac{2}{4^q} \sum_{\omega\in\Omega^\star} \left(f_\omega - f_{\overline{\omega}}\right)^2}\,; \qquad (1)$$

здесь $\Omega^\star$ — полуноситель данного частотного словаря (т. е. множество всех слов, являющихся лишь «прямыми»).

2.1 Правила Чаргаффа и модельные последовательности

Рассмотрим случайную нескоррелированную символьную последовательность из четырёхбуквенного алфавита $\aleph = \{\mathsf{A},\mathsf{C},\mathsf{G},\mathsf{T}\}$. Традиционно будем считать, что иных символов, а также пробелов и т. п. в ней нет; для реальных последовательностей, рассматриваемых ниже, это предположение не выполняется, однако мы всегда можем выбрать такой класс последовательностей, для которых влияние таких лишних символов может быть сделано сколь угодно малым.

2.1.1 Случайные независимые последовательности

Итак, будем рассматривать случайную нескоррелированную символьную последовательность из четырёхбуквенного алфавита $\aleph = \{\mathsf{A},\mathsf{C},\mathsf{G},\mathsf{T}\}$, причём будем полагать, что для неё точно выполняется первое правило Чаргаффа:

$$p(\mathsf{A}) = p(\mathsf{T}) \qquad \text{и} \qquad p(\mathsf{C}) = p(\mathsf{G})\,. \qquad (2)$$

Очевидно, что из этого немедленно следует простейшая

Теорема 1. *Для последовательности со свойствами, указанными выше, выполняется второе правило Чаргаффа с абсолютной точностью.*

Доказательство этой теоремы элементарно. В силу случайности и независимости появления символов в последовательности вероятность всякого слова длины q определяется выражением

$$p(\mathsf{A})^{l_\mathsf{A}} \cdot p(\mathsf{C})^{l_\mathsf{C}} \cdot p(\mathsf{G})^{l_\mathsf{G}} \cdot p(\mathsf{T})^{l_\mathsf{T}}\,,$$

где мультииндекс $l_\mathsf{A} + l_\mathsf{C} + l_\mathsf{G} + l_\mathsf{N} = q$ определяется структурой слова. Однако в силу равенства (2) частота слова и его комплиментарного палиндрома оказываются равными. □

Пусть теперь соотношение (2) не выполняется и $p_\mathsf{w} = \frac{1}{2}[p(\mathsf{A}) + p(\mathsf{T})]$, $p_\mathsf{s} = \frac{1}{2}[p(\mathsf{C}) + p(\mathsf{G})]$. Тогда (для определённости) $p(\mathsf{A}) = p_\mathsf{w} - \delta$, $p(\mathsf{T}) = p_\mathsf{w} + \delta$, $p(\mathsf{C}) = p_\mathsf{s} - \varepsilon$, $p(\mathsf{G}) = p_\mathsf{s} + \varepsilon$. Тогда частота (вероятность) «прямого» слова f_ω будет равна

$$(p_\mathsf{w} - \delta)^{l_\mathsf{A}} \cdot (p_\mathsf{s} - \varepsilon)^{l_\mathsf{C}} \cdot (p_\mathsf{s} + \varepsilon)^{l_\mathsf{G}} \cdot (p_\mathsf{w} + \delta)^{l_\mathsf{T}}\,, \qquad (3)$$

а частота $f_{\overline{\omega}}$ (вероятность) комплиментарного ему палиндрома

$$(p_\mathsf{w} + \delta)^{l_\mathsf{A}} \cdot (p_\mathsf{s} + \varepsilon)^{l_\mathsf{C}} \cdot (p_\mathsf{s} - \varepsilon)^{l_\mathsf{G}} \cdot (p_\mathsf{w} - \delta)^{l_\mathsf{T}}\,. \qquad (4)$$

Требуется оценить величину невязки μ (см. (1)) для частотного словаря толщины $q = l_\mathsf{A} + l_\mathsf{C} + l_\mathsf{G} + l_\mathsf{T}$. Заметим, что p_w и p_s не являются независимыми; они связаны простым и очевидным соотношением $p_\mathsf{w} + p_\mathsf{s} = 1/2$.

Рассмотрим для начала частный случай $l_\mathsf{A} = l_\mathsf{T}$ и $l_\mathsf{C} = l_\mathsf{G}$; заметим, что это ограничивает нас рассмотрением слов чётной длины, однако в дальнейшем мы откажемся от этого ограничения. Если $\delta = \varepsilon$, то частоты «прямого» слова и комплиментарного палиндрома будут равны друг другу. Если $\delta \neq \varepsilon$, то оценку можно получить, разлагая (3, 4) в ряд по степеням δ и ε; естественно, для этого следует считать эти величины малыми. Практика показывает (см. раздел 4), что такое предположение вполне приемлемо.

Разлагая каждый сомножитель в (3, 4) в ряд по степеням δ и ε и отбрасывая все члены, содержащие степени δ^2 (ε^2, соответственно) и выше, получаем

$$p_\mathsf{w}^{l_\mathsf{A}+l_\mathsf{T}} \cdot p_\mathsf{s}^{l_\mathsf{C}+l_\mathsf{G}} + \delta p_\mathsf{s}^{l_\mathsf{C}+l_\mathsf{G}} \cdot p_\mathsf{w}^{l_\mathsf{A}+l_\mathsf{T}-1}(l_\mathsf{T} - l_\mathsf{A}) + \\ + \varepsilon p_\mathsf{w}^{l_\mathsf{A}+l_\mathsf{T}} \cdot p_\mathsf{s}^{l_\mathsf{C}+l_\mathsf{G}-1}(l_\mathsf{G} - l_\mathsf{C}) \qquad (5)$$

и

$$p_\mathsf{w}^{l_\mathsf{A}+l_\mathsf{T}} \cdot p_\mathsf{s}^{l_\mathsf{C}+l_\mathsf{G}} + \delta p_\mathsf{s}^{l_\mathsf{C}+l_\mathsf{G}} \cdot p_\mathsf{w}^{l_\mathsf{A}+l_\mathsf{T}-1}(l_\mathsf{A} - l_\mathsf{T}) + \\ + \varepsilon p_\mathsf{w}^{l_\mathsf{A}+l_\mathsf{T}} \cdot p_\mathsf{s}^{l_\mathsf{C}+l_\mathsf{G}-1}(l_\mathsf{C} - l_\mathsf{G})\,. \qquad (6)$$

Окончательно, разница (5) и (6) даёт

$$\lambda = 2p_\mathsf{s}^{l_\mathsf{C}+l_\mathsf{G}-1} \cdot p_\mathsf{w}^{l_\mathsf{A}+l_\mathsf{T}-1}\Big[\delta p_\mathsf{s}(l_\mathsf{T} - l_\mathsf{A}) + \varepsilon p_\mathsf{w}(l_\mathsf{C} - l_\mathsf{G})\Big]\,,$$

что с учётом указанной выше связи приводит к

$$\mu \sim \sqrt{\lambda^2 \cdot n} = \lambda \cdot \sqrt{n}\,, \qquad (7)$$

где n — мощность частотного словаря. Подчеркнём, что формула (7) даёт оценку невязки, не зависящую от длины слов, для которых она определяется. Это означает, что данная модель генетического текста (случайная нескоррелированная последовательность) является весьма грубым приближением.

Далее, заметим, что $p_\mathsf{s} \leqslant \frac{1}{4}$ и $p_\mathsf{w} \leqslant \frac{1}{4}$. Из этого следует, что для выражения (7) справедлива следующая оценка сверху:

$$\lambda < \frac{2 \cdot \max\{\delta, \varepsilon\}}{4^{(q-1)}}$$

Таблица 1: Значения невязки (1) для девятнадцати видов (и штаммов) дрожжей; N — число исследованных последовательностей в геноме.

Организм	N	$\langle\mu(1)\rangle$	σ
A. gossypii	7	0,00248	0,00252
A. fumigatus	8	0,00073	0,00044
A. niger	19	0,00164	0,00125
A. nudilans	8	0,00089	0,00055
C. albicans	9	0,00148	0,00076
C. dublinesis	8	0,00085	0,00067
C. glabrata	13	0,07335	0,00094
C. neoformans	14	0,00124	0,00089
F. oxysporum	15	0,00096	0,00072
G. moniliformis	11	0,00082	0,00044
G. zeae	4	0,00056	0,00008
K. lactis	6	0,00178	0,00087
L. kluyveri	8	0,00228	0,00127
L. thermotolerns	8	0,00106	0,00060
P. stipitis	8	0,00120	0,00067
S. pombe 1	3	0,00093	0,00042
S. pombe (2)	3	0,00085	0,00048
Y. lipolitica	6	0,00070	0,00040
Z. rouxii	7	0,00176	0,00065

и таких слагаемых в сумме для невязки (1) не более $\frac{4^q}{2}$. Заметим, что предположение о том, что $p_{\mathsf{w}} \sim p_{\mathsf{s}} \sim \frac{1}{4}$ означает, что $l_{\mathsf{A}} \sim l_{\mathsf{T}}$ и $l_{\mathsf{C}} \sim l_{\mathsf{G}}$. Отсюда немедленно следует, что $\lambda_\omega \sim 4^{-q} \cdot \max\{\delta, \varepsilon\}$.

Полученная оценка слишком груба: предположение о том, что $p_{\mathsf{s}} \cong \frac{1}{4}$ и $p_{\mathsf{w}} \cong \frac{1}{4}$ означают, что в рассматриваемой последовательности абсолютно точно выполняется первое правило Чаргаффа и, кроме того, все нуклеотиды представлены в равных долях. На практике такого рода предположений не выполняется никогда. Более эффективной оценкой является такая, для которой[2] частота комплиментарного палиндрома имеет порядок $A \cdot N^{-1}$, где N — длина рассматриваемой последовательности, а константа A определяется структурой частотного словаря и в худшем случае имеет порядок $A \sim 4^q$, где q — длина рассматриваемого слова. Иными словами, для сколько-нибудь реалистичных случаев $A \ll N$.

3 Материалы и методы

В настоящей работе исследовалось поведение невязки (1) для различных реальных генетических текстов. Все генетические последовательности были взяты из EMBL–банка. Исследовались геномы дрожжей (19 видов и штаммов), по одной хромосоме человека (четырнадцатая хромосома), макаки и шимпанзе (первая хромосома). Для каждой последовательности определялась невязка (1) для $1 \leqslant q \leqslant 8$.

[2]Для достаточно длинных слов; но нас, собственно, и интересуют оценки поведения невязки (1) для длинных слов.

Таблица 2: Значения невязки (1) для различных групп организмов; N — число исследованных последовательностей в геноме (в семействе геномов).

Организм	N	$\langle\mu(1)\rangle$	σ
A. gambiae	5	0,00037	0,00014
B. taurus	30	0,00082	0,00065
H. sapience, ch. XIV	1	0,00062	—
M. mulatta ch. 1	1	0,00057	—
Бактерии	752	0,00234	0,00298
A. thaliana	6	0,00103	0,00036
D. melanogaster	6	0,00159	0,00224
P. traoglodites	1	0,00166	—
Митохондрии	2004	0,08281	0,04066
Хлоропласты	463	0,00654	0,00504

В случае, если исходная последовательность содержала иные, чем A, C, G и T символы, они игнорировались при составлении частотных словарей, по которым определялась невязка (1).

4 Результаты и обсуждение

В Таблицах 1 и 2 представлены результаты определения невязки (1) для различных видов. В Таблице 1 представлены результаты определения невязки для девятнадцати видов дрожжей. Следует подчеркнуть, что дрожжи являются по-своему специфическим генетическим объектом: для них характерна весьма высокая степень сложности хранения наследственной информации. Под сложностью здесь понимается то обстоятельство, что у многих видов дрожжей кодирующими являются оба стрэнда ДНК, а гены в комплиментарных стрэндах часто пересекаются.

В этой таблице представлены значения средней по геному невязки; правая колонка содержит значения стандартного отклонения невязок, определяемых для каждой хромосомы в группе либо в геноме.

В Таблице 2 представлены результаты определения невязки (1) для бактерий, хлоропластов, митохондрий, а также для некоторых геномов либо отдельных хромосом геномов высших организмов.

Митохондрии знамениты тем, что для них невязка (1) принимает самые большие — среди последовательностей иных групп — значения. Они являются своего рода «лидерами» по нарушению первого правила Чаргаффа. Более того, для них характерен и очень большой разброс значений невязки; значения стандартного отклонения ясно говорят о значительном разбросе. Данное обстоятельство хорошо иллюстрирует рисунок 1, на котором показано совместное распределение геномов митохондрий и бактерий по значениям невязки. Во-первых, следует обратить внимание на гистограмму для бактерий; она имеет вид, характерный для Пуассоновского распределе-

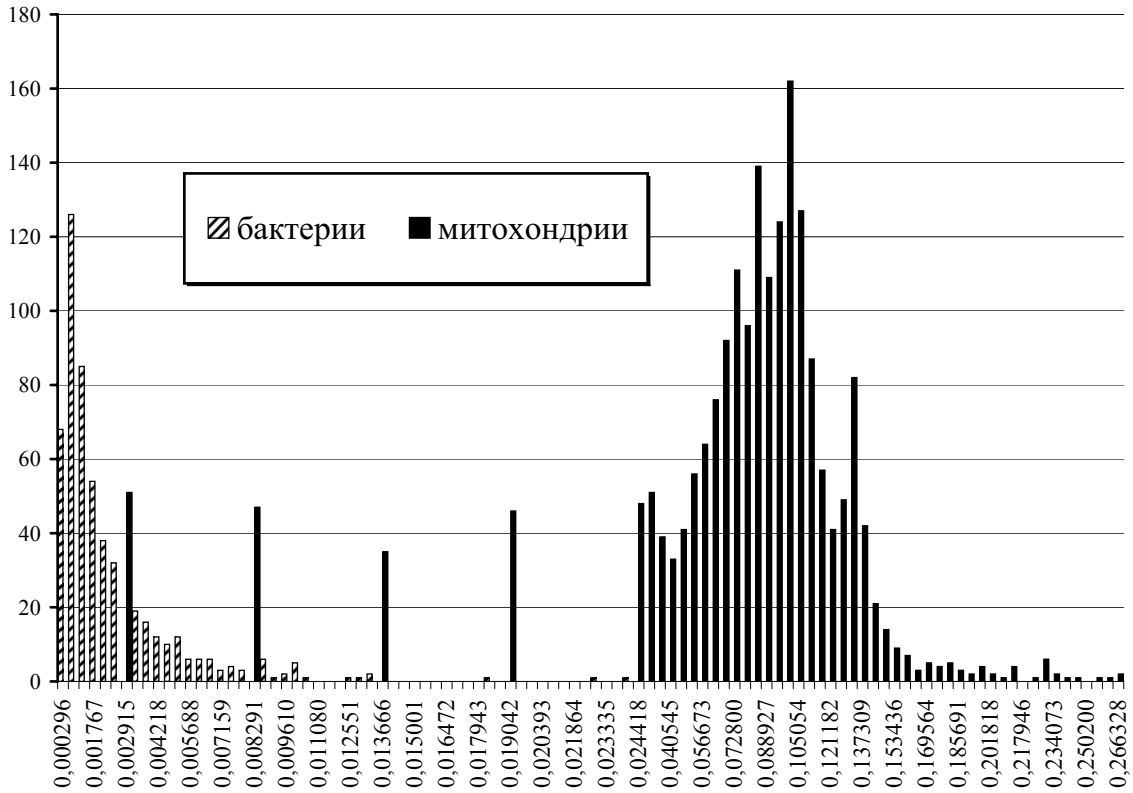


Рисунок 1: Распределение геномов митохондрий (чёрные столбцы) и бактерий (косая штриховка) по значениям невязки (1). По оси абсцисс отложены значения невязки, по оси ординат — число геномов с невязкой, попадающей в данный диапазон.

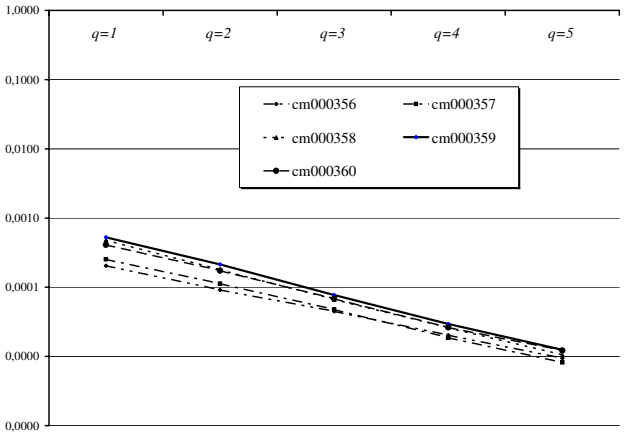


Рисунок 2: Изменение невязки (1) для генома комара. По оси абсцисс отложены значения длины слов, по оси ординат — значения невязки.

ния. Более того, такой вид гистограммы характерен для любых (в диапазоне $1 \leqslant q \leqslant 8$) значений длины слов, для которых определяется невязка. Единственным существенным исключением из этого правила являются митохондрии. Аналогичная зависимость описана в работе [3], в которой рассматривались другие комбинации геномов и видов.

Поведение невязки (1) для словарей различной толщины показывает, что она падает с ростом q толщины словаря, причём характер близко к экспоненциальному падению (см. рисунок 2). Данное наблюдение позволяет утверждать, что предложенная ранее модель [1, 2] образования структур, проявляющихся в форме второго правила Чаргаффа, не является адекватной: случайные последовательности, для которых выполняется первое правило Чаргаффа, с необходимостью обеспечивают выполнение и второго правила. Тем самым, вопрос о возникновении структурированности в геномах различных организмов, проявляющейся в виде описанной выше супер-симметрии, требует отдельного специального исследования.

Список литературы

[1] А. А. Коваль, Е. А. Темлякова и М. Г. Садовский. О некоторых комбинаторных свойствах нуклеотидных последовательностей. *VI Всесибирский конгресс женщин-математиков, 15 - 17 января 2010. Материалы конф., Красноярск, изд-во СибГТУ.*, pages 194–200., 2010.

[2] Е. А. Темлякова, Ю. А. Путинцева и М. Г. Садовский. Правила Чаргаффа, симметрия геномов и эволюция. *Труды X Межд. ФАМ-конференции, Красноярск, изд-во СФУ, (Под ред. О.Ю. Воробьчва)*, pages 339–342, 2011.

[3] G. Albrecht-Bühler. Inversions and inverted transpositions as the basis for an almost universal «format»' of genome sequences. *Genomics*, 90:297–305, 2007.

Поиск и изучение структур в символьных последовательностях, заданных случайными процессами

Михаил Георгиевич Садовский
Институт вычислительного моделирования СО РАН
Красноярск
msad@icm.krasn.ru

Ксения Андреевна Никитина
ИКИТ СФУ
Красноярск
ksklokova@yandex.ru

Аннотация. *В работе представлен новый подход к выявлению структурированности в символьных последовательностях, таких как, например, последовательность ДНК. Предметом исследования является характер распределения расстояний между короткими подпоследовательностями (словами). Показано, что реальная последовательность ДНК отличается от теоретической случайной последовательности.*

Ключевые слова. *Порядок, триплет, частотный словарь, Марковский процесс, распределение расстояний, ближайший сосед.*

1 Введение

Проблема поиска, выявления и последующего анализа различных видов структур в символьных последовательностях является универсальной для широкого круга задач, как чисто теоретических, так и прикладных. Прикладной аспект это проблемы имеет своими приложениями различные задачи в таких областях, как криптография, стеганография, теоретическое программирование, различные оптимизационные задачи. С «противоположного» конца располагаются такие науки, как филология, лингвистика, биоинформатика, биофизика и многие другие.

Поиск структуры в символьной последовательности может быть проведён в рамках двух подходов:

- для начала выделяются либо иным образом определяются те или иные образцы (набор образцов), определяется каким-то образом мера допустимого отклонения от такого образца, а затем в символьной последовательности ищутся все такие подпоследовательности, которые удовлетворяют критерию близости к заранее заданным образцам;
- в исследуемой последовательности ищутся такие фрагменты (подпоследовательности), которые являются слишком «неожиданными» для неё. Понятно, что мера такой неожиданности может существенно повлиять на результаты поиска той или иной структуры в изучаемой символьной последовательности.

© 2014 М.Г.Садовский и К.А.Никитина

Олег Воробьев (ред.), Труды XIII ФАМЭМС'2014, Красноярск: СФУ

Возможен, наконец, и в каком-то смысле промежуточный вариант: в последовательности заранее выделяются какие-то значимые фрагменты (сравнительно малой длины) и затем изучается взаимное расположение таких фрагментов. Именно в рамках такого подхода выполнена настоящая работа.

Мы изучали распределение триплетов — подпоследовательностей длины 3 — в больших символьных последовательностях. Изучались как реальные последовательности (генетические тексты), так и случайные символьные последовательности, подобные исходным генетическим текстам. Выбор триплетов в качестве тех фрагментов, взаимное расположение которых определяет структуру в символьной последовательности, не случаен: именно триплеты являются базовым элементом подавляющего большинства генетических систем. Возможно, различные генетические тексты обладают и структурами, определяемыми иными, более длинными подпоследовательностями, однако их поиск и выделение следует вести после того, как будут выявлены и описаны структуры в генетических текстах, порождаемые именно триплетами, поскольку их роль велика, а влияние, ими оказываемое, самое существенное.

2 Основные понятия

Рассмотрим бесконечную символьную последовательность из алфавита $\aleph = \{A, C, G, T\}$ мощности $m = 4$, заданную Марковским процессом порядка t. В ней вероятность появления того или иного символа алфавита $\aleph$ в определенном месте этой последовательности зависит от того, какие символы были на предыдущих t местах (рассматриваются оба варианта $t \leqslant k$ и $t > k$). **Словом** назовем любую конечную подпоследовательность, содержащуюся в исходной последовательности.

Зафиксируем два слова ω_1 и ω_2, каждое некоторой длины k. Найдём первое попавшееся вхождение слова ω_1 в рассматриваемую последовательность и некоторое вхождение слова ω_2, расположенное правее ω_1. Количество символов от начала ω_1 до начала ω_2 будем называть **расстоянием** n между словами ω_1 и ω_2. Если нигде ближе n слово ω_2 не встретилось, то в таком случае ω_2 назовем **ближайшим соседом** к

ω_1. Например,

$$k = 3, \omega_1 = ACC, \quad \omega_2 = AAA$$

$$\ldots TA\mathbf{ACC}ACTGT\mathbf{AAA}TG\ldots$$

$$n = 8$$

Рассмотрим следующие функции распределения расстояний между словами:

- $f_{\omega_1,\omega_2}(n)$ — вероятность встретить слово ω_2 на фиксированном расстоянии n справа от ω_1;
- $f^*_{\omega_1,\omega_2}(n)$ — вероятность встретить слово ω_2 на расстоянии n справа от ω_1 и не раньше, то есть вероятность того, что ω_2 будет являться ближайшим соседом ω_1 на расстоянии n.

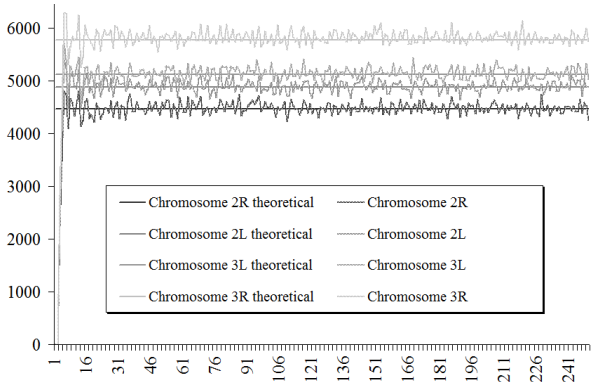


Рисунок 1: Теоретические $f_{\omega_1,\omega_2}(n)$ и экспериментальные $\tilde{f}_{\omega_1,\omega_2}(n)$ значения функции распределения фиксированнных расстояний между словами $\omega_1 = ACT$ и $\omega_2 = AGT$ для четырех последовательностей, кодирующих разные участки хромосом генома мухи *Drosophila melanogaster*. По оси абсцисс отложено расстояние между словами n.

Далее подробнее рассмотрим случай $k = t = 3$. Для него справедливы следующие утверждения. Обозначим матрицу перехода данного Марковского процесса:

$$P = \begin{pmatrix} p_{11} & p_{12} & \ldots & p_{1r} \\ p_{21} & p_{22} & \ldots & p_{2r} \\ \ldots & \ldots & \ldots & \ldots \\ p_{r1} & p_{r2} & \ldots & p_{rr} \end{pmatrix}$$

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, где 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 $r \times r = 64 \times 64$, p_{ij} — вероятность перехода из слова i в слово j. Причём для Марковского процесса сумма элементов каждой строки матрицы перехода будет равна единице:

$$\forall i = 1, 2, \ldots, r \quad \sum_{j=1}^{r} p_{ij} = 1$$

Комплементарные палиндромы

$\omega_1 - \omega_2$	f_{ω_1,ω_2}	$< \tilde{f}_{\omega_1,\omega_2} >$
CCC-GGG	545	577
ACT-AGT	3812	3852
AAA-TTT	74123	76940

Не комплементарные палиндромы

$\omega_1 - \omega_2$	f_{ω_1,ω_2}	$< \tilde{f}_{\omega_1,\omega_2} >$
GGT-CGG	916	941
TGC-GTG	1912	1934
ATT-TAT	14037	14127

Таблица 1: Теоретическая частота встречаемости пары слов ω_1 и ω_2 в последовательности f_{ω_1,ω_2}, вычисленная по формуле (3), и среднее значение функции $\tilde{f}_{\omega_1,\omega_2}(n)$ для $n \in \{5, 6, \ldots, 2000\}$, полученное в ходе вычислений над последовательностью генома *Caenorhabditis elegans*.

Стоит заметить, что из любого заранее выбранного слова за один шаг можно перейти далеко не во все возможные слова. Например, из $A\mathbf{AA}$ за один шаг можно перейти в $\mathbf{AA}A, \mathbf{AA}C, \mathbf{AA}G, \mathbf{AA}T$, но не в $\mathbf{AC}A$. Аналогично, в $\mathbf{AA}A$ можно перейти только из $A\mathbf{AA}, C\mathbf{AA}, G\mathbf{AA}, T\mathbf{AA}$. Тем самым, каждая строка (и, аналогично, каждый столбец) содержит $m = 4$ из $m^k = 64$ ненулевых элементов.

Совпадение по ближайшим $n^{max}_{\omega_1,\omega_2}$

Геном	Ch	M	N	α
Schizosaccharomyces pombe	3	12	141	0.085
Drosophila melanogaster	5	18	232	0.078
Caenorhabditis elegans	5	43	200	0.215
Random	3	16	135	0.119

Совпадение по самым дальним $n^{max}_{\omega_1,\omega_2}$

Геном	Ch	M	N	α
Schizosaccharomyces pombe	3	41	109	0.376
Drosophila melanogaster	5	38	208	0.183
Caenorhabditis elegans	5	85	146	0.582
Random	3	0	150	0

Таблица 2: Коэффициент внутривидового совпадения по самым ближним и самым дальним максимальным расстояниям $n^{max}_{\omega_1,\omega_2}$. Здесь Ch — число сравниваемых последовательностей (хромосом или их частей); Random — данные для последовательностей, не принадлежащих одному и тому же геному.

Возводя такую матрицу в степень n, получаем распределение вероятностей встречи слова j на расстоянии n от слова i. Обозначим через a_{ω_1} строку перехода из слова ω_1, выписанную из матрицы P, а через b_{ω_2} — столбец перехода в слово ω_2. Тогда вероятность встретить слово ω_2 на расстоянии n справа от слова ω_1 выражается формулой:

$$f_{\omega_1,\omega_2}(n) = a_{\omega_1} \cdot P^{n-2} \cdot b_{\omega_2}\,. \qquad (1)$$

Если же необходимо найти вероятность того, что на расстоянии n от ω_1 слово ω_2 встречается впервые, то формула (1) примет следующий вид:

$$f^*_{\omega_1,\omega_2}(n) = a_{\omega_1} \cdot P^{n-2}_{\omega_2} \cdot b_{\omega_2}\,, \qquad (2)$$

	$\tilde{f}^*_{\omega_1,\omega_2}(n)$		$f^*_{\omega_1,\omega_2}(n)$	
$\omega_1 - \omega_2$	A	λ	A	λ
AAA-TTT	13000	0,0192	10407	0,0273
ATT-TAT	7360	0,0157	7519	0,0198
ACT-AGT	4333	0,0143	6007	0,0152
TGC-GTG	4382	0,0121	7497	0,0133
CCC-GGG	1600	0,0078	2267	0,0077
GGT-CGG	1783	0,0091	1820	0,0091

Таблица 3: Значения коэффициентов в приближенной формуле $f^*_{\omega_1,\omega_2}(n) \approx A \cdot e^{-\lambda n}$ для 6 выбранных пар слов разной частоты встречаемости в последовательности генома мухи *Drosophila melanogaster*.

здесь P_{ω_2} — это матрица перехода P, в которой весь столбец перехода в слово ω_2 обнулён. Обращение всех вероятностей перехода в слово ω_2 в ноль запрещает переход в него, то есть обеспечивает то, что нигде раньше расстояния n слово ω_2 не встретится.

Понятно, что при $n \gg t$ теоретические значения функции $f_{\omega_1,\omega_2}(n)$ практически не зависят от расстояния n между словами, а зависят только от самих слов ω_1 и ω_2. График этой функции начиная с некоторого n будет представлять собой прямую, параллельную оси n. Теоретические значения второй функции $f^*_{\omega_1,\omega_2}(n)$ при больших расстояниях между словами зависят от n экспоненциально. То есть график такой функции начиная с некоторого n хорошо аппроксимируется функцией вида $A \cdot e^{-\lambda n}$. Данная зависимость подробно исследовалась в работе [3].

3 Материалы и методы

В данной работе был произведен сравнительный анализ теоретических значений функций $f_{\omega_1,\omega_2}(n)$ и $f^*_{\omega_1,\omega_2}(n)$ с данными вычислительного эксперимента, проведенного на конечных последовательностях геномов *Drosophila melanogaster*, *Homo sapiens*, *Caenorhabditis elegans* и *Schizosaccharomyces pombe*. Геномы представляют собой конечные, хотя и довольно длинные, символьные последовательности из того же 4-символьного алфавита $\aleph = \{A, C, G, T\}$, заданные однако процессом очевидно более сложным, чем Марковский процесс порядка $t = 3$.

Теоретические значения функции $f_{\omega_1,\omega_2}(n)$ при порядке Марковского процесса $t = k = 3$ для $n \gg t$ можно приближённо вычислить по формуле:

$$f_{\omega_1,\omega_2}(n) = f_{\omega_1,\omega_2} = p_{\omega_1} \cdot p_{\omega_2}\,, \qquad (3)$$

при условии, что известны вероятности появления p_{ω_1} и p_{ω_2} слов в последовательности; их в свою очередь легко вычислить по правилу $p_\omega = \frac{F_\omega}{N}$, зная частоту F_ω появления слова ω длины $k = 3$ и количество символов N в исследуемой последовательности.

Для вычисления теоретических значений функции (2) составлялась матрица перехода P для слов длины $k = 3$ по частотному словарю слов длины $k + 1 = 4$ для каждой исследуемой последовательности по следующему правилу:

$$p_{\omega_1\omega_2} = \frac{F_{\nu_1\nu_2\nu_3\nu_4}}{\sum\limits_{x\in\{A,C,G,T\}} F_{x\nu_1\nu_2\nu_3}}\,, \qquad (4)$$

где слово $\omega_1 = \nu_1\nu_2\nu_3$, $\omega_2 = \nu_2\nu_3\nu_4$ (слова пересекаются по символам ν_2 и ν_3, иначе из ω_1 невозможно перейти в ω_2, то есть $p_{\omega_1\omega_2} = 0$). Дальнейшие вычисления теоретических значений $f^*_{\omega_1,\omega_2}(n)$ производились в соответствии с формулой (2).

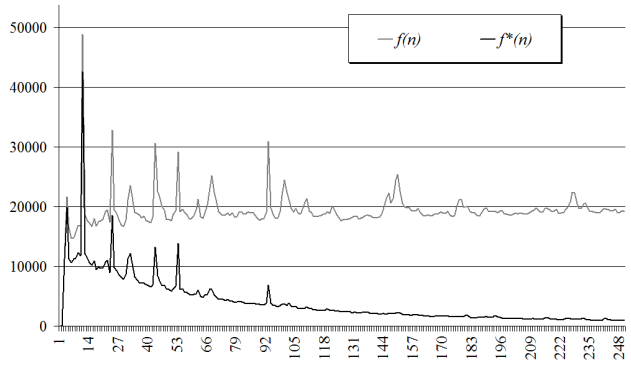


Рисунок 2: Значения функции $\tilde{f}_{\omega_1,\omega_2}(n)$ и $\tilde{f}^*_{\omega_1,\omega_2}(n)$, полученные в ходе вычислительного эксперимента для последовательности 14 хромосомы генома человека. Графики соответствуют выбранным парам слов $\omega_1 = CCC$, $\omega_2 = GGG$. По оси абсцисс откладывается расстояние n между словами.

В ходе эксперимента пришлось наложить ограничения на расстояние n между словами, выбранными в качестве соседей, в связи с ограниченными вычислительными возможностями. Исследовались значения функции распределения расстояний между ближайшими соседями для $n \leqslant 2000$.

Также для упрощения анализа полученного массива данных, из $m^{2k} = 4^6 = 4096$ всевозможных пар слов были выбраны 6 представителей по следующему принципу. Известно [4, 6, 5], что геномы большинства организмов различаются частотой нуклеотидов C и G в сравнении с нуклеотидами A и T. Поэтому на основе частотных словарей триплетов ($k = 3$) всех исследуемых последовательностей были выбраны 2 пары, состоящие из слов с относительно низкой (во всех исследуемых последовательностях) частотой встречаемости, 2 — со средней и 2 — с высокой частотой. К тому же одна из каждых двух пар представляет собой комплиментарный палиндром, а вторая — нет. Комплиментарным палиндромом называется пара слов, которые читаются одинаково в противоположных направлениях с учетом замены нуклеотидов по правилу Чаргаффа (например, AGT ⇔ ACT). Известно [2, 3, 1], что вероятности появления слов из такой пары в исходной последовательности примерно равны.

Выбранные пары, а также частоты соответствующих слов для последовательности генома 14 хромосомы

Комплементарные палиндромы

	$\omega_1 - \omega_2$	P_{ω_1}	P_{ω_2}	f_{ω_1,ω_2}
редкие	CCC-GGG	1147011	1150132	15130
средние	ACT-AGT	1403803	1414052	22767
частые	AAA-TTT	3323405	3392944	129326

Не комплементарные палиндромы

	$\omega_1 - \omega_2$	P_{ω_1}	P_{ω_2}	f_{ω_1,ω_2}
редкие	GGT-CGG	1020734	237639	2782
средние	TGC-GTG	1253556	1307919	18804
частые	ATT-TAT	2183691	1804494	45193

Таблица 4: Выбранные для анализа 6 пар слов $\omega_1 - \omega_2$ и их частоты в последовательности *Homo sapiens 14*, длина которой составляет 87191216 нуклеотидов. Частота встречаемости пары в последовательности f_{ω_1,ω_2} вычислена по формуле (3)

человека, представлены в таблице 4.

Чтобы вычислить экспериментальные значения $\tilde{f}_{\omega_1,\omega_2}(n)$ и $\tilde{f}^*_{\omega_1,\omega_2}(n)$ на конкретных последовательностях была написана программа, осуществляющая автоматизированный перебор всех встреченных пар слов на расстоянии n друг от друга.

Кроме того, изучалось распределение $n^{max}_{\omega_1,\omega_2}$ максимальных расстояний, на которых можно впервые встретить слово ω_2 справа от ω_1. Показано, что статистически достоверно эта величина заметно больше ожидаемой и достигает значения $\sim 10^4 \div 10^5$.

4 Результаты и обсуждение

Функции распределения $f_{\omega_1,\omega_2}(n)$ и $f^*_{\omega_1,\omega_2}(n)$ сильно зависят от частот тех слов ω_1 и ω_2, которые выбраны в качестве соседей. Из рисунка 3 видно, что, если приблизить значения функции распределения расстояний до ближайшего соседа экспонентой вида $\tilde{f}^*_{\omega_1,\omega_2}(n) \approx A \cdot e^{-\lambda n}$, то коэффициент затухания λ такой экспоненты тем выше, чем больше частота слов ω_1 и ω_2 в последовательности (см. таб. 3).

На рисунке 2 явно видна корреляция экспериментальных значений функций $\tilde{f}_{\omega_1,\omega_2}(n)$ и $\tilde{f}^*_{\omega_1,\omega_2}(n)$. Стоит заметить, что изучение этих отклонений теории от практики удобнее производить на основе $\tilde{f}_{\omega_1,\omega_2}(n)$, поскольку алгоритм вычисления теоретических и экспериментальных значений функции (1) проще, чем для функции (2).

Судя по рисунку 1, вероятность встретить слово ω_2 на расстоянии n справа от ω_1 действительно хорошо приближается формулой (3), то есть практически не зависит от расстояния n между словами. Это наблюдение также подтверждается таблицей 1.

Кроме того, в работе изучалось поведение второго распределения для самых дальних среди ближайших соседей. Показано, что на больших расстояниях $n > 1000$ это распределение уже не носит первоначального экспоненциального характера: вероятность

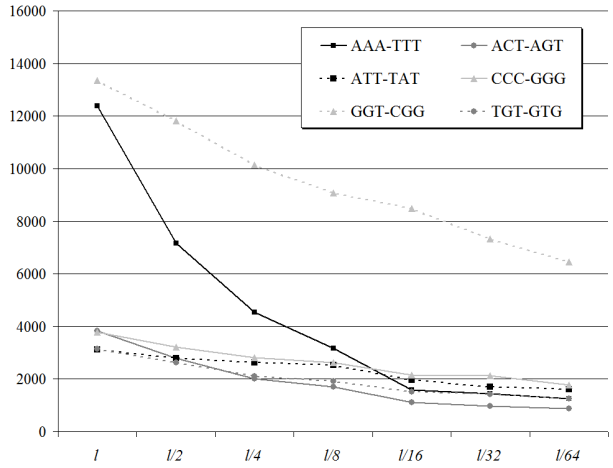


Рисунок 4: Зависимость максимального расстояния до ближайшего соседа $n^{max}_{\omega_1,\omega_2}$ от длины исходной последовательности. По оси абсцисс отмечены доли длины исходной последовательности генома *Homo sapiens 14 хромосома* $l = 87191216$ — полная длина. На оси ординат отмечены максимальное встреченное расстояние до ближайшие соседа для данной пары слов в таком участке последовательности.

существования ближайшего соседа на большом расстоянии существенно отличается от ожидаемой, оцениваемой по экспоненциальному распределению (т. н. “Long Tails”).

Анализ $n^{max}_{\omega_1,\omega_2}$ позволяет оценить внутривидовое сходство различных последовательностей. Для разных хромосом одного и того же генома выделим самые дальние и самые ближние максимальные расстояния $n^{max}_{\omega_1,\omega_2}$ среди всех пар слов ω_1 и ω_2. Введем коэффициент α, показывающий, насколько эти множества пар самых дальних $n^{max}_{\omega_1,\omega_2}$ пересекаются для разных хромосом одного генома. В работе был принят следующий способ определения α: например, рассматривая 3 хромосомы генома дрожжей *Schizosaccharomyces pombe*, отбираем для каждой из 3 последовательностей по 50 самых дальних пар; а затем подсчитываем, сколько из этих пар встретились в двух или трёх отобранных множествах — обозначим их число за M, а какие только в одном — N. Тогда $\alpha = \frac{M}{M+N}$. В таблице (2) представлены полученные результаты. Можно сделать вывод, что самые дальние максимальные расстояния действительно можно использовать для выявления внутривидового сходства; в то время как данные о самых ближних $n^{max}_{\omega_1,\omega_2}$ не слишком информативны.

Одной из важных задач здесь является определение характера зависимости наибольшего расстояния до ближайшего соседа в зависимости от длины самой последовательности. Для выявления такого ро-

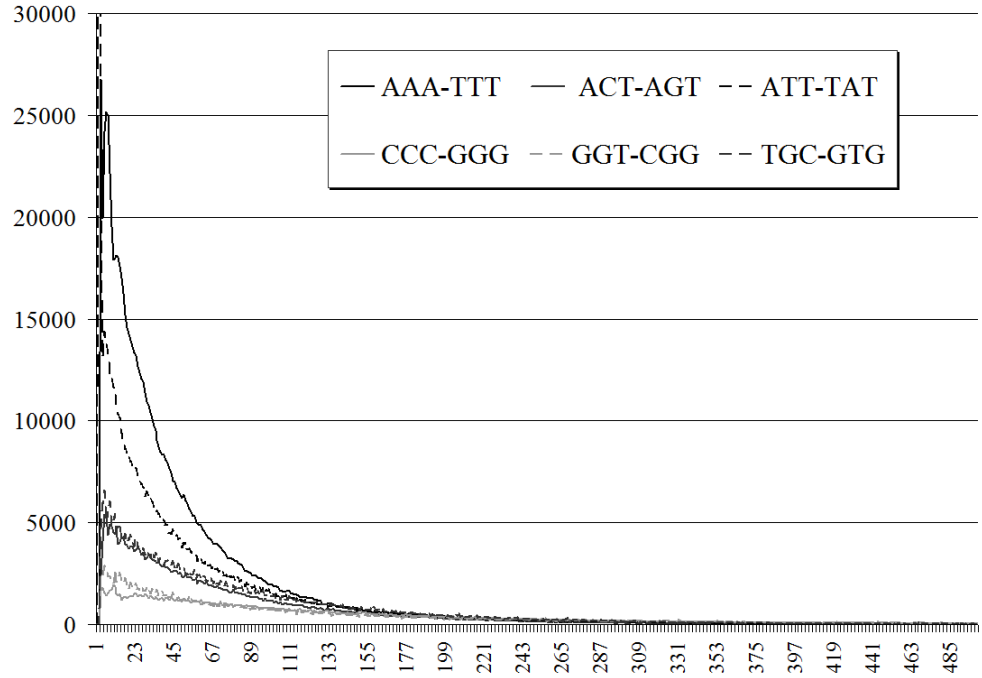


Рисунок 3: Вычисленные значения функции $\tilde{f}^{*}_{\omega_1,\omega_2}(n)$ для 6 выбранных пар слов разной частоты встречаемости в последовательности генома мухи *Drosophila melanogaster*.

да зависимости был проведён следующий вычислительный эксперимент. Значения $n^{max}_{\omega_1,\omega_2}$ определялись сперва для целой последовательности, затем для её фрагментов длиной в половину от исходной, затем для четвертей и так далее. Наименьший по длине фрагмент имел длину в $\frac{1}{64}$ от исходной. В качестве исходной последовательности бралась 14 хромосома человека.

Анализ $n^{max}_{\omega_1,\omega_2}$ в целой последовательности *Homo sapiens 14 chromosome* длиной l и отдельно в её участках длины $\frac{l}{2}, \frac{l}{4}, \frac{l}{8}, \frac{l}{16}, \ldots$ (см. рис. 4) показал, что зависимость самого дальнего ближайшего соседа $n^{max}_{\omega_1,\omega_2}$ от длины последовательности l можно грубо оценить как логарифмическую: $n^{max}_{\omega_1,\omega_2} \sim ln(l)$.

Список литературы

[1] Е. А. Темлякова, Ю. А. Путинцева и М. Г. Садовский. Правила Чаргаффа, симметрия геномов и эволюция. *Труды X Межд. ФАМ-конференции, Красноярск, изд-во СФУ, (Под ред. О. Ю. Воробьчва)*, pages 339–342, 2011.

[2] Н. А. Зайцева, Ю. А. Путинцева, М. Г. Садовский. К проблеме связи структуры и систематики на примере митохондриальных геномов. *Труды X Межд. ФАМ-конференции. Красноярск, изд-во СФУ, (Под ред. О. Ю. Воробьчва)*, pages 152–156, 2011.

[3] К. А. Клокова, М. Г. Садовский. О распределении расстояний до ближайшего соседа в Марковских цепях. *Труды XII ФАМЭБ. Красноярск: СФУ, (Под ред. О. Ю. Воробьчва)*, 2013.

[4] Varriale A.˙ Dna methylation, epigenetics, and evolution in vertebrates: Facts and challenges. *Int. J. Evol. Biol.*, 2014:ID 475981, 2014.

[5] R. Sánchez and R. Grau. An algebraic hypothesis about the primeval genetic code architecture. *Mathematical Biosciences*, 221:60–76, 2009.

[6] X. Liu, H. Diao, N. Nishi. Applied chemistry of natural dna. *Chem. Soc. Rev.*, 37(12):2745–2757, 2008.

Непрерывная стохастическая динамическая модель развития полисемии и смыслового объема ансамбля знаков естественного языка

Василий Васильевич Поддубный
Национальный исследовательский Томский государственный университет
Факультет информатики
Томск
vvpoddubny@gmail.com

Анатолий Анатольевич Поликарпов
Московский государственный университет им. М.В. Ломоносова
Филологический факультет
Москва
anatpoli@mail.ru

Аннотация. *Предлагается непрерывная стохастическая динамическая модель полисемической и объемно-смысловой эволюции знаков естественного языка. Модель основывается на предположении о диссипативном характере развития полисемии языкового знака. На основе этой модели выводятся теоретические законы одномоментных возрастно-полисемических и полисемических, возрастно-объемно-смысловых и объемно-смысловых, частотных и частотно-ранговых вероятностных распределений знаковых ансамблей. Проводится сравнение теоретических распределений с соответствующими эмпирическими полисемическими и частотными распределениями, полученными из представительных толковых и частотных словарей русского и английского языков.*

Ключевые слова. *Языковой знак, полисемия, смысловой объем, эволюция, стохастическая математическая модель, толковые и частотные словари, идентификация модели, вероятностные распределения.*

1 Введение

Предлагаемая математическая модель основывается на предположении о диссипативном характере развития полисемии языкового знака [1], [2], [3]. В упомянутых работах исследовался дискретный вариант этой модели методами имитационного моделирования. В данной работе исследуется непрерывный вариант этой модели аналитическими методами и предпринимается попытка теоретического вывода законов полисемического, возрастно-полисемического, частотного и частотно-возрастного распределений знаковых ансамблей.

2 Основные предположения модели

Основное предположение модели состоит в том, что каждый языковой знак в момент своего рождения обладает индивидуальным пределом G возможной многозначности, называемым ассоциативно-семантическим потенциалом (АСП), который постепенно реализуется (растрачивается) в процессе развития полисемии знака. При этом скорость образования у знака новых значений (как правило, все более абстрактных, чем позже они образуются) принимается пропорциональной еще неизрасходованной части АСП, вследствие чего прирост новых значений постепенно замедляется. Одновременно, но с запаздыванием на время τ_0, начинает протекать аналогичный и значительно более медленный процесс выпадения из употребления первоначальных значений знака (которые, как правило, относительно более конкретны и поэтому – относительно менее устойчивы, чем последующие значения).

Актуальная полисемия знака в любой момент времени t его жизненного цикла выражается разностью $x(t) = x_1(t) - x_2(t)$ процессов прибавки новых значений $x_1(t)$ и потери знаком ранее приобретенных значений $x_2(t)$. Непрерывная модель предполагает, что эти процессы непрерывны и подчиняются линейным дифференциальным уравнениям вида:

$$\frac{dx_1(t)}{dt} = \frac{1}{\tau_1}(G - x_1(t)),\ x_1(t_0) = 1,\ t \geq t_0; \quad (1)$$

$$\frac{dx_2(t)}{dt} = \frac{1}{\tau_2}(G - x_2(t)),\ x_2(t_0+\tau_0) = 0,\ t \geq t_0+\tau_0, \quad (2)$$

где $\tau_1 = a_1/G$, $\tau_2 = a_2/G$ – обратно пропорциональные АСП постоянные времени роста и спада полисемии с коэффициентами пропорциональности a_1 и a_2 соответственно, причем $\tau_1 \ll \tau_2$. Неравенство $\tau_1 < \tau_2$ требуется для обеспечения неотрицательности значений полисемии $x(t)$. Величина $x_1(t)$ характеризует максимальный уровень реализации АСП знака, достигнутый к моменту времени t. Величина $x_2(t)$ характеризует соответственно минимальный уровень АСП знака, оставшийся к моменту времени t.

Модель предполагает далее, что в знаковом ансамбле индивидуальные значения G и запаздывания τ_0 – независимые экспоненциально распределенные случайные величины с параметрами (математическими ожиданиями) $\langle G \rangle$ и $\langle \tau_0 \rangle$ соответственно.

Поток рождения знаков предполагается стационарным пуассоновским с интенсивностью $\lambda = 1/\langle \tau \rangle$, так

© 2014 В.В.Поддубный и А.А.Поликарпов

Олег Воробьев (ред.), Труды XIII ФАМЭМС'2014, Красноярск: СФУ

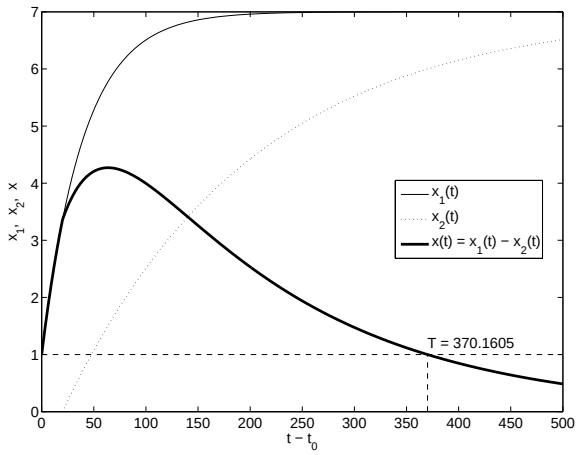


Рисунок 1: Типичная индивидуальная кривая развития полисемии ($G = 7$, $\tau_0 = 20$, $a_1 = 280$, $a_2 = 1260$), T – длительность жизни знака.

что интервалы времени между появлениями в потоке соседних знаков предполагаются независимыми экспоненциально распределенными случайными величинами с параметром (математическим ожиданием) $\langle\tau\rangle$. Моменты появления событий в стационарном пуассоновском потоке имеют равномерное распределение.

Индивидуальная кривая развития полисемии x во времени t (кривая полисемического жизненного цикла знака), зависящая от G и τ_0, получается из решения приведенных выше дифференциальных уравнений (1)-(4):

$$x(t) = (G - (G-1)e^{-G(t-t_0)/a_1}) \cdot 1(t \geq t_0) - \\ -G(1 - e^{-G(t-t_0-\tau_0)/a_2}) \cdot 1(t \geq t_0 + \tau_0),\ x(t) \leq G, \quad (3)$$

где $1(\cdot)$ – индикатор выполнения условия, заключенного в скобках (принимает значение 1, если условие выполняется, и значение 0, если иначе). Типичный ход этой кривой для отдельного языкового знака представлен на рис. 1. Прибавляемые значения знака начинаются с 1, так как каждый знак появляется в языке в каком-то одном определенном значении. Выпадающие значения начинаются с 0, так как рассматривается непрерывная модель развития полисемии, в которой полисемия изменяется непрерывно.

Как видно, кривая развития полисемии знака является унимодальной кривой с максимумом, смещенным к началу процесса эволюции знака. Полисемия знака сначала растет (с постоянным замедлением), а затем, после вступления в действие процесса выпадения из употребления начальных смысловых значений, достигает максимума и в дальнейшем убывает, пока не достигнет снова значения 1. Снижение полисемии до уровня, меньшего 1, означает, что знак теряет свое единственное смысловое значение и выходит из употребления. Это происходит в момент времени $t = t_0 + T$, где T – длительность жизни знака ($T \geq \tau_0$). Если $t \leq t_0 + T$, то разность $t - t_0$ определяет возраст знака на момент времени t. Найти аналитически длительность жизни T индивидуального языкового знака затруднительно. Она подчиняется аналитически неразрешимому трансцендентному уравнению:

$$(G-1)\exp(-GT/a_1) - G\exp(-G(T-\tau_0)/a_2) = 1.$$

В непрерывной модели можно предположить, что языковые знаки при достижении ими значений полисемии ниже 1 не выходят из употребления, а остаются в языке и живут с экспоненциально уменьшающейся до 0 полисемией при увеличении возраста за пределами T. Такое предположение не оказывает заметного влияния на получаемые результаты, но существенно упрощает вывод теоретических полисемических и возрастно-полисемических распределений. Поэтому в дальнейшем изложении мы не будем накладывать жестких ограничений на длительность жизни знака.

За начало отсчета времени удобно принять точку наблюдения, положив $t = 0$. Тогда придется отсчитывать моменты времени появления языковых знаков в потоке в обратном направлении. В результате будем иметь $t \leq 0$, а возраст знака станет равным $-t_0 \geq 0$. Обозначим теперь через t возраст знака. Формула (1) принимает более простой вид, выражающий зависимость величины полисемии знака возраста t от определяющих ее величин G и τ_0:

$$x(t) = (G - (G-1)e^{-Gt/a_1}) \cdot 1(t \geq 0) - \\ -G(1 - e^{-G(t-\tau_0)/a_2}) \cdot 1(t \geq \tau_0) = x(t, G, \tau_0) \leq G. \quad (4)$$

Поскольку те значения знака, которые соответствуют более высокому уровню реализации АСП, являются относительно более абстрактными, то они перекрывают более широкую смысловую область, имеют больший смысловой объем по сравнению со знаками более низкого уровня реализации АСП. А знаки, обладающие большим смысловым объемом, должны чаще употребляться в языковой коммуникации. Причем, можно предположить, что рост частоты употребления знака должен быть прямо пропорционален его смысловому объему. Это, в частности, должно проявляться и в том, что смысловой объем языкового знака должен быть связан с рангом знака в частотно-ранговом распределении знакового ансамбля.

Естественно предположить, что смысловой объем каждого из новых значений каждого из последующих уровней реализации АСП растет монотонно с ростом уровня реализации АСП. Такую зависимость можно представить степенной функцией этого уровня с показателем степени $\mu - 1 > 0$. Тогда суммарный смысловой объем v интервала уровней реализации

АСП от x_1 до x_2 выражается в непрерывной модели интегралом:

$$v = \mu \int_{x_1}^{x_2} x^{\mu-1} dx = x_2^{\mu} - x_1^{\mu},\ \mu > 1.$$

Следовательно, с учетом выражения (2), языковой знак возраста t с актуальной полисемией, определяемой разностью уровней семантической деривации $x_1(t)$ и $x_2(t)$, характеризуется смысловым объемом

$$\begin{gathered} v(t) = (G - (G-1)e^{-Gt/a_1})^{\mu} \cdot 1(t \geq 0) - \\ -(G(1 - e^{-G(t-\tau_0)/a_2}))^{\mu} \cdot 1(t \geq \tau_0) = \\ = v(t, G, \tau_0) \leq G^{\mu},\ \mu > 1. \end{gathered} \tag{5}$$

При $\mu = 1$ эта формула совпадает с формулой (2) для полисемии $x(t)$. Поэтому формулу (3) можно использовать как универсальную формулу, описывающую и полисемию знака (при $\mu = 1$), и его смысловой объем (при $\mu > 1$). А, следовательно, формулой (3) можно описывать и частотные характеристики языкового знака – при учете того, что смысловой его объем должен быть напрямую связан с частотой употребления знака в процессе языковой коммуникации.

3 Вывод возрастно-полисемических, полисемических, возрастно-объемно-смысловых, возрастно-частотных, частотных и частотно-ранговых распределений знаковых ансамблей

Условная плотность распределения величины v индивидуального знака в зависимости от времени t, ассоциативно-семантического потенциала G и запаздывания τ_0 представляется дельта-функцией Дирака:

$$p(v|t, G, \tau_0) = \delta(v - v(t, G, \tau_0)) \cdot 1(v \leq G). \tag{6}$$

С учетом (3) выражение (6) принимает вид:

$$\begin{gathered} p(v|t, G, \tau_0) = 1(v \leq G) \cdot \\ \cdot \delta(v - (G - (G-1)e^{-Gt/a_1})^{\mu} \cdot 1(t \geq 0) + \\ +(G - Ge^{-G(t-\tau_0)/a_2})^{\mu} \cdot 1(t \geq \tau_0)). \end{gathered} \tag{7}$$

Усредняя это распределение по распределениям τ_0 и G, можно получить условную плотность распределения $p(v|t)$ смыслового объема (и полисемии) знаков одного возраста. Интегрируя ее по v в пределах от k до $k+1$, можно получить распределение вероятностей $P(k|t)$ целочисленных значений смыслового объема (и полисемии) $k = 1, 2, \ldots$ знаков одного возраста. Удобнее, однако, вычислять не плотность распределения, а интегральную функцию распределения $F(v|t)$, используя ее затем для вычисления $P(k|t)$. Усреднение $p(v|t, G, \tau_0)$ по экспоненциальному распределению $p(\tau_0) = \exp(-\tau_0/\langle\tau_0\rangle$ запаздывания τ_0 на интервале $0 \leq \tau_0 < \infty$ проводится без затруднений и приводит к формуле:

$$\begin{gathered} p(v|t, G) = e^{-t/\langle\tau_0\rangle}\delta(v - v_1) + \frac{a_2 e^{-t/\langle\tau_0\rangle}}{G^2\langle\tau_0\rangle\mu} \cdot \\ \cdot \frac{1(v_1 - v_2 \leq v \leq v_1)}{(v_1 - v)^{1-1/\mu}(1 - (v_1 - v)^{1/\mu}/G)^{\frac{a_2}{G\langle\tau_0\rangle}+1}}, \end{gathered} \tag{8}$$

где обозначено:

$$\begin{gathered} v_1 = (G - (G-1)e^{-Gt/a_1})^{\mu}, \\ v_2 = (G - Ge^{-Gt/a_2})^{\mu}. \end{gathered} \tag{9}$$

Условная интегральная функция распределения $F(v|t, G)$ также вычисляется без труда:

$$\begin{gathered} F(v|t, G) = \int_0^v p(\theta|t, G)d\theta = 1(v \geq v_1 - v_2) - \\ -e^{-t/\langle\tau_0\rangle}\frac{1(v_1 - v_2 \leq v \leq v_1)}{(1 - (v_1 - v)^{1/\mu}/G)^{\frac{a_2}{G\langle\tau_0\rangle}}}, \end{gathered} \tag{10}$$

Нетрудно видеть, что $F(v|t, G) = 0$ при $v < v_1 - v_2$ и $F(v|t, G) = 1$ при $v > v_1$, а в интервале между этими значениями $F(v|t, G)$ монотонно возрастает от 0 до 1 с ростом v.

Формула (10) имеет сравнительно простой вид благодаря тому, что интегрирование ведется от 0, а не от 1, то есть в предположении, что знак может иметь полисемию, меньшую единицы, но при этом формально не выходить из употребления. Однако при интерпретации распределения вероятностей полисемии ансамбля знаков необходимо учитывать, что снижение уровня полисемии до величины, меньшей единицы, фактически означает выход знака из употребления.

К сожалению, усреднение по G этой условной функции распределения провести аналитически не удается. Функция распределения $F(v|t)$ остается в интегральной форме:

$$\begin{gathered} F(v|t) = 1 - \int_1^{\infty} 1(v \geq v_1 - v_2)e^{-(G-1)/\langle G\rangle}dG/\langle G\rangle - \\ -e^{-t/\langle\tau_0\rangle}\int_1^{\infty}\frac{1(v_1 - v_2 \leq v \leq v_1)e^{-(G-1)/\langle G\rangle}dG/\langle G\rangle}{(1 - (v_1 - v)^{1/\mu}/G)^{\frac{a_2}{G\langle\tau_0\rangle}}}, \end{gathered} \tag{11}$$

и может быть вычислена только численно. Разность $P(k|t) = F(k+1|t) - F(k|t)$, $k = 1, 2, \ldots$, при $\mu = 1$ определяет теоретическое возрастно-полисемическое распределение, а при $\mu > 1$ – возрастно-объемно-смысловое распределение знаков в языковом ансамбле.

Мы предполагаем пуассоновский характер потока рождения слов в языке. Интервалы времени τ_i

между соседними знаками в потоке являются при этом статистически независимыми экспоненциально распределенными случайными величинами с математическим ожиданием $\langle\tau\rangle$. Момент появления $t_{0i} = \sum_{j=1}^{i} \tau_j$ i-го знака определяется распределением Эрланга, соответствующим номеру знака в потоке. Можно показать, что среднее арифметическое распределений Эрланга при неограниченном увеличении числа слов в языке стремится к равномерному распределению на каждом интервале времени длиной $\langle\tau\rangle$. Таким образом, моменты t_{0i} появления языковых знаков можно считать реализациями равномерно распределенной случайной величины на каком угодно конечном интервале времени.

Плотность распределения полисемии, а, следовательно, и смыслового объема в любой момент времени в ансамбле n знаков выражается средним арифметическим плотностей распределения слагаемых, участвующих в усреднении, что при $n \to \infty$ и для принятого нулевого момента времени наблюдения эквивалентно усреднению распределения $F(v|t)$ по равномерному распределению t на всей полуоси возрастов:

$$F(v) = \lim_{T\to\infty} \int_0^T F(v|t)dt. \qquad (12)$$

Разность $P(k) = F(k+1) - F(k)$, $k = 1, 2, \ldots$, при $\mu = 1$ определяет теоретическое распределение полисемии ансамбля языковых знаков, а при $\mu > 1$ – теоретическое распределение их смысловых объемов. Распределение (12) является средневозрастным распределением, безусловным по отношению к возрасту.

В принятом здесь предположении частоты употребления знаков в актах коммуникации (в том числе, например, в определенном наборе текстов) пропорциональны их смысловым объемам. При этом целочисленные значения смысловых объемов знаков можно просто отождествить с абсолютными частотами употребления знаков в некотором коммуникационном поле, приняв коэффициент пропорциональности равным 1. Тогда теоретическое распределение $P(k)$ целочисленных значений k смыслового объема знаков (при $\mu > 1$) выразит долю знаков в словаре языка с абсолютной частотой k их употребления в актах коммуникации. Таким образом, $P(k)$ (при $\mu > 1$) становится теоретическим частотным распределением ансамбля знаков (долей знаков с частотой употребления k).

4 Идентификация модели

Для сопоставления теоретических распределений, полученных на основе предложенной непрерывной диссипативной стохастической динамической модели эволюции ансамбля языковых знаков, с реальными распределениями необходимо идентифицировать математическую модель с реальными эмпирическими данными, полученными из представительных толковых, частотных и исторических словарей исследуемого языка. Предлагаемая модель характеризуется шестью параметрами: $\langle G\rangle$, $\langle\tau\rangle$, $\langle\tau_0\rangle$, a_1, a_2, μ. Однако параметр $\langle\tau\rangle$ в явном виде не входит в полученные теоретические распределения, а параметр $\langle\tau_0\rangle$ можно положить равным 1, выбрав его в качестве единицы измерения времени и, следовательно, возраста. Остаются три параметра, определяющие форму вероятностного распределения полисемии: $\langle G\rangle$, a_1, a_2. Эти параметры можно определить по данным толковых словарей исследуемого языка, используя метод наименьших квадратов для подходящим образом выбранного критерия. Параметр μ можно определить затем также методом наименьших квадратов из сопоставления распределения смыслового объема знаков (теоретического частотного распределения) и эмпирического распределения частот употребления знаков языка по данным частотных словарей. Но можно определить сразу всю четверку параметров $\langle G\rangle$, a_1, a_2, μ методом наименьших квадратов, сопоставляя теоретическое и эмпирическое частотные распределения. В этом случае идентификация модели производится с использованием только объективного эмпирического частотного распределения. Можно использовать также взвешенный критерий, сопоставляющий все доступные теоретические и эмпирические распределения, как частотные, так и полисемические. Наконец, в случае необходимости, можно связать модельную единицу измерения времени с реальным масштабом времени, используя эмпирические данные о возрастном распределении знаков языка.

Для сопоставления теоретических и эмпирических полисемических и частотных распределений, изменяющихся в очень широких пределах (несколько порядков), удобно выбрать в качестве критерия близости распределений среднеквадратичное отклонение отношения их логарифмов от единицы:

$$\epsilon^2(q) = \sum_{k=1}^{k_{\max}} \left(\frac{\ln P_e(k)}{\ln P(k|q)} - 1\right)^2,\ q^* = \arg\min_q \epsilon^2(q), \quad (13)$$

где q^* – оценка вектора параметров $q = (\langle G\rangle, a_1, a_2)$, $q = \mu$ или $q = (\langle G\rangle, a_1, a_2, \mu)$, полученная по методу наименьших квадратов (13), $P(k|q)$ – теоретическое распределение, зависящее от вектора параметров q, $P_e(k)$ – эмпирическое распределение, полученное из соответствующего словаря.

Для идентификации модели и сравнения теоретических распределений с эмпирическими использовались представительные словари русского и английского языков. Для толкового и частотного словаря языка А.С. Пушкина (1958-1961) и толкового словаря Уэбстера (Webster's Collegiate 9th ed.) после идентификации модели получены значения параметров, представленные в табл. 1 (принято значение

Словари	$\langle G \rangle$	a_1	a_2	μ (для x)	μ (для v)
Пушкин	2.18	0.122	36.5	1	2.86
Уэбстер	3.28	6.12	183	1	–

Таблица 1: Параметры модели.

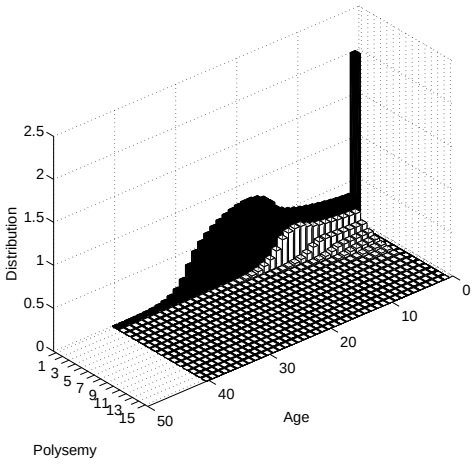


Рисунок 2: Теоретическое возрастно-полисемическое распределение с параметрами словаря Пушкина.

$\langle \tau_0 \rangle = 1$). Использовалась последовательная идентификация – сначала по параметрам $\langle G \rangle$, a_1, a_2, а затем по параметру μ.

5 Сравнительный анализ теоретических и эмпирических распределений

На рис. 2–3 представлены диаграммы вычисленных с использованием формулы (11) условных распределений $P(k|t)$ полисемии (рис. 2) и смыслового объема (рис. 3) ансамбля знаков в зависимости от их возраста при параметрах, соответствующих словарю языка Пушкина (возраст t знаков представляется в единицах $\langle \tau_0 \rangle$). Заметим, что аналогичный вид имеют и диаграммы распределений, вычисленных при параметрах, соответствующих словарю Уэбстера.

Из рис. 2 видно, что теоретически большая часть знаков неполисемична (имеет полисемию 1 – "дальняя черная стенка" на диаграмме). К этой группе относится и большинство знаков-долгожителей. Виден выход знаков преклонного возраста из употребления (знаки возраста, превышающего 40 единиц в принятой системе измерения времени, практически отсутствуют, их вероятность близка к нулю). Заметно меньше длительность жизни знаков с полисемией 2. Более высокая полисемия присуща знакам не слишком молодым и не слишком старым. Максимум полисемии приходится на знаки некоторого "зрелого возраста" – в районе нескольких единиц модельного времени.

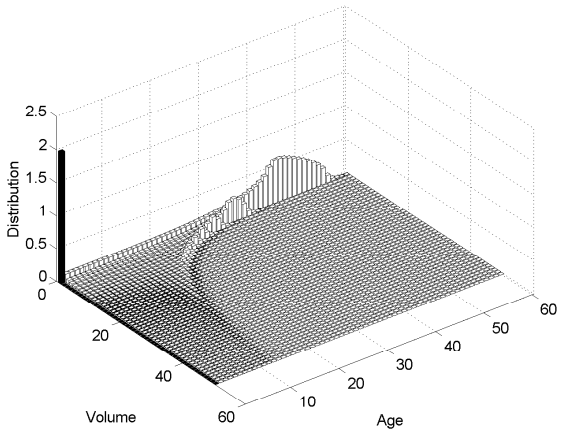


Рисунок 3: Теоретическое распределение смыслового объема ансамбля языковых знаков в зависимости от их возраста с параметрами словаря Пушкина.

На рис. 3 видно довольно любопытное теоретическое "двухребтовое" поведение распределения смыслового объема знаков в зависимости от возраста. Молодые знаки (возраста примерно до 10 единиц модельного времени) имеют распределение, медленно убывающее с ростом смыслового объема от "прямолинейного хребта" с $v = 1$. Знаки среднего возраста (от 10 до 40 единиц модельного времени) имеют двумодальное распределение смыслового объема (с максимумами на "прямолинейном хребте" при $v = 1$ и на "криволинейном хребте" при некотором $v > 1$, своем для каждого возраста, но тем меньшем, чем больше возраст). Знаки преклонного возраста (более 40 единиц модельного времени) имеют распределение смыслового объема, практически сосредоточенное на значении $v = 1$). Таким образом, малый смысловой объем теоретически присущ в основном знакам-долгожителям, тогда как наибольшим разнообразием смыслового объема теоретически обладают знаки сравнительно молодого и "зрелого" возраста.

На рис. 4–5 в билогарифмическом масштабе представлены теоретические и эмпирические полисемические распределения с параметрами словарей Пушкина и Уэбстера.

Видно хорошее согласие расчетных теоретических полисемических распределений с эмпирическими распределениями.

На рис. 6–7 в билогарифмическом масштабе представлены теоретическое распределение смыслового объема (оно же – частотное распределение знаков) и эмпирические частотное и частотно-ранговое распределения слов языка Пушкина.

Видно, что распределение смыслового объема хорошо согласуется с эмпирическими частотным и частотно-ранговым распределениями. Но распреде-

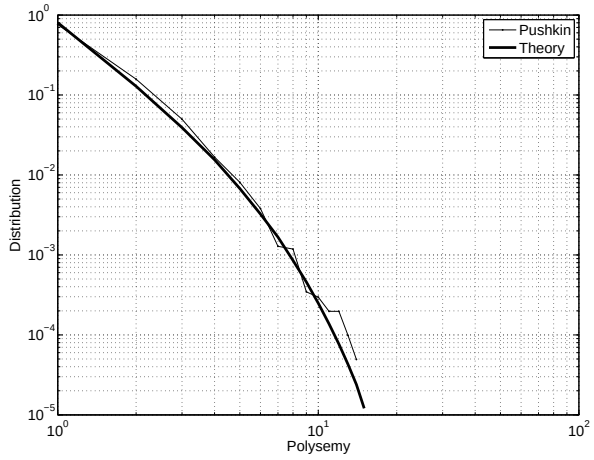


Рисунок 4: Теоретическое и эмпирическое полисемические распределения с параметрами словаря Пушкина.

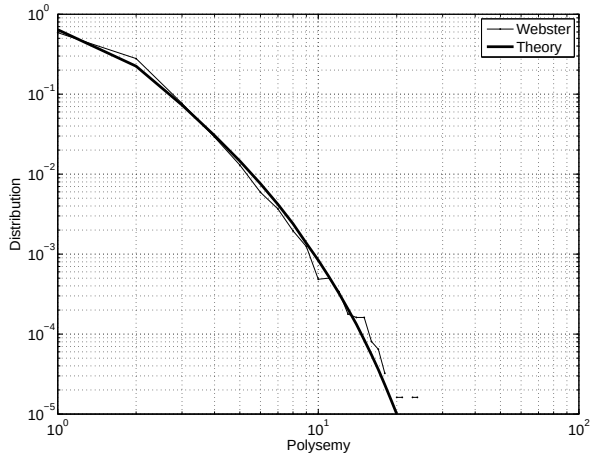


Рисунок 5: Теоретическое и эмпирическое полисемические распределения с параметрами словаря Уэбстера.

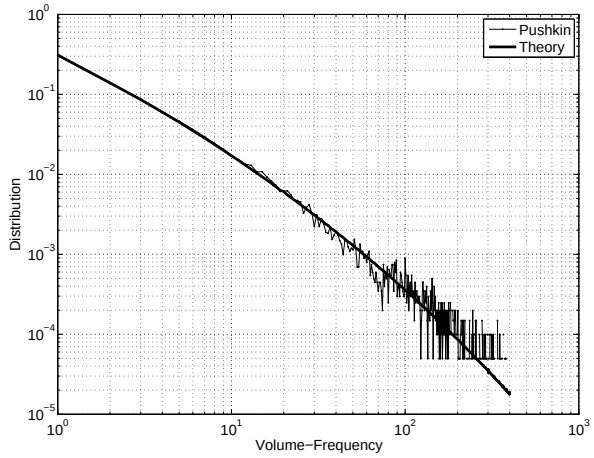


Рисунок 6: Теоретическое частотное распределение ансамбля знаков и эмпирическое частотное распределение слов языка Пушкина.

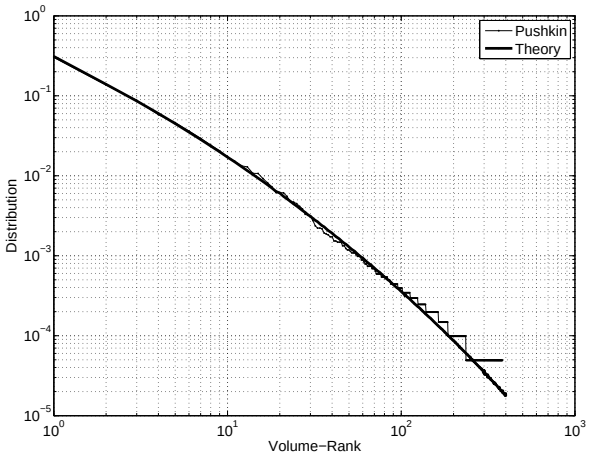


Рисунок 7: Теоретическое частотное распределение ансамбля знаков и эмпирическое частотно-ранговое распределение слов языка Пушкина.

ление целочисленных значений смыслового объема представляет собой теоретическое частотное распределение, совпадающее с теоретическим частотно-ранговым распределением. Это позволяет на основе выведенного распределения целочисленных значений смыслового объема теоретически рассчитывать и исследовать частотные и частотно-ранговые распределения знаковых ансамблей.

Благодарности

Публикация подготовлена в рамках научного проекта 14-14-70010, поддержанного РГНФ.

Список литературы

[1] В. В. Поддубный, А. А. Поликарпов. Диссипативная стохастическая динамическая модель эволюции языковых знаков. *Труды XIV Международной ЭМ конференции по эвентологической математике и смежным вопросам*, Красноярск: СФУ, 1 (под ред. О. Ю. Воробьева):182–186, 2010.

[2] В. В. Поддубный, А. А. Поликарпов. Диссипативная стохастическая динамическая модель развития языковых знаков. *Компьютерные исследования и моделирование*, 3(2):103–124, 2011.

[3] V. V. Poddubny, A. A. Polikarpov. Stochastic dynamic model of evolution of language sign ensembles. In R. Kohler I. Obradovic, E. Kelih, editor, *Methods and Applications of Quantitative Linguistics - Selected papers of the 8th International Conference on Quantitative Linguistics (QUALICO 2012)*, volume 1, pages 69–83. Academic Mind, Belgrade, Serbia, 2013.

Формирование объединенных портфелей инвесторов при различных критериях риска

Виген Геннадьевич Саркисов
Самарский государственный технический университет
Самара
vigen.sarkisov@mail.ru

Аннотация. *Исследуются новые возможности, возникающие при объединении портфелей инвесторов в один общий портфель и его последующей оптимизации. В качестве примера рассмотрено объединение портфелей в задаче Марковица. Сформулированы ограничения на критерии риска, при которых объединение целесообразно.*

Ключевые слова. *Инвестиционный портфель, доверительное управление, пул инвесторов.*

1 Введение

Множество, как классических [3], так и современных работ по управлению инвестиционными портфелями, посвящено оптимизации инвестиционного портфеля отдельного инвестора с учетом его предпочтений относительно доходности и риска. При формировании портфеля объединения (пула) инвесторов обычно стараются подобрать инвесторов с одинаковыми или близкими предпочтениями. Главным аргументом в пользу такого подхода является возможность максимально удовлетворить потребности всех инвесторов пула одновременно. Математический аппарат при этом полностью аналогичен используемому при формировании инвестиционного портфеля единичного инвестора.

В настоящей работе рассматривается возможность объединения в один пул инвесторов с различными предпочтениями и формирования для них одного общего портфеля. На первый взгляд, невозможно создать портфель, удовлетворяющий потребности всех инвесторов одновременно. Действительно, для каждого (или почти каждого) отдельного инвестора любой объединенный портфель окажется неоптимальным. Однако, для удовлетворения потребностей всех инвесторов совершенно нет необходимости создавать портфель, оптимальный одновременно для каждого из них. Достаточно сформировать портфель, который по доходности и риску не хуже, чем портфель, полученный простым объединением оптимальных портфелей всех инвесторов пула.

© 2014 В.Г.Саркисов

Олег Воробьев (ред.), Труды XIII ФАМЭМС'2014, Красноярск: СФУ

2 Основные понятия и определения

Обозначим множество портфелей через I, а множество допустимых структур портфелей x_i — через X ($x_i \in X, \forall i \in I$). Структуру инвестиционного портфеля, состоящего из N инвестиционных инструментов, будем описывать вектором-столбцом $x_i = (x_{i,1}, x_{i,2}, ..., x_{i,N})^T$.

Множество допустимых структур портфелей X описывается системой ограничений вида равенств и неравенств (1):

$$\begin{cases} v_k(x_i) = 0, & k = \overline{1,K} \\ w_l(x_i) \geq 0, & l = \overline{1,L}, \end{cases} \tag{1}$$

где v_k и w_l — линейные функции относительно долей инструментов в портфеле:

$$v_k(x_i) = \phi_{k,0} + \sum_{j=1}^{n} \phi_{k,j} x_{i,j} = \phi_{k,0} + \phi_k^T x_i, \tag{2}$$

$$w_l(x_i) = \psi_{k,0} + \sum_{j=1}^{n} \psi_{k,j} x_{i,j} = \psi_{l,0} + \psi_l^T x_i, \tag{3}$$

где $\phi_k^T = (\phi_{k,1}, \phi_{k,2}, ..., \phi_{k,N})$ и $\psi_l^T = (\psi_{l,1}, \psi_{l,2}, ..., \psi_{l,N})$ — вектора числовых коэффициентов линейных функций ($\phi_{k,1}, \psi_{l,j} \in R$).

Ограничения (1) с учетом (2) и (3) задают выпуклый многогранник в $(N-K)$–мерном пространстве.

Каждый из портфелей будем оценивать по двум критериям — средней доходности и риску. Доходности инструментов обозначим образуют вектор-столбец $r = (r_1, r_2, ..., r_N)^T$. Доходность r_p портфеля x_i:

$$r_p(x_i) = \sum_{j=1}^{n} r_j x_{i,j} = r^T x_i. \tag{4}$$

Будем полагать, что нам известна функция распределения вектора r. Обозначим вектор-столбец математических ожиданий доходностей через $m = (m_1, m_2, ..., m_n)^T$. Средняя доходность E_p портфеля x_i определяется математическим ожиданием доходности:

$$E_p(x_i) = \sum_{j=1}^{n} m_j x_{i,j} = m^T x_i. \tag{5}$$

Для оценки риска V_p существует множество критериев (обозначим множество возможных критериев через $\mathbf{V}$), большинство из которых являются функционалами от функции распределения доходности портфеля. Чаще всего в качестве критерия риска рассматривается дисперсия или квантиль доходности.

Недоминируемым будем называть такой портфель, что не существует другого портфеля, который был лучше него по одному из критериев и не хуже по другому. Множество недоминируемых портфелей обозначим $X^*(X^* \subset X)$. Множество X^* можно сформировать, находя портфель $x^* = (x_1^*, x_2^*, ..., x_N^*)$ с минимальным риском для каждого значения доходности E_0 с учетом (5), (2) и (3):

$$\begin{cases} x^* : V_p(x^*) \to \min\limits_{x^* \in X} \\ m^T x^* = E_0 \\ \phi_{k,0} + \phi_k^T x^* = 0, k = \overline{1,K} \\ \psi_{l,0} + \psi_l^T x^* \geq 0, l = \overline{1,L}. \end{cases} \tag{6}$$

Постановка задачи (6) является обобщением постановки классической задачи Марковица на случай произвольных линейных ограничений на доли инструментов в портфеле и произвольной меры риска.

3 Средняя доходность и риск объединенного портфеля пула инвесторов

Рассмотрим объединение оптимальных портфелей $x_1 \in X^*$ и $x_2 \in X^*$ двух инвесторов. Обозначим долю первого инвестора в общем портфеле через $a (a \in [0;1])$, тогда доля второго инвестора составит $1-a$. Структура объединенного портфеля x_{12}:

$$\begin{aligned} x_{12} &= ax_1 + (1-a)x_2 = \\ &= (ax_{1,1} + (1-a)x_{2,1}, ..., ax_{1,N} + (1-a)x_{2,N}). \end{aligned} \tag{7}$$

Так как портфели x_1 и x_2 являются допустимыми (удовлетворяющими ограничению (1)), то в силу линейности (2) и (3) портфель x_{12} также является допустимым. С практической точки зрения это означает, что формирование объединенного портфеля x_{12} пары инвесторов дает для каждого из них риск и доходность не хуже, чем при формировании портфелей Средняя доходность объединенного портфеля является взвешенным средним:

$$E_p(x_{12}) = aE_p(x_1) + (1-a)E_p(x_2). \tag{8}$$

В зависимости от структур портфелей x_1 и x_2 и выбранного критерия риска V портфель x_{12} может быть неоптимальным $(x_{12} \notin X^*)$. В этом случае существует оптимальный портфель $x_{12}^* \in X^*$, который является более предпочтительным, чем x_{12}. Формирование портфеля x_{12}^* вместо x_{12} (или вместо пары отдельных портфелей x_1 и x_2) дает возможность инвесторам получить лучшие значения показателей риска и доходности, чем они могли бы получить без объединения.

Интерес представляет нахождение подмножества $\mathbf{V}^+ \subset \mathbf{V}$ таких критериев риска, при которых возможно получение преимущества от объединения портфелей:

$$\begin{aligned} [V_p \in \mathbf{V}^+] \Leftrightarrow [\exists x_1, x_2 \in X^* : \\ V_p(x_{12}) > \min_{x \in X} V_p(x)|_{E_p(x)=E_p(x_{12})}] \end{aligned} \tag{9}$$

Более простой представляется задача поиска множества $\mathbf{V}^-$, являющегося дополнением для $\mathbf{V}^+$ $(\mathbf{V}^- = \mathbf{V} \setminus \mathbf{V}^+)$:

$$\begin{aligned} [V_p \in \mathbf{V}^-] \Leftrightarrow [\forall x_1, x_2 \in X^* : \\ V_p(x_{12}) = \min_{x \in X} V_p(x)|_{E_p(x)=E_p(x_{12})}] \end{aligned} \tag{10}$$

4 Необходимое условие существования возможности получения преимущества от объединения портфелей

Теорема 1. *Для неоптимальности в смысле (6) объединенного портфеля x_{12} пары инвесторов необходимо, чтобы зависимость хотя бы одной координаты вектора x^* от E_0 была нелинейной.*

Доказательство. Исходя из определения множества недоминируемых портфелей X^*, портфель x_{12} оптимален только при $x_{12} \in X^*$. Так как средняя доходность $E_p(x_{12})$ линейно зависит от долей инструментов в портфеле x_{12}, то она также линейно зависит и от долей инвесторов (a и $(1-a)$) в объединенном портфеле x_{12}. Следовательно, доли инвесторов в портфеле x_{12} линейно зависят от $E_p(x_{12})$. Если принять $E_0 = E_p(x_{12})$ и решить задачу (6), то полученный портфель x^* будет отличаться от x_{12} только в том случае, если хотя бы одна из координат вектора x^* зависит от E_0 нелинейно (в противном случае пришлось бы предположить существование двух различных прямых, соединяющих точки x_1 и x_2).

Как будет показано далее, в определенных условиях необходимое условие одновременно является и достаточным.

5 Объединение портфелей в задаче Марковица

Задача Марковица представляет собой частный случай задачи (6), где в качестве критерия риска V_p выбрана дисперсия доходности (для вычисления которой необходима матрица ковариаций доходностей C), доли всех инструментов в портфеле неотрицательны, а их сумма равна единице (отсутствует возможность

кредитования):

$$\begin{cases} x^* : V_p(x^*) = x^{*T} C x^* \to \min\limits_{x^* \in X} \\ m^T x^* = E_0 \\ \sum\limits_{j=1}^{N} x_j^* = 1 \\ x_j^* \geq 0, j = \overline{1,N}. \end{cases} \tag{11}$$

Исключим из постановки задачи (11) ограничения вида неравенств (снимем ограничение на неотрицательность долей инструментов в портфеле):

$$\begin{cases} x^* : V_p(x^*) = x^{*T} C x^* \to \min\limits_{x^* \in X} \\ m^T x^* = E_0 \\ \sum\limits_{j=1}^{N} x_j^* = 1. \end{cases} \tag{12}$$

Задача (12) содержит ограничения только вида равенств и может быть решена [1] при помощи метода множителей Лагранжа. Полученное при этом решение описывает линейные зависимости долей портфеля x^* от средней доходности E_0. Сформулированное в Теореме 1 необходимое условие существования возможности получения преимущества не выполняется. Следовательно, объединение любых двух портфелей x_1 и x_2 при ограничениях постановки задачи (12) дает недоминируемый (оптимальный) портфель $x_{12} = x^* \in X^*$.

Введение ограничений $x_j^* \geq 0, j = \overline{1,N}$ в постановку задачи (11) приводит к тому, что зависимость x^* от E_0 из линейной превращается в кусочно-линейную. Если портфели x_1 и x_2 принадлежат одному линейному участку, то портфель x_{12} вновь окажется оптимальным и дополнительное преимущество от объединения не возникнет. Если же x_1 и x_2 принадлежат разным линейным участкам, то можно строго доказать [2], что $x_{12} \notin X^*$. Формирование оптимизированного объединенного портфеля $x^* \in X^*$ позволит обоим инвесторам при сохранении доходностей портфелей x_1 и x_2 получить уровень риска более низкий, чем минимально возможный при отдельном формировании индивидуальных портфелей x_1 и x_2.

6 Применимость результатов для произвольного множества инвесторов

Выше рассматривалось объединение портфелей двух инвесторов с различными предпочтениями. Группы, состоящие из большего числа инвесторов с различными предпочтениями, также могут воспользоваться преимуществами оптимизации объединенных портфелей. Для подтверждения этого факта можно провести следующие рассуждения:

Из группы инвесторов выбирается такая пара (с портфелями x_1 и x_2), для которой объединенный портфель неоптимален ($x_{12} \notin X^*$). Полученный портфель x_{12} объединяется со следующим портфелем x_3. Последовательное присоединение очередных портфелей продолжается до тех пор, пока все портфели группы не объединятся в один общий портфель. Преимущество, полученное при оптимизации этого портфеля, может быть распределено между всеми инвесторами группы.

В общем случае распределение преимущества является вопросом договоренности участников группы. При формировании этой договоренности целесообразно учитывать, что наибольший вклад в возникновение преимущества вносят инвесторы с наиболее отличающимися от остальных предпочтениями.

7 Заключение

Проведенный анализ показал, что:

- Существуют критерии риска, при которых оптимальный для отдельного инвестора портфель $x_1 \in X^*$ может быть улучшен путем объединения его с оптимальным портфелем другого инвестора $x_2 \in X^*$ и последующей оптимизации суммарного портфеля. К таким критериям относится и дисперсия доходности (наиболее широко используемая мера риска).
- Дополнительная полезность может возникать только при нелинейном характере зависимости долей оптимального инвестиционного портфеля от математического ожидания доходности.
- Классический подход, при котором в пул объединяются инвесторы с близкими предпочтениями, менее эффективен, чем объединение инвесторов с существенно различающимися предпочтениями.

В то же время неразрешенным остался вопрос определения множества $\mathbf{V}^+$ критериев риска V_p, для которых возможно возникновение преимущества от оптимизации объединенного портфеля.

Интерес также представляет возможность распространения полученных результатов с двухкритериальной на многокритериальную задачу более общего вида.

Список литературы

[1] Д. В. Бейли У. Ф. Шарп, Г. Дж. Александер. *Инвестиции*. Инфра-М, М., 2003.

[2] В. Г. Саркисов. Система оптимального управления колективными инвестициями в модели Марковица. *Вестник Самарского государственного технического университета. Серия Технические науки*, 4(40):45–52, 2013.

[3] H. Markowitz. Portfolio selection. *The Journal Of Finance*, 1:77–91, 1952.

Нечеткая и сет-регрессионная модели распределения потребительских предпочтений между фирмами

Дарья Владиславовна Семенова
Сибирский федеральный университет
Институт математики и фундаментальной информатики
Красноярск
dariasdv@gmail.com

Анастасия Игоревна Иванова
Институт математики и фундаментальной информатики СФУ
Красноярск
nastya-18.1@mail.ru

Аннотация. *В работе проведено построение, исследование и сравнительный анализ нечетко-множественной модели И. Леунга и модели на основе случайно-множественной регрессии на примере задачи распределения потребительских предпочтений между фирмами, выпускающими косметическую продукцию.*

Ключевые слова. *Нечеткое бинарное отношение, случайное множество событий, сет-регрессия, сегментация рынка.*

Введение

Каждая фирма знает, что покупатель должен получать выгоду от приобретения товара. Тогда он будет приобретать товар еще и еще. Усилия прямого маркетинга направлены на то, чтобы увеличить знания покупателя об этих выгодах. Но какие выгоды ищут покупатели в товаре? Как понять потребности покупателей? И как много покупателей со схожими потребностями присутствует на рынке? На эти и многие другие вопросы о потребностях и желаниях покупателей дает ответы сегментирование.

Проблемы сегментирования рынка (выделение зон потребительских предпочтений) являются ключевыми при организации маркетинговой работы. Для того, чтобы добиться успеха фирме недостаточно производить продукцию высокого качества: необходимо знать и учитывать потребности покупателей, их желания, возможности. Поэтому проблема сегментации рынка является актуальной в современном мире.

Традиционные исследования торговой зоны часто опираются на понятие информации и ряд предположений об однородности рынка. Поведенческие постулаты часто оказываются или слишком упрощенными или несогласованными. Построенные модели обычно неадекватны рассматриваемым реальным ситуациям, характеризуемым неточной информацией, нечеткими процессами принятия решений. В работе проводится построение и сравнительный анализ двух математических моделей сегментации рынка косметической продукции: нечетко-множественной модели И. Леунга и модели на основе случайно-множественной регрессии.

© 2014 Д.В. Семенова, А.И. Иванова

Олег Воробьев (ред.), Труды XIII ФАМЭМС'2014, Красноярск: СФУ

1 Нечеткая модель распределения потребительских предпочтений между фирмами

1.1 Модель И.Леунга

В модели И. Леунга [5, 6] приняты следующие допущения: существование рынка; произвольная схема предпочтений населения; размещение конкурирующих товаров; фирма производитель одного типа; товары характеризуются признаками; степени важности признаков при принятии решения о покупке варьируются между индивидуумами; один товар предпочитается другому, если его признаки по своей степени важности более близки к оценке потребителя.

Пусть $\mathfrak{X} = \{x_1, x_2, \ldots, x_n\}$ — множество покупателей; $\mathfrak{Y} = \{y_1, y_2, \ldots, y_p\}$ — множество признаков товаров и $\mathfrak{Z} = \{z_1, z_2, \ldots, z_m\}$ — множество видов товаров.

Пусть $\Theta : \mathfrak{X} \times \mathfrak{Y} \to [0,1]$ есть функция принадлежности нечеткого бинарного отношения R [1, 6]. Для любого $x \in \mathfrak{X}$, $\forall y \in \mathfrak{Y}$ функция $\Theta_R(x,y)$ отражает степень важности признака y по оценке покупателя x при определении им предпочтения. Бинарное отношение R можно представить в следующем виде:

$$R = \begin{array}{c|ccc} & y_1 & \ldots & y_p \\ x_1 & \nu_R(x_1,y_1) & \ldots & \nu_R(x_1,y_p) \\ x_2 & \nu_R(x_2,y_1) & \ldots & \nu_R(x_2,y_p) \\ \vdots & \ldots & \ldots & \ldots \\ x_n & \nu_R(x_n,y_1) & \ldots & \nu_R(x_n,y_p) \end{array}. \quad (1)$$

Пусть $\Phi_S : \mathfrak{Y} \times \mathfrak{Z} \to [0,1]$ есть функция принадлежности нечеткого бинарного отношения S. $\forall y \in \mathfrak{Y}, \forall z \in \mathfrak{Z}$ функция $\Phi_S(y,z)$ отражает степени совместимости фирмы z с признаком y. Бинарное отношение S имеет вид:

$$S = \begin{array}{c|ccc} & z_1 & \ldots & z_m \\ y_1 & \nu_S(y_1,z_1) & \ldots & \nu_S(y_1,z_m) \\ y_2 & \nu_S(y_2,z_1) & \ldots & \nu_S(y_2,z_m) \\ \vdots & \ldots & \ldots & \ldots \\ y_p & \nu_S(y_p,z_1) & \ldots & \nu_S(y_p,z_m) \end{array}. \quad (2)$$

Построим бинарное отношение T:

$$T = \begin{array}{c|ccc} & z_1 & \dots & z_m \\ \hline x_1 & \mu_{A_1}(x_1,z_1) & \dots & \mu_{A_m}(x_1,z_m) \\ x_2 & \mu_{A_1}(x_2,z_1) & \dots & \mu_{A_m}(x_2,z_m) \\ \vdots & \dots & \dots & \dots \\ x_n & \mu_{A_1}(x_n,z_1) & \dots & \mu_{A_m}(x_n,z_m) \end{array}, \quad (3)$$

элементы которого определяются функцией принадлежности

$$\mu_{A_i}(x,z_i) = \frac{\sum\limits_y \nu_R(x,y)\cdot\nu_S(y,z_i)}{\sum\limits_y \nu_R(x,y)}, \quad (4)$$

для $\forall x \in \mathfrak{X},\ y \in \mathfrak{Y},\ z \in \mathfrak{Z},\ i = \overline{1,m}$. Сумма $\sum\limits_y \nu_R(x,y)$ равна степени нечеткого подмножества, указывающей число важнейших признаков y, которое потребитель x использует для оценки вида товара, а $\mu_{A_i}(x,z_i)$ можно интерпретировать как взвешенную степень предпочтения товара вида z_i индивидуумом x. Функция предпочтения, описываемая уравнением (4), удовлетворяет определению выпуклого нечеткого подмножества [1, 5]:

$$\mu_{A_i}[\lambda(x_1,z_i)+(1-\lambda)(x_2,z_i)] \geq$$

$$\geq \min[\mu_{A_i}(x_1,z_i),\mu_{A_i}(x_2,z_i)],$$

$$\forall x_1,x_2 \in V,\ \forall z_i \in Z,\ \forall\lambda \in [0,1].$$

Допущение выпуклости оправдано в том смысле, что при нем должно предполагаться монотонное увеличение значения функции принадлежности с уменьшением расстояния до фирмы. Поскольку все $\mu_{A_i}(x,z_i)$ выпуклые, их пересечения также выпуклые функции. Таким образом, можно построить матрицу D, элементы которой

$$d_{ij} = \mu_{A_j}(x_i,z_j) \wedge \mu_{A_{j+1}}(x_i,z_{j+1}), i = \overline{1,n}, j = \overline{1,m-1}$$

обозначают минимальное перекрытие торговых зон фирм.

Определим способ, с помощью которого рынок можно разделить между фирмами. Поскольку нечеткие подмножества $A_i,\ i = \overline{1,m}$ ограничены максимальными степенями $\sup\limits_x \mu_{A_i}(x),\ i = \overline{1,m}$, то и их пересечение $A_i \cap A_j,\ i,j = \overline{1,m}$ также ограничено выпуклым нечетким подмножеством и определяется функцией принадлежности, принимающей максимальное значение $\sup\limits_x \mu_{A_i\cap A_j}(x)$. Применяя теорему об отделимости [5], получаем, что наивысшая степень разделения торговой зоны, равна $1 - \sup\limits_x \mu_{A_i\cap A_j}(x)$.

В классической модели фирма монополизирует рынок. Однако, модель сбыта с доминированием более адекватна для описания торговых зон реального мира. Из этой идеи следует, что перекрытие торговой зоны скорее общий феномен, чем исключение. Для перекрытия торговых зон фирм можно использовать понятие порога разделимости. В данной модели порог разделения торговой зоны может быть ограничен условием[5]

$$l < \min_{ij}\max_x \min[\mu_{A_i}(x,z_i),\mu_{A_j}(x,z_j)] = \sup_x \mu_{A_i\cap A_j}(x),$$

где $x \in \mathfrak{X},\ z \in \mathfrak{Z},\ i,j = \overline{1,m}$.

Таким образом, для выбранного порога l торговая зона M_i фирмы определяется нечетким подмножеством уровня l. Выбирая различные значения для l, можно получить различные торговые зоны. Общее правило состоит в том, чтобы выбрать наибольшее возможное значение [5] l, меньшее $\max\limits_x \min[\mu_{A_i}(x,z_i),\mu_{A_j}(x,z_j)]$:

$$M_i = \{x|\mu_{A_i}(x) \geq \min_{ij}\max_x \min[\mu_{A_i}(x,z_i),\mu_{A_j}(x,z_j)]\}.$$

1.2 Нечеткая модель потребительских предпочтений между фирмами

На основе модели Леунга построена модель потребительских предпочтений между фирмами [4], выпускающими косметическую продукцию «Тушь для ресниц». В таблице 1 представлены признаки потребителей, участвовавших в опросе.

Таблица 1: Признаки потребителей

покупка туши	возраст покупателя	доход
1 раз в месяц (a_1)	от 18 до 25 лет (b_1)	Высокий: св. 25000 (c_1)
раз в 3 месяца (a_2)	от 26 до 38 лет (b_2)	Выше среднего: 15000-25000 (c_2)
раз в полгода (a_3)	от 39 до 45 лет (b_3)	Низкий: 6000-10000 (c_3)
		Очень низкий: 2000-6000 (c_4)

Из всего множества покупателей в результате опроса выделено следующие 12 групп:

$$\begin{array}{ll} x_1 = (a_1,b_3,c_1); & x_2 = (a_1,b_1,c_3); \\ x_3 = (a_2,b_1,c_1); & x_4 = (a_2,b_3,c_1); \\ x_5 = (a_2,b_1,c_2); & x_6 = (a_2,b_2,c_2); \\ x_7 = (a_2,b_3,c_2); & x_8 = (a_2,b_1,c_4); \\ x_9 = (a_2,b_3,c_4); & x_{10} = (a_3,b_1,c_1); \\ x_{11} = (a_3,b_2,c_3); & x_{12} = (a_3,b_2,c_4). \end{array} \quad (5)$$

В работе рассмотрены следующее множество признаков фирм: доступность; высокое качество; низкая цена; известность производителя.

Для анализа взяты фирмы: «Loreal», «Max Factor», «Bourjois», «Maybelline».

После обработки данных получены следующие результаты:

$R =$

	y_1	y_2	y_3	y_4
x_1	0.053	0.027	0.04	0.067
x_2	0.047	0.04	0.047	0.047
x_3	0.04	0.04	0.04	0.047
x_4	0.047	0.04	0.027	0.047
x_5	0.047	0.04	0.04	0.047
x_6	0.02	0.027	0.02	0.013
x_7	0.033	0.04	0.02	0.027
x_8	0.06	0.033	0.067	0.06
x_9	0.013	0.033	0.027	0.02
x_{10}	0.04	0.02	0.027	0.027
x_{11}	0.047	0.067	0.053	0.04
x_{12}	0.013	0.02	0.027	0.027

В матрице R элементы каждой строки выражают относительные степени важности признаков в принятии покупателем решения о поездки. Чем выше значение, тем более важен признак.

Элементы каждого столбца матрицы S представляют степень принадлежности или совместимости вида туши с соответствующим признаками, то есть они выражают соотношение качества признака к каждому товару.

$S =$

	z_1	z_2	z_3	z_4
y_1	0.233	0.233	0.28	0.213
y_2	0.193	0.233	0.207	0.213
y_3	0.213	0.24	0.22	0.193
y_4	0.213	0.233	0.233	0.213

Из нечетких бинарных отношений R и S получаем нечеткое бинарное отношение T:

$T =$

	z_1	z_2	z_3	z_4
x_1	0.216	0.234	0.24	0.209
x_2	0.214	0.235	0.236	0.208
x_3	0.213	0.235	0.235	0.208
x_4	0.214	0.234	0.238	0.21
x_5	0.214	0.235	0.237	0.208
x_6	0.211	0.235	0.233	0.208
x_7	0.212	0.234	0.235	0.21
x_8	0.215	0.235	0.238	0.207
x_9	0.209	0.235	0.227	0.207
x_{10}	0.217	0.235	0.242	0.208
x_{11}	0.211	0.235	0.232	0.208
x_{12}	0.211	0.235	0.23	0.207

Элементы матрицы D обозначают минимальное перекрытие торговых зон фирм между всеми рассматриваемыми фирмами.

$D =$

0.216	0.216	0.209	0.234	0.209	0.209
0.214	0.214	0.208	0.235	0.208	0.208
0.213	0.213	0.208	0.235	0.208	0.208
0.214	0.214	0.21	0.234	0.21	0.21
0.214	0.214	0.208	0.235	0.208	0.208
0.211	0.211	0.208	0.233	0.208	0.208
0.212	0.212	0.21	0.234	0.21	0.21
0.215	0.215	0.207	0.235	0.207	0.207
0.209	0.209	0.207	0.227	0.207	0.207
0.217	0.217	0.208	0.235	0.208	0.208
0.211	0.211	0.208	0.232	0.208	0.208
0.211	0.211	0.207	0.23	0.207	0.207

Найдем порог разделения.

$$\max_x \min[\mu_{A_1}(x, z_1), \mu_{A_2}(x, z_2)] = 0.217;$$
$$\max_x \min[\mu_{A_1}(x, z_1), \mu_{A_3}(x, z_3)] = 0.217;$$
$$\max_x \min[\mu_{A_1}(x, z_1), \mu_{A_4}(x, z_4)] = 0.21;$$
$$\max_x \min[\mu_{A_2}(x, z_2), \mu_{A_3}(x, z_3)] = 0.235;$$
$$\max_x \min[\mu_{A_2}(x, z_2), \mu_{A_4}(x, z_4)] = 0.21;$$
$$\max_x \min[\mu_{A_3}(x, z_3), \mu_{A_4}(x, z_4)] = 0.21;$$

Итак, 0.21 — минимальная из подсчитанных величин. Теперь из матрицы T выбираем для l, наибольшее возможное значение, которое было бы меньше 0.21, получаем $l = 0.209$. Применяя это значение в качестве порога разделения, определяем следующие торговые зоны

$$M_1 = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_{10}, x_{11}, x_{12}\};$$
$$M_2 = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}\};$$
$$M_3 = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}\};$$
$$M_4 = \{x_4, x_7\}.$$

Из полученных результатов видно, например, что

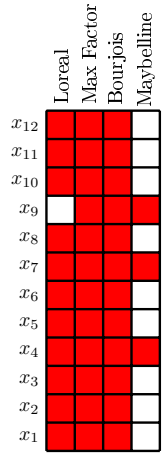


Рисунок 1: Графическое представление полученных торговых зон: по горизонтали — названия фирм, по вертикали — группы покупателей, красным цветом выделены покупатели, отдающие предпочтение той или иной фирме

фирмы «Max Factor» и «Bourjois» схожи по своим торговым зонам. Фирма «Maybelline» подходит группе покупателям с доходом выше среднего.

В рассмотренной модели такие допущения, как об одинаковом достоинстве фирм, заменяются нечетким восприятием привлекательности фирм относительно различных характерных свойств. Предпочтение, отдаваемое потребителями той или иной фирме, представляется в виде выпуклого нечеткого подмножества для исследования перекрытия торговых зон. Устанавливается порог разделения.

Отметим, что в случае максимальной неопределенности (все отношения равновозможны) получаем максимально неопределенный ответ модели. Модель является чувствительной к неполноте и к шуму в исходных данных. В [8] показано, что анализируемая модель не позволяет гарантировать качество результата как функцию качества исходных данных, что требует осторожности при ее практическом применении.

2 Сет-регрессионная модель распределения потребительских предпочтений между фирмами

Случайное множество событий [7] — это случайный элемент со значениями из множества всех подмножеств конечного множества выделенных событий $\mathfrak{X}$. Основная идея современной теории случайных множеств состоит в том, что структура статистических взаимозависимостей подмножества конечного множества полностью определяется распределением случайного множества, заданного на множестве всех его подмножеств. Распределение случайного множества [7] — это удобный математический аппарат для описания всех способов взаимодействия элементов между собой. Если дано совместное распределение двух случайных множеств событий K и L , значения которых содержаться в конечных множествах $\mathfrak{X}$ и $\mathcal{Y}$ соответственно, то регрессией K на L называется любой оператор Ψ, приближенно представляющий статистическую зависимость K от L [7, 2, 3] . Для определения зависимости между двумя случайными множествами событий в [2, 3] было предложено использовать сет-регрессию, которая устанавливает вид средней сет функциональной зависимости между этими двумя случайными множествами событий.

Задача сет-регрессии [2, 3]. *Пусть задано случайное множество событий K под $\mathfrak{X}$, т.е.*

$$K : (\Omega, \mathcal{A}, \mathbf{P}) \to (2^{\mathfrak{X}}, 2^{2^{\mathfrak{X}}}),$$

и задано случайное множество событий L под $\mathfrak{Y}$, т.е.

$$L : (\Omega, \mathcal{A}, \mathbf{P}) \to (2^{\mathfrak{Y}}, 2^{2^{\mathfrak{Y}}}).$$

Известно совместное распределение случайных множеств событий K и L под $\mathcal{Z} = \mathfrak{X} + \mathfrak{Y}$

$$\mathbf{P}(S = Z) = \mathbf{P}(K = X, L = Y),$$

где

$$S : (\Omega, \mathcal{A}, \mathbf{P}) \to (2^{\mathfrak{X}+\mathfrak{Y}}, 2^{2^{\mathfrak{X}+\mathfrak{Y}}}).$$

Пусть

$$\rho : 2^{\mathcal{Y}} \to 2^{\mathcal{Y}}$$

— некоторая метрика в $\mathcal{Y}$. Необходимо найти сет-функцию

$$\varphi : 2^{\mathfrak{X}} \to 2^{\mathcal{Y}},$$

которая доставляет минимум функционалу

$$\mathbf{E}\rho(L, \varphi(K)),$$

где $\mathbf{E}$ оператор математического ожидания.

Решение задачи сет-регрессии можно искать в виде [2, 3]

- условного сет-квантиля порядка α

$$\mathbf{Q}_\alpha \left(L = Y \,|_{K=X} \right) = \left\{ y : \mathbf{P}\left(Y \,\big|_{\mathrm{ter}(X)} \right) \geq \alpha \right\};$$

- условной сет-медианы

$$\mathrm{Med}\left(L = Y \,|_{K=X} \right) = \left\{ y : \mathbf{P}\left(Y \,\big|_{\mathrm{ter}(X)} \right) \geq \frac{1}{2} \right\};$$

- условной сет-моды

$$\mathrm{Mod}\left(L = Y \,|_{K=X} \right) = \max_{X \subseteq \mathfrak{X}} \mathbf{P}\left(L = Y \,|_{K=X} \right),$$

где $\mathrm{ter}(X) = \left(\bigcap\limits_{x \in X} \right) \bigcap \left(\bigcap\limits_{x \in \mathfrak{X} \setminus X} x^c \right)$.

2.1 Постановка задачи

Пусть дано

- множество покупателей

$$\mathfrak{X} = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}\},$$

где $x_i, i = \overline{1,12}$ определяются так же как и в нечеткой модели (5)

- множество «Признаков товара»

$$\mathfrak{Y} = \{a, b, c, d\} =$$

$$= \{\text{«Доступность»}, \text{«Высокое качество»},$$

«Низкая цена», «Известность производителя»};

- Множество «Фирм» $\mathfrak{Z} = \{\alpha, \beta, \gamma, \eta\} = \{$«Loreal», «Max Factor», «Bourjois», «Maybelline»$\}$

И пусть известны совместное распределение

$$\{p(X,Y) = \mathbf{P}(K = X, L = Y), X \subseteq \mathfrak{X}, Y \subseteq \mathfrak{Y}\}$$

двух случайных множеств событий K и L, значения которых содержаться в конечных множествах $\mathfrak{X}$ и $\mathfrak{Y}$ соответственно; совместное распределение

$$\{p(Y,Z) = \mathbf{P}(L = Y, M = Z), Y \subseteq \mathfrak{Y}, Z \subseteq \mathfrak{Z}\}$$

двух случайных множеств событий L и M, значения которых содержаться в конечных множествах $\mathfrak{Y}$ и $\mathfrak{Z}$ соответственно.

Требуется найти некоторую функцию множеств $\Psi : 2^{\mathfrak{X}} \longrightarrow 2^{\mathfrak{Z}}$, описывающую распределение потребительских предпочтений между фирмами, выпускающими косметическую продукцию.

2.2 Решение задачи

Решение состоит из трех этапов:

1. На первом этапе необходимо найти решение задачи сет-регрессии для множеств $\mathfrak{X}$ и $\mathfrak{Y}$ в виде некоторого сет-среднего $\Theta\left(L = Y\,|_{K=X}\right)$, $X \subseteq \mathfrak{X}$, $Y \subseteq \mathfrak{Y}$.

   На рис. 2 представлено решение задачи сет-регрессии множества $\mathfrak{Y}$ на $\mathfrak{X}$ в виде сет-квантиля с уровнем $\alpha = 2/5$.

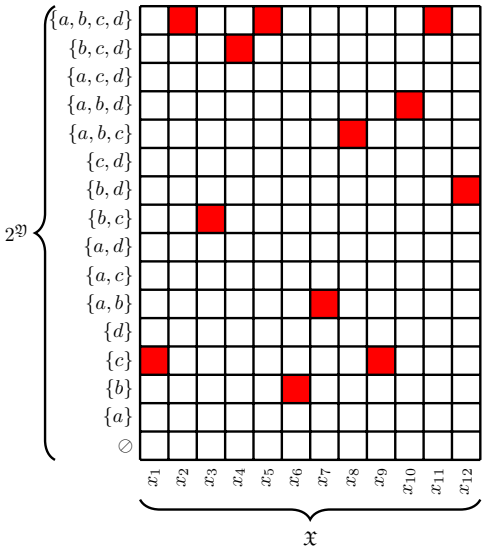


Рисунок 2: Решение задачи сет-регрессии множества $\mathfrak{Y}$ на $\mathfrak{X}$ в виде сет-квантиля с уровнем $\alpha = 2/5$.

2. На втором этапе необходимо найти решение задачи сет-регрессии для множеств $\mathfrak{Y}$ и $\mathfrak{Z}$ в виде некоторого сет-среднего $\Phi\left(M = Z\,|_{L=Y}\right)$, $Y \subseteq \mathfrak{Y}$, $Z \subseteq \mathfrak{Z}$.

   На рис. 3 представлено решение задачи сет-регрессии множества $\mathfrak{Z}$ на $\mathfrak{Y}$ в виде сет-квантиля с уровнем $\alpha = 2/5$.

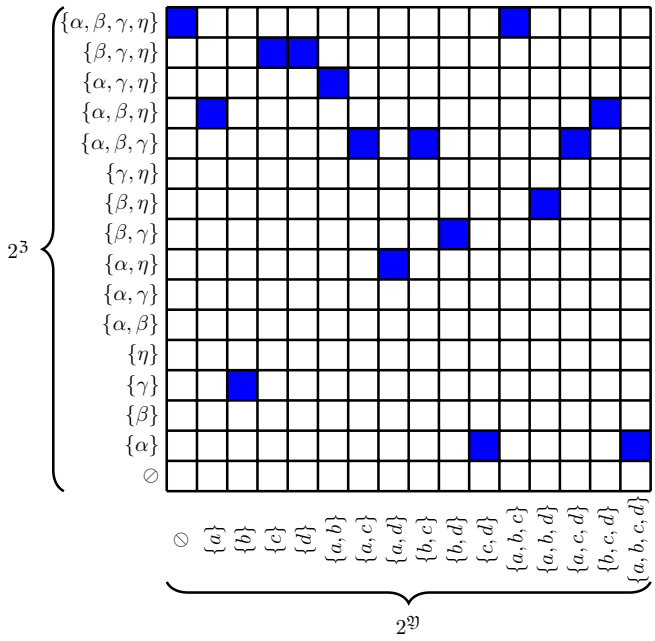


Рисунок 3: Решение задачи сет-регрессии множества $\mathfrak{Z}$ на $\mathfrak{Y}$ в виде сет-квантиля с уровнем $\alpha = 2/5$.

3. Решение задачи определяется как суперпозиция

$$\Psi(X) = \Phi\left(M = Z\,|_{\Theta(L=Y|_{K=X})}\right).$$

   На рис. 4 дано графическое представление суперпозиции сет-функций. На рис. 13 представлено решение задачи — сет-функция $\Psi(X)$, $X \subseteq \mathfrak{X}$. Окончательное решение задачи представлено на рис. 14

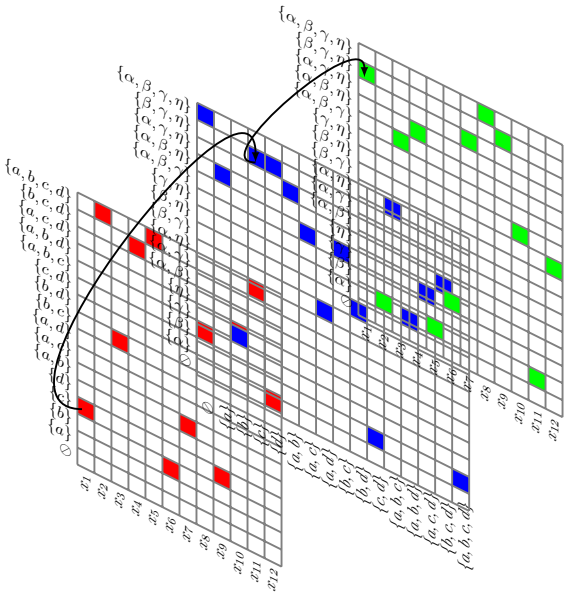


Рисунок 4: Суперпозиция сет-функций.

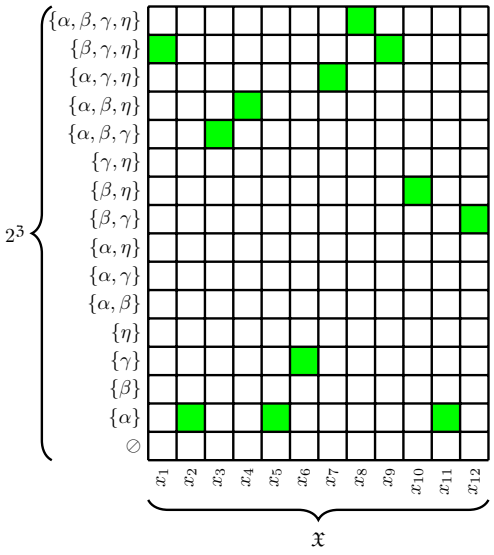


Рисунок 5: График сет-фукции $\Psi(X)$, $X \subseteq \mathfrak{X}$.

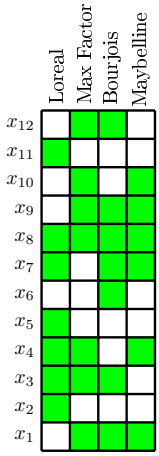


Рисунок 6: Графическое представление полученных торговых зон: по горизонтали — названия фирм, по вертикали — группы покупателей, красным цветом выделены покупатели, отдающие предпочтение той или иной фирме

Аналогично можно построить регрессионные отображения случайных множеств событий на основе других сет-средних [7, 2, 3].

Данная модель позволяет учитывать полную структуру взаимосвязей событий, в то время как нечеткая модель учитывает только лишь парные связи.

Основная сложность применения сет-регрессионной модели заключается в оценке совместных распределений случайных множеств, т.к. как в общем случае совместное распределение случайных множеств событий определяется $2^N \times 2^M$ параметрами, где N и M — мощности соответствующих избранных множеств событий. В работе [2] изложены методы преодоления проблемы размерности.

Благодарности

Авторы выражают искреннюю признательность проф. О.Ю. Воробьёву, плодотворное сотрудничество с которым явилось причиной появления основных результатов этой работы.

Список литературы

[1] С.Л. Блюмин, И.А. Шуйкова, П.В. Сараев, И.В. Черпаков. *Нечеткая логика: алгебраические основы и приложения.* Липецк: ЛЭГИ, 2002.

[2] О.Ю. Воробьев, А.Ю. Фомин. *Сет-регрессионный анализ зависимостей событий в статистических системах.* Красноярск: ИВМ СО РАН, КрасГУ, 2004.

[3] О. Ю. Тарасова. *Сеточные и регрессионные алгоритмы аппроксимации сложных систем событий: автореф. дис. на соиск. учен. степ. канд. физ.-мат. н.* Красноярск: СФУ, 2007.

[4] Семенова Д.В., Иванова А.И. Сравнение нечеткого и событийного подхода в задаче сегментации рынка. In *«Новые информационные технологии в исследовании сложных структур»: материалы Десятой российской конференции с международным участием. — Томск: Издательский Дом Томского государственного университета*, 2014.

[5] И. Леунг. Разделение на торговые зоны в нечетких условиях. In *Нечеткие множества и теория возможностей. Последние достижения / Под ред. Ягера Р.Р.* М.: Радио и связь, 1986.

[6] Семенова Д.В. *Нечеткие множества: теория и практика: учеб.пособие.* Красноярск: Краснояр. гос.ун-т., 2006.

[7] Воробьев О.Ю. *Эвентология.* Красноярск: Сибирский федеральный университет, 2007.

[8] Тимирова А.Н. Исследование нечеткой устойчивости в модели разделения на торговые зоны. *Электронный журнал «Системный анализ в науке и образовании»*, 2:1–10, 2010.

Корреляционное отношение между случайными множествами событий

Дарья Владиславовна Семенова
Сибирский федеральный университет
Институт математики и фундаментальной информатики
Красноярск
dariasdv@gmail.com

Гулиза Кушбак кызы
Институт математики и фундаментальной информатики
Базовая кафедра вычислительных и информационных технологий
guliz92@mail.ru

Аннотация. *В данной работе получена формула корреляционного отношения для измерения зависимости между двумя случайными множествами событий.*

Ключевые слова. *Корреляционное отношение, случайное множество событий, мощностной индикатор.*

1 Введение

Понятие независимости случайных величин — одно из важных понятий теории вероятностей. Случайная величина ξ называется независимой от случайной величины η, если закон распределения величины ξ не зависит от того, какое значение приняла величина η. В противном случае величины ξ и η называются зависимыми. При изучении систем случайных величин всегда следует обращать внимание на степень и характер их зависимости. Эта зависимость может быть более или менее ярко выраженной, более или менее тесной. В некоторых случаях зависимость между случайными величинами может быть настолько тесной, что, зная значение одной случайной величины, можно в точности указать значение другой. В другом крайнем случае зависимость между случайными величинами является настолько слабой и отдаленной, что их можно практически считать независимыми [4, 5].

В теории вероятностей существуют два вида зависимости: функциональная и стохастическая (вероятностная) зависимость. Функциональная зависимость величин реализуется тогда, когда каждому значению одной величины соответствует определенное значение другой величины. В большинстве случаев между переменными существуют зависимости, при которых каждому значению одной величины соответствует не какое-то определенное значение другой величины, а множество ее возможных значений. Такая зависимость называется стохастической или вероятностной. Частным случаем вероятностной зависимости является корреляционная зависимость — стохастическая зависимость между случайными величинами, при которой наблюдается функциональная зависимость между значениями одной величины и средними значениями другой величины.

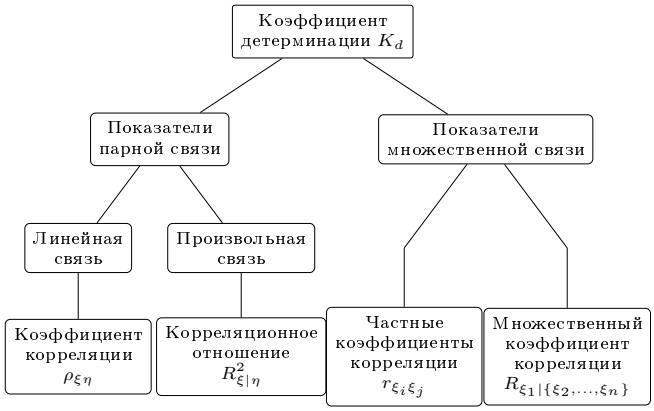


Рисунок 1: Показатели тесноты связи

В классической теории вероятностей рассматривается следующие показатели тесноты связи, которые представлены на рис. 1. Нас интересуют показатель произвольной парной связи, а именно корреляционное отношение [5]. В данной работе предлагается использовать корреляционное отношение для измерения зависимости между двумя случайными множествами событий.

1.1 Корреляционное отношение между случайными величинами

Корреляционным отношением случайной величины ξ относительно η называют число $R^2_{\xi|\eta}$, которое задаются выражением [5]

$$R^2_{\xi|\eta} = 1 - \frac{1}{\mathbf{D}\xi}\mathbf{E}[\mathbf{D}(\xi \mid \eta)] \tag{1}$$

где $\mathbf{D}(\xi \mid \eta)$ условная дисперсия случайной величины ξ относительно η.

Заметим, что, в отличие от коэффициента корреляции ρ, корреляционное отношение зависит от порядка следования случайных величин ξ и η, т.е. $R^2_{\xi|\eta}$, вообще говоря, не совпадает с $R^2_{\eta|\xi}$

Отметим основные свойства корреляционного отношения [4, 5].

© 2014 Д.В. Семенова, Г.Кушбак кызы

Олег Воробьев (ред.), Труды XIII ФАМЭМС'2014, Красноярск: СФУ

1. Корреляционное отношение $R^2_{\xi|\eta}$ можно записать в виде $R^2_{\xi|\eta} = \dfrac{1}{\mathbf{D}\xi}\mathbf{E}(\mathbf{E}(\xi \mid \eta) - \mathbf{E}\xi)^2$.
2. Корреляционное отношение $R^2_{\xi|\eta}$ удовлетворяет неравенству $0 \le R^2_{\xi|\eta} \le 1$.
3. Корреляционное отношение $R^2_{\xi|\eta} = 1$ тогда и только тогда, когда $\mathbf{E}(\xi \mid \eta) \equiv C = \mathbf{E}\xi$, т.е. линия регрессии ξ и η представляет собой горизонтальную прямую.
4. Коэффициент корреляции p и корреляционное отношение $R^2_{\xi|\eta}$ связаны неравенством $|\rho| \le R_{\xi|\eta}$.
5. Модуль $|\rho|$ коэффициента корреляции совпадает корреляционным отношением $R^2_{\xi|\eta}$ тогда и только тогда, когда линия регрессии ξ и η является прямой.

1.2 Случайное множество событий

Рассмотрим вероятностное пространство $(\Omega, \mathcal{A}, \mathbf{P})$. Пусть $\mathfrak{X} \subset \mathcal{A}$ — конечное множество событий, выбранное из алгебры $\mathcal{A}$ этого пространства.

Случайное множество событий под множеством избранных событий $\mathfrak{X}$ [2] определяется на вероятностном пространстве как случайный элемент

$$K : (\Omega, \mathcal{A}, \mathbf{P}) \to \left(2^{\mathfrak{X}}, 2^{2^{\mathfrak{X}}}\right)$$

со значениями из измеримого пространства $\left(2^{\mathfrak{X}}, 2^{2^{\mathfrak{X}}}\right)$, где

- $\mathfrak{X} \subseteq \mathcal{A}$ — выделенное конечное множество событий, состоящее из $N = |\mathfrak{X}|$ событий;
- $2^{\mathfrak{X}}$ — множество всех подмножеств множества избранных событий $\mathfrak{X}$;
- $2^{2^{\mathfrak{X}}}$ — алгебра всех его подмножеств.

Вероятностным распределением случайного множества событий K, определенного под множеством избранных событий $\mathfrak{X}$, называется каждая из сетфункций[2] на $2^{\mathfrak{X}}$

- *вероятностное распределение вероятностей I-го рода* $\{p(X/\!/\mathfrak{X}), X \subseteq \mathfrak{X}\}$, где

$$p(X/\!/\mathfrak{X}) = \mathbf{P}(K = X) = \mathbf{P}\left(\bigcap_{x\in X} x \bigcap_{x\in X^c} x^c\right);$$

- *вероятностное распределение вероятностей II-го рода* $\{p_{X/\!/\mathfrak{X}}, X \subseteq \mathfrak{X}\}$, где

$$p_{X/\!/\mathfrak{X}} = \mathbf{P}(X \subseteq K) = \mathbf{P}\left(\bigcap_{x\in X} x\right).$$

Вероятностные распределения I-го и II-го рода связаны формулами обращения Мёбиуса [2]

$$\begin{aligned} p_{X/\!/\mathfrak{X}} &= \sum_{X\subseteq Y} p(Y/\!/\mathfrak{X}), \\ p(X/\!/\mathfrak{X}) &= \sum_{X\subseteq Y} (-1)^{|Y|-|X|} p_{Y/\!/\mathfrak{X}}. \end{aligned} \tag{2}$$

Ковариация двух событий $x, y \in \mathfrak{X}$ определяется как величина

$$\mathrm{Kov}_{xy} = \mathbf{P}(x \cap y) - \mathbf{P}(x)\mathbf{P}(y) = p_{xy} - p_x p_y, \tag{3}$$

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 [2].

1.2.1 Мощностной индикатор множества событий

Определение и свойства мощностного индикатора введены в работе [3].

Случайная величина $\boldsymbol{\Upsilon}_{\mathfrak{X}}$, равная мощности $|X|$ подмножества событий, которые наступят среди N — множество событий $\mathfrak{X}$, имеющего распределение $\{p(X/\!/\mathfrak{X}), X \subseteq \mathfrak{X}\}$, называется *мощностным индикатором множеством событий* $\mathfrak{X}$, а ее вероятностное распределение, составленное из вероятностей, образующих набор $\{p_N(k), k = 0, \ldots, N\}$, где

$$p_N(k) = \mathbf{P}(\boldsymbol{\Upsilon}_{\mathfrak{X}} = k) = \sum_{|X|=k} p(X/\!/\mathfrak{X}), \tag{4}$$

называется мощностным распределением, порожденным множеством событий $\mathfrak{X}$.

Математическое ожидание мощностного индикатора $\boldsymbol{\Upsilon}_{\mathfrak{X}}$ N-множества событий $\mathfrak{X}$, имеющего распределение $\{p(X), X \subseteq \mathfrak{X}\}$, вычисляется по формуле

$$\mathbf{E}\boldsymbol{\Upsilon}_{\mathfrak{X}} = \sum_{x\in\mathfrak{X}} \mathbf{P}(x). \tag{5}$$

Дисперсия мощностного индикатора $\boldsymbol{\Upsilon}_{\mathfrak{X}}$ N-множества событий $\mathfrak{X}$, с распределением $\{p(X), X \subseteq \mathfrak{X}\}$ равна

$$\mathbf{D}\boldsymbol{\Upsilon}_{\mathfrak{X}} = \sum_{x\in\mathfrak{X}}\sum_{y\in\mathfrak{X}} \mathrm{Kov}_{xy} = \sum_{x\in\mathfrak{X}} \sigma_x^2 + \sum_{\{x,y\}\subseteq\mathfrak{X}} \mathrm{Kov}_{xy} \tag{6}$$

где

- $\sigma_x^2 = p_x(1 - p_x)$ — дисперсия события $x \in \mathfrak{X}$;
- $\mathrm{Kov}_{xy} = p_{xy} - p_x p_y$ — парная ковариация событий $\{x, y\} \in \mathfrak{X}$.

2 Постановка задачи

Пусть $(\Omega, \mathcal{A}, \mathbf{P})$ — вероятностное пространство.

Пусть $\mathfrak{X}$ — множество избранных событий $\mathfrak{X} \subseteq \mathcal{A}$ и $\mathfrak{Y}$ — множество избранных событий $\mathfrak{Y} \subseteq \mathcal{A}$. Пусть задано случайное множество событий K под $\mathfrak{X}$, т.е.

$$K : (\Omega, \mathcal{A}, \mathbf{P}) \to (2^{\mathfrak{X}}, 2^{2^{\mathfrak{X}}}),$$

и задано случайное множество событий L под $\mathfrak{Y}$, т.е.

$$L : (\Omega, \mathcal{A}, \mathbf{P}) \to (2^{\mathfrak{Y}}, 2^{2^{\mathfrak{Y}}}).$$

Известно совместное распределение случайных множеств событий K и L под $\mathcal{Z} = \mathfrak{X} \cup \mathfrak{Y}$

$$\mathbf{P}(S = Z) = \mathbf{P}(K = X, L = Y),$$

где

$$S : (\Omega, \mathcal{A}, \mathbf{P}) \to (2^{\mathfrak{X}\cup\mathfrak{Y}}, 2^{2^{\mathfrak{X}\cup\mathfrak{Y}}}).$$

Найти корреляционное отношение между случайными множествами событий $R^2_{L|K}$.

3 Решение задачи

3.1 Корреляционное отношение между двумя событиями

Рассмотрим случай $\mathfrak{X} = \{x\}$ и $\mathfrak{Y} = \{y\}$. Тогда совместное распределение случайных множеств событий K и L под $\mathcal{Z} = \mathfrak{X} + \mathfrak{Y}$ в виде таблицы 1.

Таблица 1: Совместное распределение случайных множеств событий K и L под $\mathcal{Z} = \mathfrak{X} + \mathfrak{Y}$ для случая $\mathfrak{X} = \{x\}$ и $\mathfrak{Y} = \{y\}$.

K \ L	y^c	y	$p(X//\mathfrak{X})$
x^c	$p(\emptyset//\mathfrak{X}+\mathfrak{Y})$	$p(y//\mathfrak{X}+\mathfrak{Y})$	$p(x^c//\mathfrak{X})$
x	$p(x//\mathfrak{X}+\mathfrak{Y})$	$p(xy//\mathfrak{X}+\mathfrak{Y})$	$p(x//\mathfrak{X})$
$p(Y//\mathfrak{Y})$	$p(y^c//\mathfrak{Y})$	$p(y//\mathfrak{Y})$	1

Заметим, что $p(x//\mathfrak{X}) = p_{x^c//\mathfrak{X}}$, $p(x//\mathfrak{X}) = p_{x//\mathfrak{X}}$, $p(y//\mathfrak{Y}) = p_{y^c//\mathfrak{Y}}$, $p(y//\mathfrak{Y}) = p_{y//\mathfrak{Y}}$, $p(xy//\mathfrak{X}+\mathfrak{Y}) = p_{xy//\mathfrak{X}+\mathfrak{Y}} = p_{xy}$.

Введем обозначения

$$\begin{aligned} p(\emptyset//\mathfrak{X}+\mathfrak{Y}) &= p(\emptyset), & p(y//\mathfrak{X}+\mathfrak{Y}) &= p(y),\\ p(x//\mathfrak{X}+\mathfrak{Y}) &= p(x), & p(xy//\mathfrak{X}+\mathfrak{Y}) &= p(xy),\\ p_{x^c//\mathfrak{X}} &= p_{x^c}, & p_{x//\mathfrak{X}} &= p_x,\\ p_{y//\mathfrak{Y}} &= p_{y^c}, & p_{y//\mathfrak{Y}} &= p_y. \end{aligned}$$

Сопоставим событию x случайную величину $\mathbf{1}_x(\omega)$ — индикатор события x:

$$\mathbf{1}_x(\omega) = \begin{cases} 1, & \text{если } \omega \in x;\\ 0, & \text{если } \omega \notin x. \end{cases}$$

а событию y случайную величину $\mathbf{1}_y(\omega)$ — индикатор события y:

$$\mathbf{1}_y(\omega) = \begin{cases} 1, & \text{если } \omega \in y;\\ 0, & \text{если } \omega \notin y. \end{cases}$$

Тогда совместное распределение случайного вектора $(\mathbf{1}_x, \mathbf{1}_y)$ можно представить в виде таблицы 2. Исходя

Таблица 2: Совместное распределение случайных величин $\mathbf{1}_x$ и $\mathbf{1}_y$.

$\mathbf{1}_x$ \ $\mathbf{1}_y$	0	1	$\mathbf{P}(\mathbf{1}_x)$
0	$p(\emptyset)$	$p(y)$	p_{x^c}
1	$p(x)$	$p(xy)$	p_x
$\mathbf{P}(\mathbf{1}_y)$	p_{y^c}	p_y	1

из данной таблицы найдем все числовые характеристики случайных величин:

— математические ожидания:

$$\mathbf{E}\mathbf{1}_x = p_x, \quad \mathbf{E}\mathbf{1}_y = p_y;$$

— дисперсии:

$$\mathbf{D}\mathbf{1}_x = p_x p_{x^c}, \quad \mathbf{D}\mathbf{1}_y = p_y p_{y^c},$$

— ковариация:

$$\operatorname{Cov}(\mathbf{1}_x, \mathbf{1}_y) = \mathbf{E}\mathbf{1}_x\mathbf{1}_y - \mathbf{E}\mathbf{1}_x\mathbf{E}\mathbf{1}_y =$$

$$= p_{xy} - p_x p_y = \operatorname{Kov}_{xy}.$$

Найдем корреляционное отношение $R^2_{\mathbf{1}_x|\mathbf{1}_y}$ случайной величины $\mathbf{1}_x$ относительно $\mathbf{1}_y$.

Условные распределения имеют вид

$\mathbf{1}_x \mid_{\mathbf{1}_y=0}$	0	1
$\mathbf{P}\left(\mathbf{1}_x \mid_{\mathbf{1}_y=0}\right)$	$\dfrac{p(\emptyset)}{p(\emptyset)+p(x)}$	$\dfrac{p(x)}{p(\emptyset)+p(x)}$

$\mathbf{1}_x \mid_{\mathbf{1}_y=1}$	0	1
$\mathbf{P}\left(\mathbf{1}_x \mid_{\mathbf{1}_y=1}\right)$	$\dfrac{p(y)}{p(y)+p(xy)}$	$\dfrac{p(xy)}{p(y)+p(xy)}$

Условные математические ожидания:

$$\mathbf{E}(\mathbf{1}_x \mid \mathbf{1}_y = 0) = \frac{p(x)}{p(\emptyset)+p(x)} = \frac{p(x)}{p_{y^c}},$$

$$\mathbf{E}(\mathbf{1}_x \mid \mathbf{1}_y = 1) = \frac{p(xy)}{p(y)+p(xy)} = \frac{p(xy)}{p_y}.$$

Условные дисперсии

$$\mathbf{D}(\mathbf{1}_x \mid \mathbf{1}_y = 0) = \frac{p(x)}{p_{y^c}}\left(1 - \frac{p(x)}{p_{y^c}}\right) = \frac{p(\emptyset)p(x)}{p^2_{y^c}},$$

$$\mathbf{D}(\mathbf{1}_x \mid \mathbf{1}_y = 1) = \frac{p(xy)}{p_y}\left(1 - \frac{p(xy)}{p_y}\right) = \frac{p(xy)p(y)}{p_y^2}.$$

Условную дисперсию можем рассматривать как случайную величину, которая принимает значения $\mathbf{D}(\mathbf{1}_x \mid \mathbf{1}_y = 0)$ и $\mathbf{D}(\mathbf{1}_x \mid \mathbf{1}_y = 1)$ со следующими вероятностями

$\mathbf{D}(\mathbf{1}_x \mid \mathbf{1}_y)$	$\frac{p(x)p(\emptyset)}{p_{y^c}^2}$	$\frac{p(xy)p(y)}{p_y^2}$
$\mathbf{P}(\mathbf{D}(\mathbf{1}_x \mid \mathbf{1}_y))$	p_{y^c}	p_y

Тогда математическое ожидание условной дисперсии

$$\mathbf{ED}(\mathbf{1}_x \mid \mathbf{1}_y) = \frac{p(x)p(\emptyset)}{p_{y^c}} + \frac{p(xy)p(y)}{p_y}.$$

Найденные значения $\mathbf{D1}_x$ и $\mathbf{ED}(\mathbf{1}_x \mid \mathbf{1}_y)$ подставляем в формулу корреляционного отношения (1) и получим

$$R^2_{\mathbf{1}_x|\mathbf{1}_y} = 1 - \frac{\mathbf{E}[\mathbf{D}(\mathbf{1}_x|\mathbf{1}_y)]}{\mathbf{D}(\mathbf{1}_x)} = \\ = 1 - \frac{\frac{p(x)p(\emptyset)}{p_{y^c}} + \frac{p(xy)p(y)}{p_y}}{p_x(1-p_x)}. \tag{7}$$

С помощью формулы обращения Мёбиуса (2) из известных распределений первого рода найдем распределении второго рода.

$$p(x) = p_x - p_{xy}, \qquad p(y) = p_y - p_{xy}, \tag{8}$$

$$p(\emptyset) = 1 - p_x - p_y + p_{xy}, \qquad p(xy) = p_{xy}. \tag{9}$$

Далее в (7) подставляем (8) и (9) и получим

$$R^2_{\mathbf{1}_x|\mathbf{1}_y} = \frac{1}{p_{x^c}p_xp_{y^c}p_y}\left[p_{x^c}p_xp_{y^c}p_y - \right. \\ -p_y(p_x - p_{xy})(1 - p_x - p_y + p_{xy}) - \\ \left. -p_{y^c}p_{xy}(p_y - p_{xy})\right]. \tag{10}$$

Формулы (8) – (9) можно представить в виде

$$p(x) = p_x - p_{xy} = p_x(1-p_y) - \mathrm{Kov}_{xy}, \tag{11}$$

$$p(y) = p_y - p_{xy} = p_y(1-p_x) - \mathrm{Kov}_{xy}, \tag{12}$$

$$p(\emptyset) = 1 - p_x - p_y + p_{xy} = (1-p_x)(1-p_y) + \mathrm{Kov}_{xy}, \tag{13}$$

$$p(xy) = p_{xy} = p_xp_y + \mathrm{Kov}_{xy}. \tag{14}$$

В (10) подставляем (11) – (14) и получаем

$$R^2_{\mathbf{1}_x|\mathbf{1}_y} = \frac{\mathrm{Kov}^2_{xy}}{p_xp_{x^c}p_yp_{y^c}}.$$

Корреляционным отношением событий x и y будем называть число равное

$$R^2_{x|y} = \frac{\mathrm{Kov}^2_{xy}}{p_xp_{x^c}p_yp_{y^c}}. \tag{15}$$

Из (15) видно, что корреляционное отношение событий величина симметричная, т.е. $R^2_{x|y} = R^2_{y|x}$.

Рассмотрим корреляционное отношение (15) при различных структурах зависимостей событий. В теории случайных множеств событий [2] среди всевозможных структур зависимостей выделяет три основные: вложенную, независимую и наименее пересекающуюся (непересекающуюся). Вложенные события — наиболее зависимые события, наступление одного, обеспечивает наступление все остальных. Наименее пересекающиеся события — также сильно зависимые события. Независимые события имеют характерное свойство: они всегда пересекаются и вероятность их пересечения больше нуля. Независимые события занимают среднее положение среди всех структур зависимостей, от вложенных до наименее пересекающихся, занимающих крайние противоположные места.

- Пусть x и y независимые события, тогда $\mathrm{Kov}_{xy} = 0$, и $R^2_{x|y} = 0$.
- Если случайные события x и y несовместные события (непересекающаяся структура зависимостей), тогда $p_{xy} = 0$ и $\mathrm{Kov}_{xy} = -p_xp_y$. Отсюда корреляционное отношение (15) равно

$$R^2_{x|y} = \frac{p_xp_y}{p_{x^c}p_{y^c}}. \tag{16}$$

Заметим, что в этом случае $(x) + p(y) + p(\emptyset) = p_x + p_y + p(\emptyset) = 1$. Если события x и y образуют разбиение Ω, т.е. $p(\emptyset) = 0$ и$p_x + p_y = 1$, то $R^2_{x|y} = 1$. Если $p(\emptyset) = \mathrm{const}$, $0 < \mathrm{const} < 1$, то $p_x + p_y = 1 - \mathrm{const}$ и $R^2_{x|y} = \frac{p_x(1-p_x-\mathrm{const})}{(1-p_x)(p_x+\mathrm{const})}$. На рис. 2 представлены графики корреляционного отношения $R^2_{x|y}$ для несовместных событий x и y при различных значениях $p(\emptyset) = \mathbf{P}(x^c \cap y^c)$

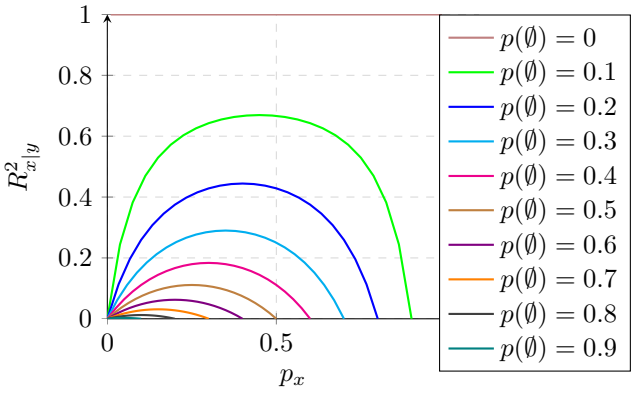


Рисунок 2: Графики корреляционного отношения $R^2_{x|y}$ для несовместных событий x и y при различных значениях $p(\emptyset) = \mathbf{P}(x^c \cap y^c)$.

- Пусть $x \subseteq y$,тогда $p_{xy} = p_x$ и

$$\mathrm{Kov}_{xy} = p_x - p_xp_y = p_xp_{y^c}.$$

Отсюда корреляционное отношение равно

$$R^2_{x|y} = \frac{p_x p_y^c}{p_{x^c} p_y}. \qquad (17)$$

На рис. 3 представлены графики корреляционного отношения $R^2_{x|y}$ для вложенных событий $x \subseteq y$ при различных значениях $p_y = \mathbf{P}(y)$.

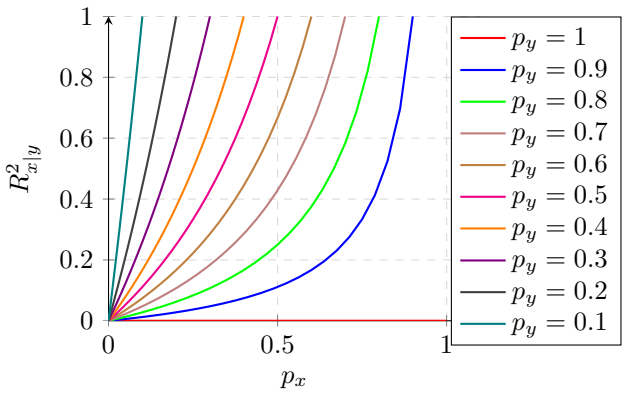


Рисунок 3: Графики корреляционного отношения $R^2_{x|y}$ для вложенных событий $x \subseteq y$ при различных значениях $p_y = \mathbf{P}(y)$.

Пусть вероятности p_x и p_y фиксированы. Тогда вероятность пересечения событий $p_{xy} = \mathbf{P}(x \cap y)$ может принимать любые значения в рамках границ Фреше [2]

$$\max\{p_x + p_y - 1, 0\} \le p_{xy} \le \min\{p_x, p_y\}.$$

На рисунке 4 представлена зависимость корреляционного отношения от p_{xy} при различных фиксированных вероятностях событий p_x и p_y.

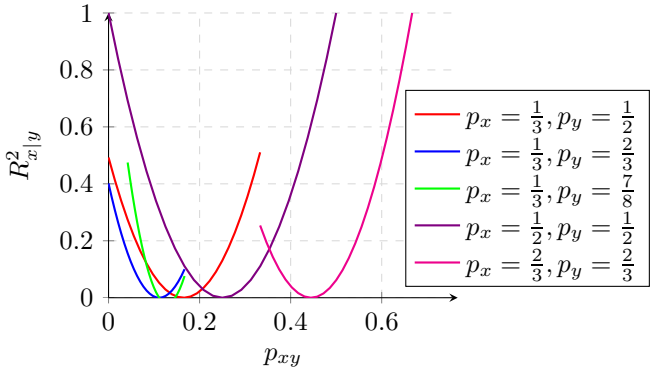


Рисунок 4: Зависимость корреляционного отношения $R^2_{x|y}$ от p_{xy} при различных фиксированных вероятностях событий p_x и p_y

3.2 Корреляционное отношение двух множеств событий

Сопоставим случайному множеству событий K под $\mathfrak{X}$ ($|\mathfrak{X}| = N$) мощностной индикатор $\boldsymbol{\Upsilon}_{\mathfrak{X}}$, а случайному множеству событий L под $\mathfrak{Y}$ ($|\mathfrak{X}| = M$) сопоставим мощностной индикатор $\boldsymbol{\Upsilon}_{\mathfrak{Y}}$, совместное распределение которых выражается через совместное распределение случайных множеств событий K и L

$$\mathbf{P}(\boldsymbol{\Upsilon}_{\mathfrak{X}} = i, \boldsymbol{\Upsilon}_{\mathfrak{Y}} = j) = \sum_{|X|=i} \sum_{|Y|=j} \mathbf{P}(K = X, L = Y).$$

Введем следующие обозначения

$$\mathbf{P}(\boldsymbol{\Upsilon}_{\mathfrak{X}} = i, \boldsymbol{\Upsilon}_{\mathfrak{Y}} = j) = p_{ij}, \quad i = \overline{0,N}, j = \overline{0,M},$$
$$\mathbf{P}(\boldsymbol{\Upsilon}_{\mathfrak{X}} = i) = r_i = \sum_{j=0}^{M} p_{ij}, \quad i = \overline{0,N},$$
$$\mathbf{P}(\boldsymbol{\Upsilon}_{\mathfrak{Y}} = j) = q_j = \sum_{i=0}^{N} p_{ij}, \quad j = \overline{0,M}.$$

Найдем корреляционное отношение $R^2_{\boldsymbol{\Upsilon}_{\mathfrak{Y}}|\boldsymbol{\Upsilon}_{\mathfrak{X}}}$.

С учетом введенных выше обозначений получаем

- $\mathbf{E}\boldsymbol{\Upsilon}_{\mathfrak{X}} = \sum\limits_{i=0}^{N} i \cdot r_i = \sum\limits_{i=0}^{N} \sum\limits_{j=0}^{M} i \cdot p_{ij},$
- $\mathbf{E}\boldsymbol{\Upsilon}_{\mathfrak{Y}} = \sum\limits_{j=0}^{M} j \cdot q_j = \sum\limits_{j=0}^{M} \sum\limits_{i=0}^{N} j \cdot p_{ij},$
- $\mathbf{D}\boldsymbol{\Upsilon}_{\mathfrak{X}} = \sum\limits_{i=0}^{N} i^2 r_i - \sum\limits_{i=0}^{N} \sum\limits_{k=0}^{N} i \cdot k \cdot r_i r_k,$
- $\mathbf{D}\boldsymbol{\Upsilon}_{\mathfrak{Y}} = \sum\limits_{j=0}^{M} j^2 q_j - \sum\limits_{j=0}^{M} \sum\limits_{k=0}^{N} j \cdot k \cdot q_j q_k,$
- условная вероятность $\mathbf{P}(\boldsymbol{\Upsilon}_{\mathfrak{Y}} = j | \boldsymbol{\Upsilon}_{\mathfrak{X}=i}) = \dfrac{p_{ij}}{r_i},$
- условное математическое ожидание

$$\mathbf{E}\boldsymbol{\Upsilon}_{\mathfrak{Y}} | \boldsymbol{\Upsilon}_{\mathfrak{X}=i} = \sum_{j=0}^{M} j \cdot \frac{p_{ij}}{r_i},$$

- условная дисперсия

$$\mathbf{D}\boldsymbol{\Upsilon}_{\mathfrak{Y}} | \boldsymbol{\Upsilon}_{\mathfrak{X}=i} = \frac{\sum\limits_{j=0}^{M} j^2 \cdot p_{ij}}{r_i} - \frac{\sum\limits_{j=0}^{M} \sum\limits_{k=0}^{M} j \cdot k \cdot p_{ik} p_{ij}}{r_i^2},$$

- математическое ожидание условной дисперсии

$$\mathbf{E}[\mathbf{D}(\boldsymbol{\Upsilon}_{\mathfrak{Y}} | \boldsymbol{\Upsilon}_{\mathfrak{X}})] = \sum_{i=0}^{N} \sum_{j=0}^{M} \sum_{k=0}^{M} \frac{j(j-k) p_{ij} p_{ik}}{r_i}.$$

Тогда корреляционное отношение имеет вид

$$R^2_{\boldsymbol{\Upsilon}_{\mathfrak{Y}}|\boldsymbol{\Upsilon}_{\mathfrak{X}}} = \frac{H - (\mathbf{E}\boldsymbol{\Upsilon}_{\mathfrak{Y}})^2}{\mathbf{D}\boldsymbol{\Upsilon}_{\mathfrak{Y}}}. \qquad (18)$$

где

$$(\mathbf{E}\boldsymbol{\Upsilon}_{\mathfrak{Y}})^2 = \sum_{j=0}^{M} \sum_{k=0}^{N} j \cdot k \cdot q_j q_k,$$

$$H = \sum_{i=0}^{N} \frac{1}{r_i} \left(\sum_{j=0}^{M} j \cdot p_{ij} \right)^2 = \\ = \sum_{i=0}^{N} \sum_{j=0}^{M} \sum_{k=0}^{M} \frac{j \cdot k \cdot p_{ij} \cdot p_{ik}}{\sum\limits_{l=0}^{M} p_{il}}. \tag{19}$$

Получим корреляционное отношение случайных множеств$R^2_{L|K}$, переходя от численного суммирование в формуле (18) к сет-суммированию [1].

Заметим, что вероятность того, что мощностной индикатор примет значение равное i равно сумме вероятностей тех подмножеств множества $\mathfrak{X}$, мощности которых равны i.

$$r_i = \mathbf{P}(\boldsymbol{\Upsilon}_{\mathfrak{Y}} = i) = \sum_{j=0}^{M} p_{ij} =$$

$$= \sum_{j=0}^{M} \sum_{|X|=i} \sum_{|Y|=j} p(X+Y//\mathfrak{X}+\mathfrak{Y}) =$$

$$= \sum_{|X|=i} \sum_{Y \subseteq \mathfrak{Y}} p(X+Y//\mathfrak{X}+\mathfrak{Y}) = \sum_{|X|=i} p(X//\mathfrak{X}).$$

Аналогично получаем, что вероятность того, что мощностной индикатор $\boldsymbol{\Upsilon}_{\mathfrak{Y}}$ примет значение равное j равно сумме вероятностей тех подмножеств множества $\mathfrak{Y}$, мощности которых равны j

$$q_j = \mathbf{P}(\boldsymbol{\Upsilon}_{\mathfrak{Y}} = j) = \sum_{|Y|=j} p(Y//\mathfrak{Y}).$$

Применим к (19) сет-суммирование

$$\sum_{j=0}^{M} j \cdot p_{ij} = \sum_{y \in Y \subseteq \mathfrak{Y}} |Y| \sum_{|X|=i} \sum_{|Y|=j} p(X+Y//\mathfrak{X}+\mathfrak{Y}) =$$

$$= \sum_{|X|=i} \sum_{Y \subseteq \mathfrak{Y}} \sum_{y \in \mathfrak{Y}} \mathbf{1}_Y(y) p(X+Y//\mathfrak{X}+\mathfrak{Y}) =$$

$$= \sum_{|X|=i} \sum_{y \in \mathfrak{Y}} \sum_{y \in Y \subseteq \mathfrak{Y}} p(X+Y//\mathfrak{X}+\mathfrak{Y}) =$$

$$= \sum_{|X|=i} \sum_{y \in \mathfrak{Y}} p_{y|X} p(X//\mathfrak{X})$$

и

$$\sum_{i=0}^{N} \frac{1}{r_i} \left(\sum_{j=0}^{M} j \cdot p_{ij} \right)^2 =$$

$$= \sum_{i=0}^{N} \frac{1}{\sum\limits_{|X|=i} p(X//\mathfrak{X})} \left(\sum_{|X|=i} \sum_{y \in \mathfrak{Y}} p_{y|X} p(X//\mathfrak{X}) \right)^2,$$

где

$$p_{y|X} = \frac{p(X+Y//\mathfrak{X}+\mathfrak{Y})}{p(X//\mathfrak{X})}.$$

Обозначим $C^k_{\mathfrak{X}} = \{X \mid X \subseteq \mathfrak{X},\ |X| = k\}$ — слои-множества, $\mathbf{P}(C^k_{\mathfrak{X}}) = \sum\limits_{|X|=k} p(X//\mathfrak{X})$ — вероятность слоя. Тогда

$$H = \sum_{i=0}^{N} \frac{1}{\mathbf{P}(C^i_{\mathfrak{X}})} \left(\sum_{X \in C^i_{\mathfrak{X}}} \sum_{y \in \mathfrak{Y}} p_{y|X} p(X//\mathfrak{X}) \right)^2,$$

$$H - (\mathbf{E}\boldsymbol{\Upsilon}_{\mathfrak{Y}})^2 = \sum_{i=0}^{N} \frac{1}{\mathbf{P}(C^i_{\mathfrak{X}})} \cdot$$

$$\cdot \left[\sum_{X \in C^i_{\mathfrak{X}}} \sum_{V \in C^i_{\mathfrak{X}}} \sum_{y \in \mathfrak{Y}} \sum_{z \in \mathfrak{Y}} p_{y|X} p(X//\mathfrak{X}) p_{z|V} p(V//\mathfrak{X}) - \right.$$

$$\left. - \sum_{y \in \mathfrak{Y}} \sum_{z \in \mathfrak{Y}} p_y p_z \right].$$

Отсюда получим формулу в общем виде корреляционного отношения для двух множеств событий, не имеющих общих элементов

$$R^2_{L|K} = \frac{1}{\sum\limits_{y \in \mathfrak{Y}} \sum\limits_{z \in \mathfrak{Y}} \mathrm{Kov}_{yz}} \cdot \\ \cdot \sum_{y,z \in \mathfrak{Y}} \left(\sum_{i=0}^{N} \sum_{X,V \in C^i_{\mathfrak{X}}} u_i(X,V) p_{y|X} p_{z|V} - p_y p_z \right), \tag{20}$$

где

$$u_i(X,V) = \frac{p(X//\mathfrak{X}) p(V//\mathfrak{X})}{\mathbf{P}(C^i_{\mathfrak{X}})},$$

и для сокращения записи используются следующие обозначения $\sum\limits_{X,V \in C^i_{\mathfrak{X}}} = \sum\limits_{X \in C^i_{\mathfrak{X}}} \sum\limits_{V \in C^i_{\mathfrak{X}}}$ и $\sum\limits_{y,z \in \mathfrak{Y}} = \sum\limits_{y \in \mathfrak{Y}} \sum\limits_{z \in \mathfrak{Y}}$.

Заметим, что если $|\mathfrak{X}| = N = 1$ и $|\mathfrak{Y}| = M = 1$, то (20) преобразуется к виду (15).

Список литературы

[1] Воробьев О. Ю. *Сет-суммирование.* Новосибирск: ВО «Наука», 1993.

[2] Воробьев О. Ю. *Эвентология.* Красноярск : Сибирский федеральный университет, 2007.

[3] Воробьев О. Ю. *Математические технологии в гуманитарных и социо-экономических науках: эвентологический взгляд: Электронный ресурс. Режим доступа: https://www.academia.edu/7293364.* Красноярск : Сибирский федеральный университет, 2014.

[4] Феллер В. *Введение в теорию вероятностей и ец приложения.* Москва : Наука, 1964.

[5] Уилкс С. *Математическая статистика : учеб. пособие.* Москва : Наука, 1967.

Семейство случайных множеств событий Франка

Дарья Владиславовна Семенова
Институт математики и фундаментальной информатики
Сибирский Федеральный университет
Красноярск
dariasdv@gmail.com

Наталья Александровна Лукьянова
Институт математики и фундаментальной информатики
Сибирский Федеральный университет
Красноярск
nata00sfu@gmail.com

Людмила Юрьевна Шангареева
Институт математики и фундаментальной информатики
Сибирский Федеральный университет
Красноярск
mizkuzy@yandex.ru.ru

Аннотация. *В работе предлагается метод, который позволяет задать вероятностное распределение множества N событий, используя N вероятностей и вид ассоциативной функции.*
Ключевые слова. *Случайное множество событий, ассоциативная функция, ассоциативное случайное множество событий Франка.*

1 Введение

Случайное множество событий K на конечном множестве событий $\mathfrak{X} \subset \mathcal{F}$ определяется как отображение $K : \Omega \to 2^{\mathfrak{X}}$, измеримое относительно пары алгебр $\left(\mathcal{F}, 2^{2^{\mathfrak{X}}}\right)$ в том смысле, что для всякого $X \in 2^{2^{\mathfrak{X}}}$ существует прообраз $K^{-1}(X) \in \mathcal{F}$, такой что $\mathbf{P}(X) = \mathbf{P}(K^{-1}(X))$.

Вероятностное распределение случайного множества событий можно представить несколькими эквивалентными распределениями вероятностей, порождённых этим множеством событий [1]:

- вероятностное распределение I-го рода — набор из $2^{|\mathfrak{X}|}$ вероятностей вида $\{p(X), X \subseteq \mathfrak{X}\}$, где
$$p(X) = \mathbf{P}(K = X) = \mathbf{P}\left(\bigcap_{x\in X} x \bigcap_{x\in X^c} x^c\right);$$
- вероятностное распределение II-го рода — набор из $2^{|\mathfrak{X}|}$ вероятностей вида $\{p_X, X \subseteq \mathfrak{X}\}$, где
$$p_X = \mathbf{P}(X \subseteq K) = \mathbf{P}\left(\bigcap_{x\in X} x\right).$$

В теории случайных событий [1] событие $\{K = \emptyset\} = \bigcap_{x\in\mathfrak{X}} x^c$ означает, что не наступило ни одно событие из $\mathfrak{X}$. В вероятностном распределении II-го рода всегда $p_\emptyset = \mathbf{P}(\Omega) = 1$.

Только распределение I-го рода обладает привычным для распределения вероятности свойством нормировки, которое очевидно следует из того, что соответствующие события $\left(\bigcap_{x\in X} x\right) \bigcap \left(\bigcap_{x\in X^c} x^c\right)$ образуют разбиение пространства элементарных событий Ω. А поскольку события $\bigcap_{x\in X} x$ образуют не разбиение, а всего лишь покрытие Ω, то соотношение нормировки для вероятностного распределения II-го рода не выполняется, а сумма вероятностей этих событий всегда больше единицы.

Вероятностное распределение II-го рода $\{p_X, X \subseteq \mathfrak{X}\}$ случайного множества событий K на $\mathfrak{X}$ удовлетворяет системе из 2^N неравенств Фреше–Хёфдинга:
$$0 \le p_X^- \le p_X \le p_X^+ \le 1,$$
где $p_X^- = \max\left\{0, 1 - \sum_{x\in X}(1 - p_x)\right\}$ — нижняя граница Фреше–Хёфдинга, $p_X^+ = \min_{x\in X} p_x$ — верхняя граница Фреше–Хёфдинга.

Вероятностные распределения I-го и II-го рода связаны взаимно-обратными формулами обращения Мёбиуса [1]
$$p_X = \sum_{Y\in 2^{\mathfrak{X}}:\ X\subseteq Y} p(Y), \tag{1}$$
$$p(X) = \sum_{Y\in 2^{\mathfrak{X}}:\ X\subseteq Y} (-1)^{|Y|-|X|} p_Y, \tag{2}$$
для всех $X \in 2^{\mathfrak{X}}$.

Если задано вероятностное распределение I-го рода $\{p(X), X \subseteq \mathfrak{X}\}$, то по формуле (1) мы всегда получим вероятностное распределение II-го рода $\{p_X, X \subseteq \mathfrak{X}\}$. Однако преобразование (2) заданного набора из 2^N чисел $\{p_X,\ X \subseteq \mathfrak{X}\}$, удовлетворяющих границам Фреше–Хёфдинга, может привести к вероятностному распределению I-рода с отрицательными значениями.

Будем говорить, что случайное множество событий K обладает легитимным вероятностным распределением, если для вероятностного распределения I-го рода выполнены следующие условия:
$$0 \le p(X) \le 1, X \subseteq \mathfrak{X}, \tag{3}$$
$$\sum_{X\subseteq\mathfrak{X}} p(X) = 1. \tag{4}$$

© 2014 Д.В.Семенова, Н.А.Лукьянова, Л.Ю.Шангареева
Олег Воробьев (ред.), Труды XIII ФАМЭМС'2014, Красноярск: СФУ

В современных теориях неопределенности широкое распространение получили классы ассоциативных функций, в частности, треугольные нормы и копулы. Алгоритмы вычисления копул и треугольных норм применимы к вероятностным распределениям случайных множеств событий.

Ассоциативная функция в теории случайных множеств событий (Associative Function — AF)

$$\mathrm{AF} : [0,1]^2 \to [0,1]$$

определяется как двуместная функция, удовлетворяющая следующим свойствам:

A1. Граничные условия

$$\begin{aligned} \mathrm{AF}(a,0) = \mathrm{AF}(0,a) = 0, \\ \mathrm{AF}(a,1) = \mathrm{AF}(1,a) = a, \end{aligned} \tag{5}$$

$a \in [0,1]$.

A2. Монотонность

$$\mathrm{AF}(a_1,b_1) \le \mathrm{AF}(a_2,b_2), \tag{6}$$

когда $a_1 \le a_2, b_1 \le b_2$.

A3. Коммутативность, т.е. для всех a, b на $[0,1]$

$$\mathrm{AF}(a,b) = \mathrm{AF}(b,a).$$

A4. Ассоциативность, т.е. для всех a, b, $c \in [0,1]$

$$\mathrm{AF}(\mathrm{AF}(a,b),c) = \mathrm{AF}(a,\mathrm{AF}(b,c)). \tag{7}$$

A5. Условие Липшиц-непрерывности

$$\mathrm{AF}(c,b) - \mathrm{AF}(a,b) \le c - a, \qquad a \le c. \tag{8}$$

Геометрически график ассоциативной функции — это поверхность (рис. 1), которая «натянута» на четырехугольник с вершинами $(0,0,0)$, $(0,1,0)$, $(1,0,0)$ и $(1,1,1)$. На рис. 1 стороны четырехугольника выделены жирными линиями и соответствуют граничным условиям (A1). Согласно свойству A2 ассоциативная функция возрастает по вертикали и по горизонтали [5], а по свойству A3 симметрична относительно плоскости $a = b$. Свойства A1 — A3, A5 гарантируют, что значение функции $\mathrm{AF}(a,b)$ будет обладать свойствами вероятности. Свойство A4 позволяет рекуррентно перейти к n-местной функции.

Заметим, что аксиомы A1-A4 соответствует определению t-нормы [4]. Таким образом, под ассоциативной функцией мы будем понимать непрерывную t-норму, удовлетворяющую условию Липшица, или, что эквивалентно, ассоциативную, коммутативную копулу [5] (Рис. 1).

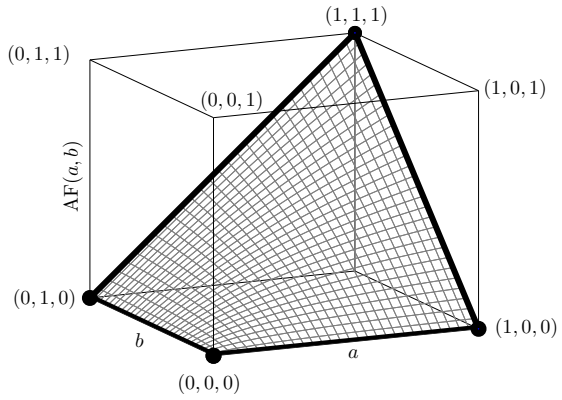


Рисунок 1: График произвольной ассоциативной функции

В работе [2] было введено определение ассоциативного случайного множества событий и его вероятностного распределения.

Ассоциативным случайным множеством событий называется случайное множество событий K на конечном множестве избранных событий $\mathfrak{X}$ с вероятностным 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 $p_x = \mathbf{P}(x)$, $x \in \mathfrak{X}$

$$p_{xy} = \mathbf{P}(x \cap y) = \mathrm{AF}(p_x, p_y),$$

$$\begin{aligned} p_{xyz} = \mathbf{P}(x \cap y \cap z) = \mathrm{AF}(p_x, \mathrm{AF}(p_y, p_z)) = \\ = \mathrm{AF}(p_x, \mathbf{P}(y \cap z)), \end{aligned}$$

$$p_X = \mathbf{P}\left(\bigcap_{x \in X} x\right) = \mathrm{AF}\left(p_x, \mathbf{P}\left(\bigcap_{y \in X \setminus \{x\}} y\right)\right),$$

при условии, что соответствующее вероятностное распределение I-го рода будет легитимным.

2 Метод ассоциативных функций в построении семейств случайных множеств событий на примере семейства Франка

Применим понятие ассоциативной функции к вероятностным распределениям случайных множеств событий. В качестве аргументов ассоциативной функции предлагается рассматривать вероятности событий, их число совпадает с мощностью базового множества. Таким образом, ассоциативные функции будут связывать вероятности пересечения множества событий $p_X = \mathbf{P}\left(\bigcap_{x \in X} x\right)$ с вероятностями самих событий $p_x = \mathbf{P}(x)$, $x \in X$, $X \subseteq \mathfrak{X}$.

Рекуррентный метод построения ассоциативных множеств событий

Вход:

- множество событий $\mathfrak{X}$, $|\mathfrak{X}| = N$;
- N вероятностей событий p_x, $x \in \mathfrak{X}$;
- ассоциативная функция $\mathrm{AF}(a, b)$.

Выход: вероятностное распределение II-го рода $\{p_X, X \subseteq \mathfrak{X}\}$ случайного множества событий K на $\mathfrak{X}$.

Основная идея: исходя из известных вероятностей событий $p_x = \mathbf{P}(x)$, $x \in \mathfrak{X}$, формирование вероятностей пересечения множеств событий p_X, $X \subseteq \mathfrak{X}, |X| > 1$ осуществлять последовательно согласно рекуррентной формуле:

$$p_X = \mathbf{P}\left(\bigcap_{x\in X} x\right) = \mathrm{AF}\left(p_x, \mathbf{P}\left(\bigcap_{y\in X\setminus\{x\}} y\right)\right). \quad (9)$$

Например,

$$p_{xy} = \mathbf{P}(x \cap y) = \mathrm{AF}(p_x, p_y),$$
$$p_{xyz} = \mathbf{P}(x \cap y \cap z) = \mathrm{AF}(p_x, \mathrm{AF}(p_y, p_z))$$

и далее аналогичным образом.

Формула (9) позволяет построить вероятностное распределение II-го рода случайного множества событий, где в качестве входных параметров выступают N вероятностей событий и вид ассоциативной функции. В результате формируются $2^N - N - 1$ вероятностей II-го рода, удовлетворяющих границам Фреше–Хёфдинга, которых не хватало до полного набора. Однако полученные распределения могут быть нелигитимными в смысле определения 3. Следовательно, для каждого семейства ассоциативных функций необходимо определять условия легитимности построенных распределений.

В [1] был предложен удобный инструмент анализа структур вероятностной зависимости — $|X|$-арная ковариация произвольного множества событий $X \subset \mathcal{F}$, $|X| > 1$:

$$\mathrm{Kov}_X = \mathbf{P}\left(\bigcap_{x\in X} x\right) - \prod_{x\in X} \mathbf{P}(x) = p_X - \prod_{x\in X} p_x. \quad (10)$$

$|X|$-арная ковариация Kov_X обращается в ноль, когда события из X независимы; больше нуля, когда события из множества X статистически притягиваются; и меньше нуля, когда события из множества X статистически отталкиваются. Известно [1], что для множества $\mathfrak{X}$, состоящего из N событий, существует набор из $2^N - N - 1$ арных ковариаций, которые определяются вероятностным распределением II-го рода случайного множества событий на $\mathfrak{X}$.

С учетом (9) формула (10) примет вид

$$\mathrm{Kov}_X = \mathrm{AF}\left(p_x, \mathbf{P}\left(\bigcap_{y\in X\setminus\{x\}} y\right)\right) - \prod_{x\in X} p_x. \quad (11)$$

2.1 Семейство Франка

Проиллюстрируем предложенный рекуррентный подход на примере t-нормы (копулы) Франка (Рис.2), которая является ассоциативной функцией [3] при $\alpha \neq 0$:

$$\mathrm{AF}_\alpha(a, b) = \ln\left(1 + \frac{(e^{-\alpha\cdot a} - 1)(e^{-\alpha\cdot b} - 1)}{(e^{-\alpha} - 1)}\right)^{-\frac{1}{\alpha}}. \quad (12)$$

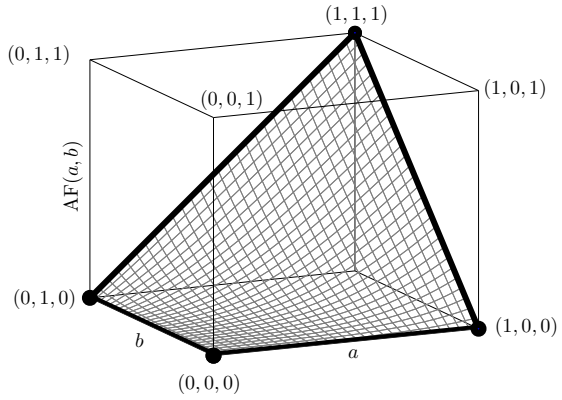


Рисунок 2: График ассоциативной функции Франка с параметром $\alpha = -5$.

Пример 1. Пусть $\mathfrak{X} = \{x, y\}$ и пусть известны вероятности событий $p_x = \frac{1}{6}$ и $p_y = \frac{2}{3}$. Тогда из (9) и (12) получаем

$$p_{xy} = \ln\left(1 + \frac{(e^{-\alpha\cdot\frac{1}{6}} - 1)(e^{-\alpha\cdot\frac{2}{3}} - 1)}{(e^{-\alpha} - 1)}\right)^{-\frac{1}{\alpha}}.$$

По формулам обращения Мёбиуса (2) перейдем к вероятностному распределению I-го рода (Рис. 3):

$$p(\varnothing) = \ln\left[e^{\frac{1}{6}} \cdot \left(1 + \frac{(e^{-\alpha\cdot\frac{1}{6}} - 1)(e^{-\alpha\cdot\frac{2}{3}} - 1)}{(e^{-\alpha} - 1)}\right)^{-\frac{1}{\alpha}}\right];$$

$$p(x) = \ln\left[e^{\frac{1}{6}} \cdot \left(1 + \frac{(e^{-\alpha\cdot\frac{1}{6}} - 1)(e^{-\alpha\cdot\frac{2}{3}} - 1)}{(e^{-\alpha} - 1)}\right)^{\frac{1}{\alpha}}\right];$$

$$p(y) = \ln\left[e^{\frac{2}{3}} \cdot \left(1 + \frac{(e^{-\alpha\cdot\frac{1}{6}} - 1)(e^{-\alpha\cdot\frac{2}{3}} - 1)}{(e^{-\alpha} - 1)}\right)^{\frac{1}{\alpha}}\right];$$

$$p(xy) = \ln\left(1 + \frac{(e^{-\alpha\cdot\frac{1}{6}} - 1)(e^{-\alpha\cdot\frac{2}{3}} - 1)}{(e^{-\alpha} - 1)}\right)^{-\frac{1}{\alpha}}.$$

Заметим, что

$$\lim_{\alpha\to-\infty} p(\varnothing) = \frac{1}{6} \quad \lim_{\alpha\to\infty} p(\varnothing) = \frac{1}{3};$$

$$\lim_{\alpha\to-\infty} p(x) = \frac{1}{6} \quad \lim_{\alpha\to\infty} p(x) = 0;$$

$$\lim_{\alpha\to-\infty} p(y) = \frac{2}{3} \quad \lim_{\alpha\to\infty} p(y) = \frac{1}{2};$$

$$\lim_{\alpha\to-\infty} p(xy) = 0 \quad \lim_{\alpha\to\infty} p(xy) = \frac{1}{6}.$$

Парная ковариация будет иметь вид (Рис. 4)

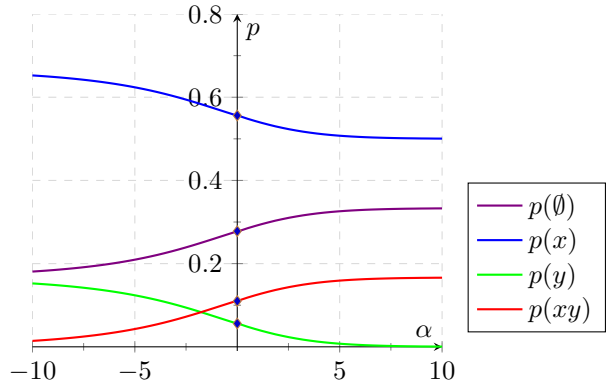


Рисунок 3: Зависимость вероятностей I-го рода от α, $\alpha \neq 0$ при $p_x = \frac{1}{6}$ и $p_y = \frac{2}{3}$.

$$\mathrm{Kov}_{xy} = \ln\left(1 + \frac{(e^{-\alpha\cdot\frac{1}{6}} - 1)(e^{-\alpha\cdot\frac{2}{3}} - 1)}{(e^{-\alpha} - 1)}\right)^{-\frac{1}{\alpha}} - \frac{1}{9}.$$

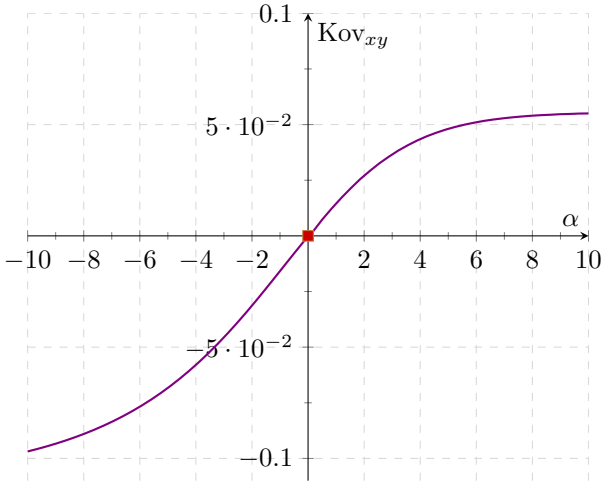


Рисунок 4: Зависимость парной ковариации семейства вероятностных распределений Франка от параметра $\alpha \in [-10; 10] \setminus \{0\}$ при $p_x = \frac{1}{6}$ и $p_y = \frac{2}{3}$.

Пример 2. Пусть $\mathfrak{X} = \{x, y, z\}$ и пусть известны вероятности событий $p_x = \frac{3}{8}$, $p_y = \frac{1}{8}$ и $p_z = \frac{5}{8}$. Тогда из (9) и (12) получаем

$$p_{xy} = \ln\left(1 + \frac{(e^{-\alpha\cdot\frac{3}{8}} - 1)(e^{-\alpha\cdot\frac{1}{8}} - 1)}{(e^{-\alpha} - 1)}\right)^{-\frac{1}{\alpha}};$$

$$p_{xz} = \ln\left(1 + \frac{(e^{-\alpha\cdot\frac{3}{8}} - 1)(e^{-\alpha\cdot\frac{5}{8}} - 1)}{(e^{-\alpha} - 1)}\right)^{-\frac{1}{\alpha}};$$

$$p_{yz} = \ln\left(1 + \frac{(e^{-\alpha\cdot\frac{1}{8}} - 1)(e^{-\alpha\cdot\frac{5}{8}} - 1)}{(e^{-\alpha} - 1)}\right)^{-\frac{1}{\alpha}};$$

$$p_{xyz} = \ln\left(1 + \frac{(e^{-\alpha\cdot\frac{3}{8}} - 1)(e^{-\alpha\cdot\frac{1}{8}} - 1)(e^{-\alpha\cdot\frac{5}{8}} - 1)}{(e^{-\alpha} - 1)^2}\right)^{-\frac{1}{\alpha}}.$$

По формулам обращения Мёбиуса (2) перейдем к вероятностному распределению I-го рода (Рис. 5):

$$p(x) = \ln\left(\frac{e^{\alpha\frac{3}{8}}(E_\varnothing + E_xE_y)(E_\varnothing + E_xE_z)}{(E_\varnothing^2 + E_xE_yE_z)}\right)^{\frac{1}{\alpha}};$$

$$p(y) = \ln\left(\frac{e^{\alpha\frac{1}{8}}(E_\varnothing + E_xE_y)(E_\varnothing + E_yE_z)}{(E_\varnothing^2 + E_xE_yE_z)}\right)^{\frac{1}{\alpha}};$$

$$p(z) = \ln\left(\frac{e^{\alpha\frac{5}{8}}(E_\varnothing + E_xE_z)(E_\varnothing + E_yE_z)}{(E_\varnothing^2 + E_xE_yE_z)}\right)^{\frac{1}{\alpha}};$$

$$p(xy) = \ln\left(\frac{E_\varnothing^2 + E_\varnothing E_xE_y}{E_\varnothing^2 + E_xE_yE_z}\right)^{-\frac{1}{\alpha}};$$

$$p(xz) = \ln\left(\frac{E_\varnothing^2 + E_\varnothing E_xE_z}{E_\varnothing^2 + E_xE_yE_z}\right)^{-\frac{1}{\alpha}};$$

$$p(yz) = \ln\left(\frac{E_\varnothing^2 + E_\varnothing E_yE_z}{E_\varnothing^2 + E_xE_yE_z}\right)^{-\frac{1}{\alpha}};$$

$$p(xyz) = \ln\left(\frac{E_\varnothing^2}{E_\varnothing^2 + E_xE_yE_z}\right)^{\frac{1}{\alpha}}; \quad p(\varnothing) =$$

$$= \ln\left(\frac{(E_\varnothing + E_xE_y)(E_\varnothing + E_xE_z)(E_\varnothing + E_yE_z)}{e^{-\alpha\frac{1}{8}}E_\varnothing(E_\varnothing^2 + E_xE_yE_z)}\right)^{-\frac{1}{\alpha}}.$$

где $E_x = e^{-\alpha\cdot p_x} - 1$ для $x \in \mathfrak{X}$; $E_\varnothing = e^{-\alpha} - 1$.

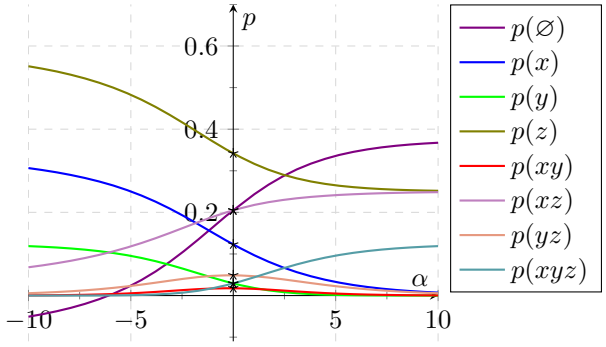


Рисунок 5: Зависимость вероятностей I-го рода от α, $\alpha \neq 0$ при $p_x = \frac{3}{8}$, $p_y = \frac{1}{8}$ и $p_z = \frac{5}{8}$.

Заметим, что пределы

$$\begin{aligned}
&\lim_{\alpha\to-\infty} p(\varnothing) = -\frac{1}{8}; && \lim_{\alpha\to\infty} p(\varnothing) = \frac{3}{8};\\
&\lim_{\alpha\to-\infty} p(x) = \frac{3}{8}; && \lim_{\alpha\to\infty} p(x) = 0;\\
&\lim_{\alpha\to-\infty} p(y) = \frac{1}{8}; && \lim_{\alpha\to\infty} p(y) = 0;\\
&\lim_{\alpha\to-\infty} p(z) = \frac{5}{8}; && \lim_{\alpha\to\infty} p(z) = \frac{1}{4};\\
&\lim_{\alpha\to-\infty} p(xy) = 0; && \lim_{\alpha\to\infty} p(xy) = 0;\\
&\lim_{\alpha\to-\infty} p(xz) = 0; && \lim_{\alpha\to\infty} p(xz) = \frac{1}{4};\\
&\lim_{\alpha\to-\infty} p(yz) = 0; && \lim_{\alpha\to\infty} p(yz) = 0;\\
&\lim_{\alpha\to-\infty} p(xyz) = 0; && \lim_{\alpha\to\infty} p(xyz) = \frac{1}{8};\\
&\sum_{X\subseteq\mathfrak{X}} \lim_{\alpha\to-\infty} p(X) = 1; && \sum_{X\subseteq\mathfrak{X}} \lim_{\alpha\to\infty} p(X) = 1.
\end{aligned}$$

Поскольку, $\lim\limits_{\alpha\to-\infty} p(\varnothing) = -\frac{1}{8}$, то решим неравенство $p(\varnothing) \geq 0$ относительно параметра α. Получаем $\alpha > -6.077558351$. Таким образом, получили диапазон[1] для $\alpha \in [-6; 0) \cup (0; +\infty)$, при которых получаются легитимные вероятностные распределения.

По формуле (11) найдем арные ковариации, которые будут иметь вид (Рис. 6).

$$\mathrm{Kov}_{xy} = p_{xy} - \frac{3}{64}; \quad \mathrm{Kov}_{xz} = p_{xz} - \frac{15}{64};$$

$$\mathrm{Kov}_{yz} = p_{yz} - \frac{5}{64}; \quad \mathrm{Kov}_{xyz} = p_{xyz} - \frac{15}{64}.$$

Итак, для семейства случайных множеств событий Франка общая формулы вероятностных распределений II-го рода и I-го рода имеют следующий вид:

$$p_X = -\frac{1}{\alpha}\ln\left(1 + \frac{\prod\limits_{x\in X}(e^{-\alpha p_x} - 1)}{(e^{-\alpha} - 1)^{|X|-1}}\right), X \subseteq \mathfrak{X},$$

$$p(X) = \ln \prod_{Y\supseteq X}\left(1 + \frac{\prod\limits_{x\in Y}(e^{-\alpha p_x} - 1)}{(e^{-\alpha} - 1)^{|Y|-1}}\right)^{\frac{(-1)^{|Y|-|X|+1}}{\alpha}}.$$

[1] сделали округление до -6

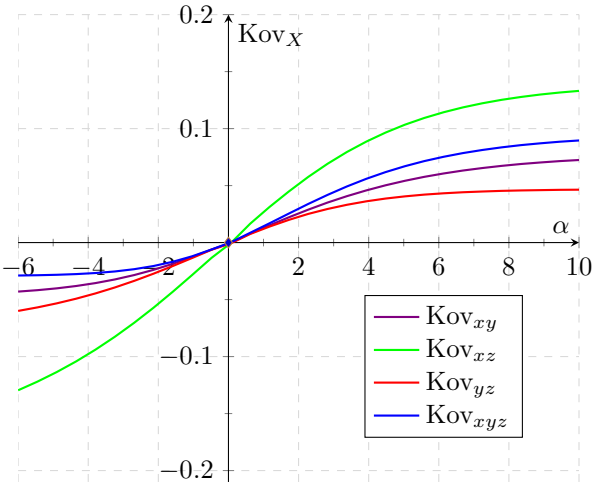


Рисунок 6: Зависимость арной ковариации семейства вероятностных распределений Франка от параметра $\alpha \in [-6; 10] \setminus \{0\}$ при $p_x = \frac{3}{8}$, $p_y = \frac{1}{8}$ и $p_z = \frac{5}{8}$.

3 Заключение

В работе предложен метод построения ассоциативных множеств событий для семейства Франка, который продемонстрирован на примерах дуплета и триплета событий. В этих примерах найдены арные ковариации для анализа структуры вероятностной зависимости, полученных вероятностных распределений.

Список литературы

[1] О. Ю. Воробьев. Эвентология. *учебник /О.Ю.Воробьев*, Красноярск: Сибирский федеральный университет:435, 2007.

[2] Ю.С. Кочанова. Эвентологические распределения случайных множеств событий на основе ассоциативных функций. *дис. ... магистра мат. и комп. наук*, 010200.68.03 Кочанова Юлия Сергеевна:52, 2013.

[3] M.J.Frank. *An Introduction to Copulas (Second Edition)*. Aequationes Math. 19, 1979.

[4] E.P. Klement, R. Mesiar. *Logical, Algebraic, Analytic and Probabilistic Aspects of Triangular Norms*. Dordrecht: Kluwer, 2005.

[5] R.B.Nelsen. *An Introduction to Copulas (Second Edition)*. New York: Springer Science+Business Media, Inc., 2006.

Оценка параметра ассоциативной функции методом наименьших квадратов для семейств Франка и Али-Михаэля-Хака случайных множеств событий

Дарья Владиславовна Семенова
Институт математики и фундаментальной информатики
Сибирский Федеральный университет
Красноярск
dariasdv@gmail.com

Наталья Александровна Лукьянова
Институт математики и фундаментальной информатики
Сибирский Федеральный университет
Красноярск
nata00sfu@gmail.com

Людмила Юрьевна Шангареева
Институт математики и фундаментальной информатики
Сибирский Федеральный университет
Красноярск
mizkuzy@yandex.ru.ru

Аннотация. *В статье рассматривается метод наименьших квадратов для нахождения параметра ассоциативной функции применительно к семействам Франка и Али-Михаэля-Хака случайных множеств событий, построение легитимного вероятностного распределения для каждого найденного параметра.*

Ключевые слова. *Случайные множества событий, метод наименьших квадратов, ассоциативное случайное множество событий Франка, ассоциативное случайное множество событий Али-Михаэля-Хака, вероятностное распределение.*

1 Введение

Чтобы статистически оценить вероятностное распределение случайного множества N событий на основе наблюдений приходится сталкиваться с проблемой хронической недостаточности имеющегося количества наблюдений для надлежащей оценки всех 2^N вероятностей $p(X)$, образующих данное вероятностное распределение. В общей ситуации преодолеть эту проблему прямым увеличением статистики наблюдений не представляется возможным, поскольку для качественной статистической оценки 2^N вероятностей требуется, по крайней мере, не меньшее число наблюдений. Типичная особенность статистиких наблюдений за множествами событий заключается в том, что даже при не очень больших $N = |\mathfrak{X}|$ число оцениваемых вероятностей, равное 2^N, оказывается больше, а часто и гораздо больше числа наблюдений: $n \ll 2^N$. Однако, статистическая оценка N вероятностей событий $\mathbf{P}(x)$ из множества $\mathfrak{X}$ может быть проведена с заданным уровнем надежности без особых проблем [4]. В работах [1, 3, 2, 4] были предложены методы получения вероятностных распределений, имея в распоряжении только N вероятностей событий. В данной работе для задания вероятностных распределений, имея на руках Nвероятностей событий, предлагается использовать ассоциативные функции(t-нормы, копулы). В статье рассматривается метод наименьших квадратов для нахождения параметра ассоциативной функции применительно к семействам Франка и Али-Михаэля-Хака случайных множеств событий, построение легитимного вероятностного распределения для каждого найденного параметра.

© 2014 Д.В.Семенова, Н.А.Лукьянова, Л.Ю.Шангареева
Олег Воробьев (ред.), Труды XIII ФАМЭМС'2014, Красноярск: СФУ

2 Предварительные сведения

Случайное множество событий K на конечном множестве событий $\mathfrak{X} \subset \mathcal{F}$ определяется как отображение $K : \Omega \to 2^{\mathfrak{X}}$, измеримое относительно пары алгебр $\left(\mathcal{F}, 2^{2^{\mathfrak{X}}}\right)$ в том смысле, что для всякого $X \in 2^{2^{\mathfrak{X}}}$ существует прообраз $K^{-1}(X) \in \mathcal{F}$, такой что $\mathbf{P}(X) = \mathbf{P}(K^{-1}(X))$.

Вероятностное распределение случайного множества событий можно представить несколькими эквивалентными распределениями вероятностей, порождённых этим множеством событий [1]:

- вероятностное распределение I-го рода — набор из $2^{|\mathfrak{X}|}$ вероятностей вида $\{p(X), X \subseteq \mathfrak{X}\}$, где
$$p(X) = \mathbf{P}(K = X) = \mathbf{P}\left(\bigcap_{x \in X} x \bigcap_{x \in X^c} x^c\right);$$
- вероятностное распределение II-го рода — набор из $2^{|\mathfrak{X}|}$ вероятностей вида $\{p_X, X \subseteq \mathfrak{X}\}$, где
$$p_X = \mathbf{P}(X \subseteq K) = \mathbf{P}\left(\bigcap_{x \in X} x\right).$$

В теории случайных событий обозначение $\emptyset$ используется для $\mathfrak{X}^c = \Omega \setminus \mathfrak{X}$. Событие $\emptyset = \bigcap\limits_{x \in \mathfrak{X}} x^c$ означает, что не наступило ни одно событие из $\mathfrak{X}$. В вероятностном распределении II-го рода всегда $p_\emptyset = 1$.

Только распределение I-го рода обладает привычным для распределения вероятности свойством нормировки, которое очевидно следует из того, что со-

ответствующие события $\left(\bigcap\limits_{x\in X} x\right)\bigcap\left(\bigcap\limits_{x\in X^c} x^c\right)$ образуют разбиение пространства элементарных событий Ω. А поскольку события $\bigcap\limits_{x\in X} x$ образуют не разбиение, а всего лишь покрытие Ω, то соотношение нормировки для вероятностного распределения II-го рода не выполняется, сумма вероятностей этих событий всегда больше единицы.

Вероятностное распределение II-го рода $\{p_X, X\subseteq\mathfrak{X}\}$ случайного множества событий K на $\mathfrak{X}$ удовлетворяет системе из 2^N неравенств Фреше–Хёвдинга:

$$0\le p_X^-\le p_X\le p_X^+\le 1,$$

где $p_X^- = \max\left\{0, 1-\sum\limits_{x\in X}(1-p_x)\right\}$ — нижняя граница Фреше–Хёвдинга, $p_X^+ = \min\limits_{x\in X} p_x$ — верхняя граница Фреше–Хёвдинга.

Вероятностные распределения I-го и II-го рода связаны взаимно-обратными формулами обращения Мёбиуса [1]

$$p_X = \sum_{Y\in 2^{\mathfrak{X}}:\, X\subseteq Y} p(Y), \tag{1}$$

$$p(X) = \sum_{Y\in 2^{\mathfrak{X}}:\, X\subseteq Y} (-1)^{|Y|-|X|} p_Y, \tag{2}$$

для всех $X\in 2^{\mathfrak{X}}$.

Если задано вероятностное распределение I-го рода $\{p(X), X\subseteq\mathfrak{X}\}$, то по формуле (1) мы всегда получим вероятностное распределение II-го рода $\{p_X, X\subseteq\mathfrak{X}\}$. Однако преобразование (2) заданного набора из 2^N чисел $\{p_X,\ X\subseteq\mathfrak{X}\}$, удовлетворяющих границам Фреше–Хёвдинга, может привести к вероятностному распределению I-рода с отрицательными значениями.

Будем говорить, что случайное множество событий K обладает легитимным вероятностным распределением, если для вероятностного распределения I-го рода выполнены следующие условия:

$$0\le p(X)\le 1, X\subseteq\mathfrak{X}, \tag{3}$$

$$\sum_{X\subseteq\mathfrak{X}} p(X) = 1. \tag{4}$$

В современных теориях неопределенности широкое распространение получили классы ассоциативных функций, в частности, треугольные нормы и копулы. Алгоритмы вычисления копул и треугольных норм применимы к вероятностным распределениям случайных множеств событий.

Определение 1. *Ассоциативная функция в теории случайных множеств событий (Associative Function — AF)*

$$\mathrm{AF}: [0,1]^2\to[0,1]$$

определяется как двуместная функция, удовлетворяющая следующим аксиомам:

A1. Граничные условия, для $a\in[0,1]$

$$\begin{aligned}\mathrm{AF}(a,0) = \mathrm{AF}(0,a) = 0,\\ \mathrm{AF}(a,1) = \mathrm{AF}(1,a) = a.\end{aligned} \tag{5}$$

A2. Монотонность, когда $a_1\le a_2, b_1\le b_2$

$$\mathrm{AF}(a_1,b_1)\le\mathrm{AF}(a_2,b_2). \tag{6}$$

A3. Коммутативность, т.е. для всех a, b *на* $[0,1]$

$$\mathrm{AF}(a,b) = \mathrm{AF}(b,a).$$

A4. Ассоциативность, т.е. для всех a, b, $c\in[0,1]$

$$\mathrm{AF}(\mathrm{AF}(a,b),c) = \mathrm{AF}(a,\mathrm{AF}(b,c)). \tag{7}$$

A5. Условие Липшиц-непрерывности

$$\mathrm{AF}(c,b)-\mathrm{AF}(a,b)\le c-a,\qquad a\le c. \tag{8}$$

Геометрически график ассоциативной функции — это поверхность (рис. 1), которая «натянута» на четырехугольник с вершинами $(0,0,0)$, $(0,1,0)$, $(1,0,0)$ и $(1,1,1)$. На рис. 1 стороны четырехугольника выделены жирными линиями и соответствуют граничным условиям (A1). Согласно свойству A2 ассоциативная функция возрастает по вертикали и по горизонтали [7], а по свойству A3 симметрична относительно плоскости $a=b$. Свойства A1 — A3, A5 гарантируют, что значение функции $\mathrm{AF}(a,b)$ будет обладать свойствами вероятности. Свойство A4 позволяет рекуррентно перейти к n-местной функции.

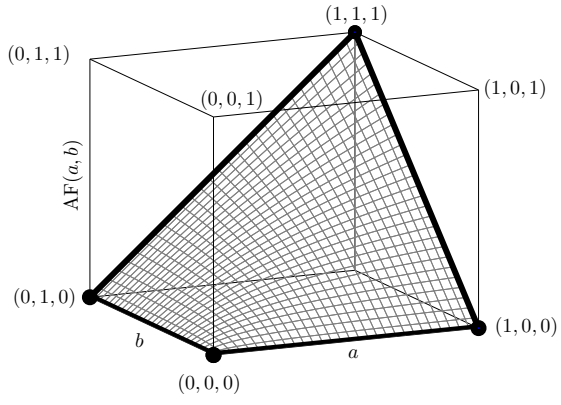


Рисунок 1: График произвольной ассоциативной функции

Заметим, что аксиомы A1-A4 соответствует определению t-нормы [6]. Таким образом, под ассоциативной функцией мы будем понимать непрерывную t-норму, удовлетворяющую условию Липшица, или,

что эквивалентно, ассоциативную, коммутативную копулу [7] (Рис. 1).

Ассоциативным случайным множеством событий называется случайное множество событий K под конечным множеством избранных событий $\mathfrak{X}$ с вероятностным 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 $p_x = \mathbf{P}(x)$, $x \in \mathfrak{X}$

$$p_{xy} = \mathbf{P}\left(x \bigcap y\right) = \mathrm{AF}\left(p_x, p_y\right),$$

$$p_{xyz} = \mathbf{P}\left(x \bigcap y \bigcap z\right) = \mathrm{AF}\left(p_x, \mathrm{AF}(p_y, p_z)\right) =$$

$$= \mathrm{AF}\left(p_x, \mathbf{P}\left(y \bigcap z\right)\right),$$

$$p_X = \mathbf{P}\left(\bigcap_{x \in X} x\right) = \mathrm{AF}\left(p_x, \mathbf{P}\left(\bigcap_{y \in X \setminus \{x\}} y\right)\right),$$

при условии, что соответствующее вероятностное распределение I-го рода будет легитимным.

Теорема 1. *Пусть вероятности событий $p_x = \mathbf{P}(x) > 0$, $x \in \mathfrak{X}$, тогда ассоциативная функция*

$$\mathrm{AF}_\alpha(a, b) = Frank(a, b, \alpha) =$$

$$= -\frac{1}{\alpha} \ln\left(1 + \frac{(e^{-\alpha \cdot a} - 1)(e^{-\alpha \cdot b} - 1)}{(e^{-\alpha} - 1)}\right),$$

где $\alpha \in (-\infty; \infty) \setminus \{0\}$, определяет ассоциативное случайное множество Франка с вероятностным распределением II-го рода:

$$p_X = -\frac{1}{\alpha} \ln\left(1 + \frac{\prod\limits_{x \in X} (e^{-\alpha p_x} - 1)}{(e^{-\alpha} - 1)^{|X|-1}}\right), X \subseteq \mathfrak{X}, \quad (9)$$

и с вероятностным распределением I-го рода:

$$p(X) = \ln \prod_{Y \supseteq X} \left(1 + \frac{\prod\limits_{x \in Y} (e^{-\alpha p_x} - 1)}{(e^{-\alpha} - 1)^{|Y|-1}}\right)^{\frac{(-1)^{|Y|-|X|+1}}{\alpha}}.$$

Ассоциативное случайное множество Франка будет иметь легитимное вероятностное распределение, если все $p_x, x \in \mathfrak{X}$ и параметр $\alpha \neq 0$ удовлетворяют следующей системе из $2^{|\mathfrak{X}|}$ неравенств для всех $X \subseteq \mathfrak{X}$

$$1 \leq \prod_{Y \supseteq X} \left(1 + \frac{\prod\limits_{x \in Y} (e^{-\alpha p_x} - 1)}{(e^{-\alpha} - 1)^{|Y|-1}}\right)^{\frac{(-1)^{|Y|-|X|+1}}{\alpha}} \leq e.$$

Теорема 2. *Пусть вероятности событий $p_x = \mathbf{P}(x)$, $x \in \mathfrak{X}$ удовлетворяют неравенствам $0 < p_x < 1$. Тогда ассоциативная функция*

$$\mathrm{AF}_\alpha(a, b) = \frac{ab}{1 - \alpha(1 - a)(1 - b)}, \quad \alpha \in [0, 1]$$

определяет семейство $\mathcal{R}_{AMH}$ ассоциативных случайных множеств Али-Михаэля-Хака с вероятностным распределением 2-го рода для всех $X \subseteq \mathfrak{X}$, $|X| > 1$

$$p_X = \frac{\prod\limits_{x \in X} p_x}{(1 - \alpha)^{|X|-1} + \sum\limits_{k=1}^{|X|-1} \alpha^k \cdot \left[S - \prod\limits_{x \in X} p_x\right]}, \quad (10)$$

где $S = \sum\limits_{Y \subseteq C_X^k} \left[(-1)^{k-|Y|} \cdot \delta_k(Y) \cdot \prod\limits_{x \in Y} p_x\right]$,

$C_X^k = \{Y : Y \subseteq X, |Y| = k\}$,

$$\delta_k(Y) = \begin{cases} 1, & k = 1, \text{ или } k = |X| - 1; \\ k - |Y| + 1, & 1 < k < |X| - 1. \end{cases}$$

3 Оценка параметра ассоциативной функции методом наименьших квадратов

Метод наименьших квадратов (МНК) — метод оценки параметров модели на основании экспериментальных данных, содержащих случайные ошибки. В основе метода лежат следующие рассуждения: при замене точного (неизвестного) параметра модели приблизительным значением необходимо минимизировать разницу между экспериментальными данными и теоретическими (вычисленными при помощи предложенной модели). Это позволяет рассчитать параметры модели с помощью МНК с минимальной погрешностью. Мерой разницы в методе наименьших квадратов служит сумма квадратов отклонений действительных (экспериментальных) значений от теоретических. Выбираются такие значения параметров модели, при которых сумма квадратов разностей будет наименьшей.

Рассмотрим следующую задачу. Пусть $(\Omega, \mathcal{F}, \mathbf{P})$ — вероятностное пространство. Рассмотрим множество событий $\mathfrak{X}$, выбранных из алгебры событий $\mathcal{F}$. Пусть K ассоциативное случайное множество событий, определяемое однопараметрической функцией $\mathrm{AF}_\alpha(a, b)$, $a, b \in [0; 1]$ и $\alpha \in \mathcal{D}$, где $\mathcal{D} \subseteq \mathbb{R}$ — область допустимых значений α.

Пусть имеется статистика наблюдений, по результатам которой получены

- оценки вероятностей событий $\hat{p}_x$, $x \in \mathfrak{X}$;
- и оценки для некоторых вероятностей пересечений событий $\hat{p}_X$, где $X \in \mathfrak{Y} \subseteq 2^{\mathfrak{X}}$, $|\mathfrak{Y}| \ll 2^{|\mathfrak{X}|}$.

Таким образом, по существующим статистическим наблюдениям из множества $2^{\mathfrak{X}}$ мы выбираем некоторое подмножество $\mathfrak{Y}$, элементами которого являются некоторые пересечения событий $\bigcap\limits_{x \in X} x$, вероятности которых известны. Исходя из этих данных получим

оценку параметра α по методу наименьших квадратов.

Оценка параметра ассоциативной функции по МНК.

Дано:

- множество событий $\mathfrak{X}$, выбранных из алгебры событий $\mathcal{F}$;
- оценка вероятностей событий $\hat{p}_x$, $x \in \mathfrak{X}$;
- вид однопараметрической ассоциативной функции $\mathrm{AF}_\alpha(a,b)$, $a,b \in [0;1]$ и $\alpha \in \mathcal{D}$, где $\mathcal{D} \subseteq \mathbb{R}$ — область допустимых значений α;
- множество $\mathfrak{Y} \subseteq 2^{\mathfrak{X}}$, для элементов которого получены оценки $\{\hat{p}_X, \quad X \in \mathfrak{Y}, |X| > 1\}$, удовлетворяющие неравенству Фреше-Хёвдинга

$$0 \le p_X^- \le \hat{p}_X \le p_X^+ \le 1,$$

где $p_X^- = \max\left\{0, 1 - \sum\limits_{x \in X}(1 - \hat{p}_x)\right\}$ — нижняя граница Фреше–Хёвдинга, $p_X^+ = \min\limits_{x \in X} \hat{p}_x$ — верхняя граница Фреше–Хёвдинга.

Найти: значения параметра $\alpha \in \mathcal{D}$, доставляющего минимум функционалу

$$\Phi(\hat{p}_x, x \in X; \alpha) =$$

$$= \left(\mathrm{AF}_\alpha(\hat{p}_x, x \in X) - \hat{p}_X\right)^2 \longrightarrow \min_{\alpha \in \mathcal{D}}, \quad X \in \mathfrak{Y}, \qquad (11)$$

для каждого элемента множества $\mathfrak{Y}$ Запишем для (11) необходимое условие экстремума

$$\frac{\partial \Phi(\hat{p}_x, x \in X; \alpha)}{\partial \alpha} = 0, \quad X \in \mathfrak{Y}$$

т.е. необходимо найти решение уравнения

$$\left(\mathrm{AF}_\alpha(\hat{p}_x, x \in X) - \hat{p}_X\right) \cdot \frac{\partial \mathrm{AF}_\alpha(\hat{p}_x, x \in X)}{\partial \alpha} = 0$$

при ограничениях $\alpha \in \mathcal{D}$.

Рассмотрим решение задачи на примере однопараметрических семейств Франка и Али-Михаэля-Хака, взяв для анализа входные данные из примера рассмотренного в [5].

Пример 1.Оценка параметра для семейства Франка.

Дано:

- $\mathfrak{X} = \{x, y, z\}$;
- $\hat{p}_x = \dfrac{3}{8}$, $\hat{p}_y = \dfrac{1}{8}$, $\hat{p}_z = \dfrac{5}{8}$;
- Ассоциативная функция Франка

$$\mathrm{AF}_\alpha(a,b) = Frank(a,b,\alpha) =$$

$$= -\frac{1}{\alpha}\ln\left(1 + \frac{(e^{-\alpha \cdot a} - 1)(e^{-\alpha \cdot b} - 1)}{(e^{-\alpha} - 1)}\right);$$

- $\mathfrak{Y} = \{\{xy\}, \{xz\}\}$, $\hat{p}_{xy} = 0.1$, $\hat{p}_{xz} = 0.25$, где оценки вероятностей пересечения событий удовлетворяют неравенству Фреше-Хёвдинга

$$p_{xy}^- = \max\{0, 1 - (1 - \hat{p}_x) - (1 - \hat{p}_y)\} = 0,$$
$$p_{xy}^+ = \min\{\hat{p}_x, \hat{p}_y\} = 0.125,\ p_{xy}^- \le \hat{p}_{xy} \le p_{xy}^+,$$

$$p_{xz}^- = \max\{0, 1 - (1 - \hat{p}_x) - (1 - \hat{p}_z)\} = 0,$$
$$p_{xz}^+ = \min\{\hat{p}_x, \hat{p}_z\} = 0.375,\ p_{xz}^- \le \hat{p}_{xz} \le p_{xz}^+.$$

Найти:

- параметр $\alpha \in (-\infty, +\infty) \setminus \{0\}$.

Решение:

Пусть

$$\mathrm{AF}_\alpha(\hat{p}_x, \hat{p}_y) = Frank(\hat{p}_x, \hat{p}_y, \alpha) =$$

$$= -\frac{1}{\alpha}\ln\left(1 + \frac{(e^{-\alpha \cdot \frac{3}{8}} - 1)(e^{-\alpha \cdot \frac{1}{8}} - 1)}{(e^{-\alpha} - 1)}\right);$$

$$\mathrm{AF}_\alpha(\hat{p}_x, \hat{p}_z) = Frank(\hat{p}_x, \hat{p}_z, \alpha) =$$

$$= -\frac{1}{\alpha}\ln\left(1 + \frac{(e^{-\alpha \cdot \frac{3}{8}} - 1)(e^{-\alpha \cdot \frac{5}{8}} - 1)}{(e^{-\alpha} - 1)}\right);$$

Находим значения α отдельно для каждого пересечения событий, для которых получены оценки вероятностей.

Решаем уравнение:

$$\left[\left(\mathrm{AF}_\alpha(\hat{p}_x, \hat{p}_y) - \frac{1}{10}\right)^2\right]'_\alpha = 0.$$

Используя инструментарий Mathcad, получаем значение $\alpha = 4.871143485$.

По формуле (9) и формулам обращения Мёбиуса получаем легитимное распределение I-го рода для этого триплета:

$$\{p(\varnothing); p(x); p(y); p(z); p(xy); p(xz); p(yz); p(xyz)\} =$$

$$= \{0.334; 0.034; 0.002; 0.266; 0.005; 0.241; 0.023; 0.095\}.$$

Решаем уравнение:

$$\left[\left(\mathrm{AF}_\alpha(\hat{p}_x, \hat{p}_z) - \frac{1}{4}\right)^2\right]'_\alpha = 0.$$

Используя инструментарий Mathcad, получаем значение $\alpha \approx 0.5745$ (сделали округление до 10^{-4}).

По формуле (9) и формулам обращения Мёбиуса получаем легитимное распределение I-го рода для этого триплета:

$$\{p(\varnothing); p(x); p(y); p(z); p(xy); p(xz); p(yz); p(xyz)\} =$$

$$= \{0.227; 0.108; 0.023; 0.327; 0.017; 0.213; 0.048; 0.037\}.$$

В примере 2. [5] для триплета событий $\mathfrak{X} = \{x, y, z\}$ с известными вероятностями событий $\hat{p}_x = \frac{3}{8}$, $\hat{p}_y = \frac{1}{8}$, $\hat{p}_z = \frac{5}{8}$ с помощью ассоциативной функции Франка был получен диапазон для $\alpha \in [-6; 0) \cup (0; +\infty)$, при которых получаются легитимные вероятностные распределения. Имея дополнительные оценки входных данных $\hat{p}_{xy} = 0.1$, $\hat{p}_{xz} = 0.25$, получаем распределения для $\alpha = 0.5745$; $\alpha = 4.871$.

Пример 2.Оценка параметра для семейства Али-Михаэля-Хака

Дано:

- $\mathfrak{X} = \{x, y, z\}$;
- $\hat{p}_x = \frac{3}{8}$, $\hat{p}_y = \frac{1}{8}$, $\hat{p}_z = \frac{5}{8}$;
- Ассоциативная функция Али-Михаэля-Хака

$$\mathrm{AF}_\alpha(a, b) = \frac{ab}{1 - \alpha(1 - a)(1 - b)};$$

- $\mathfrak{Y} = \{\{xy\}, \{xz\}\}$, $\hat{p}_{xy} = 0.1$, $\hat{p}_{xz} = 0.25$.

Найти:

- параметр $\alpha \in [0, 1]$.

Решение:

Пусть

$$\mathrm{AF}_\alpha(\hat{p}_x, \hat{p}_y) = \frac{\hat{p}_x \cdot \hat{p}_y}{1 - \alpha(1 - \hat{p}_x)(1 - \hat{p}_y)} = \frac{3}{64 - 35\alpha};$$

$$\mathrm{AF}_\alpha(\hat{p}_x, \hat{p}_z) = \frac{\hat{p}_x \cdot \hat{p}_z}{1 - \alpha(1 - \hat{p}_x)(1 - \hat{p}_z)} = \frac{15}{64 - 15\alpha}.$$

Находим значения α отдельно для каждого пересечения событий, для которых получены оценки вероятностей.

Решаем уравнение:

$$\left[\left(\mathrm{AF}_\alpha(\hat{p}_x, \hat{p}_y) - \frac{1}{10}\right)^2\right]'_\alpha = 0.$$

Используя инструментарий Mathcad, получаем значение $\alpha \approx 0.971$ (сделали округление до 10^{-3}).

По формуле (10) и формулам обращения Мёбиуса получаем легитимное распределение I-го рода для этого триплета:

$$\{p(\varnothing); p(x); p(y); p(z); p(xy); p(xz); p(yz); p(xyz)\} =$$

$$= \{0.3; 0.064; 0.003; 0.3; 0.007; 0.21; 0.022; 0.093\}.$$

Решаем уравнение:

$$\left[\left(\mathrm{AF}_\alpha(\hat{p}_x, \hat{p}_z) - \frac{1}{4}\right)^2\right]'_\alpha = 0.$$

Используя инструментарий Mathcad, получаем значение $\alpha \approx 0.267$ (сделали округление до 10^{-3}).

По формуле (10) и формулам обращения Мёбиуса получаем легитимное распределение I-го рода для этого триплета:

$$\{p(\varnothing); p(x); p(y); p(z); p(xy); p(xz); p(yz); p(xyz)\} =$$

$$= \{0.228; 0.108; 0.022; 0.327; 0.017; 0.212; 0.048; 0.038\}.$$

Список литературы

[1] О. Ю. Воробьев. Эвентология. *учебник /О.Ю.Воробьев*, Красноярск: Сибирский федеральный университет:435, 2007.

[2] О. Ю. Воробьев. Обзор теории копулы и эвентологическая теория копулы. *Труды XIII Всероссийской ФАМ'2006 конференции*, Красноярск: СФУ, 1 (под ред. О. Ю. Воробьева):56–94, 2007.

[3] О. Ю. Воробьев. Широкая зависимость событий и аппроксимация эвентологических распределений широкомультипликативными сет-функциями. *Труды VIII международной ФАМ'2009 конференции*, Красноярск: СФУ, 1 (под ред. О. Ю. Воробьева):101–122, 2009.

[4] О. Ю. Воробьев. Фреше-граничные эвентологические распределения и их применение. *Труды XIV международной ЭМ'2010 конференции*, Красноярск: СФУ, 1 (под ред. О. Ю. Воробьева):57–69, 2010.

[5] Д.В.Семенова, Н.А.Лукьянова, Л.Ю.Шангареева. Семейство случайных множеств событий Франка. *Труды XIII междунар. ФАМЭМС'2014 конференции*, Красноярск (под ред. О. Ю. Воробьева):в этом сборнике, 2014.

[6] E.P. Klement, R. Mesiar. *Logical, Algebraic, Analytic and Probabilistic Aspects of Triangular Norms.* Dordrecht: Kluwer, 2005.

[7] R.B.Nelsen. *An Introduction to Copulas (Second Edition).* New York: Springer Science+Business Media, Inc., 2006.

Список авторов

Абдушукуров Абдурахим Ахмедович 11, 13, 23
Алманова Дильноза Султановна 19
Бабенышев Сергей Валерьевич 158
Баранова Ирина Владимировна 27, 35, 40, 170
Богодухов Дмитрий Михайлович 35
Воробьев Олег Юрьевич 44, 95, 112, 122
Голденок Елена Евгеньевна 131
Горный Борис Эмануилович 158
Гребнев Ярослав Владимирович 204
Григорьев Юрий Дмитриевич 136
Демиденко Виталий Анатольевич 178
Демочкин Валерий 15
Джамирзаев Абдусуннат Абдурахмонович 142
Додонова Мария Михайловна 143
Есин Роман Витальевич 147
Жумакулов Хуршид Кодиралиевич 21
Зупаров Талат Маруфович 153
Иванова Анастасия Игоревна 222
Имомов Аъзам Абдурахимович 19, 21
Какаджанова Лейла Рэшитовна 11
Клоков Андрей Владимирович 155
Князь Димитрия Владимировна 40
Крупкина Татьяна Валерьевна 158
Кустицкая Татьяна Алексеевна 163
Кушбак кызы Гулиза 228
Лукьянова Наталья Александровна 234, 239
Мажаров Владимир Федорович 158
Максимова Кристина Игоревна 170
Махонин Игорь Владиславович 178
Мурадов Рустамжон Собитхонович 23
Нартов Яков Викторович 180
Никитина Ксения Андреевна 208
Нифонтов Александр Сергеевич 186
Новоселов Аркадий Арсеньевич 193, 196, 200
Поддубный Василий Васильевич 213
Поликарпов Анатолий Анатольевич 213
Распопов Виталий Евгеньевич 155
Рахимова Гулноза Гафуровна 202
Сагидуллаев Калмурза Сапарбаевич 202
Садовский Михаил Георгиевич 204, 208
Саркисов Виген Геннадьевич 219
Сафаров Ойбек 153
Семенова Дарья Владиславовна . 222, 228, 234, 239
Тухтаев Эркин Эгамбердиевич 21
Шангареева Людмила Юрьевна 234, 239
Чуянов Хуршид 13

List of authors

Abdushukurov Abdurakhim 11, 13, 23
Almanova Dilnoza 19
Babenyshev Sergey 158
Baranova Irina 27, 35, 40, 170
Bogodukhov Dmitriy 35
Chuyanov Khurshid 13
Demidenko Vitaliy 178
Demotchkine Val 15
Djamirzayev Abdusunnat 142
Dodonova Maria 143
Esin Roman 147
Goldenok Ellen 131
Gornyi Boris 158
Grebnev Yaroslav 204
Grigoriev Yuri 136
Imomov Azam 19, 21
Ivanova Anastasiya 222
Jumakulov Khurshid 21
Klokov Andrey 11
Knyaz' Dimitriya 40
Krupkina Tatiana 158
Kushbak kyzy Guliza 228
Kustitskaya Tatyana 163
Lukyanova Natalia 234, 239
Makhonin Igor 178
Maksimova Kristina 170
Mazharov Vladimir 158
Muradov Rustamjon 23
Nartov Yakov 180
Nifontov Aleksandr 186
Nikitina Ksenya 208
Novosyolov Arcady 193, 196, 200
Poddubny Vasily 213
Polikarpov Anatoly 213
Rakhimova Gulnoza 202
Raspopov Vitaly 155
Sadovsky Michael 208, 204
Sagidullaev Kalmurza 202
Safarov Oybek 153
Sarkisov Vigen 219
Shangareeva Ludmila 234, 239
Semenova Daria 222, 228, 234, 239
Tukhtaev Erkin 21
Tursunov Gafurjon 202
Vorobyev Oleg 44, 95, 112, 122
Zuparov Talat 153

Abstracts

★ **Имомов Аъзам и Дильноза Алманова (Карши, Узбекистан)** 19

Об оценке основного параметра долгоживущего Марковского ветвящегося процесса

Аннотация. *В работе рассматривается задача оценки основного параметра Марковского Q- процесса. Предлагается несмещенная оценка, и исследуются асимптотические свойства ее дисперсий.*
Ключевые слова. *Марковские Q-процессы, Марковские ветвящиеся процессы, оценка основного параметра.*

★ **Имомов Аъзам (Карши, Узбекистан), Жумакулов Хуршид (Коканд, Узбекистан) и Эркин Тухтаев (Карши, Узбекистан)** 21

Об оценке основного параметра в Q-процессе

Аннотация. *В работе рассматривается задача оценки основного параметра Q-процесса. Предлагается несмещенная оценка, и исследуются асимптотические свойства ее дисперсий.*
Ключевые слова. *Q-процессы, ветвящиеся процессы, оценка основного параметра.*

★ **Абдушукуров Абдурахим и Лейла Какаджанова (Ташкент, Узбекистан)** 11

О специальном эмпирическом процессе и его применении

Аннотация. *В работе исследованы асимптотические свойства одного класса эмпирических процессов по определенным классам интегрируемых функций.*
Ключевые слова. *Эмпирические процессы, классы Гливенко-Кантелли, метрическая энтропия.*

★ **Abdushukurov Abdurahim and Rustamjon Muradov (Tashkent, Uzbekistan)** 23

The asymptotic Gaussian properties of copula estimating the survival function for dependent random censoring

Abstract. *In this paper we consider the problem of estimating the survival function for dependent random censoring observations on the right. It is proved the asymptotic Gaussian properties of estimates.*
Keywords. *Random censoring, survival function, Archimedean copula*

★ **Abdushukurov Abdurahim and Khurshid Chuyanov (Uzbekistan, Tashkent)** 13

Estimating of survival function in cox model under competing risks

Abstract. *We propose three types of and corresponding semiparametric estimators for survival functions in Cox model in presence of competing risks.*
Keywords. *Survival function, Cox model, Competing risks.*

★ **Baranova Irina (Krasnoyarsk, Russia)** 27

Covariance analysis of polytypic data using the bipartite sets of events method

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 by correlation.*
Keywords. *Event, probability, covariance, bipartite set of events, set of bipartite set of events*

★ **Baranova Irina and Dmitriy Bogodukhov (Krasnoyarsk, Russia)** 35

Application of genetic algorithms to solve the problem of data clustering

Abstract. *The aim of paper is developing a genetic algorithm for finding solution to the problem of data clustering with a given number of clusters. It is considered the basic concepts, formulation of the problem of data clustering and popular methods of data clustering. This paper describes one of the most popular method of data clustering: k-means method. The principles of genetic algorithms are described. A detailed description of the developed genetic algorithm to solve the problem described clustering are shown. We fulfilled a comparison of the genetic algorithm and k-means method.*
Keywords. *Cluster, data, clustering, metric, functional, genetic algorithm*

★ **Baranova Irina and Dimitriya Knyaz' (Krasnoyarsk, Russia)** 40

Clustering methods of multivariate statistical data

Abstract. *The article discusses the basic concepts, algorithms and methods of clustering. The purpose of cluster analysis is the choice of the necessary tools and techniques for data processing and analysis. In relation to specific research project selected, the statistical procedures that are most relevant for the decision of tasks in relation to specific topics.*
Keywords. *Cluster analysis, object, typology, cluster, k–means, shortest unlocked path, Lance–Williams algorithm.*

★ **Demotchkine Val (London, United Kingdom)** 15

Hedge Fund Strategies, Financial Instability, Fat Tails and Leverage Cycles

Abstract. *The article discusses hedge funds in connection to the latest financial crisis. Hedge funds will continue to use leverage to enhance their alpha generation capabilities, and while we do not believe that leverage is a requirement, we do think that those funds that can obtain and use it appropriately may have a competitive advantage going forward. Indeed, recent market turbulence highlights the importance of focusing on high quality hedge fund firms that have developed well-structured, robust investment processes to address both investment risk and operational risks that can easily translate into investment risk, all while achieving their investment objective through a variety of market environments.*
Keywords. *Hedge Funds, Performance fees, Fat tails, leverage, speculators, risk, volatility, financial markets, Currency speculations, Bank Run, Global Financial Crisis, eventology, Financial Instability, Nash Equilibrium, Government intervention.*

★ **Djamirzayev Abdusunnat (Tashkent, Uzbekistan)** . 142

About the limiting distribution sequences of random variables with random indices

Abstract. *In this paper we generalize a result of the limiting distribution F.J.Anskombe sequences of random variables with random indices.*
Keywords. *Random variables, distribution function, convergence, mixing.*

★ **Dodonova Maria (Krasnoyarsk, Russia)** 143

Research of various formulations of the knapsack problem and methods of problem solution

Abstract. *In paper we study different variants of the knapsack problem formulation. It is considered and implemented the four*

most popular algorithms of solving this problem: exhaustive search, greedy algorithm, genetic algorithm and branch and bound method. We carried out the classification and comparison of these methods in speed and accuracy of finding solutions. Also in article practical knapsack problem is considered.
Keywords. *Knapsack problem, optimization, combinatorics, exhaustive search, greedy algorithm, genetic algorithm, branch and bound method.*

★

Goldenok Ellen (Krasnoyarsk, Russia) 131

Assessment of an association of consumer segments for check market statistics

Abstract. *In this paper is an attempt to build contingency tables known in mathematical statistic, based on a check statistic for eventological analysis of the market, in particular, to analyze the interaction of consumer segments. Some indicators of association of these segments were calculated.*
Keywords. *Probability, set of events, probability distribution, contingency tables, categorical data.*

Esin Roman (Krasnoyarsk, Russia) 147

Solve the inverse problem for the theory of risk measures distorted probability

Abstract. *In this paper we describe a method for solving the inverse problem of risk measures distorted probability, based on the calculation of the functional distorted probability, risk aversion and on the comparison of individual decisions with decisions made using different types of perturbing functions.*
Keywords. *Theory of risk, coherent risk measures, risk aversion, distorted probability functional.*

★ **Grigoriev Yuri (St. Petersburg, Russia)** 136

Polar equation and experimental ϕ_p-designs for polynomial regression

Abstract. *In this paper criteria of D- E- and A-optimality as elements of a set of Pukelshajm's ϕ_p-functions are considered. Modern approaches to construction of corresponding ϕ_p-designs for polynomial regression are discussed.*
Keywords. *Polynomial regression, polar equation, ϕ_p-designs, Pukelsheim's information function, elimination theory, Gröbner's basis.*

★ **Klokov Andrey and Vitaly Raspopov (Krasnoyarsk, Russia))** 155

Numerical identification of the leading coefficients of the system of parabolic equations

Abstract. *In this paper, we consider the inverse problem for a system of parabolic equations. Inverse problems can be associated with the interpretation and processing of the results of real thermal processes. In fact, the rate of temperature measurement in the remote interior points may be below the rate of temperature measurement outer surface. Thus violated the continuous dependence of the results on the input temperature data, so the inverse problems are usually incorrect. A lot of works are devoted to study inverse problems. The Cauchy inverse problem for an equation in this paper is generalized to a system of two equations.*
Keywords. *The inverse problem, parabolic equation, conditions override, numerical identification, Cauchy problem*

★ **Krupkina Tatiana, Mazharov Vladimir, Gornyi Boris (Krasnoyarsk, Russia) and Babenyshev Sergey (Zheleznogorsk, Russia)** 158

About various approaches to the construction of integral indicators

Abstract. *We consider three general types of problems of establishing dependencies of various output indicators from a set of features characterizing the public health conditions. We define methods for constructing integral indicators for each type of problem. Results are based on a series of works in which health indicators were used as output indicators, and the environment was characterized by socio-economic conditions.*
Keywords. *Socio-economic indicators, integral indicators, factor analysis, principal components.*

★ **Kustitskaya Tatyana (Krasnoyarsk, Russia)** 163

Coherent risk measures. Examples, variations, generalizations

Abstract. *The article presents a review of the main examples, variations and generalizations of coherent risk measures. Properties of described risk measures are examined form the viewpoint of the opportunities they afford to represent individual preferences.*
Keywords. *Coherent risk measures, convex risk measures, generalized risk measures, CVaR, measure of disturbed probability.*

★ **Maksimova Kristina and Irina Baranova (Krasnoyarsk, Russia)** 170

The problem of finding the association rules method in Data Mining

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 is described an association rules method. It is suggested the probabilistic algorithm. It is solved the practical problem of finding the association rules in products selling statistics.*
Keywords. *Data Mining, tasks, association rules method, products selling statistics*

★ **Makhonin Igor and Vitaliy Demidenko (Krasnoyarsk, Russia)** 178

Problems of balance of sign graphs

Abstract. *We formulate the main objectives of the study of signed graphs. The algorithm of check of balance of sign graphs is offered. The main formulations of a problem of sign balance are provided.*
Keywords. *Signed graph, balance.*

★ **Nartov Yakov (Krasnoyarsk, Russia)** 180

Regression as a Conditional Mean Probability Event

Abstract. *The article are devoted event view on regression by the conditional mean probability event concept.*
Keywords. *Probability, regression, eventology, mean probability event, conditional mean probability event.*

★ **Nifontov Aleksander (Krasnoyarsk, Russia)** 186

Mean-phenomenon portfolio analysis

Abstract. *Considered eventological Markowitz portfolio problem. Mean-phenomenon formulation of the problem.*
Keywords. *Eventology, probability theory, event, probability, portfolio analysis, portfolio analysis of events, Markowits, mean-phenomenon.*

★ **Novosyolov Arcady (Krasnoyarsk, Russia)** 193

Stress-testing of risk models: some pitfalls

Abstract. *The paper describes a method of stress testing of risk models, and a problem one can encounter when applying the method. Approaches to dealing with the problem are discussed.*

Keywords. *Risk, model, conditional distribution, stress testing, scenario.*

★ **Novosyolov Arcady (Krasnoyarsk, Russia)** 196

A generalization of Frechet bounds concept

Abstract. *A problem of Frechet bounds calculation for an arbitrary set theoretic operation is formulated and studied in the current paper. We introduce the concept of dual operations, study the relations between bounds for dual operations, define the valence (covalence) of operations, and calculate Frechet bounds for operations on two-element sets.*
Keywords. *Event, distribution function, Frechet bound, terrace, valence, duality, generator.*

★ **Novosyolov Arcady (Krasnoyarsk, Russia)** 200

Measurability of random set of events

Abstract. *The paper is devoted to establishing unconditional measurability of a random set of events, which follows directly from its structure without any additional requirements.*
Keywords. *Event, set of events, algebra, measurability, terrace.*

★ **Poddubny Vasily (Tomsk, Russia) and Anatoly Polikarpov (Moscow, Russia)** 213

Continuous Stochastic Dynamic Model for the Evolution of Polysemy and Sense Volume of Signs Ensembles of Natural Language

Abstract. *A continuous stochastic dynamic model for the evolution of polysemy and semantic volume of natural language signs is offered. The model is based on the assumption of the dissipative nature of the polysemy development of linguistic signs. On the basis of this model theoretical laws for synchronous (simultaneous) probability distributions for signs' ensembles are derived (age-and-polysemy, polysemy, age-and-sense-volume, sense-volume, frequency and frequency-rank distributions). Theoretically derived conclusions are compared with the corresponding empirical polysemy and sense volume distributions for lexical signs obtained from representative explanatory and frequency dictionaries of Russian and English.*
Keywords. *Evolution, linguistic sign, polysemy, semantic volume of a sign, stochastic mathematical model, explanatory and frequency dictionaries, identification of the model, the probability distributions.*

★ **Rakhimova Gulnoza, Sagidullaev Kalmurza and Gafurjon Tursunov (Tashkent, Uzbekistan)** 202

The rate of convergence in the sequential estimation of fixed-width intervals of asymptotic variance of rank estimators of shift

Abstract. *We studied the rate of convergence of a sequence of stopping time and the width of the confidence intervals.*
Keywords. *Rank estimator, asymptotic normality, stopping moment, fixed-width intervals.*

★ **Sadovsky Michael and Yaroslav Grebnev (Krasnoyarsk, Russia)** 204

Chargaff's rules and super-symmetry in genomes

Abstract. *Some preliminary results are presented exploring the problem of the genome super-symmetry. That latter is the equivalence of the frequencies of two strings making a complimentary palindrome observed within the same strand. Some simple models of DNA are studied that yield such super-symmetry. The models are compared to the real data observed on mitochondria, chloroplasts, yeasts and some other genomes of higher animals.*
Keywords. *DNA, string, frequency, palindrome, complimentary rule, random process, Markov process, orderness.*

★ **Sadovsky Michael and Ksenya Nikitina (Krasnoyarsk, Russia)** 208

Search and study of structures in the character sequences defined by random processes

Abstract. *This paper presents a novel approach to the identification of a structured symbol sequences, such as, for example, a DNA sequence. The subject of this study is the nature of the distribution of distances between the short subsequences (words). It is shown that the actual DNA sequence differs from the theoretical random sequence.*
Keywords. *Order, triplet, frequency dictionary, Markov process, the distribution of distances, the nearest neighbor.*

★ **Sarkisov Vigen (Samara, Russia)** 219

Creating Investors' Pooled Portfolios at Various Risk Criteria

Abstract. *The new opportunities arising from optimization of consolidated investors portfolios are investigated. The consolidation of Markowitz's portfolios is considered as an example. The restrictions on risk criteria at which consolidation is profitable are also formulated.*
Keywords. *Investment portfolio, trust management, pool of investors.*

★ **Semenova Daria and Anastasiya Ivanova (Krasnoyarsk, Russia)** 222

Fuzzy and set-regression model of distribution of consumer preferences between firms

Abstract. *In the paper the construction, study and comparative analysis of fuzzy model of I. Leung and a model based on a regression of random set on the example of the distribution of consumer preferences between companies manufacturing cosmetic products.*
Keywords. *Fuzzy binary relation, a random set of events, set-regression, segmentation of the market.*

★ **Semenova Daria and Guliza Kushbak kyzy (Krasnoyarsk, Russia)** 228

Correlation ratio between random sets of events

Abstract. *In this paper, we obtain a formula for measuring the correlation relationship of dependence between two random sets of events.*
Keywords. *Correlation ratio, random sets of events, cardinality indicator.*

★ **Semenova Daria, Lukyanova Natalia and Ludmila Shangareeva (Krasnoyarsk, Russia)** 234

Family of random sets of events of Frank

Abstract. *In paper a method which allows to set probabilistic distribution of a set of N events, using N probabilities and a type of associative function, is offered.*
Keywords. *Random set of events, associative function, Frank's associative random set of events.*

★ **Semenova Daria, Lukyanova Natalia and Ludmila Shangareeva (Krasnoyarsk, Russia)** 239

Estimate of the parameter of associative function by the least squares method for associative random sets of events of Frank and Ali-Mikhail-Haq

Abstract. *In paper the method of least squares to find the parameter of associative function with respect to the families of random set of events of Frank and Ali-Mikhail-Haq is considered. Legitimate probabilistic distributions for each found parameter are calculated.*

Keywords. *Random set of events, associative function, Frank's associative random set of events. Random set of events, the method of least squares, Frank's associative random set of events, Ali-Mikhail-Haq's associative random set of events, probabilistic distribution.*

★ **Vorobyev Oleg (Krasnoyarsk, Russia)** 44

Beyond usual measuring statistical interconnections: An eventologically revised approach to measurement for agreement and association (statistical dependence) of a set of events

Abstract. *The author has developed a general theory for measuring the statistical interconnections between events, called "mathematical eventology". The paper represents fresh "eventological" ideas that radically rethinking the basis of previous approaches to measuring statistical interconnections of events beyond the traditional measurements in the probability theory and mathematical statistics, adding to the already well-known methods of measuring association, or statistical dependency, another new group of methods of measuring agreement of a set of events. It shows by useful "total" quantities (called* ***agreement****,* ***association****,* ***covariance****,* ***set-distance****,* ***set-nearness****, and* ***set-closeness*** *of a set of events) formulas in this context can be simplified and elegantly summarized for many applications, including for the correct generalization of the traditional indicators of statistical interconnections of events as Pearson correlation, Yule association, as well as a popular index of agreement — Cohen kappa. Proposed new methods demonstrate the growing potential of "eventological" technologies in the study of structures of interconnections (agreement, association, or statistical dependency) of events, which applies to the measurement in the nominal (categorical), and the ordered and numeric (difference, ratio) scales.*
Keywords. *Categorical data, association, dissociation, dependency, independency, agreement, disagreement, contingencies, many-way contingency tables, interrater agreement measures, multiple raters, multiple categories, two-way contingency table, many-way contingency tables, covariance, set-distance, set-closeness, set-grouping.*

★ **Vorobyev Oleg (Krasnoyarsk, Russia)** 95

Set-*d*istogram measurers of disagreement of events

Abstract. *An extension of the notion of set-distance of an N-set of events characterizing its disagreement to $(N+1)$-family of set-distances of different orders is considered. This family of disagreement indexes called set-****d****istogram of a set of events characterizes the distribution of disagreement of events relative to the power of subsets of the given set of events and serves as a convenient tool for measuring the detailed disagreement of the set of events. Examples of set-****d****istogram for some types of probability distributions of sets of events with different structures of agreement and disagreement.*
Keywords. *Probability, multivariate statistical analysis, eventology, set of events, agreement, disagreement, set-distance, set-distogram, set theory operation, attic, cellar.*

★ **Vorobyev Oleg (Krasnoyarsk, Russia)** 112

On the foundations of the eventological method

Abstract. *Some foundations of the eventological method are discussed. New notions of the universal K-event (the universal Kolmogorov event) and the có-event (the name of universal K-event) are defined. Two new eventological axioms of* ***có-event*** *and of* ***a set of có-events*** *are formulated from which proposed earlier axioms VI and VII follow. New concepts of a set-product and an ordered set-product of sets of có-events are introduced that differ from the usual concepts of direct and Cartesian products significantly.*
Keywords. *Eventology, probability, event, universal K-event, có-event, axiom of có-event, axiom of a set of có-events, set-product of sets.*

★ **Vorobyev Oleg (Krasnoyarsk, Russia)** 122

Eventological generalization of Fréchet bounds for a set of có-events

Abstract. *An eventological glance at the generalization of the concept of Fréchet bounds for the probability of arbitrary set-theoretic operation on a finite set of có-events is proposed.*
Keywords. *Probability, multivariate statistics, event, có-event, eventology, Fréchet bound, arbitrary set-theoretic operation, layer, attic, cellar.*

★ **Zuparov Talat and Oybek Safarov (Tashkent, Uzbekistan)** 153

The central limit theorem for random sums of a linear process generated sequence of m-dependent random variables

Abstract. *In this work we prove the central limit theorem for random sums of a linear process generated sequence of m-dependent random variables.*
Keywords. *Stationary sequence of random variables, linear process, the central limit theorem, m-dependent random variables, random amount.*

П р и г л а ш е н и е

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

XIV
конференция
ФИНАНСОВО∼АКТУАРНАЯ МАТЕМАТИКА и ЭВЕНТОЛОГИЯ МНОГОМЕРНОЙ СТАТИСТИКИ
ФАМЭМС'2015

24 ∼ 25 апреля 2015 года, Красноярск

Т е м ы

- Финансово-актуарная математика
- Математика в гуманитарных, социо-экономических и естественнонаучных областях.
- Эвентология многомерной статистики
- Эвентология безопасности
- Эвентология риска и принятия решений в условиях риска и неопределенности
- Эвентология и математическая эвентология
- Философская эвентология и философия вероятности
- Практическая эвентология
- Эвентологическая экономика и психология
- Эвентологические проблемы искусственного интеллекта
- Эвентоконвергенция наук и технологий
- Теория вероятностей и математическая статистика
- Многомерный статистический анализ
- Принятие решений в условиях риска и неопределенности
- Измерение и модели риска
- Теория нечетких событий, обобщенная теория неопределенности
- Математические начала хаоса в экономике
- Системный анализ и управление случайными событиями

Важнейшие даты:

- 22 марта 2015 — завершение приема заявок, докладов, секций
- 29 марта 2015 — 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
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

XIV
Conference
FINANCIAL and ACTUARIAL MATHEMTAICS and EVENTOLOGY of MULTIVARIATE STATISTICS
FAMEMS'15

24 ~ 25 April 2015, Krasnoyarsk

Themes

- Financial and actuarial mathematics
- Mathematics in the humanities, socio-economic and natural sciences
- Eventology of multivariate statistics
- Eventology of safety
- Eventology of risk and decision-making under risk and uncertainty
- Eventology and mathematical eventology
- Philosophical eventology and philosophy of probability
- Eventology and the new humanity
- Practical eventology
- Eventological economics and psychology
- Eventological problems of artificial intelligence
- Converging sciences and technologies
- Probability theory and statistics
- Multivariate statistical analysis
- 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

Important dates:

- 22 March 2015 — deadline for papers, and sessions
- 29 March 2015 — notification about acceptance/rejection; distribution of invitations

Instructions:

- http://fam.conf.sfu-kras.ru/submission-e.php

Научное издание

Т р у д ы
XIII ФАМЭМС'2014 конференции
по финансово-актуарной математике
и эвентологии многомерной статистики

Под редакцией Олега Воробьёва

ISBN 978-5-9903358-4-4

Сибирский федеральный университет
Институт математики и фундаментальной информатики
Свободный пр., 79, Красноярск, 660041
Электронное издание • http://fam.conf.sfu-kras.ru

ЭМ и ФАМ семинары

Сибирский федеральный университет
Институт математики
Торгово-экономический институт
Институт проблем пожарной безопасности
Институт вычислительного моделирования СО РАН

http://fam.conf.sfu-kras.ru
http://em.conf.sfu.kras.ru
famemconf@gmail.com