М. Б. Лагутин

НАГЛЯДНАЯ МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Учебное пособие

5-е издание (электронное)

Рекомендовано
Учебно-методическим объединением
по классическому университетскому образованию
в качестве учебного пособия
для студентов высших учебных заведений,
обучающихся по направлению «Математика»
и «Математика. Прикладная математика»

Москва
БИНОМ. Лаборатория знаний
2015

УДК 519.22
ББК 22.17
Л14

Р е ц е н з е н т ы:

кандидат физ.-мат. наук Э. М. Кудлаев,
зав. каф. матем. статистики ф-та ВМиК МГУ
академик РАН Ю. В. Прохоров,
доктор физ.-мат. наук, проф. Ю. Н. Тюрин

Лагутин М. Б.

Л14 Наглядная математическая статистика [Электронный ресурс] : учебное пособие / М. Б. Лагутин. — 5-е изд. (эл.). — Электрон. текстовые дан. (1 файл pdf : 475 с.). — М. : БИНОМ. Лаборатория знаний, 2015. — Систем. требования: Adobe Reader XI ; экран 10".

ISBN 978-5-9963-2955-7

Основы теории вероятностей и математической статистики излагаются в форме примеров и задач с решениями. Книга также знакомит читателя с прикладными статистическими методами. Для понимания материала достаточно знания начал математического анализа. Включено большое количество рисунков, контрольных вопросов и числовых примеров.

Для студентов, изучающих математическую статистику, исследователей и практиков (экономистов, социологов, биологов), применяющих статистические методы.

УДК 519.22
ББК 22.17

Деривативное электронное издание на основе печатного аналога: Наглядная математическая статистика : учебное пособие / М. Б. Лагутин. — 3-е изд., испр. — М. : БИНОМ. Лаборатория знаний, 2013. — 472 с. : ил. — ISBN 978-5-9963-1530-7.

ISBN 978-5-9963-2955-7

ПРЕДИСЛОВИЕ

Что за польза от книги без картинок и разговоров?

Льюис Кэрролл, «Приключения Алисы в стране чудес»

Перед Вами, уважаемый читатель, итог размышлений автора о содержании начального курса математической статистики. Настоящая книга — это, в первую очередь, множество занимательных примеров и задач, собранных из различных источников. Задачи предназначены для активного освоения понятий и развития у читателя навыков квалифицированной статистической обработки данных. Для их решения достаточно знания элементов математического анализа и теории вероятностей (краткие сведения по теории вероятностей и линейной алгебре даны в приложении).

Акцент делается на наглядном представлении материала и его неформальном пояснении. Теоремы, как правило, приводятся без доказательств (со ссылкой на источники, где их можно найти). Наша цель — и осветить практически наиболее важные идеи математической статистики, и познакомить читателя с прикладными методами.

Первая часть книги (гл. 1–5) может служить введением в теорию вероятностей. Особенностью этой части является подход к освоению понятий теории вероятностей через решение ряда задач, относящихся к области статистического моделирования (имитации случайности на компьютере). Ее материал, в основном, доступен школьникам старших классов и студентам 1-го курса.

Вторая и третья части (гл. 6–13) посвящены, соответственно, оценкам параметров статистических моделей и проверке гипотез. Они могут быть особенно полезны студентам при подготовке к экзамену по математической статистике.

Четвертая и пятая части (гл. 14–21) предназначаются, в первую очередь, лицам, желающим применить статистические методы для анализа экспериментальных данных.

Наконец, шестая часть (гл. 22–26) включает в себя ряд более специальных тем, обобщающих и дополняющих содержание предыдущих глав.

Собранный в книге материал неоднократно использовался на занятиях по математической статистике на механико-математическом факультете МГУ им. М. В. Ломоносова.

Автор будет считать свой труд небесполезным, если, перелистав книгу, читатель не потеряет к ней интереса, а захочет ознакомиться

Ей сна нет от французских книг, а мне от русских больно спится!

Фамусов в «Горе от ума» А. С. Грибоедова

Никогда не теряй из виду, что гораздо легче многих не удовлетворить, чем удовольствовать.

Козьма Прутков, «Мысли и афоризмы»

с теорией и приложениями статистики как по этому, так и по другим учебникам.

При работе над книгой образцом для автора была популярная серия книг для школьников Я. И. Перельмана. Хотелось, по возможности, использовать живую форму изложения и стиль, характерный для этой серии.

Я благодарен моим коллегам по лаборатории Математической статистики МГУ им. М. В. Ломоносова М. В. Козлову и Э. М. Кудлаеву за прочтение рукописи этой книги и полезные замечания.

М. Лагутин

К ЧИТАТЕЛЮ

В книге Д. Пойа «Математическое открытие» (см. [62] в списке литературы) выделены *три принципа обучения*. Первым (и важнейшим) из них является

Стимулирование

Надо заинтересовать учащегося, убедить в полезности изучения предмета. Для успешности учебы необходимо четкое представление о том, зачем нужна сообщаемая информация.

Приведем мнение по этому вопросу известного героя детективного жанра (ведь восстановление по частностям общей картины есть также и задача математической статистики).

> «Мне представляется, что человеческий мозг похож на маленький пустой чердак, который вы можете обставить, как хотите. Дурак натащит туда всякой рухляди, какая попадется под руку, и полезные, нужные вещи уже некуда будет всунуть, или в лучшем случае до них среди всей этой завали и не докопаешься. А человек толковый тщательно отбирает то, что он поместит в свой мозговой чердак. Он возьмет лишь инструменты, которые понадобятся ему для работы, но зато их будет множество, и все он разложит в образцовом порядке. Напрасно люди думают, что у этой маленькой комнатки эластичные стены и их можно растягивать сколько угодно. Уверяю вас, придет время, когда, приобретая новое, вы будете забывать что-то из прежнего. Поэтому страшно важно, чтобы ненужные сведения не вытесняли собой нужных.»
>
> *А. Конан Дойл, «Этюд в багровых тонах»*

Математическая статистика — один из наиболее часто используемых в приложениях разделов математики. На результаты практически любого научного эксперимента влияют неучтенные в модели факторы, накладывается случайный шум. Методы математической статистики, как правило, позволяют наиболее полно и надежно извлекать полезную информацию из зашумленных данных. В книгу включены многочисленные примеры применения статистических методов для решения практических задач.

Чтобы побудить читателя глубже изучить теорию вероятностей, на языке которой формулируются статистические теоремы, многие главы завершаются вероятностным парадоксом или занимательным экспериментом.

Основа, подлинное содержание всякого познания доставляется именно наглядной концепцией мира, которая может быть добыта лишь нами самими и отнюдь не может быть как-либо преподана извне.

Артур Шопенгауэр, «Афоризмы житейской мудрости»

Студент — это не гусь, которого надо нафаршировать, а факел, который нужно зажечь.

То, что вы были вынуждены открыть сами, оставляет в вашем уме дорожку, которой вы можете снова воспользоваться, когда в этом возникнет необходимость.

Г. Лихтенберг, «Aphorismen», Berlin, 1902–1906

Следующим принципом обучения является

Активность

По-настоящему разобраться в некоторой теории можно лишь самостоятельно решая задачи из данной области. Пассивного чтения даже хорошего учебника, увы, недостаточно для подлинного овладения предметом.

Каждая глава этой книги (за исключением дополнительных глав 22–26) содержит задачи (с решениями). Они обычно упорядочены по сложности, самые трудные отмечены звездочкой. Автор надеется, что читатель попробует решить некоторые из заинтересовавших его задач или, хотя бы, разберет решения, так как в них содержится значительная часть материала. Кроме того, по ходу изложения встречаются контрольные вопросы, ответы на которые приведены в конце соответствующей главы.

Возможность активного усвоения материала во многом определяется стилем его изложения.

При изложении математического рассуждения мастерство заключается в умении дать образованному читателю возможность сразу, не заботясь о деталях, схватить основную идею; последовательные дозы должны быть такими, чтобы их можно было глотать «с ходу»; в случае неудачи или если бы читатель захотел что-либо проверить, перед ним должна стоять четко ограниченная маленькая задача (например, проверить тождество; две пропущенные тривиальности могут в совокупности образовать непреодолимое препятствие).

Дж. Литлвуд, «Математическая смесь»

Наконец, третий принцип — это соблюдение последовательности фаз обучения

Исследование → формализация → усвоение

Важно начинать новую тему с содержательных примеров, чтобы можно было «потрогать руками», прочувствовать ситуацию. Можно попробовать придумать какой-нибудь способ решения проблемы лишь на основе здравого смысла. Если он на самом деле окажется бесполезным, то это лишь подтвердит важность теории, позволяющей получить приемлемое решение.

Всякое человеческое познание начинается с созерцаний, переходит от них к понятиям и заканчивается идеями.

И. Кант, «Критика чистого разума»

Абстрактные определения становятся по-настоящему понятны лишь тогда, когда они используются при решении конкретных задач в различных моделях. В книге «Теория катастроф» В. И. Арнольд пишет:

> «Абстрактные определения возникают при попытках обобщить «наивные» понятия, сохраняя их основные свойства. Теперь, когда мы знаем, что эти попытки не приводят к реальному расширению круга объектов (для многообразий это установил Уитни, для групп — Кэли, для алгоритмов — Черч), не лучше ли в преподавании вернуться к «наивным» определениям? (...) Пуанкаре подробно обсуждает методические преимущества наивных определений окружности и дроби в «Науке и методе»: невозможно усвоить правило сложения дробей, не разрезая, хотя бы мысленно, яблоко или пирог.»

При написании этой книги автор старался следовать указанным принципам обучения. Вероятно, какие-то методические приемы окажутся полезными преподавателям статистики, хотя, безусловно справедливо утверждал Козьма Прутков, что

У всякого портного свой взгляд на искусство!

Часть I

ВЕРОЯТНОСТЬ И СТАТИСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

Глава 1

ХАРАКТЕРИСТИКИ СЛУЧАЙНЫХ ВЕЛИЧИН

В основе математической статистики лежит теория вероятностей. Аксиоматика теории вероятностей была разработана А. Н. Колмогоровым (опубликована в 1933 г.). Читателю, возможно, известны такие основные понятия этой теории, как независимость событий или математическое ожидание случайной величины. Тем не менее, будет полезно напомнить самое необходимое для дальнейшего изложения (см. также приложение П1*) и учебники [19], [39], [90] в списке литературы).

Вероятность — это важнейшее понятие в современной науке особенно потому, что никто совершенно не представляет, что оно означает.

Бертран Рассел, из лекции, 1929 г.

§ 1. ФУНКЦИИ РАСПРЕДЕЛЕНИЯ И ПЛОТНОСТИ

Пример 1. Измерим время ξ от первого включения до перегорания электрической лампочки.

Пример 2. Подбросим монетку. Если она упадет гербом вверх, будем считать, что $\xi = 1$, иначе положим $\xi = 0$.

Обобщая эти примеры, представим, что проводится эксперимент, результат которого (действительное число ξ) зависит от случая. Как охарактеризовать *случайную величину* ξ, дать вероятностный закон ее поведения?

Допустим, что возможно повторить эксперимент несколько раз. Обозначим через $\xi_1, \ldots, \xi_n$ полученные при этом значения. Тогда для произвольной точки x на прямой можно подсчитать ν_n — количество значений, попавших левее x (рис. 1).

Предположим, что существует некоторое число, к которому будет приближаться частота ν_n/n при неограниченном увеличении n. Естественно рассматривать это число как *вероятность того, что* ξ *не больше, чем* x. Обозначим эту вероятность через $\mathbf{P}(\xi \leqslant x)$. (Формальные определения понятий вероятности и случайной величины приведены в П1.)

Читал ли что-нибудь? Хоть мелочь?

Репетилов в «Горе от ума» А. С. Грибоедова

Сперва аз да буки, а там и науки.

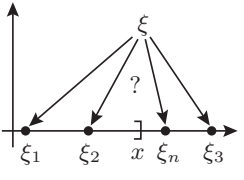


Рис. 1

P: Probabilitas (*лат.*) — вероятность.

Пример 3. На рис. 2 показан график частоты появлений буквы «а» в стихотворении М. Ю. Лермонтова «Бородино». Размах

*) П1 обозначает ссылку на раздел 1 приложения.

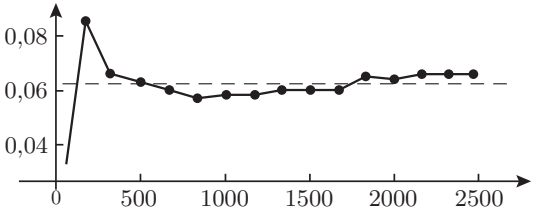


Рис. 2

колебаний частоты быстро уменьшается, она стабилизируется на уровне чуть большем, чем 0,06. В таблице приведены вероятности, с которыми встречаются в большом по объему тексте буквы русского алфавита, включая «пробел» между словами (данные взяты из [92, с. 238]). Отметим, что итоговая частота появлений буквы «а» в стихотворении «Бородино», равная $162/2461 \approx 0{,}066$, лишь незначительно отличается от соответствующей вероятности 0,062.

—	о	е, ё	а	и	т	н	с
0,175	0,090	0,072	0,062	0,062	0,053	0,053	0,045
р	в	л	к	м	д	п	у
0,040	0,038	0,035	0,028	0,026	0,025	0,023	0,021
я	ы	з	ь, ъ	б	г	ч	й
0,018	0,016	0,016	0,014	0,014	0,013	0,012	0,010
х	ж	ю	ш	ц	щ	э	ф
0,009	0,007	0,006	0,006	0,004	0,003	0,003	0,002

Зафиксируем n и рассмотрим поведение частоты ν_n/n при изменении «границы» x (см. рис. 1). При сдвиге точки x вправо, количество значений $\xi_1, \ldots, \xi_n$, оказавшихся левее x, будет увеличиваться. Поэтому вероятность $\mathbf{P}(\xi \leqslant x)$ (как предел частоты) будет неубывающей функцией от x, которая стремится к 1 при $x \to +\infty$ и стремится к 0 при $x \to -\infty$.

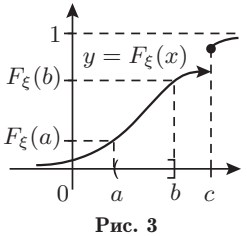


Рис. 3

Определение. Функция $F_\xi(x) = \mathbf{P}(\xi \leqslant x)$ называется *функцией распределения* случайной величины ξ.

Зная $F_\xi(x)$, можно найти вероятность попадания ξ в любой промежуток $(a,b]$ на прямой (рис. 3):

$$\mathbf{P}(a < \xi \leqslant b) = \mathbf{P}(\xi \leqslant b) - \mathbf{P}(\xi \leqslant a) = F_\xi(b) - F_\xi(a).$$

Вопрос 1. Как это доказать формально, используя свойство непрерывности из П1?

Если функция распределения $F_\xi(x)$ имеет разрыв в точке c, то величина скачка $F_\xi(c) - F_\xi(c-)$ равна

$$\mathbf{P}(\xi = c) = \mathbf{P}(\xi \leqslant c) - \mathbf{P}(\xi < c).$$

Случайные величины мы будем задавать с помощью функций распределения.

Определение. Случайная величина η *равномерно распределена на отрезке* $[0,1]$, если

$$F_\eta(x) = \begin{cases} 0 & \text{при } x \leqslant 0, \\ x & \text{при } 0 < x < 1, \\ 1 & \text{при } x \geqslant 1. \end{cases}$$

Такое распределение соответствует *выбору точки наудачу* из отрезка $[0,1]$, поскольку для любых $0 \leqslant a < b \leqslant 1$ вероятность попадания значения η в отрезок $[a,b]$ равна его длине $b-a$ (рис. 4).

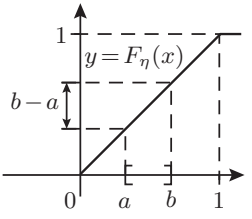


Рис. 4

Определение. Случайная величина τ называется *показательной с параметром* $\lambda > 0$, если

$$F_\tau(x) = \begin{cases} 0 & \text{при } x \leqslant 0, \\ 1 - e^{-\lambda x} & \text{при } x > 0. \end{cases}$$

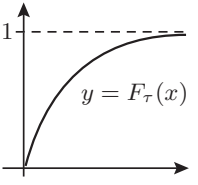


Рис. 5

График функции распределения $F_\tau(x)$ приведен на рис. 5.

Показательное распределение можно использовать для описания времени эксперимента из примера 1.

Вопрос 2.
Чему равна $\mathbf{P}(\tau > 3/\lambda)$ точно и приближенно?

Определение. Если существует такая функция $p_\xi(x) \geqslant 0$, что для произвольных $a < b$

$$\mathbf{P}(a \leqslant \xi \leqslant b) = \int_a^b p_\xi(x)\,dx,$$

то говорят, что случайная величина ξ (или ее распределение вероятностей) имеет *плотность* $p_\xi(x)$ (рис. 6).

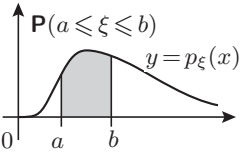


Рис. 6

Когда плотность существует, ее можно найти дифференцированием функции распределения:

$$p_\xi(x) = \frac{d}{dx}\,F_\xi(x) = \lim_{\Delta x \to 0} \frac{F_\xi(x+\Delta x) - F_\xi(x)}{\Delta x}\,.$$

Таким образом, плотностью равномерной величины η является функция $I_{[0,1]}$ (здесь и далее I_A обозначает *индикатор множества* A: $I_A(x) = 1$ при $x \in A$, $I_A(x) = 0$ при $x \notin A$), а плотностью показательной величины τ служит $p_\tau(x) = \lambda e^{-\lambda x} I_{[0,+\infty)}$ (рис. 7).

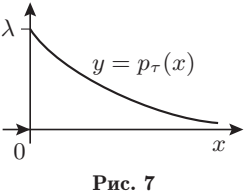


Рис. 7

Не у всякой случайной величины есть плотность. Например, ее нет у *дискретных* (принимающих конечное или счетное*) число значений) величин. Такова определяемая ниже бернуллиевская случайная величина.

Я. Бернулли (1654–1705), швейцарский математик.

*) Множество называют *счетным*, если его элементы можно перенумеровать натуральными числами.

Определение. Случайная величина ζ имеет *распределение Бернулли с вероятностью «успеха»* p $(0 \leqslant p \leqslant 1)$, если она принимает значения 0 и 1 с такими вероятностями: $\mathbf{P}(\zeta = 0) = 1 - p$ и $\mathbf{P}(\zeta = 1) = p$.

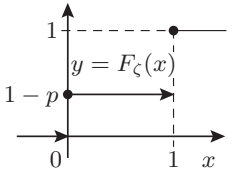


Рис. 8

График функции распределения $F_\zeta(x)$ бернуллиевской случайной величины ζ приведен на рис. 8. Распределение Бернулли при $p = 1/2$ годится как вероятностная модель эксперимента из примера 2. Значение $p \neq 1/2$ отвечает случаю несимметричной монеты.

Вопрос 3.
Как выглядит график функции распределения дискретной случайной величины ξ, принимающей значения $x_1 < x_2 < \ldots$ с соответствующими вероятностями $p_1, p_2, \ldots$?

§ 2. МАТЕМАТИЧЕСКОЕ ОЖИДАНИЕ И ДИСПЕРСИЯ

Не всегда требуется полная информация о случайной величине ξ, выражающаяся в ее функции распределения $F_\xi(x)$. Иногда достаточно знать, где располагается область «типичных» значений ξ. Одной из важных характеристик «центра» этой области является математическое ожидание.

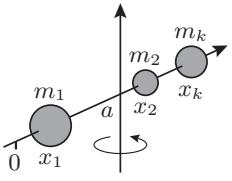


Рис. 9

Проблема. На тонком стержне (числовой прямой) в точках с координатами x_k находятся массы m_k (рис. 9). Где следует выбрать точку a крепления стержня к вертикальной оси, чтобы минимизировать *момент инерции* относительно нее $I_a = \sum (x_k - a)^2 m_k$?

Оказывается, точку крепления стержня надо поместить в *центр масс* $c = \sum x_k m_k / \sum m_k$ (см. задачу 1). Вероятностными аналогами центра масс c и момента инерции относительно него I_c служат математическое ожидание и дисперсия.

Определение. Для дискретной случайной величины ξ, принимающей значения $x_1, x_2, \ldots$ с соответствующими вероятностями $p_1, p_2, \ldots$, математическим ожиданием называется число

$$\mathbf{M}\xi = \sum_k x_k p_k. \tag{1}$$

Например, для бернуллиевской случайной величины ζ имеем

$$\mathbf{M}\zeta = 0 \cdot (1 - p) + 1 \cdot p = p.$$

Определение. Когда у случайной величины ξ есть плотность $p_\xi(x)$, ее математическое ожидание вычисляется по формуле

$$\mathbf{M}\xi = \int\limits_{-\infty}^{+\infty} x\, p_\xi(x)\, dx. \tag{2}$$

Для показательной случайной величины τ нетрудно подсчитать, интегрируя по частям, что

$$\mathbf{M}\tau = \int\limits_0^\infty x\, \lambda e^{-\lambda x} dx = \frac{1}{\lambda} \int\limits_0^\infty y\, e^{-y} dy = \frac{1}{\lambda} \left[0 + \int\limits_0^\infty e^{-y} dy \right] = \frac{1}{\lambda}\,.$$

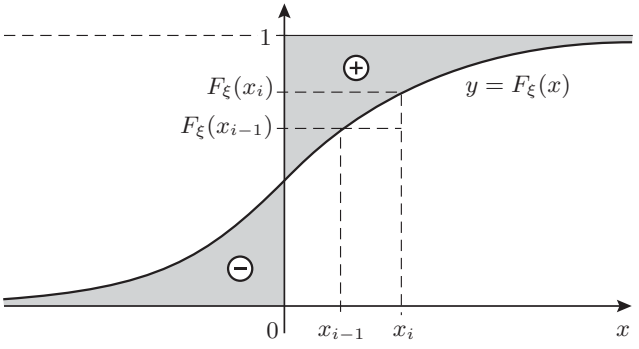


Рис. 10

Оба приведенных выше определения являются частными случаями следующего определения математического ожидания как интеграла Стильтьеса.

Определение. Для случайной величины ξ с функцией распределения $F_\xi(x)$ математическим ожиданием называется

$$\mathbf{M}\xi = \int\limits_{-\infty}^{+\infty} x\, F_\xi(dx) = \lim_{D\to 0} \sum_i x_i \,[F_\xi(x_i) - F_\xi(x_{i-1})],$$

где $D = \max |x_i - x_{i-1}|$ — *диаметр разбиения.*

Общее определение $\mathbf{M}\xi$ как интеграла Лебега приведено в приложении П2.

Рисунок 10 иллюстрирует *геометрическое представление* математического ожидания как разности площадей закрашенных областей со знаком «+» и знаком «−». Действительно, интегральная сумма в определении $\mathbf{M}\xi$ совпадает с суммой площадей (с учетом знака x_i) прямоугольников с шириной x_i и высотой $F_\xi(x_i) - F_\xi(x_{i-1})$. При измельчении разбиения она приближается к площади (с учетом знака) закрашенной области.

Геометрическое представление дает другой способ подсчета математического ожидания $\mathbf{M}\tau$ показательной случайной величины (см. рис. 5):

$$\mathbf{M}\tau = \int\limits_0^\infty \mathbf{P}(\tau > x)\, dx = \int\limits_0^\infty [1 - F_\tau(x)]\, dx = \int\limits_0^\infty e^{-\lambda x}\, dx = \frac{1}{\lambda}\,.$$

Для случайной величины ξ, принимающей только целые неотрицательные значения: $\mathbf{P}(\xi = k) = p_k$, $k \geqslant 0$, геометрическое представление величины $\mathbf{M}\xi$ (рис. 11) объясняет следующую формулу:

$$\mathbf{M}\xi = \sum_{k=0}^{\infty} \mathbf{P}(\xi > k). \qquad (3)$$

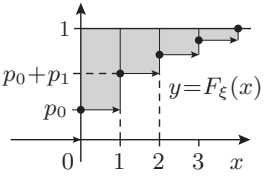


Рис. 11

Замечание. Математическое ожидание определено *не для всякой* случайной величины. Возможна ситуация, когда на рис. 10 и площадь области со знаком «+», и площадь области со знаком «−»

О. Коши (1798–1857), французский математик.

равны ∞. В этом случае возникает неопределенность вида $\infty - \infty$. Например, для *закона Коши* с плотностью $p_\xi(x) = 1/[\pi(1+x^2)]$ каждая из площадей есть

$$\int\limits_0^\infty x\, p_\xi(x)\, dx = \frac{1}{\pi}\int\limits_0^\infty \frac{x}{1+x^2}\, dx = \frac{1}{2\pi}\int\limits_0^\infty \frac{dy}{1+y} = \frac{1}{2\pi}\,\ln(1+y)\Big|_0^\infty = \infty.$$

Следовательно, $\mathbf{M}\xi$ не существует, несмотря на то, что 0 — центр распределения (плотность $p_\xi(x)$ симметрична относительно 0).

Обсудим теперь понятие **дисперсии** случайной величины.

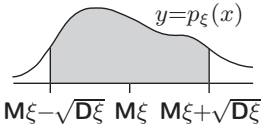


Рис. 12

Как правило, помимо $\mathbf{M}\xi$ бывает важно знать величину типичного «разброса» значений ξ вокруг среднего. Мерой этого «разброса» может служить *стандартное отклонение* $\sqrt{\mathbf{D}\xi}$ (рис. 12), где *дисперсия* $\mathbf{D}\xi$ определяется формулой

$$\mathbf{D}\xi = \mathbf{M}\,(\xi - \mathbf{M}\xi)^2\,,$$

т. е. $\mathbf{D}\xi$ — это среднее квадрата отклонения ξ от $\mathbf{M}\xi$.

Для вычисления дисперсии полезно равенство

$$\mathbf{D}\xi = \mathbf{M}\xi^2 - (\mathbf{M}\xi)^2\,. \tag{4}$$

Вопрос 4.
Как получить формулу (4) с помощью свойств математического ожидания из приложения П2?

Для примера вычислим дисперсию бернуллиевской случайной величины ζ. Прежде всего, заметим, что ζ^2 и ζ одинаково распределены. Поэтому $\mathbf{M}\zeta^2 = \mathbf{M}\zeta = p$ и $\mathbf{D}\zeta = p - p^2 = p(1-p)$.

Вы давиче его мне исчисляли свойства, но многие забыли? — Да?

Чацкий в «Горе от ума» А. С. Грибоедова [В слове «давеча» сохранена авторская орфография.]

§ 3. НЕЗАВИСИМОСТЬ СЛУЧАЙНЫХ ВЕЛИЧИН

Обычно вероятностную модель необходимо построить не для одного эксперимента, а для серии опытов. В этом случае нередко можно предполагать отсутствие взаимного влияния разных опытов друг на друга, их независимость.

Определение. Случайные величины $\xi_1, \ldots, \xi_n$ называются *независимыми*, если для любых $a_i < b_i$ $(i = 1, \ldots, n)$

$$\mathbf{P}(a_i < \xi_i \leqslant b_i,\, i = 1, \ldots, n) = \prod_{i=1}^{n} \mathbf{P}(a_i < \xi_i \leqslant b_i).$$

В частности, если все $a_i = -\infty$, то для произвольных $x_1, \ldots, x_n$

$$\mathbf{P}(\xi_1 \leqslant x_1, \ldots, \xi_n \leqslant x_n) = F_{\xi_1}(x_1) \cdot \ldots \cdot F_{\xi_n}(x_n). \tag{5}$$

Независимые равномерно распределенные на отрезке $[0,1]$ случайные величины $\eta_1, \ldots, \eta_n$ можно считать координатами случайного вектора, равномерно распределенного в n-мерном единичном кубе. Действительно, равенство

$$\mathbf{P}(a_i \leqslant \eta_i \leqslant b_i,\ i = 1, \ldots, n) = (b_1 - a_1) \cdot \ldots \cdot (b_n - a_n),$$

где $0 \leqslant a_i < b_i \leqslant 1$, означает, что вероятность попадания точки $(\eta_1, \ldots, \eta_n)$ в произвольный параллелепипед с параллельными осям координат ребрами и находящийся целиком внутри единичного куба равна его объему (рис. 13 при $n = 3$). На самом деле, параллелепипед можно заменить на любое множество A, для которого определено понятие n-мерного объема.

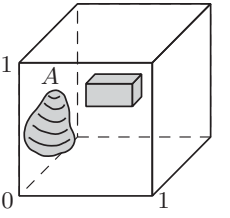


Рис. 13

Говорят, что бесконечная последовательность $\{\xi_i\}$ *образована независимыми случайными величинами*, если свойство независимости выполняется для любого конечного набора из них.

Определение. Последовательность независимых бернуллиевских случайных величин $\zeta_1, \zeta_2, \ldots$ с одинаковой вероятностью «успеха» p называют *испытаниями (или однородной схемой) Бернулли.*[*)]

В заключение параграфа приведем интуитивно понятное утверждение, которое часто применяется при доказательстве статистических теорем.

Лемма о независимости. Пусть $\xi_1, \ldots, \xi_{n+m}$ — независимые случайные величины; f и g — борелевские функции (см. приложение П2) на $\mathbb{R}^n$ и $\mathbb{R}^m$ соответственно. Тогда случайные величины $\eta_1 = f(\xi_1, \ldots, \xi_n)$ и $\eta_2 = g(\xi_{n+1}, \ldots, \xi_{n+m})$ независимы.

Доказательство этой леммы можно найти, например, в [48, с. 53].

§ 4. ПОИСК БОЛЬНЫХ

Применим элементарную теорию вероятностей к решению одной проблемы выявления больных (см. [82, с. 254]).

Во время второй мировой войны всех призывников в армию США подвергали медицинскому обследованию. Реакция Вассермана позволяет обнаруживать в крови больных сифилисом определенные антитела. Р. Дорфманом была предложена простая методика, на основе которой необходимое для выявления всех больных число проверок удалось уменьшить в 5 раз!

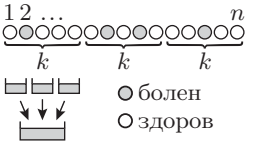


Рис. 14

Методика. Смешиваются пробы крови k человек и анализируется полученная смесь (рис. 14). Если антител нет, то этой одной проверки достаточно для k человек. В противном случае кровь каждого человека из этой группы нужно исследовать отдельно, и для k человек всего потребуется $k + 1$ раз провести анализ.

Вероятностная модель. Предположим, что вероятность обнаружения антител p одна и та же для всех n обследуемых, и результаты анализов для различных людей независимы, т. е. моделью является последовательность из n испытаний Бернулли с вероятностью «успеха» p.

Допустим для простоты, что n делится нацело на k. Тогда надо проверить n/k групп обследуемых. Пусть X_j — количество

[*)] Если у каждой случайной величины ζ_i своя вероятность «успеха» p_i, то схему называют *неоднородной*.

проверок, потребовавшихся в j-й группе, $j = 1, \ldots, n/k$. Тогда

$$X_j = \begin{cases} 1 & \text{с вероятностью } (1-p)^k \text{ (все } k \text{ человек здоровы)}, \\ k+1 & \text{с вероятностью } 1-(1-p)^k \text{ (есть больные)}. \end{cases}$$

Обозначим *общее число проверок* $X_1 + \ldots + X_{n/k}$ через Z. Задача заключается в том, как для заданного значения p[*)] определить размер группы $k_0 = k_0(p)$, минимизирующий $\mathbf{M}Z$.

Вопрос 5. Чем плох слишком большой размер группы?

Согласно формуле (1) находим

$$\mathbf{M}X_j = 1 \cdot (1-p)^k + (k+1) \cdot \left[1-(1-p)^k\right] = k+1-k\,(1-p)^k.$$

Отсюда по свойствам математического ожидания (П2) имеем

$$\mathbf{M}Z = \mathbf{M}X_1 + \ldots + \mathbf{M}X_{n/k} = \frac{n}{k}\ \mathbf{M}X_1 = n\left[1+1/k-(1-p)^k\right].$$

Вопрос 6. Какая ошибка допущена на рис. 15 в изображении графика функции $H(x)$ при малых p?

Положим $H(x) = 1 + 1/x - (1-p)^x$ при $x > 0$.

Для близких к нулю значений p минимум функции $H(x)$ достигается в точке x_0, где x_0 — наименьший из корней уравнения $H'(x) = 0$, т. е. уравнения

$$1/x^2 + (1-p)^x \ln(1-p) = 0. \tag{6}$$

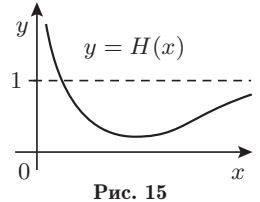


Рис. 15

Его нельзя разрешить явно относительно x. Поэтому, используя формулу $(1-p)^x \approx 1 - px$ при малых p, заменим $H(x)$ на функцию $\widetilde{H}(x) = 1 + 1/x - 1 + px = 1/x + px$, имеющую точку минимума $\widetilde{x}_0 = 1/\sqrt{p}$, причем $\widetilde{H}(\widetilde{x}_0) = 2\sqrt{p}$. Для $p = 0{,}01$ получаем $\widetilde{x}_0 = 10$ и $\widetilde{H}(\widetilde{x}_0) = 1/5$, т. е. $\mathbf{M}Z \approx n/5$.[**)]

Не пропускайте их, они еще не раз пригодятся в дальнейшем!

ЗАДАЧИ

Я занимался до сих пор решением ряда задач, ибо при изучении наук примеры полезнее правил.

И. Ньютон, «Всеобщая арифметика»

1. Докажите, используя свойства математического ожидания (П2), что функция $f(a) = \mathbf{M}(\xi - a)^2$ при $a = \mathbf{M}\xi$ имеет минимум, равный $\mathbf{D}\xi$.

2. Случайные величины $\eta_1, \ldots, \eta_n$ независимы и равномерно распределены на отрезке $[0, 1]$. Вычислите $\mathbf{M}\overline{\eta}$ и $\mathbf{D}\overline{\eta}$ среднего арифметического $\overline{\eta} = \dfrac{1}{n}(\eta_1 + \ldots + \eta_n)$.

3. Для случайных величин из задачи 2 найдите функцию распределения $F_{\eta_{(n)}}(x)$, $\mathbf{M}\eta_{(n)}$ и $\mathbf{D}\eta_{(n)}$, где $\eta_{(n)} = \max\{\eta_1, \ldots, \eta_n\}$.

4. Обозначим через ν число «неудач» до появления первого «успеха» в схеме Бернулли с параметром p. Вычислите $\mathbf{M}\nu$.
Указание. Примените формулу (3).

5. Рассмотрим следующую стратегию поиска больных. Все обследуемые разбиваются на пары. Если объединенная проба крови не содержит антител, то оба здоровы. В противном случае исследуется кровь первого из них. Если этот человек здоров, то другой должен быть болен, и в таком случае достаточно двух

[*)] Это значение можно оценить с помощью частоты выявления заболевания в предыдущих обследованиях.

[**)] Асимптотика x_0 и $H(x_0)$ при $p \to 0$ исследуется в задаче 6.

тестов. Если же первый оказался больным, то кровь второго также должна быть подвергнута анализу, и поэтому потребуется три теста. Выясните, при каких значениях вероятности p обнаружения заболевания у отдельного обследуемого данная стратегия будет в среднем экономичнее индивидуальной проверки.

6* Пусть $x_0 = x_0(p)$ — наименьший из корней уравнения (6). Докажите, что $x_0 \sim 1/\sqrt{p}$ и $H(x_0) \sim 2\sqrt{p}$ при $p \to 0$.[*)]

РЕШЕНИЯ ЗАДАЧ

1. С учетом свойств математического ожидания (см. приложение П2) и формулы (4) находим, что функция

Растолковать прошу.

Репетилов в «Горе от ума» А. С. Грибоедова

$$f(a) = \mathbf{M}\left[\xi^2 - 2a\xi + a^2\right] = \mathbf{M}\xi^2 - 2a\mathbf{M}\xi + a^2 = (a - \mathbf{M}\xi)^2 + \mathbf{D}\xi$$

есть квадратный трехчлен с минимумом в точке $a = \mathbf{M}\xi$.

2. Согласно формуле (2) $\mathbf{M}\eta_1 = \int\limits_0^1 x\,dx = 1/2$. (Это можно понять и без вычислений: плотность $p_{\eta_1}(x) = I_{[0,\,1]}$ симметрична относительно прямой $x = 1/2$.)

Далее, в силу следствия из П2 имеем $\mathbf{M}\eta_1^2 = \int\limits_0^1 x^2 dx = 1/3$. Применяя формулу (4), получаем, что $\mathbf{D}\eta_1 = 1/3 - 1/4 = 1/12$.

Наконец, согласно свойствам математического ожидания и дисперсии из приложения П2 запишем:

$$\mathbf{M}\overline{\eta} = \frac{1}{n}\left(\mathbf{M}\eta_1 + \ldots + \mathbf{M}\eta_n\right) = \mathbf{M}\eta_1 = \frac{1}{2},$$

$$\mathbf{D}\overline{\eta} = \frac{1}{n^2}\left(\mathbf{D}\eta_1 + \ldots + \mathbf{D}\eta_n\right) = \frac{1}{n}\,\mathbf{D}\eta_1 = \frac{1}{12\,n}$$

(во второй строке использована независимость случайных величин $\eta_1, \ldots, \eta_n$).

Обратим внимание на то, что случайные величины η_1 и $\overline{\eta}$ имеют одинаковое математическое ожидание, но дисперсия у $\overline{\eta}$ в n раз меньше. Эти соотношения, очевидно, выполняются и для произвольных независимых одинаково распределенных случайных величин $\varepsilon_1, \ldots, \varepsilon_n$ с конечной дисперсией. Такая модель используется для описания ошибок измерения.

Семь раз отмерь, а один — отрежь.

3. Максимум из случайных величин $\eta_1, \ldots, \eta_n$ не превосходит x тогда и только тогда, когда все η_i не больше, чем x (рис. 16), поэтому

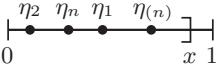


Рис. 16

$$F_{\eta_{(n)}}(x) = \mathbf{P}\left(\eta_{(n)} \leqslant x\right) = \mathbf{P}\left(\eta_1 \leqslant x, \ldots, \eta_n \leqslant x\right).$$

[*)] Здесь $f(p) \sim g(p)$ означает, что $f(p)/g(p) \to 1$.

В силу независимости случайных величин η_i из формулы (5) для $x \in [0,1]$ выводим, что

$$F_{\eta_{(n)}}(x) = \mathbf{P}(\eta_1 \leqslant x) \cdot \ldots \cdot \mathbf{P}(\eta_n \leqslant x) = [\mathbf{P}(\eta_1 \leqslant x)]^n = x^n.$$

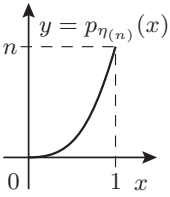


Рис. 17

График соответствующей плотности

$$p_{\eta_{(n)}}(x) = dF_{\eta_{(n)}}(x)/dx = nx^{n-1} I_{[0,1]}$$

изображен на рис. 17 (для $n > 2$).

Применяя формулу (2), вычисляем

$$\mathbf{M}\eta_{(n)} = \int_0^1 x\, nx^{n-1} dx = n \int_0^1 x^n dx = \frac{n}{n+1}.$$

Замечание. Интуитивно ясно, что длины отрезков, на которые делят $[0,1]$ взятые наудачу n точек, распределены одинаково (см. задачу 7 из гл. 10). Поэтому самая правая из точек будет находиться в среднем на расстоянии $1/(n+1)$ от 1. [Однако, *наименьший* из отрезков разбиения имеет длину порядка $1/n^2$ (задача 4 из гл. 4).]

Наконец, $\mathbf{M}\eta_{(n)}^2 = \int\limits_0^1 x^2\, nx^{n-1} dx = n/(n+2)$, откуда в силу соотношения (4) находим, что

$$\mathbf{D}\eta_{(n)} = n/(n+2) - [n/(n+1)]^2 = n/[(n+1)^2(n+2)].$$

Замечание. Дисперсия $\mathbf{D}\eta_{(n)}$ с ростом n убывает намного быстрее, чем дисперсия $\mathbf{D}\overline{\eta}$: порядок малости первой есть $1/n^2$, второй — $1/n$. Это связано с тем, что плотность $p_{\eta_1}(x) = I_{[0,1]}$ имеет разрыв в точке $x = 1$.

4. Вероятность p_k того, что до первого «успеха» в схеме Бернулли будет ровно k «неудач», в силу независимости испытаний равна $q^k p$, где $q = 1 - p$ (рис. 18). Это так называемое *геометрическое распределение.*[*)] Случайная величина ν дает пример дискретной случайной величины, имеющей счетное множество значений: $\mathbf{P}(\nu = k) = p_k$, $k \geqslant 0$. Суммируя геометрическую прогрессию, находим $\mathbf{P}(\nu > k) = p_{k+1} + p_{k+2} + \ldots = q^{k+1} p\,(1 + q + \ldots) = q^{k+1}$. Применяя формулу (3), получаем $\mathbf{M}\nu = q + q^2 + q^3 + \ldots = q/p$.

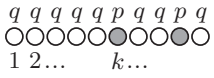


Рис. 18

5. Пусть Y_j — число проверок, потребовавшихся для j-й пары обследуемых ($j = 1,2,\ldots,n/2$), $q = 1 - p$. Тогда

$$Y_j = \begin{cases} 1 & \text{с вероятностью } q^2 \quad \text{(нет больных)}, \\ 2 & \text{с вероятностью } qp \quad \text{(первый здоров, второй болен)}, \\ 3 & \text{с вероятностью } (pq + p^2) = p \quad \text{(в противном случае)}. \end{cases}$$

[*)] Вероятности p_k образуют геометрическую прогрессию.

Согласно формуле (1), $\mathbf{M}Y_j = 1 \cdot q^2 + 2 \cdot qp + 3 \cdot p = 1 + 3p - p^2$. Отсюда находим *ожидаемое общее число проверок*

$$\mathbf{M}Z = \mathbf{M}Y_1 + \ldots + \mathbf{M}Y_{n/2} = (n/2)\,\mathbf{M}Y_1 = n\,(1+3p-p^2)/2.$$

Следовательно, «парная» стратегия в среднем эффективней индивидуальной проверки, когда $1+3p-p^2 < 2$, т. е. при условии, что $p < (3-\sqrt{5})/2 = 1-\varkappa \approx 0{,}382$. Здесь $\varkappa = (\sqrt{5}-1)/2 \approx \approx 0{,}618$ обозначает *«золотое сечение»* — пропорцию, почитавшуюся в древнегреческом искусстве и архитектуре, при которой «меньшее» относится к «большему», как «большее» к «целому»: $(1-\varkappa) : \varkappa = \varkappa : 1$ (рис. 19).

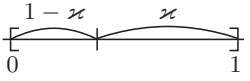


Рис. 19

Любопытно, что при $p \geqslant 1-\varkappa$ вообще не существует стратегии проверки, которая экономичнее индивидуальной. Этот красивый результат получил в 1960 г. П. Ангар (см. [82, с. 147]).

6. Положим $\varepsilon = \varepsilon(p) = -\ln(1-p)$. Разложение логарифма в ряд при $p \to 0$ дает эквивалентность $\varepsilon \sim p$. Подставив ε в уравнение (6), получим

$$1/x^2 = \varepsilon\, e^{-\varepsilon x}. \tag{7}$$

Покажем, что при достаточно малых ε это уравнение имеет два корня: $x_0 = x_0(\varepsilon)$ и $x_1 = x_1(\varepsilon)$ — соответственно точка минимума и точка локального максимума функции $H(x)$ (рис. 22).

Пусть $y = \varepsilon x$. Легко видеть, что (7) равносильно уравнению

$$y^2 e^{-y} = \varepsilon. \tag{8}$$

Дифференцированием устанавливается, что левая часть уравнения (8) на множестве $\{y \geqslant 0\}$ имеет максимум $M = 4e^{-2}$ в точке $y_+ = 2$.

Из рис. 20 очевидно, что при $\varepsilon < M$ уравнение (8) имеет два корня: $y_0 = y_0(\varepsilon) \to 0$ и $y_1 = y_1(\varepsilon) \to \infty$ при $\varepsilon \to 0$. Покажем, что $x_0 = y_0/\varepsilon \sim 1/\sqrt{\varepsilon}$.

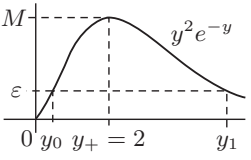


Рис. 20

Действительно,

$$y_0^2/\varepsilon = e^{y_0} \to 1 \Rightarrow y_0/\sqrt{\varepsilon} \to 1 \Leftrightarrow y_0 \sim \sqrt{\varepsilon}.$$

Отсюда $x_0 \sim 1/\sqrt{\varepsilon} \sim 1/\sqrt{p}$. Для доказательства эквивалентности $H(x_0) \sim 2\sqrt{p}$ остается подставить найденную асимптотику в формулу, определяющую функцию $H(x)$.

ОТВЕТЫ НА ВОПРОСЫ

1. Возьмем $A_k = (c-1/k, c]$, $k = 1,2,\ldots$. Эти события вложены: $A_k \supset A_{k+1}$, причем $c = \bigcap A_k$. По свойству непрерывности (П1) $\mathbf{P}(\xi = c) = \lim\limits_{k\to\infty} \mathbf{P}(c-1/k < \xi \leqslant c) = F_\xi(c) - F_\xi(c-)$. Отсюда заключаем, что если функция $F_\xi(x)$ непрерывна, то вероятность попадания ξ в любую фиксированную точку на прямой равна 0.

Прошу мне дать ответ.

Софья
в «Горе от ума»
А. С. Грибоедова

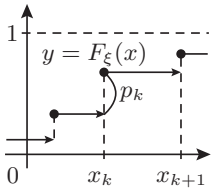


Рис. 21

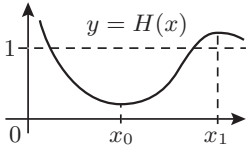


Рис. 22

Для меня давно уже аксиома, что мелочи — это самое важное.

А. Конан Дойл, «Установление личности»

2. $\mathsf{P}(\tau > 3/\lambda) = 1 - \mathsf{P}(\tau \leqslant 3/\lambda) = 1 - F_\tau(3/\lambda) = e^{-3} \approx 0{,}05$. Те, кто знаком с усиленным законом больших чисел (П6), могут отсюда заметить, что только 5% приборов с показательным временем работы до поломки служат более трех средних сроков $1/\lambda$ (согласно примеру к формуле (2)).

3. См. рис. 21.

4. Используем свойства 1 и 2 математического ожидания из П2:

$$\mathsf{M}(\xi - \mathsf{M}\xi)^2 = \mathsf{M}[\xi^2 - 2\,\xi\,\mathsf{M}\xi + (\mathsf{M}\xi)^2] =$$
$$= \mathsf{M}\xi^2 - 2\,(\mathsf{M}\xi)^2 + (\mathsf{M}\xi)^2 = \mathsf{M}\xi^2 - (\mathsf{M}\xi)^2.$$

5. В группе очень большого размера k почти обязательно будут присутствовать больные, и поэтому объединенная проверка станет излишней.

6. На рис. 22 изображен правильный график $H(x)$ в случае малого p. Действительно, $H(x) \to +\infty$ при $x \to 0$ и $H(x) \to 1$ при $x \to +\infty$, причем в последнем случае функция $1/x$ убывает медленнее, чем $(1-p)^x$, и поэтому график $H(x)$ приближается к асимптоте $y = 1$ сверху.

Глава 2

ДАТЧИКИ СЛУЧАЙНЫХ ЧИСЕЛ

Пусть $\eta_1,\eta_2,\ldots$ — координаты точек, взятых наудачу из отрезка $[0,1]$, т. е. независимые и равномерно распределенные на отрезке $[0,1]$ случайные величины.

Бросая в воду камешки, смотри на круги, ими образуемые; иначе такое бросание будет пустою забавою.

Козьма Прутков

Проблема. Как построить числовую последовательность $y_1,y_2,\ldots$, которую можно было бы рассматривать как реализацию случайных величин $\eta_1,\eta_2,\ldots$?

Элементы такой последовательности называются *псевдослучайными числами*, а устройства (или алгоритмы) для их получения — *датчиками*.

§ 1. ФИЗИЧЕСКИЕ ДАТЧИКИ

Простейшим физическим датчиком является, вероятно, рулетка и подобные ей устройства. Рассмотрим вращающуюся с малым трением вокруг оси стрелку, конец которой описывает окружность единичной длины (рис. 1). Раскручивая повторно стрелку, будем получать в качестве $y_1,y_2,\ldots$ координаты конца стрелки в местах остановок.

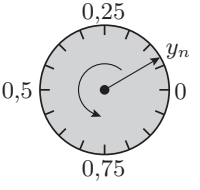


Рис. 1

Другой датчик основан на следующем утверждении из теории вероятностей (см. [12, с. 242], [39, с. 49]).

Утверждение. Для того, чтобы случайная величина η была равномерно распределена на $[0,1]$ необходимо и достаточно, чтобы разряды ζ_i ее двоичной записи (т. е. $\eta = \sum\limits_{i=1}^{\infty} 2^{-i}\zeta_i$) образовывали схему Бернулли с вероятностью «успеха» $p = 1/2$ (см. § 3 гл. 1).

Вопрос 1. Как с помощью $\zeta_1,\zeta_2,\ldots$ построить бесконечную последовательность независимых равномерно распределенных на $[0,1]$ случайных величин $\eta_1,\eta_2,\ldots$?

Таким образом, для получения одного псевдослучайного числа с точностью до 2^{-n} можно подбросить симметричную монетку n раз и сложить 2^{-i} для тех i $(i = 1,\ldots,n)$, при которых выпадал герб.

Вместо монетки можно использовать шум в электроприборах (см. [58, с. 269]). Обозначим через T_i моменты времени, когда шум переходит некоторый пороговый уровень C снизу вверх или сверху

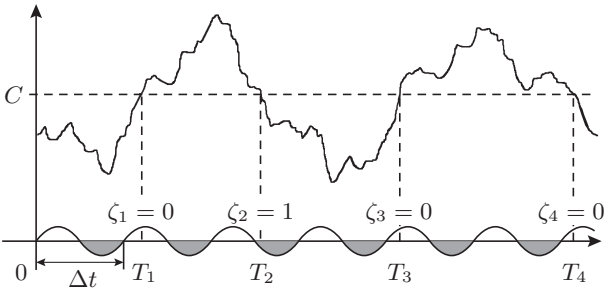


Рис. 2

вниз (рис. 2). Положим ζ_i равной 0 или 1 в зависимости от того, перейден ли порог во время первой или во время второй половины цикла электронных часов, у которых длина цикла Δt намного меньше, чем среднее время между переходами шума через уровень C.

Вопрос 2. Обратно, как из $\eta_1,\eta_2,\ldots$ получить схему Бернулли $\zeta_1,\zeta_2,\ldots$ с заданной вероятностью «успеха» p?

Этим и другим физическим датчикам свойственны следующие общие **недостатки**:

1) для работы датчиков необходимо специальное оборудование, которое обычно требует тщательной настройки;

2) опыт, использующий генерируемые физическим датчиком числа, не воспроизводим в том смысле, что нельзя получить те же самые $y_1,y_2,\ldots$ при его повторном проведении;

3) физические датчики плохо совместимы с компьютерами, так как время получения псевдослучайных чисел несоизмеримо велико по сравнению со скоростью расчетов.

Для преодоления этих недостатков используют таблицы случайных чисел и математические датчики.

§ 2. ТАБЛИЦЫ СЛУЧАЙНЫХ ЧИСЕЛ

Таблица случайных чисел представляет собой зафиксированные результаты работы некоторого датчика. Обычно она имеет вид последовательности псевдослучайных цифр, разбитых на группы для удобства использования (см. Т1).

Т1 обозначает табл. 1 в конце книги.

Рис. 3

Каждый может составить собственную таблицу, вынимая из шляпы бумажки с номерами от 0 до 9 или подбрасывая правильный икосаэдр, у которого каждая из цифр нанесена на 2 из 20 граней (рис. 3).

Как с помощью такой таблицы получать псевдослучайные числа?

Сначала выберем наугад первое число; для чего можно, не глядя в таблицу, загадать номера строки и столбца. Соответствующий набор цифр принимается в качестве знаков после запятой в десятичном представлении y_1. Например, загадав в таблице Т1 строку 1 и столбец 2, получим $y_1 = 0{,}09$.

Далее, начиная с выбранного места, будем считывать таблицу по столбцу (по строке или в любом другом порядке, который не зависит от содержания таблицы) и получать y_2, $y_3, \ldots$. Так, считывая Т1 вниз по столбцу, генерируем

$y_1 = 0{,}09; \quad y_2 = 0{,}54; \quad y_3 = 0{,}42; \quad y_4 = 0{,}01; \ldots$

Если требуются псевдослучайные числа с точностью не до двух, а до четырех знаков после запятой, то нужно считывать также и пары цифр, расположенные в соседнем столбце:

$y_1 = 0{,}0973; \quad y_2 = 0{,}5420; \quad y_3 = 0{,}4226; \quad y_4 = 0{,}0190; \ldots$

При всей своей простоте использование таблицы случайных чисел может приводить к неверным заключениям (см. § 5), да и сама таблица может оказаться недостаточно качественной.

В книге [72] в качестве случайных цифр предлагаются 20 000 знаков после запятой в десятичном представлении числа π. Однако среди первых 10 000 цифра 0 встречается только 937 раз. Согласно центральной предельной теореме (П6) для независимых равновероятных цифр такое может наблюдаться не чаще, чем в двух случаях из ста (см. также задачу 2 гл. 18).

Таблицы случайных чисел неудобны для использования в компьютерных программах тем, что требуют для своего хранения довольно много оперативной памяти. По этой причине для генерации псевдослучайных чисел чаще применяют так называемые *математические датчики*.

§ 3. МАТЕМАТИЧЕСКИЕ ДАТЧИКИ

Эти датчики обычно представляют собой рекуррентные алгоритмы, генерирующие число y_n по предыдущему числу y_{n-1}.

Рассмотрим вначале простой, но довольно «плохой» датчик — *метод середины квадрата* (Дж. фон Нейман, 1946). Зададим произвольное четырехзначное*) число k_0. Например, пусть $k_0 = 8473$. Вычислим $k_0^2 = 71\mathbf{7917}29$. Выделив средние 4 цифры, получим $k_1 = 7917$. Положим $y_1 = k_1 \cdot 10^{-4} = 0{,}7917$. Затем вычислим $k_1^2 = 62\mathbf{6788}89$. Тогда $k_2 = 6788$, $y_2 = 0{,}6788$ и т. д. Ясно, что 1) выбор числа k_0 полностью определяет всю последовательность $y_1, y_2, \ldots$; 2) процесс зациклится не позднее, чем через 10^4 шагов (существует число, которое сразу воспроизводит самое себя: $3792^2 = 14\mathbf{3792}64$); 3) можно так неудачно задать k_0 (например, 1000 или 0085), что $y_n = 0$, начиная с некоторого n.

Более сложный и часто используемый на практике *мультипликативный датчик* работает по следующей схеме. Задаются стартовое число k_0 (например, 1), множитель m и делитель d. Далее

*) Чтобы можно было использовать 8-разрядный калькулятор.

последовательно вычисляются $y_1, y_2, \ldots$ по формулам

Здесь запись «$a \mod b$» обозначает остаток от деления a на b.

$$\begin{cases} k_n = (m \cdot k_{n-1}) \mod d, \\ y_n = k_n/d. \end{cases} \tag{1}$$

Какие значения можно рекомендовать для чисел m и d? Выбор простого числа $d = 2^{31} - 1 = 2\,147\,483\,647$ предпочтителен для тех компьютеров, которые позволяют использовать 32 двоичных разряда для представления целых чисел. Множитель m выбирают так, чтобы последовательность $k_1, k_2, \ldots$, прежде чем зациклиться, пробегала все возможные значения от 1 до $d - 1$. В результате изучения статистических свойств датчика для различных множителей Дж. Фишман и Л. Мур предложили использовать, в частности, $m = 630\,360\,016$ или $m = 764\,261\,123$ (см. [58, с. 271]).

В программном обеспечении иногда встречаются быстро работающие, но недостаточно качественные датчики. Так, датчик RAND из библиотеки STDLIB Borland C++ зацикливается всего через 232 шага. В [31, с. 190] анализируется датчик RANDU ($d = 2^{31}, m = 2^{16} + 3 = 65\,539$), вошедший в SSP — библиотеку научных программ для IBM-360. Оказывается, что все точки с координатами $(y_{3n-2}, y_{3n-1}, y_{3n})$ располагаются в точности на одной из 15 плоскостей вида $9y_{3n-2} - 6y_{3n-1} + y_{3n} = k$, где $k = -5, \ldots, 9$, вместо того, чтобы равномерно плотно заполнять единичный трехмерный куб (рис. 4)*).

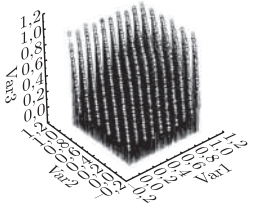


Рис. 4

В заключение, рассмотрим датчик (см. [58, с. 272]), который был исследован Б. Уичманом и И. Хиллом в 1982 г. Чтобы получить y_n, запустим одновременно три мультипликативных датчика с параметрами

$$d_1 = 30\,269, \qquad m_1 = 171;$$
$$d_2 = 30\,307, \qquad m_2 = 172;$$
$$d_3 = 30\,323, \qquad m_3 = 170.$$

Каждый из них на n-м шаге генерирует y'_n, y''_n и y'''_n соответственно. Положим $y_n = \{y'_n + y''_n + y'''_n\}$, где $\{\cdot\}$ обозначает *дробную часть* действительного числа.

Этот датчик имеет период около $3 \cdot 10^{13}$, что значительно превосходит период датчика Фишмана и Мура $2^{31} - 2 \approx 2 \cdot 10^9$, и на компьютере он работает в несколько раз быстрее.

§ 4. СЛУЧАЙНОСТЬ И СЛОЖНОСТЬ

Природе разума свойственно рассматривать вещи не как случайные, но как необходимые.

Б. Спиноза, «Этика», часть 2, теорема XLIV

Проблема построения псевдослучайных чисел волновала в XX веке многие умы. Фон Мизес рассматривал бесконечные последовательности, у которых частоты символов стабилизируются по подпоследовательностям. Какими подпоследовательностями при этом

*) Легко проверить, что $9k_{3n-2} - 6k_{3n-1} + k_{3n} = 0 \mod 2^{31}$

разумно ограничиваться — вопрос, который уточнял Черч. Принципиально иной подход предложил А. Н. Колмогоров. Он провел параллель между случайностью и алгоритмической сложностью: случайным выглядит то, что очень сложно получить. Кстати, на волновавший одно время общественность вопрос о возможности получения кодированной информации из других миров А. Н. Колмогоров отвечал, что если уровень развития иных космических цивилизаций намного выше земного, то сообщения от них будут восприниматься как случайный сигнал.

Что кажется подчас лишь случаем слепым, то рождено источником глубоким.

Ф. Шиллер

Приведем отрывок из [72, с. 177] о связи между случайностью и сложностью.

> «В связи с псевдослучайными числами возникает следующий вопрос. В каком смысле их можно считать случайными, если они получены с помощью детерминированных (неслучайных) алгоритмов? В 1965–66 гг. Колмогоров и Мартин-Леф представили понятие случайности в новом свете. Они определили, когда последовательность из 0 и 1 можно считать случайной. Основная идея состоит в следующем. Чем сложнее описать последовательность (т. е. чем длиннее «самая короткая» программа, конструирующая эту последовательность), тем более случайной ее можно считать. Длина «самой короткой» программы, естественно, различна для разных компьютеров. По этой причине выбирают стандартную машину, называемую машиной Тьюринга. Мерой сложности последовательности является длина наиболее короткой программы на машине Тьюринга, которая генерирует эту последовательность. Сложность — мера иррегулярности. Последовательности, длина которых равна N, называются случайными, если их сложность близка к максимальной. (Можно показать, что большинство последовательностей именно таковы.) Мартин-Леф доказал, что эти последовательности можно считать случайными, так как они удовлетворяют всем статистическим тестам на случайность. Таким образом, сложность и случайность тесно взаимосвязаны. Если программист собирается получать «настоящие» случайные числа, то в силу результатов Колмогорова и Мартин-Лефа он сможет это сделать только с помощью достаточно длинной программы. В то же время на практике генераторы случайных чисел очень короткие. Как совместить эти два факта?»

На практике в отношении к математическим датчикам в основном господствует «презумпция случайности»: алгоритм используют, если не установлено, что он «плохой». Почти каждый датчик выдает приемлемые по качеству псевдослучайные числа в количестве нескольких десятков или сотен. Однако при моделировании случайных процессов порой приходится генерировать многие тысячи чисел. Непросто найти датчик, чтобы на таких длинных последовательностях существующие методы проверки (см. § 2, гл. 12) его не забраковали.

§ 5. ЭКСПЕРИМЕНТ «НЕУДАЧИ»

В [82, с. 29] приведен пример задачи, при попытке решения которой с помощью таблицы случайных чисел возникает интересный парадокс.

Задача. Пусть X_0 обозначает величину моей «неудачи» (скажем, время ожидания в очереди, сумму штрафа или других финансовых потерь). Предположим, что мои знакомые подвергли себя опыту того же типа. Обозначим размеры их «неудач» через $X_1, X_2, \ldots$. Сколько (в среднем) знакомых придется мне опросить, пока не встретится человек, размер неудачи которого не меньше, чем у меня?

Формализуем задачу. Допустим, что $X_0, X_1, \ldots$ — независимые величины с одной и той же непрерывной функцией распределения. Введем случайную величину $N = \min\{n \geqslant 1 : \ X_n \geqslant X_0\}$. Чему равно математическое ожидание $\mathbf{M}N$?

Как будет показано ниже, ответ не зависит от того, какое именно *непрерывное* распределение имеют случайные величины X_n, поэтому будем считать, что они равномерно распределены на отрезке $[0, 1]$.

Имея в виду усиленный закон больших чисел (П6), попытаемся эмпирически оценить $\mathbf{M}N$ средним арифметическим значений n_i, получаемых при моделировании ситуации с помощью таблицы Т1.

Сначала разыграем значение x_0, выбирая наугад некоторое число в таблице. Пусть, скажем, это будет третье число в первой строке. Тогда $x_0 = 0{,}73$. Для моделирования $x_1, x_2, \ldots$ будем считывать таблицу от выбранного числа вниз по столбцу. Получим $x_1 = 0{,}20$, $x_2 = 0{,}26$, $x_3 = 0{,}90$. Этого достаточно, так как $0{,}90 \geqslant 0{,}73$, поэтому $n_1 = 3$. Повторив опыт k раз, можно оценить $\mathbf{M}N$ с помощью $\overline{n} = (n_1 + \ldots + n_k)/k$.

Вопрос 3. Повторите опыт 10 раз. Чему равно $\overline{n}$?

А теперь найдем ответ теоретически. Из непрерывности распределения величины X_n следует, что $\mathbf{P}(X_0 = X_n) = 0$ при $n \geqslant 1$ (см. вопрос 1 гл. 1). Поэтому неравенство в определении случайной величины N можно заменить на строгое. Далее,

$$\mathbf{P}(N > n) = \mathbf{P}(X_0 = \max\{X_0, X_1, \ldots, X_n\}). \tag{1}$$

Если под знаком вероятности заменить X_0 (слева от равенства) на любую из X_i, $i = 1, \ldots, n$, то вероятность, очевидно, не изменится. Поэтому вероятность того, что именно X_0 окажется наибольшей среди $X_0, X_1, \ldots, X_n$, равна $1/(1+n)$ (и не зависит от распределения при условии его непрерывности).

Поскольку случайная величина N принимает только целые неотрицательные значения, то, согласно формуле (3) гл. 1, получаем

$$\mathbf{M}N = \sum_{n=0}^{\infty} \frac{1}{n+1} = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \ldots = \infty, \tag{2}$$

так как гармонический ряд расходится (см. [46, с. 14]).

Почему же попытка оценить $\mathbf{M}N$ с помощью моделирования приводит к результату, совершенно не похожему на теоретический ответ? Этому можно дать несколько объяснений.

Прежде всего, используя псевдослучайные числа, округленные до двух знаков после запятой, мы неявно *непрерывную модель заменяем дискретной*: $\mathbf{P}(\widetilde{X}_n = i/100) = 0{,}01$, $i = 0, \ldots, 99$. Поэтому $\mathbf{P}(\widetilde{X}_0 = \widetilde{X}_n) \neq 0$ при $n \geqslant 1$ и

$$\mathbf{M}\widetilde{N} = \sum_{n=1}^{100} \frac{1}{n}, \tag{3}$$

где, в отличие от формулы (2), суммирование членов гармонического ряда идет до 100, а не до ∞ (см. задачу 4 ниже).

Воспользовавшись тем, что

$$\lim_{m\to\infty} \left(\sum_{n=1}^{m} \frac{1}{n} - \ln m \right) = \gamma = -\int_0^\infty e^{-x} \ln x \, dx \approx 0{,}577,$$

где γ обозначает *постоянную Эйлера,* получаем, что

$$\mathbf{M}\widetilde{N} \approx \ln 100 + \gamma \approx 4{,}605 + 0{,}577 \approx 5{,}2.$$

Однако обычно моделирование дает еще меньшее значение. Это происходит потому, что иногда экспериментатор, неудачно выбрав x_0 (например, 0,98), не желает долго ждать появления еще большего псевдослучайного числа и выбирает другое x_0 — поменьше. Тем самым он производит *подгонку данных* и, отбрасывая большие значения n_i, занижает результат.

Еще одной причиной несоответствия теории и моделирования является *малый размер выборки.* Дело в том, что на результат эксперимента сильно влияют редкие события — появления очень близких к 1 значений x_0, которые обычно не происходят при малом числе испытаний.

Замечание. Используя в эксперименте псевдослучайные числа, округленные до k знаков после запятой, получим $\mathbf{M}\widetilde{N} \approx k \ln 10 + \gamma$, т. е. результат зависит от точности представления чисел $x_0, x_1, \ldots$. Эта ситуация напоминает тот факт, что длина береговой линии, измеряемая по карте, зависит от ее масштаба (рис. 5). Таблица случайных чисел аналогична так называемым *фракталам* — геометрическим объектам, сколь угодно малые части которых подобны целому.*)

Куда как чуден создан свет!

Фамусов в «Горе от ума» А. С. Грибоедова

Л. Эйлер (1707–1783), швейцарский математик, механик, физик и астроном. В 1727–1741, 1766–1783 гг. работал в России.

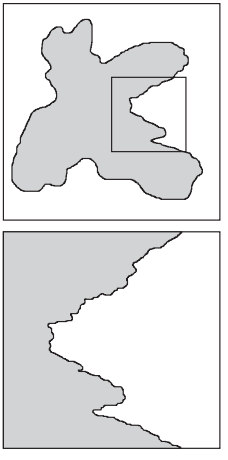


Рис. 5

Правильно в философии рассматривать сходство даже в вещах, далеко отстоящих друг от друга.

Аристотель

*) См., например, А. Д. Морозов «*Введение в теорию фракталов*», Москва—Ижевск: Институт компьютерных исследований, 2002.

§ 6. ТЕОРЕМЫ СУЩЕСТВОВАНИЯ И КОМПЬЮТЕР

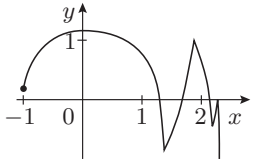


Рис. 6

Верится с трудом...

Чацкий в «Горе от ума»
А. С. Грибоедова

Приведем пример из области численного решения дифференциальных уравнений из [5, с. 85], показывающий, что нужно с осторожностью относиться к компьютерным вычислениям.

Для приближенного решения *задачи Коши*

$$\begin{cases} y'(x) = f(x,y), \\ y(x_0) = y_0 \end{cases}$$

можно использовать *метод Эйлера*: $y_{i+1} = y_i + h\,f(x_i,y_i)$, $x_{i+1} = x_i + h$, где $h > 0$ — некоторый малый шаг (см. [6, с. 430]).

Рассмотрим пример. Пусть $f(x,y) = -x/y$, $y(-1) = 0{,}21$. График численного решения с шагом $h = 0{,}1$ приведен на рис. 6.

На самом деле, правая часть уравнения $-x/y$ имеет разрыв при $y = 0$, поэтому теоретическое решение $y = \sqrt{1{,}0441 - x^2}$ (переменные разделяются) не может быть продолжено в полуплоскость $y < 0$. При $y_i \approx 0$ касательная имеет большой наклон и метод Эйлера «перепрыгивает» на другую интегральную кривую.

Сегодня это действительно слишком просто: вы можете подойти к компьютеру и практически без знания того, что вы делаете, создавать разумное и бессмыслицу с поистине изумительной быстротой.

Дж. Бокс

Книга книгой, а своим умом двигай.

ЗАДАЧИ

1. Попробуйте придумать «свой» датчик случайных чисел (важно не качество датчика, а оригинальность идеи).

2. Случайные величины $\eta_1,\eta_2,\ldots$ — независимы и равномерно распределены на $[0,1]$. Положим

$$K = \{n \geqslant 2\colon \eta_1 > \eta_2 > \ldots > \eta_{n-1} < \eta_n\},$$

т. е. $(K-1)$ — длина «нисходящей серии».

а) Смоделируйте 20 значений случайной величины K с помощью таблицы Т1 и оцените $\mathbf{M}K$ их средним арифметическим.

б) Найдите $\mathbf{M}K$ теоретически.

Указание. Вычислите вероятность $\mathbf{P}(K > n)$ и примените формулу (3) гл. 1.

3. Выполните то же самое для случайной величины L, где

$$L = \min\{n \geqslant 2\colon \eta_1 + \ldots + \eta_n > 1\}$$

(рис. 7).

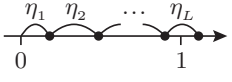


Рис. 7

Указание. Для нахождения вероятностей $\mathbf{P}(L > n)$ используйте *формулу свертки* (см. П3).

4. Докажите формулу (3) с помощью свойств условного математического ожидания (П7):

а) вычислите $\mathbf{M}(\widetilde{N} \mid \widetilde{X}_0 = x_0) = \sum\limits_{n=0}^{\infty} \mathbf{P}(\widetilde{N} > n \mid \widetilde{X}_0 = x_0)$,

б) найдите $\mathbf{M}\widetilde{N}$ по свойству 1 из П7.

5* Пусть случайная величина M равномерно распределена на множестве $\{1,2,\ldots,n\}$: $\mathbf{P}(M = m) = 1/n$, $m = 1,\ldots,n$. Случайная величина J равна остатку от деления n на M. Найдите $\lim\limits_{n\to\infty} \mathbf{P}(J \geqslant M/2)$.

Рис. 8

Замечание. Ввиду рис. 8 кажется правдоподобным, что при увеличении n распределение J будет приближаться к равномерному на множестве $\{0,1,\ldots,M-1\}$, и, следовательно, искомая вероятность должна стремиться к $1/2$. Однако на самом деле этот предел равен примерно 0,386.

Вопрос 4. В чем заключается «обман» рис. 8?

РЕШЕНИЯ ЗАДАЧ

Если на клетке слона прочтешь надпись «буйвол», не верь глазам своим.

Козьма Прутков

1. По-видимому, можно считать случайными последние четыре цифры номеров телефонов из записной книжки.

2. Используя строки таблицы Т1, начиная с первой, получим $k_1 = 3$, $k_2 = 2,\ldots$, $k_1 + \ldots + k_{20} = 64$, оценка для $\mathbf{M}K$ равна 3,2.

Ввиду симметрии любой порядок $\eta_1,\ldots,\eta_n$ равновозможен. Поэтому $\mathbf{P}(K > n) = \mathbf{P}(\eta_1 > \eta_2 > \ldots > \eta_n) = 1/n!$. Отсюда по формуле (3) гл. 1 имеем $\mathbf{M}K = \sum\limits_{n=0}^{\infty} \mathbf{P}(K > n) = \sum\limits_{n=0}^{\infty} \frac{1}{n!} = e \approx 2{,}718$.

3. Так же, как и в предыдущей задаче, для моделирования используем строки таблицы Т1. Получим $l_1 = 4$, $l_2 = 3\ldots$, $l_1 + \ldots + l_{20} = 55$, оценка для $\mathbf{M}L$ равна 2,75.

Пусть $S_n = \eta_1 + \ldots + \eta_n$. Тогда $\mathbf{P}(L > n) = \mathbf{P}(S_n \leqslant 1) = F_{S_n}(1)$. Докажем по индукции с помощью формулы свертки (П3), что функция распределения $F_{S_n}(x) = x^n/n!$ при $0 \leqslant x \leqslant 1$. (При произвольных x функция распределения $F_{S_n}(x)$ задается формулой (4) гл. 4.)

База. При $n = 1$ функция распределения $F_{S_1}(x) = F_{\eta_1}(x) = x$, $0 \leqslant x \leqslant 1$.

Шаг. Так как $S_n = S_{n-1} + \eta_n$, где S_{n-1} и η_n независимы, то

$$F_{S_n}(x) = \int\limits_{-\infty}^{+\infty} F_{S_{n-1}}(x-y)\, F_{\eta_n}(dy) = \int\limits_{0}^{x} \frac{(x-y)^{n-1}}{(n-1)!}\, dy = \frac{x^n}{n!}\,.$$

Подставив $x = 1$, находим $\mathbf{P}(L > n) = 1/n!$. Следовательно, случайные величины L и K из задачи 2 одинаково распределены, $\mathbf{M}L = \mathbf{M}K = e$.

Иное (геометрическое) решение вытекает из того, что подмножества точек n-мерного единичного куба, удовлетворяющие неравенствам $x_1 + \ldots + x_n \leqslant 1$ и $x_1 \geqslant x_2 \geqslant \ldots \geqslant x_n$, представляют собой n-мерные симплексы[*)]: первому из них,

[*)] Вершинами произвольного n-мерного симплекса служат $(n+1)$ точек из пространства $\mathbb{R}^n$, не лежащие ни в какой $(n-1)$-мерной гиперплоскости.

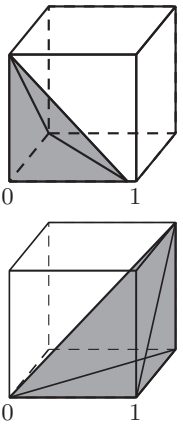


Рис. 9

помимо начала координат $(0,\ldots,0)$, принадлежат вершины куба вида $(0,\ldots,0,1,0,\ldots,0)$, а второму — вида $(1,\ldots,1,0,\ldots,0)$ (рис. 9 для $n=3$).

Линейное преобразование с верхнетреугольной матрицей

$$\begin{pmatrix} 1 & 1 & \ldots & 1 \\ 0 & 1 & \ldots & 1 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \ldots & 0 & 1 \end{pmatrix}$$

отображает первый симплекс на второй. Якобиан преобразования (П9) равен 1, поэтому объемы не изменяются. При решении задачи 2 было установлено, что второй симплекс имеет объем $1/n!$.

4. Используем независимость и равномерную распределенность на множестве $\{0;\ 0{,}01;\ \ldots;\ 0{,}99\}$ случайных величин $\widetilde{X}_0,\ldots,\widetilde{X}_n$:

$$\begin{aligned}\mathbf{P}(\widetilde{N}>n\,|\,\widetilde{X}_0=x_0) &= \\ &= \mathbf{P}(\widetilde{X}_1<x_0,\ldots,\widetilde{X}_n<x_0\,|\,\widetilde{X}_0=x_0) = \\ &= \mathbf{P}(\widetilde{X}_1<x_0,\ldots,\widetilde{X}_n<x_0) = \prod_{i=1}^{n}\mathbf{P}(\widetilde{X}_i<x_0) = x_0^n.\end{aligned}$$

Отсюда находим $\mathbf{M}(\widetilde{N}\,|\,\widetilde{X}_0=x_0)=\sum\limits_{n=0}^{\infty}x_0^n=\dfrac{1}{1-x_0}$. Наконец, усредняя по x_0, докажем формулу (3):

$$\begin{aligned}\mathbf{M}\widetilde{N} &= \sum_{x_0}\mathbf{M}(\widetilde{N}\,|\,\widetilde{X}_0=x_0)\,\mathbf{P}(\widetilde{X}_0=x_0) = \\ &= \frac{1}{100}\sum_{x_0=0}^{0{,}99}\frac{1}{1-x_0} = \sum_{n=1}^{100}\frac{1}{n}.\end{aligned}$$

5. Пусть $I_M=2n/M-2\,[n/M]$, где $[\,\cdot\,]$ — *целая часть* числа. Так как $n=M\,[n/M]+J$, то $0\leqslant I_M=2J/M<2$. Отсюда

$$[I_M]=\left[\frac{2n}{M}\right]-2\left[\frac{n}{M}\right]=\left[\frac{2J}{M}\right]=\begin{cases}1, & \text{если } J\geqslant M/2,\\ 0, & \text{если } J<M/2.\end{cases}$$

В соответствии с формулой полной вероятности (П7) запишем

$$\mathbf{P}(J\geqslant M/2)=\frac{1}{n}\sum_{m=1}^{n}[I_m]=\frac{1}{n}\sum_{m=1}^{n}\left(\left[\frac{2n}{m}\right]-2\left[\frac{n}{m}\right]\right).$$

При $n \to \infty$ эта сумма стремится к интегралу

$$\int\limits_0^1 \left(\left[\frac{2}{x}\right] - 2\left[\frac{1}{x}\right] \right) dx = \lim_{n\to\infty} \sum_{m=1}^{n-1} \int\limits_{1/(m+1)}^{1/m} \left(\left[\frac{2}{x}\right] - 2\left[\frac{1}{x}\right] \right) dx =$$

$$= \lim_{n\to\infty} \sum_{m=1}^{n-1} \left(\frac{1}{m+1/2} - \frac{1}{m+1} \right) = 2\left(\frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \ldots \right).$$

Из разложения $\ln(1+x) = x - x^2/2 + x^3/3 - x^4/4 + \ldots$ получаем, что искомый предел равен $2\ln 2 - 1 \approx 0{,}386$.

ОТВЕТЫ НА ВОПРОСЫ

1. Запишем $\zeta_1, \zeta_2, \ldots$ по диагоналям и по каждой i-й строке построим свою случайную величину η_i:

$\eta_1 \leftarrow \zeta_1, \zeta_2, \zeta_4, \zeta_7, \ldots$

$\eta_2 \leftarrow \zeta_3, \zeta_5, \zeta_8, \ldots$

$\eta_3 \leftarrow \zeta_6, \zeta_9, \ldots$

. . . . . .

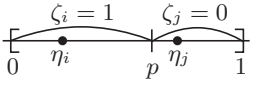


Рис. 10

2. Если $\eta_i \leqslant p$, то положим $\zeta_i = 1$, иначе $\zeta_i = 0$ (рис. 10).

3. Обычно получается результат около 2 или 3.

4. На рис. 8 изображен случай, когда M много меньше n. Однако, для всех $M > 2n/3$ (что происходит с вероятностью $1/3$) справедливо неравенство $J < M/2$ (рис. 11).

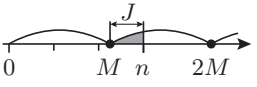


Рис. 11

Глава 3

МЕТОД МОНТЕ-КАРЛО

— Думаю, вам не стоит беспокоиться, — сказал я. — До сих пор всегда оказывалось, что в его безумии есть метод.
— Лучше бы было сказать, что есть безумие в его методе, — пробормотал инспектор.

А. Конан Дойл, «Записки о Шерлоке Холмсе»

В широком смысле методом Монте-Карло называется численный метод решения математических задач при помощи псевдослучайных чисел. Его название происходит от города Монте-Карло в княжестве Монако, знаменитого своими игорными домами.

§ 1. ВЫЧИСЛЕНИЕ ИНТЕГРАЛОВ

Замечательно, что науке, начинавшейся с рассмотрения азартных игр, суждено было стать важнейшим объектом человеческого знания.

Лаплас, «Аналитическая теория вероятностей»

П. Лаплас (1749–1827), французский математик.

Знакомство с методом начнем с рассмотрения задачи численного интегрирования функции $\varphi(x)$, заданной на отрезке $[0,1]$. Как вычислить приближенно интеграл $I = \int\limits_0^1 \varphi(x)\,dx$? Пожалуй, простейший способ — *метод прямоугольников*. Он состоит в замене I на интегральную сумму $I_n = \dfrac{1}{n}\sum\limits_{i=1}^{n} \varphi(x_i)$, где $x_i = \dfrac{i-1/2}{n}$ — это *«узлы» равномерной сетки*, т. е. середины интервалов разбиения отрезка $[0,1]$ на n равных частей (рис. 1).

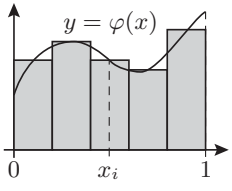


Рис. 1

Б. Тейлор (1685–1731), английский математик.

При условии, что $\varphi(x)$ дважды непрерывно дифференцируема, можно оценить *погрешность метода прямоугольников* $\delta_n = |I - I_n|$ так:

$$\delta_n \leqslant \frac{M}{24\,n^2}\,, \quad \text{где } M = \max_{0\leqslant x\leqslant 1} |\varphi''(x)|. \tag{1}$$

Доказательство. Положим $a_i = x_i - \dfrac{1}{2n}$, $\quad b_i = x_i + \dfrac{1}{2n}$. По формуле Тейлора $\varphi(x) = \varphi(x_i) + \varphi'(x_i)(x - x_i) + \dfrac{1}{2}\,\varphi''(\xi)\,(x - x_i)^2$, где $x \in [a_i, b_i]$ и $\xi = \xi(x) \in [a_i, b_i]$. Поскольку $\int\limits_{a_i}^{b_i} \varphi'(x_i)(x - x_i)\,dx = 0$,

$$R_i = \int_{a_i}^{b_i} [\varphi(x) - \varphi(x_i)]\,dx = \frac{1}{2}\int_{a_i}^{b_i} \varphi''[\xi(x)]\,(x - x_i)^2 dx,$$

$$|R_i| \leqslant \frac{M}{2}\int_{a_i}^{b_i} (x - x_i)^2\,dx = \frac{M}{6}\,(x - x_i)^3\Big|_{a_i}^{b_i} = \frac{M}{24\,n^3}\,.$$

Остается применить неравенство $\left|\sum\limits_{i=1}^{n} R_i\right| \leqslant \sum\limits_{i=1}^{n} |R_i|$. ■

Таким образом, для гладких функций погрешность метода прямоугольников имеет порядок малости $1/n^2$.

Вопрос 1. Каков порядок малости при выборе в качестве «узлов» не середин, а правых концов отрезков разбиения?

Метод Монте-Карло для вычисления интеграла I отличается от метода прямоугольников тем, что в качестве «узлов» используются псевдослучайные числа $y_1, \ldots, y_n$.

Обоснование. Пусть случайные величины $\eta_1, \eta_2, \ldots$ — независимы и равномерно распределены на $[0,1]$. Положим $\xi_i = \varphi(\eta_i)$. По теореме о замене переменных из П2

$$\mathbf{M}\xi_1 = \int\limits_{-\infty}^{\infty} \varphi(y)\, p_{\eta_1}(y)\, dy = \int\limits_{-\infty}^{\infty} \varphi(y)\, I_{[0,\,1]}\, dy = \int\limits_{0}^{1} \varphi(y)\, dy = I.$$

Согласно усиленному закону больших чисел (см. П6) имеем

$$\widehat{I}_n = \frac{1}{n}\sum_{i=1}^{n} \xi_i = \frac{1}{n}\sum_{i=1}^{n} \varphi(\eta_i) \xrightarrow{\text{п. н.}} \mathbf{M}\xi_1 = I \quad \text{при } n \to \infty.$$

Таким образом, если рассматривать псевдослучайные числа $y_1, \ldots, y_n$ как реализацию $\eta_1, \ldots, \eta_n$, то с ростом n погрешность приближения должна стремиться к нулю.

§ 2. «ПРАВИЛО ТРЕХ СИГМ»

Получим оценку для величины погрешности метода Монте-Карло на основе центральной предельной теоремы (П6). Если $0 < \mathbf{D}\xi_1 = \int\limits_0^1 \varphi^2(y)\, dy - I^2 < \infty$, то при $n \to \infty$

$$\frac{\sqrt{n}\,(\widehat{I}_n - I)}{\sqrt{\mathbf{D}\xi_1}} = \frac{\sum\limits_{i=1}^{n} \xi_i - nI}{\sqrt{n\,\mathbf{D}\xi_1}} \xrightarrow{d} Z,$$

где $\xrightarrow{d}$ обозначает сходимость по распределению (см. П5), а предельная случайная величина Z имеет функцию распределения

$$\Phi(x) = \mathbf{P}(Z \leqslant x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-u^2/2}\, du$$

(рис. 2). Такая случайная величина называется *стандартной нормальной* (обозначение: $Z \sim \mathcal{N}(0,1)$).

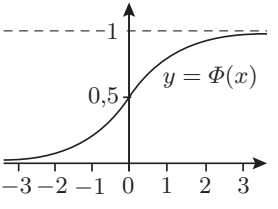


Рис. 2

В силу указанной сходимости для любого $x \geqslant 0$ при $n \to \infty$

$$\mathbf{P}\left(\frac{\sqrt{n}\,|\widehat{I}_n - I|}{\sqrt{\mathbf{D}\xi_1}} \leqslant x\right) \to \mathbf{P}(-x \leqslant Z \leqslant x) = \Phi(x) - \Phi(-x).$$

В силу симметрии имеем равенства

$$\Phi(-x) = 1 - \Phi(x) \quad \text{и} \quad \Phi(x) - \Phi(-x) = 2\,\Phi(x) - 1.$$

Для получения численных оценок рассмотрим подробнее поведение функции $\Phi(x)$. К сожалению, ее нельзя выразить через элементарные функции. Для приближенного вычисления $\Phi(x)$ можно

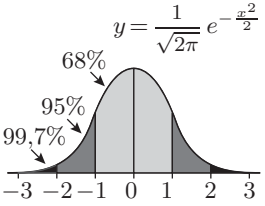


Рис. 3

воспользоваться таблицей Т2. Рисунок 3 иллюстрирует некоторые табличные значения.

Определение. Случайная величина $X = \mu + \sigma Z$, где $Z \sim \mathcal{N}(0,1)$, называется *нормальной с параметрами* μ *и* σ^2 (обозначение: $X \sim \mathcal{N}(\mu,\sigma^2)$). При этом

$$F_X(x) = \mathbf{P}\left(Z \leqslant \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right),$$

$$p_X(x) = \frac{1}{\sigma}\,\Phi'\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma\sqrt{2\pi}}\; e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Вопрос 2.
Как вы думаете, какие три ошибки чаще всего допускают студенты на экзамене, пытаясь написать формулу плотности закона $\mathcal{N}(\mu,\sigma^2)$?

Какой вероятностный смысл имеют характеристики μ и σ^2? Нетрудно убедиться, что $\mu = \mathbf{M}X$, а $\sigma^2 = \mathbf{D}X$. Геометрический смысл параметров μ и σ заключается в том, что прямая $x = \mu$ является осью симметрии плотности $p_X(x)$, а $\mu \pm \sigma$ — точками перегиба $p_X(x)$. Графики плотностей при $\mu = 0$ и нескольких значениях σ приведены на рис. 4.

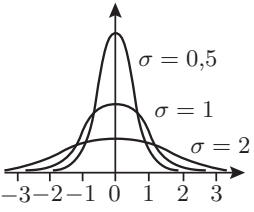


Рис. 4

Согласно определению случайной величины X и с учетом рис. 3 имеем

$$\mathbf{P}(\mu - 3\sigma \leqslant X \leqslant \mu + 3\sigma) = \mathbf{P}(|Z| \leqslant 3) \approx 0{,}997.$$

Другими словами, случайная величина $X \sim \mathcal{N}(\mu,\sigma^2)$ принимает значения из отрезка $[\mu - 3\sigma, \mu + 3\sigma]$ с вероятностью 0,997, которую зачастую не отличают от 1. Это утверждение известно на практике как **«правило трех сигм»**.

Возвращаясь к оценке погрешности метода Монте-Карло, заключаем, что при достаточно больших n в соответствии с «правилом трех сигм» выполняется неравенство

$$|\widehat{I}_n - I| \leqslant 3\sqrt{\mathbf{D}\xi_1}/\sqrt{n} \text{ с вероятностью близкой к 1.} \qquad (2)$$

§ 3. КРАТНЫЕ ИНТЕГРАЛЫ

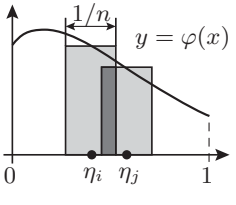


Рис. 5

Неразумно использовать метод Монте-Карло для вычисления одномерных интегралов — для этого существуют квадратурные формулы (см., например, [6, с. 375]), простейшая из которых — рассмотренная выше формула метода прямоугольников. Дело в том, что метод Монте-Карло имеет ряд существенных **недостатков**.

1) Оценка погрешности (2) имеет порядок малости $1/\sqrt{n}$ в отличие от порядка $1/n^2$ оценки (1). Это можно связать с тем, что метод Монте-Карло нерационально использует информацию: прямоугольники «случайной» интегральной суммы частично перекрываются (рис. 5). (Вообще же, для математической статистики это обычно, что оценка отстоит от оцениваемого параметра на величину порядка $1/\sqrt{n}$, где n — число наблюдений.) Из-за мед-

ленной сходимости метод обычно применяют для решения тех задач, где результат достаточно получить с небольшой точностью (5–10%).

2) Для вычисления правой части формулы (2) надо знать, чему равна $\mathbf{D}\xi_1 = \int\limits_0^1 \varphi^2(y)dy - I^2$, или хотя бы уметь ее оценивать.

3) В отличие от метода прямоугольников оценка погрешности метода Монте-Карло справедлива лишь с некоторой вероятностью.

Тем не менее, этот метод (или его модификации) часто оказывается единственным численным методом, позволяющим решить задачу. Особенно он бывает полезен для вычисления интегралов большой кратности. Дело в том, что число «узлов» сетки возрастает как n^k, где k — кратность интеграла (так называемое «проклятие размерности»). Так, чтобы найти интеграл по десятимерному кубу, используя в качестве «узлов» только его вершины, надо $2^{10} = 1024$ раза вычислить значение интегрируемой функции. В практических задачах эти вычисления могут оказаться довольно долгими, например, когда для расчета значений требуется численное решение систем нелинейных или дифференциальных уравнений.

Напротив, метод Монте-Карло не зависит от размерности: чтобы найти приближенное значение интеграла

$$I = \int\limits_0^1 \dots \int\limits_0^1 \varphi(x_1, \dots, x_k)\, dx_1 \dots dx_k$$

с точностью порядка $1/\sqrt{n}$ достаточно случайно набросать n точек в k-мерный единичный куб (разбив псевдослучайные числа на группы из k элементов) и вычислить среднее арифметическое значений φ в этих точках. В частности, если функция — индикатор некоторой области D, то с помощью метода Монте-Карло можно приближенно определить объем этой области. Например, частота случайных точек, попавших под дугу окружности на рис. 6 будет служить приближением к $\pi/4$.

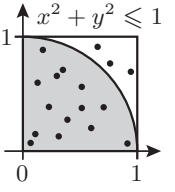


Рис. 6

Отметим, что формула (2) оценки погрешности, полученная для одномерного случая, сохраняется и для $k > 1$.

Рассмотрим модификацию метода Монте-Карло (см. [29, с. 124]), которая называется *расслоенной выборкой (выборкой по группам)*. Каждую из сторон единичного k-мерного куба разобьем на N равных частей. При этом куб разбивается на $n = N^k$ «кубиков» Δ_i со стороной $1/N$. В каждом из Δ_i $(i = 1, \dots, n)$ выбирается независимо равномерно распределенная k-мерная «точка» $\boldsymbol{\eta}_i = (\eta_i^{(1)}, \dots, \eta_i^{(k)})$, и интеграл оценивается с помощью случайной суммы

$$\tilde{I}_n = \frac{1}{n} \sum_{i=1}^{n} \varphi(\boldsymbol{\eta}_i). \qquad (3)$$

Найдем порядок малости *стандартного отклонения* $\sqrt{\mathsf{D}\tilde{I}_n}$ при увеличении n. Так как слагаемые в сумме из (3) независимы, то

$$\mathsf{D}\tilde{I}_n = \frac{1}{n^2}\sum_{i=1}^{n}\mathsf{D}\varphi(\boldsymbol{\eta}_i). \tag{4}$$

Ввиду формулы (4) гл. 1 верно неравенство $\mathsf{D}\varphi(\boldsymbol{\eta}_i) \leqslant \mathsf{M}\varphi^2(\boldsymbol{\eta}_i)$. Отсюда для ограниченной функции φ получаем оценку

$$\sqrt{\mathsf{D}\tilde{I}_n} \leqslant \sqrt{\mathsf{M}\tilde{I}_n^2} \leqslant M/\sqrt{n}\,,$$

где $M = \max|\varphi(x)|$. Это — обычный порядок точности. Однако точность увеличивается, если колебания функции φ в пределах каждого из Δ_i имеют порядок линейных размеров «кубиков». Это выполняется, когда функция φ имеет ограниченные частные производные по каждой из переменных x_j: $|\partial\varphi/\partial x_j| \leqslant L$, $j = 1, \ldots, k$ (рис. 7).

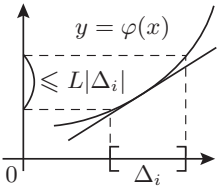


Рис. 7

Ж. Лагранж (1736–1813), французский математик.

Для получения оценки стандартного отклонения для таких функций выберем внутри каждого Δ_i произвольную неслучайную точку $\boldsymbol{y}_i$ (например, центр «кубика»), и, перемещаясь от $\boldsymbol{y}_i$ к $\boldsymbol{\eta}_i$ вдоль осей координат, по теореме Лагранжа о среднем (см. [45, с. 277]) запишем

$$\varphi(\boldsymbol{\eta}_i) - \varphi(\boldsymbol{y}_i) = \sum_{j=1}^{k}\alpha_i^{(j)}\left(\eta_i^{(j)} - y_i^{(j)}\right), \tag{5}$$

где $\alpha_i^{(j)}$ — значения производных $\partial\varphi/\partial x_j$ в некоторых точках из Δ_i (зависящих от $\boldsymbol{\eta}_i$ и $\boldsymbol{y}_i$). Положим $\xi_i^{(j)} = \alpha_i^{(j)}\left(\eta_i^{(j)} - y_i^{(j)}\right)$.

Для произвольных случайных величин $\xi_1, \ldots, \xi_k$, имеющих конечный второй момент (П2) справедливо неравенство (задача 3)

$$\sqrt{\mathsf{D}\sum_{j=1}^{k}\xi_j} \leqslant \sum_{j=1}^{k}\sqrt{\mathsf{D}\xi_j}. \tag{6}$$

Из него и формулы (5) следует, что

$$\sqrt{\mathsf{D}\varphi(\boldsymbol{\eta}_i)} \leqslant \sum_{j=1}^{k}\sqrt{\mathsf{D}\xi_i^{(j)}} \leqslant \sum_{j=1}^{k}\sqrt{\mathsf{M}\left(\xi_i^{(j)}\right)^2}.$$

Так как точки $\boldsymbol{\eta}_i$ и $\boldsymbol{y}_i$ принадлежат Δ_i, то $\left|\xi_i^{(j)}\right| \leqslant L/N = L\,n^{-1/k}$. Поэтому $\mathsf{D}\varphi(\boldsymbol{\eta}_i) \leqslant L^2k^2n^{-2/k}$. Наконец, из равенства (4) выводим оценку

$$\sqrt{\mathsf{D}\tilde{I}_n} \leqslant L\,k\,n^{-1/2-1/k} = C\,n^{-1/2-1/k}.$$

Таким образом, порядок малости $\sqrt{\mathsf{D}\tilde{I}_n}$ меньше, чем порядок $n^{-1/2}$ погрешности обычного метода Монте-Карло. Однако при $k = 1$ погрешность убывает со скоростью $n^{-3/2}$, что медленнее скорости сходимости n^{-2} метода прямоугольников на гладких функциях.

Замечание. Для некоторого класса функций k переменных Холтоном и Соболем были построены (неслучайные) последовательности «узлов», для которых погрешность имеет порядок $\ln^k n/n$, что на практике эквивалентно $n^{-1+\varepsilon}$, где $\varepsilon > 0$ сколь угодно мало. Свойства, формулы и программы для их расчета содержатся в [74].

§ 4. ШАР, ВПИСАННЫЙ В k-МЕРНЫЙ КУБ

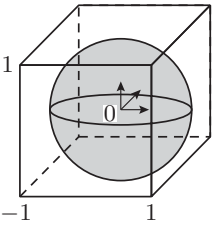


Рис. 8

В многомерном случае при использовании метода Монте-Карло могут возникнуть неожиданности, связанные с тем, что наша обычная геометрическая интуиция может навести на неверное представление о k-мерных множествах.

Рассмотрим k-мерный шар $\{\boldsymbol{x}\colon x_1^2 + \ldots + x_k^2 \leqslant 1\}$, вписанный в k-мерный куб $\{\boldsymbol{x}\colon |x_j| \leqslant 1, \quad j = 1, \ldots, k\}$ (рис. 8 для $k = 3$). Вероятность p_k того, что выбранная случайно в кубе точка окажется внутри шара, равна отношению объема k-мерного шара радиуса $r = 1$ к объему k-мерного куба со стороной 2. Очевидно, $p_1 = 1$, $p_2 = \pi r^2/2^2 = \pi/4 \approx 0{,}785$, $p_3 = \frac{4}{3}\,\pi r^3/2^3 = \pi/6 \approx 0{,}524$. Оказывается, в общем случае

Вопрос 3.
Что подсказывает вам интуиция о порядке этой вероятности при $k = 10$?

$$p_k = \left(\sqrt{\pi}/\,2\right)^k \Big/ \Gamma\left(\frac{k+2}{2}\right).$$

Здесь $\Gamma(x) = \int\limits_0^\infty t^{x-1}e^{-t}dt$ — известная из анализа *гамма-функция Эйлера*, график которой приведен на рис. 9.

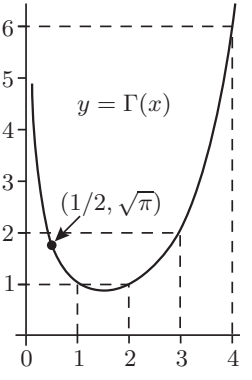


Рис. 9

Интегрированием по частям легко установить *основное свойство гамма-функции*: $\Gamma(x+1) = x\,\Gamma(x)$. Из него вытекает, что $\Gamma(n) = (n-1)!$ при $n = 1, 2, \ldots$, т. е. гамма-функция интерполирует точки плоскости с координатами $(n, (n-1)!)$.

Отметим, кстати, что постоянная Эйлера γ, появившаяся в § 5 гл. 2, равна $-\Gamma'(1)$. В отличие от чисел π и e, неизвестно (на сегодняшний день), является ли эта постоянная иррациональным числом.

Выведем формулу для вероятности p_k. Ввиду симметрии она равна объему части k-мерного шара, находящейся в области $\{\boldsymbol{x}\colon x_j \geqslant 0,\ j = 1, \ldots, k\}$ (см. рис. 6 для $k = 2$). Перейдем от переменных $(x_1, x_2, \ldots, x_k)$ к *полярным координатам* $(r, \varphi_1, \ldots, \varphi_{k-1})$:

$$\begin{aligned}
&x_1 = r\cos\varphi_1,\\
&x_2 = r\sin\varphi_1\cos\varphi_2,\\
&\ldots\ldots\ldots\\
&x_{k-1} = r\sin\varphi_1\ldots\sin\varphi_{k-2}\cos\varphi_{k-1},\\
&x_k = r\sin\varphi_1\ldots\sin\varphi_{k-2}\sin\varphi_{k-1}.
\end{aligned} \tag{8}$$

Любая формула, включенная в книгу, уменьшает число ее покупателей вдвое.

Стивен Хокинг

Якобиан замены $J = r^{k-1}\sin^{k-2}\varphi_1\sin^{k-3}\varphi_2\ldots\sin\varphi_{k-2}$ ([22, с. 440]).

В результате замены получаем представление

$$p_k = \int\limits_0^1 \int\limits_0^{\pi/2} \ldots \int\limits_0^{\pi/2} |J|\,dr\,d\varphi_1\ldots d\varphi_{k-1}. \tag{9}$$

Поскольку $\sin\varphi_j \geqslant 0$ при $0 \leqslant \varphi_j \leqslant \pi/2$, знак модуля в формуле (9) можно убрать, и кратный интеграл распадается в произведение

$$\int\limits_0^1 r^{k-1}dr \cdot I_{k-2} \cdot \ldots \cdot I_1, \quad \text{где } I_m = \int\limits_0^{\pi/2} \sin^m \varphi \, d\varphi, \;\; m = 1, \ldots, k-2.$$

Первый сомножитель равен $1/k$. Для вычисления I_m сделаем замену $y = \sin^2\varphi$. Тогда $I_m = \frac{1}{2}\int\limits_0^1 y^{(m-1)/2}(1-y)^{-1/2}\,dy = \frac{1}{2}B(\frac{m+1}{2}, \frac{1}{2})$, где

$$B(r,s) = \int\limits_0^1 y^{r-1}(1-y)^{s-1}\,dy = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)}, \quad r>0, s>0, \tag{10}$$

— *бета-функция Эйлера*. Так как $I_0 = \pi/2 = \frac{1}{2}\,B\left(\frac{1}{2}, \frac{1}{2}\right)$, то из формулы (10) следует, что $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$ и $I_m = (\sqrt{\pi}/2)\Gamma\left(\frac{m+1}{2}\right)/\Gamma\left(\frac{m+2}{2}\right)$. При перемножении I_m, $m = 1, \ldots, k-2$, гамма-функции перекрестно сокращаются, что и приводит к указанному выше ответу, который можно записать в явном виде:

$$p_k = \begin{cases} (\pi/2)^i \,/\, (2\cdot 4\cdot 6\cdot \ldots \cdot k), & \text{если } k = 2i, \\ (\pi/2)^i \,/\, (3\cdot 5\cdot 7\cdot \ldots \cdot k), & \text{если } k = 2i+1. \end{cases} \tag{11}$$

В частности, по первой из этих формул находим, что $p_{20} \approx 2{,}5 \cdot 10^{-8}$.

Рассмотренный пример наглядно показывает, что может потребоваться очень много псевдослучайных точек, чтобы получить удовлетворительное приближение методом Монте-Карло для интеграла от функции, принимающей большие значения на «тощих» многомерных областях. При этом «тощими» могут оказаться области, которые на первый взгляд таковыми не представляются.

§ 5. РАВНОМЕРНОСТЬ ПО ВЕЙЛЮ

Г. Вейль (1885–1955), немецкий математик.

Определение. Числовая последовательность $x_1, x_2, \ldots$, где $x_i \in [0,1]$, называется *равномерной по Вейлю* , если частота попаданий точек x_i на любой отрезок $[a,b] \subseteq [0,1]$ стремится к его длине $b-a$ при $n \to \infty$.

Согласно усиленному закону больших чисел (П6) реализация $\eta_1, \eta_2, \ldots$ последовательности независимых и равномерно распределенных на отрезке $[0,1]$ случайных величин обладает этим свойством с вероятностью 1.

Существуют и *неслучайные* последовательности, равномерные по Вейлю (см. задачу 5). Такой пример дает

Доказательство этой теоремы можно найти в [29, с. 95]

Теорема 1. Пусть $x_n = \{\alpha\, n\}$, где $\{\,\cdot\,\}$ обозначает *дробную часть* числа, α — любое иррациональное число. Тогда последовательность $x_1, x_2, \ldots$ является равномерной по Вейлю.

Если последовательность $x_1,x_2,\ldots$ равномерна по Вейлю, то для произвольной *интегрируемой по Риману* на отрезке $[0,1]$ функции $\varphi(x)$ осуществляется сходимость (см. [45, с. 561])

Б. Риман (1826–1866), немецкий математик.

$$I_n = \frac{1}{n}\sum_{i=1}^{n}\varphi(x_i) \to I = \int_0^1 \varphi(x)\,dx \quad \text{при } n\to\infty. \tag{12}$$

При использовании равномерной по Вейлю последовательности для вычисления кратных (в частности, двойных) интегралов возможно такое (см. задачу 4), что при $n\to\infty$

$$\frac{1}{n}\sum_{i=1}^{n}\varphi(x_{2i-1},x_{2i}) \to C \neq \int_0^1\int_0^1 \varphi(x,y)\,dxdy.$$

Вопрос 4. Будет ли (12) выполняться для $\widetilde{\varphi}(x)=\sum\limits_{i=1}^{\infty} I_{\{x=x_i\}}$, при условии, что последовательность $x_1,x_2,\ldots$ равномерна по Вейлю, причем все x_i различны?

Определение. Числовая последовательность $x_1,x_2,\ldots$, где $x_i\in[0,1]$, называется *вполне равномерной*, если для произвольного натурального числа k частота попаданий k-мерных точек $\bigl(x_{(n-1)k+1},\ldots,x_{nk}\bigr)$ в любой находящийся внутри единичного k-мерного куба параллелепипед с параллельными координатным осям ребрами стремится к его объему при $n\to\infty$.

В начале тридцатых годов Д. Шамперноун доказал, что число

0,1234567891011121314151617181920212223 ...

(т. е. десятичные знаки являются последовательными натуральными числами) обладает следующим свойством: частота, с которой любая группа из k цифр встречается в его записи, стремится к 10^{-k}. (Неизвестно, обладают ли этим свойством числа π и e.)

Отметим, что вполне равномерность — асимптотическое понятие. Это означает, что конечный отрезок последовательности можно как угодно «испортить», скажем, заменив все элементы нулями, и тем самым сделать ее непригодной для вычислительных целей.

§ 6. ПАРАДОКС ПЕРВОЙ ЦИФРЫ

Как вы думаете, с какой вероятностью число 2^n начинается с цифры 7? (Под вероятностью здесь понимается предел частоты появления 7 при $n\to\infty$, если он существует.) Если вы полагаете, что эта вероятность равна $1/9$, то ошибаетесь!

Действительно, 2^n начинается с цифры m, $1\leqslant m\leqslant 9$, если найдется такое l, что $m\cdot 10^l\leqslant 2^n<(m+1)\cdot 10^l$. Логарифмируя, получаем

$$\log_{10} m + l \leqslant n\log_{10}2 < \log_{10}(m+1)+l.$$

Все три числа, участвующие в этом двойном неравенстве, принадлежат отрезку $[l, l+1]$. Поэтому, переходя к дробным частям, имеем

$$\log_{10} m \leqslant \{n \log_{10} 2\} < \log_{10}(m+1).$$

Так как $\log_{10} 2$ — иррациональное число, то в силу теоремы 1 последовательность $x_n = \{n \log_{10} 2\}$ является равномерной по Вейлю. Отсюда *искомая вероятность*

$$p_m = \log_{10}(m+1) - \log_{10} m. \tag{13}$$

В частности, $p_7 = \log_{10} 8 - \log_{10} 7 \approx 0{,}058$, что почти вдвое меньше, чем $1/9 \approx 0{,}111$.

Можно предложить пари (см. [72, с. 187]), что первая цифра некоторого взятого наугад «большого числа» N окажется не больше, чем 4. Вероятность выигрыша в нем для $N = 2^n$ — вовсе не $4/9 \approx 0{,}444$, а с учетом формулы (13) имеет значение

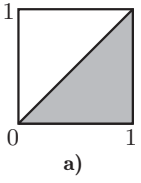
$$p_1 + p_2 + p_3 + p_4 = \log_{10} 5 - \log_{10} 1 \approx 0{,}699.$$

Любопытно, что подсчитанная по таблице 7.6 из [10] частота выигрышей в этом пари для $N = n!$ ($n = 1, \ldots, 100$) равна 0,68.

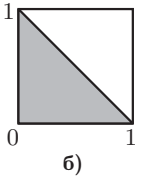
ЗАДАЧИ

Как раз четыреста!... Нет! триста...

Фамусов и Хлестова в «Горе от ума» А. С. Грибоедова



а)



б)

Рис. 10

1. Сколько случайных точек надо бросить в единичный квадрат (см. рис. 6), чтобы получить две верные цифры после запятой у числа π с вероятностью 0,997?

2. Используя разложение числа e в ряд, докажите, что $x_n = \{e\, n!\} \to 0$ при $n \to \infty$.

3. С помощью неравенства Коши—Буняковского (П4) получите неравенство (6).

4. Для датчика $x_n = \{\alpha\, n\}$, где $\alpha > 0$ — произвольное иррациональное число, найдите предел частоты попаданий точек (x_{2i-1}, x_{2i}) в подмножество единичного квадрата

а) $y \leqslant x$,

б) $x + y \leqslant 1$ (рис. 10).

Указание. Для $0 < \alpha < 1$ выразите x_{n+1} через x_n.

5*. Рассмотрим два числовых треугольника:

а)						**б)**					
		1/2						1/2			
	1/3		2/3				1/4	2/4	3/4		
1/4		2/4		3/4		1/8	2/8	3/8	...	7/8	
		...						...			

Либо дождик, либо снег, либо будет, либо нет.

Будут ли получаться равномерные по Вейлю последовательности, если считывать эти треугольники по строкам?

РЕШЕНИЯ ЗАДАЧ

1. Если «успехом» считать попадание точки под дугу окружности $x^2 + y^2 = 1$ (см. рис. 6), то результаты бросаний образуют схему Бернулли с вероятностью «успеха» $p = \pi/4$. Поскольку дисперсия отдельного испытания в этой схеме равна $p(1-p)$, то для частоты «успехов» I_n согласно формуле (2) имеем

$$|I_n - \pi/4| \leqslant C_n = \frac{3\sqrt{(\pi/4)(1-\pi/4)}}{\sqrt{n}} \text{ с вероятностью } 0{,}997.$$

Оценка для погрешности вычисления самого π равна $4C_n$. Приравнивая ее к требуемой точности 0,005, находим $n \approx 970\,835$. Таким образом потребуется почти миллион точек!

Доход не бывает без хлопот.

2. Как известно из математического анализа,

$$e = 1 + \frac{1}{1!} + \frac{1}{2!} + \ldots + \frac{1}{n!} + \varepsilon_n,$$

где

$$\varepsilon_n = \frac{1}{(n+1)!} + \frac{1}{(n+2)!} + \ldots =$$

$$= \frac{1}{(n+1)!}\left(1 + \frac{1}{n+2} + \frac{1}{(n+2)(n+3)} + \ldots\right) <$$

$$< \frac{1}{(n+1)!}\left(1 + \frac{1}{2} + \frac{1}{2^2} + \ldots\right) = \frac{2}{(n+1)!}\ .$$

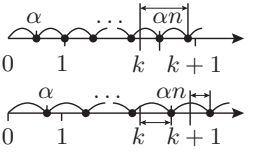


Рис. 11

Поэтому $\{e\,n!\} = n!\,\varepsilon_n < \dfrac{2}{n+1} \to 0$ при $n \to \infty$. Таким образом, подпоследовательность равномерной по Вейлю последовательности сама может и не обладать этим свойством.

3. Преобразуем квадрат левой части неравенства (6):

$$\mathbf{D}\sum_{j=1}^{k}\xi_j = \mathbf{M}\left(\sum_{j=1}^{k}\xi_j - \mathbf{M}\sum_{j=1}^{k}\xi_j\right)^2 = \mathbf{M}\left(\sum_{j=1}^{k}\xi'_j\right)^2 =$$

$$= \mathbf{M}\sum_{i,j=1}^{k}\xi'_i\xi'_j = \sum_{i,j=1}^{k}\mathbf{M}\xi'_i\xi'_j, \quad \text{где } \xi'_j = \xi_j - \mathbf{M}\xi_j.$$

Квадрат правой части неравенства (6) представляется в виде

$$\left(\sum_{j=1}^{k}\sqrt{\mathbf{D}\xi_j}\right)^2 = \left(\sum_{j=1}^{k}\sqrt{\mathbf{M}{\xi'_j}^2}\right)^2 = \sum_{i,j=1}^{k}\sqrt{\mathbf{M}{\xi'_i}^2}\sqrt{\mathbf{M}{\xi'_j}^2}.$$

Остается применить свойство (4) математического ожидания из П2 и неравенство Коши—Буняковского (П4).

4. Прежде всего, можно считать, что $0 < \alpha < 1$, ибо иначе $\alpha' = \{\alpha\}$ задает ту же самую последовательность x_n.

Для таких α справедливо представление $x_{n+1} = \{x_n + \alpha\}$ (два возможных случая приведены на рис. 11). Таким образом, $x_{n+1} = f(x_n)$, где $f(x) = \{x + \alpha\}$ (рис. 12). Поэтому все точки (x_{2i-1}, x_{2i}) будут лежать на графике этой функции, вместо того, чтобы равномерно плотно заполнять весь квадрат.

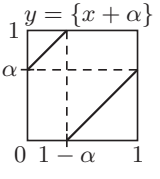


Рис. 12

Вопрос 5.
Почему это верно?

Далее, из равномерности по Вейлю последовательность точек x_i вытекает, что ординаты x_{2i} тоже обладают этим свойством.

Так как $x_{2i} \leqslant x_{2i-1}$ тогда и только тогда, когда x_{2i} попадает на отрезок $[0, \alpha]$, то частота в случае а) будет стремиться к α, а не к $1/2$ (площади треугольника), как должно быть для «настоящего» датчика случайных чисел.

В случае б) из-за симметрии графика $f(x)$ относительно диагонали квадрата $x + y = 1$ предел частоты равен $1/2$.

Замечание. Мультипликативный датчик из § 3 гл. 2

$$\begin{cases} k_n = (m \cdot k_{n-1}) \mod d, \\ y_n = k_n/d \end{cases} \iff \begin{cases} y_0 = k_0/d, \\ y_{n+1} = \{m \cdot y_n\} \end{cases}$$

также удовлетворяет соотношению $y_{n+1} = f(y_n)$ с $f(x) = \{mx\}$ (рис. 13).

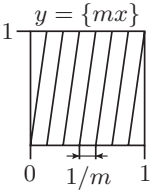


Рис. 13

В [29] приведен ряд утверждений, из которых следует, что при $m \to \infty$ этот датчик позволяет правильно вычислять интегралы любой кратности.

5. В случае а) обозначим через $m_k = k(k+1)/2$ — общее число элементов последовательности в k первых строках треугольника. Пусть $m_{k-1} < n \leqslant m_k$ (т. е. x_n принадлежит k-й строке). Тогда $x_n = (n - m_{k-1})/(k+1)$ (рис. 14). Оценим сверху ν_n — количество попаданий точек x_i, $i = 1, \ldots, n$, в промежуток $(0, x]$ при $0 < x < 1$ (здесь $[\,\cdot\,]$ — целая часть числа):

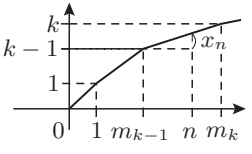


Рис. 14

$$\nu_n \leqslant \sum_{j=1}^{k} [(j+1)x] \leqslant \sum_{j=2}^{k+1} (jx+1) = x\left(\frac{k(k+1)}{2} + k\right) + k = b_k.$$

Аналогично оценим ν_n снизу:

$$\nu_n \geqslant \sum_{j=1}^{k-1} [(j+1)x] \geqslant \sum_{j=2}^{k} (jx-1) = x\left(\frac{k(k+1)}{2} - 1\right) - k + 1 = a_k.$$

Отсюда $\dfrac{a_k}{m_k} \leqslant \dfrac{\nu_n}{n} \leqslant \dfrac{b_k}{m_{k-1}}$, где обе границы стремятся к x при $k \to \infty$. Поэтому последовательность $x_1, x_2, \ldots$ равномерна по Вейлю.

В случае б) количество попаданий точек x_i в промежуток $(0, 1/2]$ для подпоследовательности вида

$$n'_k = 2^{k+1} - k - 2 \quad (k \text{ полных строк})$$

и для подпоследовательности вида

$$n''_k = 2^k - k - 1 + 2^{k-1} \quad (k - 1/2 \text{ строк})$$

одинаково и равно $l_k = 2^k - 1$. При $k \to \infty$ получаем, что $l_k/n'_k \to 1/2$, в то время как $l_k/n''_k \to 2/3$, что противоречит равномерности по Вейлю.

Замечание. Рассмотрим частоту появлений произвольной цифры на j-м месте после запятой в десятичном представлении **а)** $\sqrt{n}$, **б)** $\log n$. Оказывается, что в первом случае предел этой частоты равен $\dfrac{1}{10}$, что следует из равномерности по Вейлю последовательности $x_n = \{\sqrt{n}\}$ (рис. 15). Во втором же случае предела нет, а множество предельных точек представляет собой некоторый отрезок внутри отрезка $[0, 1]$.

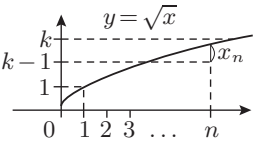


Рис. 15

ОТВЕТЫ НА ВОПРОСЫ

1. По теореме Лагранжа $\varphi(x) = \varphi(x_i) + \varphi'(\xi)(x - x_i)$. Рассуждая так же, как при выводе формулы (1), получаем оценку: $\delta_n \leqslant \frac{1}{2} L/n$, где $L = \max\limits_{0 \leqslant x \leqslant 1} |\varphi'(x)|$. Таким образом, при использовании в качестве «узлов» правых концов отрезков разбиения обычным порядоком малости погрешности является $1/n$. Причиной тому служит отсутствие взаимной компенсации площадей со знаками «+» и «−» на рис. 16, происходящей при выборе «узлов» посередине.

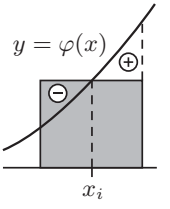


Рис. 16

2. В следующей формуле допущены все три ошибки:

$$\frac{1}{\sqrt{2\pi\sigma}}\, e^{-\frac{(x-\mu)^2}{\sigma}}.$$

Человек редко ошибается дважды — обычно раза три или больше.

Джон Перри Барлоу

3. Согласно формуле (11), вероятность $p_{10} \approx 2{,}5 \cdot 10^{-3}$. Таким образом, почти весь «объем» десятимерного куба сосредоточен в его 1024 «углах». В результате в среднем только одна из 400 случайных точек попадает внутрь десятимерного шара.

4. Для интегрируемости по Риману на отрезке согласно *критерию Лебега* необходимо и достаточно (см. [59, с. 34]), чтобы функция была ограниченной, а множество ее точек разрыва имело *лебегову меру нуль* (т. е. чтобы существовало покрытие этого множества конечной или счетной системой интервалов, сумма длин которых сколь угодно мала).

А. Лебег (1875–1941), французский математик.

Равномерная по Вейлю последовательность обязана, очевидно, быть *всюду плотной* в $[0, 1]$ (в любом интервале из $[0, 1]$ должны присутствовать ее точки). Поэтому множеством точек разрыва функции $\widetilde{\varphi}(x)$ является весь отрезок $[0, 1]$. Следовательно, эта функция не интегрируема по Риману. Однако $\widetilde{\varphi}(x)$ отличается от 0 только на счетном множестве точек x_i, в силу чего интеграл Лебега от нее существует и равен 0. В свою очередь, левая часть формулы (12) при каждом n равна 1.

5. Подпоследовательность ординат x_{2i} генерируется датчиком $\{\beta i\}$ с иррациональным $\beta = 2\alpha$.

Глава 4

ПОКАЗАТЕЛЬНЫЕ И НОРМАЛЬНЫЕ ДАТЧИКИ

Нормальные герои всегда идут в обход.

Из кинофильма «Айболит-66»

В этой главе мы расскажем, как с помощью псевдослучайных чисел имитировать реализации случайных величин с заданным распределением. Прежде всего, познакомимся со стандартным способом моделирования — методом обратной функции. Затем рассмотрим специальные датчики для показательного и нормального законов.

§ 1. МЕТОД ОБРАТНОЙ ФУНКЦИИ

Допустим, что функция распределения $F(x)$ непрерывна и строго возрастает. Тогда на интервале $(0,1)$ существует непрерывная и монотонная обратная функция $F^{-1}(y)$ и справедливо

Утверждение 1. Если случайная величина η равномерно распределена на отрезке $[0,1]$, то случайная величина $\xi = F^{-1}(\eta)$ имеет функцию распределения $F(x)$.

Доказательство. Так как $0 \leqslant F(x) \leqslant 1$, то (см. рис. 1)

$$F_\xi(x) = \mathbf{P}(\xi \leqslant x) = \mathbf{P}(F(\xi) \leqslant F(x)) = \mathbf{P}(\eta \leqslant F(x)) = F(x). \qquad \blacksquare$$

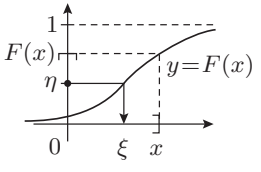


Рис. 1

Определение. *Выборкой* размера n из распределения F называется случайный вектор $(\xi_1,\ldots,\xi_n)$, компоненты которого независимы и одинаково распределены с функцией распределения $F(x)$.

Ввиду утверждения 1 для того, чтобы моделировать реализацию $x_1,\ldots,x_n$ выборки $(\xi_1,\ldots,\xi_n)$ из распределения F, достаточно преобразовать псевдослучайные числа $y_1,\ldots,y_n$ с помощью обратной функции F^{-1}.

Применим метод обратной функции для моделирования выборки из *показательного закона* с функцией распределения $F(x) = (1 - e^{-\lambda x})\, I_{\{x>0\}}$. Найдем обратную функцию:

$$y = 1 - e^{-\lambda x} \iff x = -\frac{1}{\lambda}\ln(1-y).$$

Поэтому формула для моделирования выглядит так:

$$x_i = -\frac{1}{\lambda}\ln(1-y_i),$$

где y_i $(i = 1,\ldots,n)$ — псевдослучайные числа.

Вопрос 1. Нельзя ли предложить похожую, но более простую формулу, дающую тот же результат?

Еще один пример применения метода обратной функции содержится в задаче 1.

Замечание 1. Многие утверждения, выполняющиеся для последовательности независимых и равномерно распределенных на $[0,1]$ случайных величин $\eta_1,\eta_2,\ldots$ (например, усиленный закон больших чисел или центральная предельная теорема из П6) остаются верными и после преобразования, если *«хвосты распределения»* $1-F(x)$ и $F(-x)$ достаточно быстро убывают при $x\to+\infty$. С другой стороны, в § 2 показано, что выборки из распределений с «тяжелыми хвостами» ведут себя иначе.

Хвост сзади, спереди
какой-то чудный выем.

Чацкий в «Горе от ума»
А. С. Грибоедова

§ 2. РАСПРЕДЕЛЕНИЯ ЭКСТРЕМАЛЬНЫХ ЗНАЧЕНИЙ

Рассмотрим пример из [17, с. 18]. Пусть элементы выборки $(X_1,\ldots,X_n)$ имеют функцию распределения

$$F(x)=\begin{cases}1-\dfrac{1}{\ln x} & \text{при } x>e,\\ 0 & \text{при } x\leqslant e.\end{cases} \tag{1}$$

Обозначим $\max\{X_1,\ldots,X_n\}$ через $X_{(n)}$ и вычислим приближенно $\gamma=\mathbf{P}(X_{(n)}>10^7)$ при $n=4$.*) С учетом независимости и одинаковой распределенности случайных величин $X_1,\ldots,X_n$ запишем

$$\mathbf{P}(X_{(n)}\leqslant x)=\mathbf{P}(X_1\leqslant x,\ldots,X_n\leqslant x)=[F(x)]^n.$$

Поскольку $\ln 10\approx 2{,}3$, а $(1-\varepsilon)^n\approx 1-\varepsilon n$ при малых ε, получаем

$$\gamma=1-\left(1-\frac{1}{7\ln 10}\right)^4\approx 1-\left(1-\frac{4}{7\cdot 2{,}3}\right)=\frac{4}{16{,}1}\approx 1/4$$

(более точный подсчет γ на калькуляторе дает значение 0,226). Таким образом, примерно *в каждом четвертом случае* (!) значение $X_{(4)}$ будет превышать 10^7.

Оказывается, из-за того, что функция распределения $F(x)$ имеет «сверхтяжелый» правый «хвост» (рис. 2), распределение случайной величины $X_{(n)}$ чрезвычайно быстро с ростом n «уходит» на $+\infty$, и никаким линейным преобразованием не удается «вернуть» его в конечную область. Точнее: невозможно подобрать такие «центрирующие» константы a_n и «нормирующие» константы $b_n>0$, чтобы последовательность $(X_{(n)}-a_n)/b_n$ сходилась бы по распределению (см. П5) к невырожденному закону**) (см. [17, с. 82]).

Вопрос 2.
Можно ли указать такую последовательность констант b_n, чтобы $X_{(n)}/b_n\xrightarrow{d}0$ при $n\to\infty$?

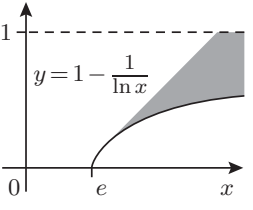


Рис. 2

*) Для выборки, скажем, из закона $\mathcal{N}(0,1)$ ввиду правила «трех сигм» (см. § 2 гл. 3) такая вероятность ничтожно мала.

**) Распределение случайной величины ξ *вырождено*, если $\mathbf{P}(\xi=\text{const})=1$.

Если такой закон для максимума выборки из *некоторого* распределения существует, то он (с точностью до сдвига и масштаба) обязательно принадлежит (см. [17, с. 66]) одному из трех типов **распределений экстремальных значений**:

$$\text{Тип I:}\quad F(x)=\exp\left\{-e^{-x}\right\},\qquad -\infty<x<\infty;$$

$$\text{Тип II:}\quad F(x)=\begin{cases}0, & x\leqslant 0,\\ \exp\left\{-x^{-\alpha}\right\}, & x>0;\end{cases}$$

$$\text{Тип III:}\quad F(x)=\begin{cases}\exp\left\{-(-x)^{\alpha}\right\}, & x\leqslant 0,\\ 1, & x>0.\end{cases}$$

Здесь типы II и III представляют собой однопараметрические классы распределений с параметром $\alpha>0$.

В приведенной ниже теореме даются простые *достаточные* условия «притяжения» (т. е. сходимости $(X_{(n)}-a_n)/b_n$ для некоторых a_n и b_n) к каждому из трех возможных типов.

Теорема 1. Допустим, что

1) $\lim\limits_{x\to\infty} e^x[1-F(x)]=\beta>0$, тогда $X_{(n)}-\ln(\beta n)\xrightarrow{d}\xi$, где случайная величина ξ имеет функцию распределения типа I;

2) для некоторого $\alpha>0$ существует $\lim\limits_{x\to\infty} x^{\alpha}[1-F(x)]=\beta>0$, тогда $X_{(n)}/(\beta n)^{1/\alpha}\xrightarrow{d}\eta$, где случайная величина η имеет функцию распределения типа II;

3) $F(c)=1$, где $c<\infty$, и для некоторого $\alpha>0$ существует $\lim\limits_{x\to c-}(c-x)^{-\alpha}[1-F(x)]=\beta>0$, тогда $(\beta n)^{1/\alpha}(X_{(n)}-c)\xrightarrow{d}\zeta$, где случайная величина ζ имеет функцию распределения типа III.

Доказательство. В первом случае

$$\begin{aligned}&\mathbf{P}\left(X_{(n)}-\ln(\beta n)\leqslant x\right)=\\ &\quad=[F(x+\ln(\beta n))]^n=[1-\beta\,\exp\{-x-\ln(\beta n)\}+o(1/n)]^n=\\ &\quad=\left[1-e^{-x}/n+o(1/n)\right]^n\to\exp\{-e^{-x}\}\text{ при }n\to\infty.\end{aligned}$$

В остальных случаях доказательство аналогично. ■

Необходимые и достаточные условия «притяжения» были получены Б. В. Гнеденко в 1943 г. (см. [17, с. 49]).

Возвращаясь к примеру функции распределения $F(x)$, заданной формулой (1), покажем, что неосуществимое с помощью линейного преобразования, можно в этом случае осуществить с помощью нелинейного. Положим $\widetilde{X}_i=\ln X_i$. Очевидно, что функцией распределения случайной величины $\widetilde{X}_i$ будет $\widetilde{F}(x)=\left(1-\dfrac{1}{x}\right)I_{\{x>1\}}$.

В силу теоремы 1 последовательность $\widetilde{X}_{(n)}/n$ сходится по распределению к закону экстремальных значений II-го типа с $\alpha = 1$.*)

Так как $\min\{X_1, \ldots, X_n\} = -\max\{-X_1, \ldots, -X_n\}$, то предельные распределения для минимума выборки получаются из законов трех экстремальных типов с помощью преобразования

$$G(x) = 1 - F(-x).$$

Например, закону III-го типа соответствует функция распределения $1 - \exp\{-x^\alpha\}$, $x \geqslant 0$. Это распределение известно в теории прочности материалов под именем *закона Вейбулла—Гнеденко* («принцип слабейшего звена»). Его частным случаем при $\alpha = 1$ является показательный закон с параметром $\lambda = 1$ (см. задачу 2).

§ 3. ПОКАЗАТЕЛЬНЫЙ ДАТЧИК БЕЗ ЛОГАРИФМОВ

Выборку из показательного закона можно моделировать и без вычисления логарифмов (см. [82, с. 59]). Для этого рассмотрим таблицу из независимых равномерно распределенных на отрезке $[0, 1]$ случайных величин

$$\begin{array}{ccccc}
\eta_{11}, & \eta_{12}, & \ldots & \eta_{1j}, & \ldots \\
\eta_{21}, & \eta_{22}, & \ldots & \eta_{2j}, & \ldots \\
\ldots & \ldots & \ldots & \ldots & \ldots \\
\eta_{i1}, & \eta_{i2}, & \ldots & \eta_{ij}, & \ldots \\
\ldots & \ldots & \ldots & \ldots & \ldots
\end{array}$$

Для i-й строки таблицы определим случайную величину K_i как $\{j \geqslant 2\colon \eta_{i1} > \eta_{i2} > \ldots > \eta_{i(j-1)} < \eta_{ij}\}$, т. е. $(K_i - 1)$ — это *длина «нисходящей серии»*, начиная от начала i-й строки (см. задачу 2 гл. 2). С последовательностью независимых и одинаково распределенных случайных величин K_i свяжем схему Бернулли, считая «успехом» событие $\{K_i$ четно$\}$ и «неудачей» — $\{K_i$ нечетно$\}$.

Утверждение 2. Обозначим через ν число «неудач» до первого «успеха». Тогда $\mathbf{P}(\nu + \eta_{(\nu+1)1} \leqslant x) = 1 - e^{-x}$, $x > 0$, т. е. сумма числа «неудач» и первой величины в строке «успеха» показательно распределена с параметром $\lambda = 1$.

Доказательство. Прежде всего, вычислим

$$\mathbf{P}(K_1 > n, \eta_{11} \leqslant y) = \mathbf{P}(y \geqslant \eta_{11} \geqslant \eta_{12} \geqslant \ldots \geqslant \eta_{1n}).$$

Последнее событие происходит тогда и только тогда, когда все n точек попадают на отрезок $[0, y]$ и оказываются упорядоченными

*) Переход от X_i к $\widetilde{X}_i$ сильно сжимает прямую с распределенной на ней вероятностной массой, благодаря чему становится возможным получить предельный закон дальнейшим линейным сжатием прямой.

по убыванию. Каждому порядку точек соответствует одно из $n!$ равных по объему подмножеств n-мерного куба со стороной y (см. задачу 3 гл. 2). Таким образом, $\mathbf{P}(K_1 > n, \eta_{11} \leqslant y) = y^n/n!$. Отсюда имеем

$$\mathbf{P}(K_1 = n, \eta_{11} \leqslant y) = \mathbf{P}(K_1 > n-1, \eta_{11} \leqslant y) - \mathbf{P}(K_1 > n, \eta_{11} \leqslant y) =$$
$$= \frac{y^{n-1}}{(n-1)!} - \frac{y^n}{n!}\,.$$

Суммируя по четным значениям n, находим

$$\mathbf{P}(K_1 \text{ четно}, \eta_{11} \leqslant y) = \frac{y}{1!} - \frac{y^2}{2!} + \frac{y^3}{3!} - \frac{y^4}{4!} + \ldots = 1 - e^{-y}.$$

Если положить в этом равенстве $y = 1$, то получим, что вероятность «успеха» $p = \mathbf{P}(K_1 \text{ четно}) = 1 - e^{-1}$.

Вычислим совместное распределение случайных величин ν и $\eta_{(\nu+1)1}$:

$$\begin{aligned}\mathbf{P}(\nu = i, \eta_{(\nu+1)1} \leqslant y) &= \\ &= \mathbf{P}\left(K_j \text{ нечетно}, j = 1, \ldots, i, K_{i+1} \text{ четно}, \eta_{(i+1)1} \leqslant y\right) = \\ &= \mathbf{P}\left(K_j \text{ нечетно}, j = 1, \ldots, i\right) \mathbf{P}\left(K_{i+1} \text{ четно}, \eta_{(i+1)1} \leqslant y\right) = \\ &= q^i \left(1 - e^{-y}\right), \quad \text{где } q = 1 - p = e^{-1}.\end{aligned}$$

Для завершения доказательства утверждения 2 потребуется следующее утверждение (проверить которое предлагается в задаче 3).

Утверждение 3. Целая $[\,\cdot\,]$ и дробная $\{\,\cdot\,\}$ части показательной случайной величины τ с параметром $\lambda = 1$ независимы, причем $[\tau]$ распределена по геометрическому закону: $\mathbf{P}([\tau] = i) = q^i p$, где $p = 1 - e^{-1}, q = 1 - p$, а $\{\tau\}$ имеет функцию распределения $F_{\{\tau\}}(x) = \frac{1}{p}\left(1 - e^{-x}\right)$ при $0 \leqslant x \leqslant 1$ (рис. 3).

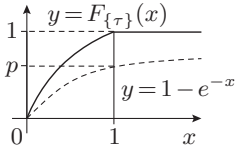


Рис. 3

Отсюда находим совместное распределение $[\tau]$ и $\{\tau\}$:

$$\mathbf{P}([\tau] = i, \{\tau\} \leqslant y) = \mathbf{P}([\tau] = i)\, \mathbf{P}(\{\tau\} \leqslant y) = q^i \left(1 - e^{-y}\right).$$

Оно совпадает с распределением ν и $\eta_{(\nu+1)1}$. Следовательно, $\nu + \eta_{(\nu+1)1}$ и $[\tau] + \{\tau\} = \tau$ распределены одинаково. ■

§ 4. БЫСТРЫЙ ПОКАЗАТЕЛЬНЫЙ ДАТЧИК*)

Вычисление $\ln x$ на компьютере обычно основано на разложении логарифма в ряд Тейлора, и для получения достаточной точности надо выполнить около 20 арифметических действий (см. [6, с. 362]). Из-за этого моделирование показательной выборки с помощью метода обратной функции происходит довольно медленно.

*) Материал этого параграфа технически более сложен, но важен: леммы 1–3 будут неоднократно использоваться в дальнейшем.

При расчете математических моделей (скажем, при определении надежности системы, состоящей из элементов со «случайными» временами работы и ремонта) показательный датчик обычно является глубоко вложенной подпрограммой. Ввиду этого ускорение его работы представляет значительный интерес. Оказывается, на основе приводимой ниже теоремы 2 время моделирования можно *существенно уменьшить*. Для ее доказательства потребуются несколько новых понятий.

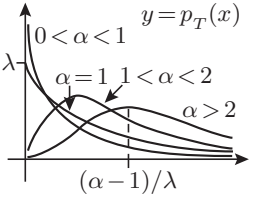


Рис. 4

Определение. Случайная величина T имеет *гамма-распределение* с параметрами $\alpha > 0$ и $\lambda > 0$ (обозначение: $T \sim \Gamma(\alpha,\lambda)$), если ее плотность (см. графики на рис. 4) задается формулой

$$p_T(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)}\, x^{\alpha-1} e^{-\lambda x}\, I_{\{x>0\}},$$

где $\Gamma(\alpha)$ — гамма-функция Эйлера (см. § 4 гл. 3).

Показательный закон является частным случаем гамма-распределения при $\alpha = 1$.

Момент k-го порядка $\mathbf{M}T^k$ гамма-распределения равен

$$\frac{\lambda^\alpha}{\Gamma(\alpha)} \int\limits_0^\infty x^{k+\alpha-1} e^{-\lambda x} dx = \frac{\Gamma(\alpha+k)}{\lambda^k \Gamma(\alpha)} = \frac{\alpha(\alpha+1)\dots(\alpha+k-1)}{\lambda^k}\,. \qquad (2)$$

Вопрос 3. Чему равна дисперсия показательного закона?

Лемма 1. Если случайные величины $T_1 \sim \Gamma(\alpha_1,\lambda)$ и $T_2 \sim \Gamma(\alpha_2,\lambda)$ независимы, то их сумма $T_1 + T_2 \sim \Gamma(\alpha_1 + \alpha_2, \lambda)$.

Доказательство. Без ограничения общности докажем эту лемму при $\lambda = 1$. По формуле свертки (П3) запишем плотность

$$p_{T_1+T_2}(x) = \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \int\limits_0^x t^{\alpha_1-1} e^{-t} (x-t)^{\alpha_2-1} e^{-(x-t)} dt.$$

Сделав замену $t = xy$, приведем выражение для $p_{T_1+T_2}(x)$ к виду

$$C\, x^{\alpha_1+\alpha_2-1} e^{-x}, \quad \text{где } C = \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \int\limits_0^1 y^{\alpha_1-1} (1-y)^{\alpha_2-1} dy.$$

Поскольку интеграл от плотности равен 1, без вычислений находим, что константа $C = 1/\Gamma(\alpha_1 + \alpha_2)$.

Попутно была выведена формула (10) гл. 3, связывающая бета- и гамма-функции:

$$B(r,s) = \int\limits_0^1 y^{r-1}(1-y)^{s-1} dy = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)}\,, \quad r > 0, s > 0. \qquad \blacksquare$$

Определение. Набор $\xi_{(1)} \leqslant \xi_{(2)} \leqslant \ldots \leqslant \xi_{(n)}$ упорядоченных по возрастанию значений компонент выборки $(\xi_1,\ldots,\xi_n)$ называется *вариационным рядом*, а сами случайные величины $\xi_{(k)}$ — *порядковыми статистиками*:

$$\xi_{(1)} = \min\{\xi_1,\ldots,\xi_n\},$$
$$\xi_{(2)} = \max\{\min\{\xi_1,\ldots,\xi_{i-1},\xi_{i+1},\ldots,\xi_n\}\},$$
$$\ldots$$
$$\xi_{(n)} = \max\{\xi_1,\ldots,\xi_n\}.$$

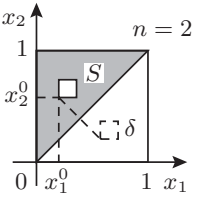


Рис. 5

Лемма 2. Пусть случайные величины $\eta_1,\ldots,\eta_n$ независимы и равномерно распределены на отрезке $[0,1]$. Тогда плотность вектора $(\eta_{(1)},\ldots,\eta_{(n)})$ (см. П8) равна $n!$ на множестве $S = \{\boldsymbol{x}\colon 0 < x_1 < \ldots < x_n < 1\}$ и равна нулю вне этого множества.

Доказательство. Для произвольной точки $(x_1^0,\ldots,x_n^0)$ из S построим «кубик» $\{\boldsymbol{x}\colon x_i^0 \leqslant x_i \leqslant x_i^0+\delta,\ \ i=1,\ldots,n\}$ с ребрами достаточно малой длины δ, целиком лежащий в S (рис. 5 для $n=2$). Так как все перестановки $\eta_{i_1} < \eta_{i_2} < \ldots < \eta_{i_n}$ равновероятны, то

$$\mathsf{P}\left(x_i^0 \leqslant \eta_{(i)} \leqslant x_i^0+\delta,\ i=1,\ldots,n\right) =$$
$$= n!\,\mathsf{P}\left(x_i^0 \leqslant \eta_i \leqslant x_i^0+\delta,\ i=1,\ldots,n,\ \eta_1 < \ldots < \eta_n\right) = n!\,\delta^n.$$

Переход к пределу при $\delta \to 0$ (см. П8) завершает доказательство. ■

Вопрос 4.
Какую плотность имеет вектор $(\xi_{(1)},\ldots,\xi_{(n)})$, если компоненты выборки распределены с плотностью $p(x)$?

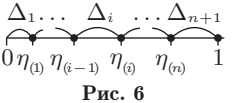


Рис. 6

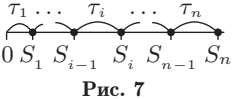


Рис. 7

Следствие. Рассмотрим так называемые *равномерные спейсинги* $\Delta_i = \eta_{(i)} - \eta_{(i-1)}$, $i=1,\ldots,n+1$, $\eta_{(0)}=0$, $\eta_{(n+1)}=1$ (рис. 6). Вектор $(\Delta_1,\ldots,\Delta_n)$ получается из $(\eta_{(1)},\ldots,\eta_{(n)})$ с помощью линейного преобразования с верхнетреугольной матрицей (см. П10), на диагонали которой стоят единицы. Якобиан преобразования равен 1. По формуле преобразования из П8 плотность вектора $(\Delta_1,\ldots,\Delta_n)$ равна $n!$ в области $\{\boldsymbol{x}\colon x_1+\ldots+x_n<1, x_i>0, i=1,\ldots,n\}$ и равна нулю вне этой области.

Лемма 3. Пусть $\boldsymbol{\tau} = (\tau_1,\ldots,\tau_n)$ — выборка из показательного распределения с параметром λ, $S_i = \tau_1+\ldots+\tau_i$ (рис. 7). Тогда вектор $(S_1/S_n,\ldots,S_{n-1}/S_n)$ распределен так же, как вектор порядковых статистик $(\eta_{(1)},\ldots,\eta_{(n-1)})$ для выборки размера $n-1$ из равномерного распределения на $[0,1]$.

Доказательство. То же самое линейное преобразование, что и в следствии леммы 2, переводит случайный вектор $\boldsymbol{S} = (S_1,\ldots,S_n)$ в вектор $\boldsymbol{\tau}$. Используя независимость τ_i, по формуле преобразования (П8) находим плотность $\boldsymbol{S}$:

$$p_{\boldsymbol{S}}(s_1,\ldots,s_n) = \prod_{i=1}^{n} \lambda e^{-\lambda(s_i-s_{i-1})} = \lambda^n e^{-\lambda s_n},\ 0 = s_0 < s_1 < \ldots < s_n.$$

Положим $X_i = S_i/S_n$, $i=1,\ldots,n-1$, $X_n = S_n$. Тогда $S_i = X_i X_n$, $i=1,\ldots,n-1$. Нетрудно убедиться, что якобиан J этого преобра-

зования равен x_n^{n-1}. Отсюда плотность вектора $\boldsymbol{X} = (X_1, \ldots, X_n)$

$$p_{\boldsymbol{X}}(x_1, \ldots, x_n) = |J|\, p_{\boldsymbol{S}}(x_1 x_n, \ldots, x_{n-1} x_n, x_n) = \lambda^n x_n^{n-1} e^{-\lambda x_n} \qquad (3)$$

на множестве $\{\boldsymbol{x}\colon 0 < x_1 < \ldots < x_{n-1} < 1,\ x_n > 0\}$. Проинтегрировав по последней координате (см. П8), получим

$$p_{(X_1, \ldots, X_{n-1})}(x_1, \ldots, x_{n-1}) = \lambda^n \int\limits_0^\infty x_n^{n-1} e^{-\lambda x_n} dx_n = \Gamma(n) = (n-1)!$$

на множестве $\{\boldsymbol{x}\colon 0 < x_1 < \ldots < x_{n-1} < 1\}$. ■

Теперь все готово, чтобы сформулировать и доказать основной результат этого параграфа.

Теорема 2. Пусть случайные величины $\eta_1, \ldots, \eta_{2n-1}$ независимы и равномерно распределены на $[0,1]$, $\xi_1, \ldots, \xi_{n-1}$ — расставленные в порядке возрастания величины $\eta_{n+1}, \ldots, \eta_{2n-1}$, $\xi_0 = 0$, $\xi_n = 1$. Тогда вектор $\boldsymbol{\tau}'$ с компонентами $\tau_i' = -\dfrac{1}{\lambda}\,(\xi_i - \xi_{i-1}) \ln(\eta_1 \eta_2 \ldots \eta_n)$, $i = 1, \ldots, n$, представляет собой выборку из показательного закона с параметром λ.

Например, в случае $n = 2$ имеем

$$\tau_1' = -\frac{1}{\lambda}\,\eta_3 \ln(\eta_1 \eta_2), \quad \tau_2' = -\frac{1}{\lambda}\,(1 - \eta_3) \ln(\eta_1 \eta_2).$$

Экономия времени при моделировании возникает из-за того, что для получения показательной выборки размера n требуется только одно вычисление логарифма. Однако при этом приходится генерировать $2n-1$ псевдослучайных чисел и упорядочивать по возрастанию $n-1$ из них. Расчеты на компьютерах разных типов показали, что практически оптимальным является вариант алгоритма при $n = 3$ (см. [29, с. 27]). Такой датчик работает примерно *вдвое быстрее* метода обратной функции.

Доказательство. Положим $X_i' = \xi_i$, $i = 1, \ldots, n-1$; $X_n' = -\dfrac{1}{\lambda} \ln(\eta_1 \eta_2 \ldots \eta_n) = \sum\limits_{i=1}^{n} \tau_i$, где $\tau_i = -\dfrac{1}{\lambda}\,\ln \eta_i$. Согласно методу обратной функции, случайные величины τ_i показательно распределены с параметром λ, т. е. $\tau_i \sim \Gamma(1, \lambda)$. При этом $\tau_1, \ldots, \tau_n$ независимы, будучи функциями от независимых величин η_i (см. лемму о независимости из § 3 гл. 1). Отсюда в силу леммы 1 выводим, что

$$X_n' \sim \Gamma(n, \lambda),$$

т. е.

$$p_{X_n'}(x) = \frac{\lambda^n}{(n-1)!}\; x^{n-1} e^{-\lambda x} I_{\{x > 0\}}.$$

Очевидно, случайный вектор $(X_1', \ldots, X_{n-1}')$ распределен так же, как и вектор порядковых статистик $(\eta_{(1)}, \ldots, \eta_{(n-1)})$ для выборки

размера $n-1$ из равномерного распределения на отрезке $[0,1]$. По лемме 2 его плотностью является

$$p_{(X'_1,\ldots,X'_{n-1})}(x_1,\ldots,x_{n-1}) = (n-1)!$$

в области $\{\boldsymbol{x}\colon 0 < x_1 < \ldots < x_{n-1} < 1\}$.

При этом $(X'_1,\ldots,X'_{n-1})$ и X'_n независимы как функции от независимых векторов $(\eta_{n+1},\ldots,\eta_{2n-1})$ и $(\eta_1,\ldots,\eta_n)$. Следовательно, плотность вектора $\boldsymbol{X}' = (X'_1,\ldots,X'_n)$ имеет вид

$$p_{\boldsymbol{X}'}(x_1,\ldots,x_n) = p_{(X'_1,\ldots,X'_{n-1})}(x_1,\ldots,x_{n-1})\, p_{X'_n}(x_n) = \lambda^n x_n^{n-1} e^{-\lambda x_n}$$

в области $\{\boldsymbol{x}\colon 0 < x_1 < \ldots < x_{n-1} < 1,\ x_n > 0\}$, т. е. она совпадает с задаваемой формулой (3) плотностью вектора $\boldsymbol{X}$ из леммы 3.

Осталось заметить, что $\boldsymbol{X}'$ преобразуется в $\boldsymbol{\tau}'$ с помощью того же взаимно однозначного отображения, которое в лемме 3 переводит $\boldsymbol{X}$ в $\boldsymbol{\tau}$: $\tau_i = (X_i - X_{i-1})X_n$, $i = 1,\ldots,n-1$; $\tau_n = (1 - X_{n-1})X_n$ (убедитесь!). Поэтому векторы $\boldsymbol{\tau}'$ и $\boldsymbol{\tau}$ одинаково распределены. ■

За один раз дерева не срубишь.

§ 5. НОРМАЛЬНЫЕ СЛУЧАЙНЫЕ ЧИСЛА*)

Для моделирования реализации выборки из распределения $\mathcal{N}(0,1)$ с помощью метода обратной функции надо уметь вычислять значения обратной функции $\Phi^{-1}(y)$ к функции распределения $\Phi(x)$ стандартного нормального закона. Приближенно это можно делать, например, интерполируя достаточно подробную таблицу $\Phi^{-1}(y)$ или заменяя ее «близкой» функцией. Так, Хамакер (см. [58, с. 281]) предложил следующую *аппроксимацию*:

$$\Phi^{-1}(y) \approx \operatorname{sign}(y-0{,}5)(1{,}238z\,(1+0{,}0262z)),$$

где

$$z = \sqrt{-\ln(4y\,(1-y))},$$

$$\operatorname{sign}(x) = \begin{cases} -1, & \text{если } x < 0, \\ 0, & \text{если } x = 0, \\ 1, & \text{если } x > 0. \end{cases} \quad \text{— } \textit{знак числа } x.$$

Она обеспечивает две верные цифры после запятой для $|\Phi^{-1}(y)| \leqslant 4$.

Другой метод приближенного моделирования основывается на *центральной предельной теореме* (П6). Для независимых и равномерно распределенных на $[0,1]$ случайных величин $\eta_1,\eta_2\ldots$ согласно задаче 2 гл. 1 имеем $\mathbf{M}\eta_i = 1/2$, $\mathbf{D}\eta_i = 1/12$. Центральная предельная теорема для суммы $S_n = \eta_1 + \ldots + \eta_n$ дает сходимость

$$(S_n - n/2) \Big/ \sqrt{n/12} \xrightarrow{d} Z \sim \mathcal{N}(0,1) \quad \text{при } n \to \infty.$$

*) В табл. Т1 приведен фрагмент таблицы таких чисел из [10].

Взяв $n = 12$, получим случайную величину $S_{12} - 6$, распределение которой мало отличается от стандартного нормального закона $\mathcal{N}(0,1)$.

Замечание 2. Обратим внимание на то, что, в отличие от Z, при любом n случайная величина S_n ограничена. Используя формулу свертки (П3), можно доказать (см. [82, с. 42]), что функция распределения суммы S_n имеет вид

$$F_{S_n}(x) = \frac{1}{n!} \sum_{k=0}^{n} (-1)^k C_n^k (x-k)_+^n, \tag{4}$$

где $f_+ = \max\{0, f\}$. На каждом из отрезков $[i-1, i]$, $i = 1, \ldots, n$, она является многочленом степени n ($F_{S_n}(x)$ на $[0,1]$ была вычислена при решении задачи 3 гл. 2). На концах отрезков графики гладко «состыкованы». На рис. 8 изображены соответствующие плотности $p_{S_n}(x) = (d/dx)\, F_{S_n}(x)$ для $n = 1,2,3$.

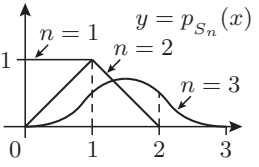


Рис. 8

В заключение рассмотрим способ *точного* моделирования, базирующийся на *нелинейном преобразовании* пары независимых и равномерно распределенных на $[0,1]$ случайных величин η_1, η_2 в пару независимых $\mathcal{N}(0,1)$ случайных величин X, Y:

$$X = \sqrt{-2\ln\eta_1}\,\cos(2\pi\eta_2), \quad Y = \sqrt{-2\ln\eta_1}\,\sin(2\pi\eta_2). \tag{5}$$

Доказательство. Для независимых $\mathcal{N}(0,1)$ случайных величин X и Y плотностью вектора (X,Y) служит

$$p_{(X,Y)}(x,y) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}\, \frac{1}{\sqrt{2\pi}}\, e^{-\frac{y^2}{2}} = \frac{1}{2\pi}\, e^{-\frac{x^2+y^2}{2}}.$$

Обозначим через R и Φ *полярные координаты* точки (X,Y): $X = R\cos\Phi$, $Y = R\sin\Phi$. Используя формулу преобразования из П8 (якобиан замены равен r), находим плотность вектора (R,Φ):

$$p_{(R,\Phi)}(r,\varphi) = \frac{1}{2\pi}\, e^{-\frac{r^2}{2}}\, r, \qquad r > 0,\ 0 < \varphi < 2\pi.$$

Так как она распадается в произведение плотностей

$$p_R(r) = re^{-\frac{r^2}{2}} I_{\{r>0\}} \quad \text{и} \quad p_\Phi(\varphi) = \frac{1}{2\pi}\, I_{\{0<\varphi<2\pi\}},$$

то R и Φ независимы. Интегрируя плотности, вычисляем функцию распределения $F_R(r) = 1 - e^{-r^2/2}$ при $r \geqslant 0$ и $F_\Phi(\varphi) = \varphi/(2\pi)$ при $0 \leqslant \varphi \leqslant 2\pi$.

Отсюда методом обратной функции получаем формулы для моделирования случайных величин R и Φ: $R = \sqrt{-2\ln\eta_1}$, $\Phi = 2\pi\eta_2$, которые только остается подставить в формулы замены координат. ■

Задача 6 дает способ моделирования выборки из закона Коши на основе датчика *нормальных* случайных чисел.

§ 6. НАИЛУЧШИЙ ВЫБОР

Сказка. В некотором царстве, некотором государстве жила-была царевна. И приехали к ней свататься добры-молодцы, один другого лучше. Заходили женихи в палаты царские по очереди да кланялись царевне. Все бы хорошо, да вот беда — добры-молодцы уж больно обидчивы! Коли сразу не давала своего согласия царевна, садились на коня да и подавались восвояси. А ей-то о женихах заранее ничего не ведомо, известно только, сколько всего их пожаловало. Как тут царевне быть, как выбрать из них самого достойного?

Иногда приходится полагаться на случай; ни в чем нельзя быть вполне уверенным в морском сражении.

Г. Нельсон

Вероятностная модель. Предполагая, что женихи становятся в очередь в случайном порядке, будем считать, что пространством элементарных событий (см. П1) является множество всех перестановок из n элементов: $\omega = (i_1, \ldots, i_n)$, где i_k — различные числа от 1 до n, а вероятностная мера — равномерная: $p(\omega) = 1/n!$. Удобно наглядно представлять перестановку в виде точек на действительной прямой с координатами $X_1, \ldots, X_n$, такими что X_k находится правее X_l, если $i_k > i_l$ (см. [69, с. 15]). Точки появляются одна за другой, и желательно остановиться на точке с наибольшей координатой.

Обсуждение. Если царевна выберет первого «попавшегося» жениха, то вероятность, что именно он окажется наилучшим, равна всего лишь $1/n$. Можно, пропустив m первых женихов (чтобы осмотреться, какие вообще бывают женихи), затем остановиться на том, кто понравится больше всех предыдущих, а если такого не окажется — на последнем. Конечно, при этом не исключена возможность упустить самого лучшего!

Вопрос 5.
Почему при любом четном n и $m = n/2$ вероятность наилучшего выбора будет не менее $1/4$?

Оказывается, для произвольного n можно подобрать m_n^*, при котором эта вероятность будет больше, чем $1/e \approx 0{,}368$.

Докажем это. Рассмотрим процесс появления точек X_i на прямой.*) Для следующей после первой точки имеются две возможности — быть левее или правее (мы исключаем возможность равноценных женихов). Вообще, после i-й точки имеется $i+1$ промежутков, куда может попасть следующая $(i+1)$-я точка. Понятно, что число всех возможных размещений равно $2 \cdot 3 \cdot \ldots \cdot n = n!$.

Предположим, что рассматриваемая стратегия приводит к выбору k-й по счету точки (событие A_k), $k = m+1, \ldots, n$. Такой выбор означает, что все точки при $i = m+1, \ldots, k-1$ располагаются левее крайней правой из первых пробных m точек, а k-я — правее всех. На расположение первых m точек это не налагает никаких ограничений, но для $(m+1)$-й точки имеется не $m+1$ промежутков, а лишь m (она не может быть крайней правой), для $(m+2)$-й — лишь

*) Возьмите лист бумаги и карандаш и нарисуйте точки в соответствии с тем, что написано ниже.

$m+1$ промежутков,..., для $(k-1)$-й — лишь $k-2$ промежутков, для k-й имеется единственная возможность — попасть правее всех предшествующих, затем для $(k+1)$-й имеется уже $k+1$ промежутков и т. д. Таким образом, при $k = m+1, \ldots, n$ вероятность

$$\mathbf{P}(A_k) = \frac{1\cdot 2\cdot \ldots \cdot m\cdot m\cdot \ldots \cdot (k-2)\cdot (k+1)\cdot \ldots \cdot n}{1\cdot 2\cdot 3\cdot \ldots \cdot n} = \frac{m}{(k-1)k}\,.$$

Конечно, максимальная точка может попасть и в число пропускаемых первых m. Обозначим это событие через A_m.

Пусть событие B означает остановку на максимуме. Рассмотрим событие $B\cap A_k$, $k = m+1, \ldots, n$, состоящее в том, что остановка произошла на k-м шаге и привела к наилучшему выбору. Для случая, когда k-я по счету точка оказывается абсолютно максимальной, для следующей $(k+j)$-й точки ($j = 1,2,\ldots,n-k$) имеется не $k+j$ промежутков, куда она может попасть, а лишь $k+j-1$ (она не может попасть правее наибольшей). Это означает, что вместо последних сомножителей $(k+1)\cdot \ldots \cdot n$ в формуле для $\mathbf{P}(A_k)$ надо взять $k\cdot \ldots \cdot (n-1)$, что влечет равенство $\mathbf{P}(B\cap A_k) = m/[n(k-1)]$.

Пользуясь формулой полной вероятности (П7) для разбиения A_k, $k = m,\ldots,n$, находим вероятность выбора «наилучшего жениха» при заданном m:

$$\mathbf{P}_m(B) = \sum_{k=m+1}^{n} \mathbf{P}(B\cap A_k) = \frac{m}{n}\sum_{k=m}^{n-1}\frac{1}{k}\,.$$

Обозначим через m_n^* то значение m, при котором $\mathbf{P}_m(B)$ будет наибольшей. В следующей таблице, взятой из [9, с. 38], для некоторых значений n приведены соответствующие числа m_n^* и $p_n^* = \mathbf{P}_{m_n^*}(B)$.

n	2	3	4	5	7	10	20	50	100	1000
m_n^*	0	1	1	2	2	3	7	18	37	368
p_n^*	0,5	0,5	0,458	0,433	0,414	0,399	0,384	0,374	0,371	0,368

Можно доказать, что p_n^* убывает с ростом n. Из неравенства

$$\frac{m}{n}\sum_{k=m}^{n-1}\frac{1}{k+1} \leqslant \frac{m}{n}\int_m^n \frac{dx}{x} = -\frac{m}{n}\ln\frac{m}{n} \leqslant \frac{m}{n}\sum_{k=m}^{n-1}\frac{1}{k}$$

и того, что функция $f(x) = -x\ln x$ имеет максимум в точке $x^* = e^{-1}$, вытекает, что $p_n^* \to f(x^*) = e^{-1} \approx 0{,}368$.

Таким образом, предложенная стратегия является *оптимальной* (в смысле максимума вероятности сделать наилучший выбор), когда количество пропускаемых женихов приблизительно равно целой части числа n/e.

В заключение отметим, что если $X_1,\ldots,X_n$ — выборка из равномерного распределения на $[0,1]$ (или, в силу метода обратной функции, — из любого непрерывного распределения), благодаря тому,

что X_i ограничены сверху 1, существует стратегия, при которой остановка на максимуме происходит с вероятностью не менее 0,58 (см. [9, с. 60]).

ЗАДАЧИ

Навык мастера ставит.

1. Пусть случайная величина Z имеет *плотность Коши* $p_Z(x) = \dfrac{1}{\pi(1+x^2)}$. Получите формулу для моделирования выборки из этого распределения методом обратной функции.

2. Выясните, к какому закону «притягивается»

а) максимум,

б) минимум выборки из показательного распределения с функцией распределения $F(x) = 1 - e^{-x}$ при $x \geqslant 0$.

Указание. Запишите функцию распределения минимума и подберите a_n и b_n.

3. Докажите утверждение 3.

4*. Пусть $\Delta_{(1)} = \min\{\Delta_1, \ldots, \Delta_{n+1}\}$, где Δ_k — k-й спейсинг выборки из равномерно распределенных на $[0,1]$ случайных величин (см. следствие в § 4). Установите на основе леммы 3 сходимость распределения $n^2\Delta_{(1)}$ при $n \to \infty$ к показательному закону с $\lambda = 1$.

Указание. Используйте свойства сходимости (П5) и закон больших чисел (П6).

5*. Для выборки $\tau_1, \ldots, \tau_n$ из показательного распределения

а) убедитесь, что спейсинги $\widetilde{\Delta}_i = \tau_{(i)} - \tau_{(i-1)}$, $i = 1, \ldots, n$, независимы (здесь $\tau_0 = 0$);

б) найдите распределение минимального показательного спейсинга.

6*. Докажите, что если X и Y — независимые $\mathcal{N}(0,1)$-случайные величины, то $Z = X/Y$ распределена по закону Коши.

Указание. Примените формулу преобразования плотности из П8 при замене $(X,Y) \to (X/Y, Y)$.

РЕШЕНИЯ ЗАДАЧ

1. Прежде всего, вычислим функцию распределения случайной величины Z:

$$F_Z(x) = \frac{1}{\pi} \int\limits_{-\infty}^{x} \frac{du}{1+u^2} = \frac{1}{2} + \frac{1}{\pi} \operatorname{arctg} x.$$

Отсюда получаем искомую формулу для моделирования

$$z_i = \operatorname{tg}[\pi(y_i - 1/2)] = -\operatorname{ctg}(\pi y_i), \tag{6}$$

где y_i — псевдослучайные числа.

2. **а)** Согласно теореме 1 для показательного закона с $\lambda = 1$ имеем: $\tau_{(n)} - \ln n \xrightarrow{d} \xi$, где $\mathbf{P}(\xi \leqslant x) = \exp\{-e^{-x}\}$ (тип I).

б) Так как показательное распределение является частным случаем закона Вейбулла—Гнеденко при $\alpha = 1$, то предельный закон для минимума выборки $\tau_{(1)}$ также должен быть показательным. Действительно, используя независимость и одинаковую распределенность случайных величин τ_i, $i = 1, \ldots, n$, запишем

$$\mathbf{P}(\tau_{(1)} > a_n + b_n x) = \prod_{i=1}^{n} \mathbf{P}(\tau_i > a_n + b_n x) = \mathrm{e}^{-n(a_n + b_n x)}.$$

Взяв $a_n = 0$ и $b_n = 1/n$, видим, что случайная величина $n\,\tau_{(1)}$ показательно распределена с $\lambda = 1$.

3. Для $i = 0,1,\ldots$, $0 \leqslant y \leqslant 1$, $q = 1 - p = e^{-1}$ имеем:

$$\mathbf{P}([\tau]=i, \{\tau\}\leqslant y)=\mathbf{P}(i\leqslant\tau\leqslant i+y)=e^{-i}-e^{-i-y}=q^i(1-e^{-y}).$$

При $y = 1$ находим, что $\mathbf{P}([\tau] = i) = q^i p$. Сложим вероятности несовместных событий $\{[\tau] = i, \{\tau\} \leqslant y\}$, $i = 0,1,\ldots,n$ (П7):

$$\mathbf{P}([\tau] \leqslant n, \{\tau\} \leqslant y) = \frac{1}{p}\,(1 - e^{-y}) \sum_{i=0}^{n} q^i p =$$

$$= \frac{1}{p}\,(1 - e^{-y})\,\mathbf{P}([\tau] \leqslant n).$$

Устремляя здесь $n \to \infty$, находим $F_{\{\tau\}}(y)$ и, тем самым, устанавливаем независимость случайных величин $[\tau]$ и $\{\tau\}$.

4. Пусть $(\tau_1, \ldots, \tau_{n+1})$ — показательная выборка с параметром $\lambda = 1$, $S_{n+1} = \tau_1 + \ldots + \tau_{n+1}$. Вектор равномерных спейсингов $(\Delta_1, \ldots, \Delta_{n+1})$ по лемме 3 распределен так же, как вектор $(\tau_1/S_{n+1}, \ldots, \tau_{n+1}/S_{n+1})$. Отсюда $\Delta_{(1)} \sim \tau_{(1)}/S_{n+1}$.

Согласно решению задачи 2, случайная величина $\tau_{(1)}$ имеет показательное распределение с параметром $n+1$. В силу закона больших чисел (П6) $S_{n+1}/(n+1) \xrightarrow{\mathsf{P}} \mathbf{M}\tau_1 = 1$. Учитывая непрерывность при $x > 0$ функции $\varphi(x) = 1/x$, из представления

$$n^2\,\frac{\tau_{(1)}}{S_{n+1}} = [n/(n+1)]^2\,\frac{1}{S_{n+1}/(n+1)}\,(n+1)\tau_{(1)}$$

по свойствам сходимости (П5) получаем в качестве предельного закона показательное распределение с параметром $\lambda = 1$.

Доказанное утверждение также легко выводится из следующего изящного результата Б. де Финетти (1964 г.): для произвольных $x_1 \geqslant 0, \ldots, x_{n+1} \geqslant 0$

$$\mathbf{P}\,(\Delta_1 > x_1, \ldots, \Delta_{n+1} > x_{n+1}) = (1 - x_1 - \ldots - x_{n+1})_+^n,$$

где $f_+ = \max\{0, f\}$.[*)] Если положить $h = x_1 = \ldots = x_{n+1}$, то

$$\mathbf{P}(\Delta_{(1)} > h) = [1 - (n+1)h]_+^n. \qquad (7)$$

[*)] Как его доказать, можно узнать из учебника [82, с. 57].

Взяв $h = x/n^2$ для произвольного $x \geqslant 0$, видим, что правая часть формулы (7) сходится к e^{-x} при $n \to \infty$.

5. **а)** Пусть для краткости $\lambda = 1$. С учетом ответа на вопрос 4 запишем плотность распределения вектора показательных порядковых статистик:

$$p_{(\tau_{(1)},\ldots,\tau_{(n)})}(x_1,\ldots,x_n) = n!\prod_{i=1}^{n} e^{-x_i}, \quad 0 < x_1 < \ldots < x_n.$$

Отсюда (аналогично следствию из § 4) по формуле преобразования плотности (П8) для линейного отображения с якобианом 1, находим плотность вектора спейсингов:

$$p_{(\widetilde{\Delta}_1,\ldots,\widetilde{\Delta}_n)}(u_1,\ldots,u_n) = n!\prod_{i=1}^{n} e^{-(u_1+\ldots+u_i)} = \prod_{i=1}^{n} i\, e^{-iu_{n+1-i}}$$

в области $\{\boldsymbol{u}\colon u_i > 0,\ i = 1,\ldots,n\}$. Таким образом, плотность представляется в виде произведения плотностей. Поэтому случайные величины $\widetilde{\Delta}_i$, $i = 1,\ldots,n$, независимы и показательно распределены с параметрами $n+1-i$ соответственно.

б) Используем этот результат для нахождения распределения *минимального спейсинга* $\widetilde{\Delta}_{(1)} = \min\{\widetilde{\Delta}_1,\ldots,\widetilde{\Delta}_n\}$:

$$\mathsf{P}(\widetilde{\Delta}_{(1)} > x) = \prod_{i=1}^{n} \mathsf{P}(\widetilde{\Delta}_i > x) = \prod_{i=1}^{n} e^{-ix} = e^{-[n(n+1)/2]\,x}, \quad x > 0,$$

т. е. случайная величина $\widetilde{\Delta}_{(1)}$ распределена по показательному закону с параметром $n(n+1)/2$. Ввиду свойств сходимости (П5) отсюда следует, что предельным распределением для $n^2\widetilde{\Delta}_{(1)}$ при $n \to \infty$ служит показательной закон с $\lambda = 1/2$.

6. Совместной плотностью независимых $\mathcal{N}(0,1)$-случайных величин X и Y является

$$p_{(X,Y)}(x,y) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}\, \frac{1}{\sqrt{2\pi}}\, e^{-\frac{y^2}{2}} = \frac{1}{2\pi}\, e^{-\frac{x^2+y^2}{2}}.$$

Сделаем замену $s = x/y$, $t = y$. При этом якобиан обратного преобразования равен t. Согласно формуле преобразования плотности получаем

$$p_{(S,T)}(s,t) = |t|\, p_{(X,Y)}(st,t) = |t|\, \frac{1}{2\pi}\, e^{-\frac{(st)^2+t^2}{2}}.$$

Интегрируя по t, находим

$$p_S(s) = \frac{1}{2\pi}\int\limits_{-\infty}^{+\infty} |t|\, e^{-\frac{(1+s^2)t^2}{2}}\, dt = \frac{1}{\pi(1+s^2)}\int\limits_{0}^{+\infty} e^{-u}\, du = \frac{1}{\pi(1+s^2)}.$$

Другое доказательство вытекает из формул (5), периодичности функции ctg и формулы (6) для моделирования Z:

$$X/Y = \operatorname{ctg}(2\pi\eta_2) \sim \operatorname{ctg}(\pi\eta_2) \sim -\operatorname{ctg}(\pi\eta_2).$$

ОТВЕТЫ НА ВОПРОСЫ

1. Если случайная величина η равномерно распределена на $[0,1]$, то $1-\eta$ имеет, очевидно, такое же распределение. Поэтому для моделирования можно использовать следующую формулу: $x_i = -\frac{1}{\lambda}\ln y_i$.

2. Возьмем $b_n = e^{nc_n}$, где $0 < c_n \to \infty$ при $n \to \infty$. Для любого $x > 0$ при достаточно больших n выполняется условие $xe^{nc_n} > e$. Тогда $\mathsf{P}(X_{(n)}/b_n \leqslant x) = \mathsf{P}(X_{(n)} \leqslant xe^{nc_n}) = [1-(nc_n+\ln x)^{-1}]^n \to 1$ при $n \to \infty$.

3. Для показательной случайной величины τ из формулы (2) при $\alpha = 1$ находим $\mathsf{M}\tau = 1/\lambda$, $\mathsf{M}\tau^2 = 2/\lambda^2$, откуда

$$\mathsf{D}\tau = \mathsf{M}\tau^2 - (\mathsf{M}\tau)^2 = 2/\lambda^2 - 1/\lambda^2 = 1/\lambda^2.$$

4. Вектор $(\xi_{(1)},\ldots,\xi_{(n)})$ получается из вектора равномерных порядковых статистик $(\eta_{(1)},\ldots,\eta_{(n)})$ с помощью преобразования $x_i = F^{-1}(y_i)$, где $F^{-1}(y)$ — обратная функция к функции распределения элементов выборки. Очевидно, что якобианом обратного преобразования служит $J = \prod\limits_{i=1}^{n} p(x_i)$. Следовательно, плотностью вектора порядковых статистик $(\xi_{(1)},\ldots,\xi_{(n)})$ является функция

$$p_{(\xi_{(1)},\ldots,\xi_{(n)})}(x_1,\ldots,x_n) = n!\prod_{i=1}^{n} p(x_i) \quad \text{при } x_1 < x_2 < \ldots < x_n.$$

5. Обратим внимание на места в перестановке чисел $n-1$ и n. В том случае, когда число $n-1$ попадает в первую половину перестановки, а n — во вторую, рассматриваемая стратегия приводит к наилучшему выбору. Из симметрии вероятность такого расположения чисел $n-1$ и n равна $1/4$.

Глава 5

ДИСКРЕТНЫЕ И НЕПРЕРЫВНЫЕ ДАТЧИКИ

§ 1. МОДЕЛИРОВАНИЕ ДИСКРЕТНЫХ ВЕЛИЧИН

Simulation (*англ.*) — имитация, моделирование.

Начнем с рассмотрения *общего метода* моделирования реализации $x_1,\ldots,x_n$ выборки $(\xi_1,\ldots,\xi_n)$ из произвольного дискретного распределения. Положим $p_k = \mathbf{P}(\xi_1 = c_k)$, $k = 1,2,\ldots$. Разобьем отрезок $[0,1]$ на части длины p_k и обозначим через r_m сумму $\sum\limits_{k=1}^{m} p_k$ $(r_0 = 0)$. Если псевдослучайное число y_i (см. гл. 2) попадает в промежуток $(r_{j-1}, r_j]$, то полагаем $x_i = c_j$ (рис. 1).*)

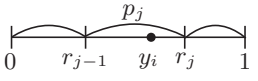


Рис. 1

Этот метод на самом деле является методом обратной функции (см. § 1 гл. 4), если определить $F^{-1}(y) = \inf\{x\colon F(x) \geqslant y\}$ (рис. 2).

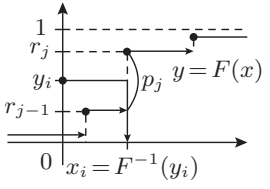


Рис. 2

Для некоторых дискретных случайных величин можно предложить специальные (более простые или более быстрые) датчики. Так, для моделирования случайной величины ν, имеющей геометрическое распределение: $p_k = \mathbf{P}(\nu = k) = (1-p)^k p$, $k = 0,1,\ldots$, достаточно подсчитать число «неудач» до первого «успеха» в схеме Бернулли. Другой пример дает биномиальный закон.

Определение. Случайная величина Z_n имеет *биномиальное распределение* с параметрами n и p $(n = 1,2,\ldots, 0 \leqslant p \leqslant 1)$, если $b(k,n,p) = \mathbf{P}(Z_n = k) = C_n^k p^k (1-p)^{n-k}$, $k = 0,1,\ldots,n$ (рис. 3). (Здесь $C_n^k = n!/[k!(n-k)!]$ — *число сочетаний из n по k*).

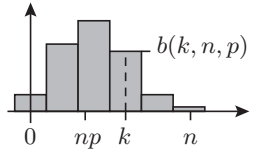


Рис. 3

Способ моделирования случайных величин с таким распределением основан на следующем утверждении.

Утверждение 1. Пусть $\zeta_1,\zeta_2,\ldots$ — схема Бернулли с вероятностью «успеха» p. Тогда число «успехов» в n испытаниях $Z_n = \zeta_1 + \ldots + \zeta_n$ имеет биномиальное распределение.

Доказательство. Докажем его по индукции, используя непосредственно проверяемое тождество $C_n^{k-1} + C_n^k = C_{n+1}^k$. В силу леммы о независимости из § 3 гл. 1, случайные величины

*) В частности, для бернуллиевской случайной величины получаем способ моделирования, предложенный в ответе на вопрос 2 гл. 2.

$Z_n = \zeta_1 + \ldots + \zeta_n$ и ζ_{n+1} независимы. Поэтому

$$b(k,n+1,p) =$$
$$= \mathbf{P}(Z_n = k)\,\mathbf{P}(\zeta_{n+1} = 0) + \mathbf{P}(Z_n = k-1)\,\mathbf{P}(\zeta_{n+1} = 1) =$$
$$= \sum_{i=0}^{1} C_n^{k-i} p^{k-i} (1-p)^{n-k+i} p^i (1-p)^{1-i} =$$
$$= \left(C_n^k + C_n^{k-1}\right) p^k (1-p)^{n+1-k}. \quad \blacksquare$$

На основе утверждения 1 с помощью свойств математического ожидания и дисперсии из П2 получаем, что $\mathbf{M}Z_n = np$ и $\mathbf{D}Z_n = np(1-p)$.

Вопрос 1.
К какому предельному закону сходится при $n \to \infty$ $(Z_n - np)/\sqrt{np(1-p)}$?

(Здесь вероятность $p \in (0,1)$ и не зависит от n.)

Определение. Случайная величина N имеет *распределение Пуассона* с параметром $\lambda > 0$, если для $k = 0,1,\ldots$

$$p(k,\lambda) = \mathbf{P}(N = k) = \frac{\lambda^k}{k!}\, e^{-\lambda}.$$

Рисунок 4 дает представление о поведении $p(k,\lambda)$ при $\lambda > 1$. Легко вычислить математическое ожидание закона Пуассона:

$$\mathbf{M}N = \sum_{k=1}^{\infty} k\, p(k,\lambda) = \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}\, e^{-\lambda} = \lambda.$$

С. Пуассон (1781–1840), французский физик и математик.

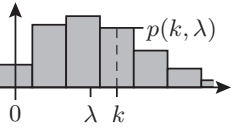


Рис. 4

Замечание 1. Пуассоновское распределение получается предельным переходом из биномиального при $n \to \infty$, $p \to 0$, $np \to \lambda$:

$$b(k,n,p) = \frac{n(n-1)\cdot\ldots\cdot(n-k+1)}{k!}\, p^k (1-p)^{n-k} =$$
$$= \frac{(np)^k}{k!}\, (1-p)^n \left[\left(1 - \frac{1}{n}\right)\cdot\ldots\cdot\left(1 - \frac{k-1}{n}\right)(1-p)^{-k}\right],$$

где $(1-p)^n \to e^{-\lambda}$, а выражение в квадратных скобках стремится к 1. Пуассоновское приближение биномиального распределения при больших n и малых p иногда называют *законом редких событий*.

Вопрос 2.
Чему равна $\mathbf{D}N$?
а) Угадайте.
б) Вычислите.

Для моделирования пуассоновской выборки понадобится

Утверждение 2. Пусть $\tau_1, \tau_2, \ldots$ — независимые показательно распределенные с параметром λ случайные величины. Положим $S_n = \tau_1 + \ldots + \tau_n$, N — число значений S_n, $n = 1,2,\ldots$, на отрезке $[0,1]$. Тогда случайная величина N имеет распределение Пуассона с параметром λ.

Доказательство. Согласно лемме 1 гл. 4 случайная величина $S_n \sim \Gamma(n,\lambda)$, т. е. плотность $p_{S_n}(x) = \lambda^n x^{n-1} e^{-\lambda x}/(n-1)!$ при $x > 0$. Дифференцированием нетрудно проверить, что соответствующей этой плотности функцией распределения является

$$F_{S_n}(x) = 1 - e^{-\lambda x} \sum_{i=0}^{n-1} (\lambda x)^i / i! \quad \text{при } x > 0. \qquad (1)$$

Положим $S_0 = 0$. Тогда с учетом положительности τ_{n+1} находим

$$\mathbf{P}(N=n)=\mathbf{P}(S_n\leqslant 1, S_{n+1}>1)=\mathbf{P}(S_n\leqslant 1)-\mathbf{P}(S_{n+1}\leqslant 1)=\frac{\lambda^n}{n!}\ e^{-\lambda},$$

что и требовалось установить.

Докажем утверждение 2 другим способом. Для $n \geqslant 1$ запишем равенство

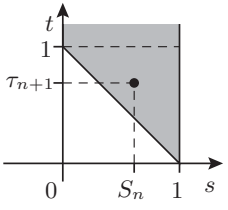


Рис. 5

$$\mathbf{P}(N = n) = \mathbf{P}(0 \leqslant S_n \leqslant 1, \tau_{n+1} > 1 - S_n).$$

Поскольку S_n и τ_{n+1} независимы, их совместная плотность имеет вид $p_{(S_n,\tau_{n+1})}(s,t) = p_{S_n}(s)\, p_{\tau_{n+1}}(t)$. Проинтегрируем ее по множеству $\{0 \leqslant s \leqslant 1, t > 1 - s\}$ (рис. 5):

$$\mathbf{P}(N=n)=\int\limits_0^1\int\limits_{1-s}^{\infty} p_{S_n}(s)\, p_{\tau_{n+1}}(t)\, dt\, ds=\int\limits_0^1 p_{S_n}(s)\, \mathbf{P}(\tau_{n+1}>1-s)\, ds=$$

$$=\int\limits_0^1 \frac{\lambda^n}{(n-1)!}\ s^{n-1}e^{-\lambda s}\cdot e^{-\lambda(1-s)}\, ds=\frac{\lambda^n}{(n-1)!}\ e^{-\lambda}\int\limits_0^1 s^{n-1}ds=\frac{\lambda^n}{n!}\ e^{-\lambda}.$$

Так как $\sum\limits_{n=0}^{\infty} \mathbf{P}(N = n) = 1$, то

$$\mathbf{P}(N = 0) = 1 - \sum_{n=1}^{\infty} \mathbf{P}(N = n) = e^{-\lambda}. \qquad \blacksquare$$

На основе утверждения 2 и метода обратной функции (см. § 1 гл. 4) получаем формулу для моделирования пуассоновских случайных величин:

$$n_i = \min\left\{k \geqslant 0\colon \prod_{j=0}^{k} y_{ij} < e^{-\lambda}\right\},$$

где y_{ij} — псевдослучайные числа ($i = 1,2,\ldots$; $j = 0,1,\ldots$).

§ 2. ПОРЯДКОВЫЕ СТАТИСТИКИ И СМЕСИ

Для выборки $(\eta_1, \ldots, \eta_n)$ из равномерного распределения на $[0, 1]$ найдем распределение k-й порядковой статистики $\eta_{(k)}$ (см. § 4 гл. 4).

Утверждение 3. $F_{\eta_{(k)}}(x) = \sum\limits_{i=k}^{n} C_n^i\, x^i(1-x)^{n-i}$, $0 \leqslant x \leqslant 1$.

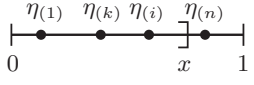


Рис. 6

Доказательство. «Успехом» будем называть попадание точки η_i левее x (рис. 6). При этом случайные величины $\zeta_i = I_{\{\eta_i \leqslant x\}}$ образуют схему Бернулли с $p = x$, а число «успехов» имеет биномиальное распределение. Другими словами, вероятность того, что *в точности* i из n точек попадут левее (и, следовательно,

$n-i$ точек правее) x, равна $C_n^i x^i(1-x)^{n-i}$. Но $\mathsf{P}(\eta_{(k)} \leqslant x)$ — это вероятность того, что *по крайней мере* k точек окажутся слева от x $(i = k, k+1, \ldots, n)$. ■

Вопрос 3. Какая функция распределения у случайной величины $\xi_{(k)}$, если компоненты выборки ξ_i имеют непрерывную функцию распределения $F(x)$?

Дифференцированием функции $F_{\eta_{(k)}}(x)$ вычисляем плотность:

$$p_{\eta_{(k)}}(x) = \sum_{i=k}^{n} C_n^i\, i\, x^{i-1}(1-x)^{n-i} - \sum_{i=k}^{n} C_n^i\,(n-i)\, x^i(1-x)^{n-i-1}.$$

Так как последнее слагаемое во второй сумме равно нулю, запишем

$$p_{\eta_{(k)}}(x) = \sum_{i=k}^{n} n\, C_{n-1}^{i-1}\, x^{i-1}(1-x)^{n-i} - \sum_{i=k}^{n-1} n\, C_{n-1}^{i}\, x^{i}(1-x)^{n-1-i}.$$

Все слагаемые в первой сумме, кроме первого, сокращаются:

$$p_{\eta_{(k)}}(x) = n\, C_{n-1}^{k-1} x^{k-1}(1-x)^{n-k}. \tag{2}$$

Поскольку $(n-1)! = \Gamma(n)$, из (2) и формулы (10) гл. 3 выводим при $0 < x < 1$, что

$$p_{\eta_{(k)}}(x) = \frac{1}{B(k, n-k+1)}\, x^{k-1}(1-x)^{n-k}.$$

Определение. Случайная величина U имеет *бета-распределение* с параметрами $r > 0$ и $s > 0$, если ее плотность задается формулой $p_U(x) = \dfrac{1}{B(r,s)}\, x^{r-1}(1-x)^{s-1} I_{\{0<x<1\}}$.

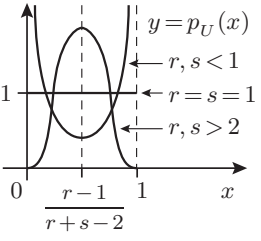


Рис. 7

Графики на рис. 7 дают представление о плотности $p_U(x)$ при разных значениях r и s. В частности, при $r = s = 1$ бета-распределение сводится к равномерному распределению на $[0,1]$.

Другим частным случаем (при $r = s = 1/2$) является *арксинус-распределение* с функцией распределения $F(x) = \dfrac{2}{\pi} \arcsin\sqrt{x}$ на $[0,1]$. Оно возникает в качестве предельного закона для времени, в течение которого находился в выигрыше первый из двух равных по силам игроков: $\delta_n = \sum\limits_{k=1}^{n} I_{\{S_{k-1}\geqslant 0, S_k \geqslant 0\}}$, где $S_k = X_1 + \ldots + X_k$, случайные величины X_i независимы, $\mathsf{P}(X_i = -1) = \mathsf{P}(X_i = 1) = 1/2$, $i = 1,2,\ldots$. Тогда $\mathsf{P}(\delta_n/n \leqslant x) \to F(x)$ при $n \to \infty$ (см. § 4 гл. 16).

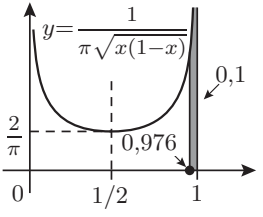


Рис. 8

Так как $F(0{,}976) \approx 0{,}9$ (рис. 8), то в среднем *в каждом пятом случае* один из игроков будет лидировать на протяжении не менее 97,6% времени игры.

Легко подсчитать *момент k-го порядка* бета-распределения:

$$\mathsf{M}U^k = \frac{1}{B(r,s)} \int_0^1 x^{k+r-1}(1-x)^{s-1}dx = \frac{B(r+k,s)}{B(r,s)}. \tag{3}$$

Для того чтобы познакомиться с еще одним методом моделирования (так называемым *методом суперпозиции*) потребуется ввести понятие смеси распределений.

Определение. Пусть $p_k \geqslant 0$, $\sum\limits_k p_k = 1$, $F_k(x)$ — некоторые функции распределения. Тогда функция распределения $F(x) = \sum p_k F_k(x)$ называется *смесью распределений* $F_k(x)$ с *весами* p_k.

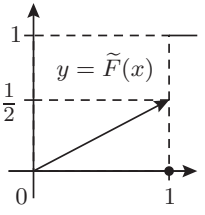


Рис. 9

Пример 1. Функция $\widetilde{F}(x)$, приведенная на рис. 9, представляет собой смесь с весами $1/2$ функций распределения $F_\xi(x) = I_{\{x \geqslant 1\}}$ (т. е. $\xi = 1$ с вероятностью 1) и $F_\eta(x)$ случайной величины η, равномерно распределенной на отрезке $[0,1]$.

Для моделирования случайной величины, функция распределения которой является смесью, используется основанный на формуле полной вероятности (П7)

Вопрос 4.
Можно ли $\widetilde{F}(x)$ представить в виде смеси с положительными весами других функций распределения?

Метод суперпозиции

1) Разыгрывается значение дискретной случайной величины, принимающей значения $k = 1,2,\ldots$ с вероятностями p_k (см. § 1). Обозначим полученное в результате розыгрыша значение через k_0.

2) Моделируется случайная величина с функцией распределения $F_{k_0}(x)$ некоторым способом (скажем, методом обратной функции из § 1 гл. 4).

Пример 2. Пусть ξ распределена на $[0,1]$ с плотностью $p_\xi(x)$, представимой в виде *степенного ряда* $\sum\limits_{k=0}^{\infty} a_k x^k$ с $a_k \geqslant 0$ ([29, с. 20]). Положим $p_k = a_k/(k+1)$. Тогда $p_\xi(x) = \sum\limits_{k=0}^{\infty} p_k\,(k+1)\,x^k$. Но случайная величина $\eta_{(k+1)} = \max\{\eta_1,\ldots,\eta_{k+1}\}$, где η_i независимы и равномерно распределены на $[0,1]$, обладает плотностью $p_{\eta_{(k+1)}}(x) = (k+1)\,x^k$ (см. задачу 3 гл. 1), т. е. $\eta_{(k+1)}$ имеет бета-распределение с параметрами $r = k+1$ и $s = 1$, и для моделирования ξ можно применить метод суперпозиции.

Вопрос 5.
Почему в этом примере $\sum\limits_{k=0}^{\infty} p_k = 1$?

К сожалению, рассмотренный подход нельзя использовать, если среди коэффициентов a_k есть отрицательные. Однако для *непрерывных* плотностей можно предложить способ приближенного моделирования, основанный на аппроксимации их *полиномами Бернштейна* $f_n(x) = \sum\limits_{k=0}^{n} f\left(\frac{k}{n}\right) C_n^k\, x^k (1-x)^{n-k}$.

С. Н. Бернштейн (1880–1966), российский математик.

К. Вейерштрасс (1815–1897), немецкий математик.

Теорема Вейерштрасса. Если функция $f(x)$ непрерывна на отрезке $[0,1]$, то $f_n(x) \to f(x)$ равномерно по x при $n \to \infty$.

Доказательство. Известно, что всякая непрерывная на отрезке $[0,1]$ функция *ограничена*: $|f(x)| \leqslant M < \infty$ (см. [45, с. 193]). Кроме

того, она является *равномерно непрерывной*: для всякого $\varepsilon > 0$ найдется $\delta > 0$ такое, что $|f(x) - f(y)| \leqslant \varepsilon$, коль скоро $|x - y| \leqslant \delta$ (см. [45, с. 446]).

Пусть Z_n — количество «успехов» в n испытаниях Бернулли с вероятностью «успеха» x. Обозначим для краткости через p_k биномиальную вероятность $b(k,n,x) = \mathbf{P}(Z_n = k)$, т. е.

$$p_k = C_n^k\, x^k (1-x)^{n-k}, \quad k = 0,1,\ldots,n.$$

Отметим, что в силу теоремы о замене переменных (П2), полином Бернштейна $f_n(x)$ равен $\mathbf{M}\, f(Z_n/n)$.

Используя неравенство Чебышева (П4), получаем

$$\sum_{\left\{k:\, \left|\frac{k}{n}-x\right|>\delta\right\}} p_k = \mathbf{P}\left(\left|\frac{Z_n}{n} - x\right| > \delta\right) \leqslant \frac{\mathbf{D}Z_n}{n^2\delta^2} = \frac{x(1-x)}{n\delta^2} \leqslant \frac{1}{4n\delta^2}\,.$$

Отсюда выводим равномерную по $x \in [0,1]$ оценку погрешности приближения:

$$\begin{aligned}
|f(x) - f_n(x)| &= \left|\sum_{k=0}^{n} \left[f(x) - f\left(\frac{k}{n}\right)\right] p_k\right| \leqslant \\
&\leqslant \sum_{\left\{k:\, \left|\frac{k}{n}-x\right|\leqslant\delta\right\}} \left|f(x) - f\left(\frac{k}{n}\right)\right| p_k + \sum_{\left\{k:\, \left|\frac{k}{n}-x\right|>\delta\right\}} \left|f(x) - f\left(\frac{k}{n}\right)\right| p_k \leqslant \\
&\leqslant \varepsilon \sum_{\left\{k:\, \left|\frac{k}{n}-x\right|\leqslant\delta\right\}} p_k + 2M \sum_{\left\{k:\, \left|\frac{k}{n}-x\right|>\delta\right\}} p_k \leqslant \varepsilon + \frac{M}{2n\delta^2}\,.
\end{aligned}$$

Переход к пределу при $\varepsilon \to 0$ и $n \to \infty$ завершает доказательство. ■

Следствие. Непрерывную на отрезке $[0,1]$ *плотность* $f(x)$ можно равномерно приблизить *плотностями*, в качестве которых годятся *нормированные полиномы Бернштейна*

$$\widetilde{f}_n(x) = f_n(x)/d_n, \quad \text{где } d_n = \int_0^1 f_n(x)\,dx.$$

В самом деле, из равномерной сходимости $f_n(x)$ к $f(x)$ вытекает, что $d_n = \int\limits_0^1 f_n(x)\,dx \to \int\limits_0^1 f(x)\,dx = 1$, т. е. $\widetilde{f}_n(x) \to f(x)$ равномерно.

Для моделирования случайной величины $\widetilde{\xi}_n$ с плотностью $\widetilde{f}_n(x)$ методом суперпозиции представим нормированный полином $\widetilde{f}_n(x)$ в виде смеси:

$$\widetilde{f}_n(x) = \sum_{k=0}^{n} \frac{f\,(k/n)}{d_n(n+1)} \left[(n+1)\, C_n^k\, x^k (1-x)^{n-k}\right], \tag{4}$$

где в квадратных скобках стоит (см. формулу (2)) плотность $(k+1)$-й порядковой статистики для выборки размера $n+1$ из

равномерного распределения на $[0,1]$. Из формулы (4) интегрированием получаем, что

$$d_n = \frac{1}{n+1}\sum_{k=0}^{n} f\left(\frac{k}{n}\right) \quad \text{и} \quad p_k = f\left(\frac{k}{n}\right) \Big/ \sum_{k=0}^{n} f\left(\frac{k}{n}\right).$$

При больших n равномерные порядковые статистики разбивают отрезок $[0,1]$ на примерно равные по длине части, так что приближенное моделирование случайной величины ξ с плотностью $f(x)$ методом суперпозиции, основанное на формуле (4), по существу является заменой ξ на дискретную случайную величину, принимающую значения $\frac{k}{n}$ с вероятностями, пропорциональными $f\left(\frac{k}{n}\right)$.

§ 3. МЕТОД НЕЙМАНА (МЕТОД ИСКЛЮЧЕНИЯ)*)

Дж. фон Нейман (1903–1957), американский математик.

Предположим , что случайная величина ξ распределена на отрезке $[a,b]$, причем ее плотность ограничена: $\max\limits_{x\in[a,b]} p_\xi(x) = C < \infty$. Случайные величины $\eta_1,\eta_2,\ldots$ — независимы и равномерно распределены на $[0,1]$, $X_i = a + (b-a)\,\eta_{2i-1}$, $Y_i = C\,\eta_{2i}$ $(i = 1,2,\ldots)$. Таким образом, пары (X_i,Y_i) независимы и равномерно распределены в прямоугольнике $[a,b]\times[0,C]$ (рис. 10). Обозначим через *ν номер первой точки с координатами (X_i,Y_i), попавшей под график плотности* $p_\xi(x)$, т. е. $\nu = \min\{i\colon Y_i \leqslant p_\xi(X_i)\}$. Положим $X_\nu = \sum\limits_{n=1}^{\infty} X_n I_{\{\nu=n\}}$.

Утверждение 4. При выполнении приведенных выше условий случайная величина X_ν распределена так же, как ξ.

Доказательство. Пусть p — это *вероятность попадания точки (X_i,Y_i) под график плотности*, $q = 1-p$. Тогда вероятность $p = \mathbf{P}(Y_1 \leqslant p_\xi(X_1))$ есть отношение площади под графиком $y = p_\xi(x)$ к площади прямоугольника:

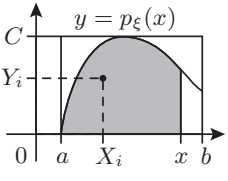


Рис. 10

$$p = \frac{\int\limits_a^b p_\xi(x)\,dx}{C(b-a)} = \frac{1}{C(b-a)}.$$

По формуле полной вероятности (см. П7) представим функцию распределения X_ν:

$$F_{X_\nu}(x) = \mathbf{P}(X_\nu \leqslant x) = \sum_{n=1}^{\infty} \mathbf{P}(\nu = n, X_n \leqslant x).$$

Принимая во внимание, что

$$\{\nu = n\} = \{Y_i > p_\xi(X_i), i = 1,\ldots,n-1, Y_n \leqslant p_\xi(X_n)\},$$

*) Материал этого параграфа не используется в дальнейшем.

и что события $\{Y_i > p_\xi(X_i), i = 1, \ldots, n-1\}$ и $\{Y_n \leqslant p_\xi(X_n), X_n \leqslant x\}$ независимы согласно лемме из § 3 гл. 1, получаем (см. рис. 10):

$$F_{X_\nu}(x) =$$
$$= \sum_{n=1}^{\infty} \mathbf{P}(Y_i > p_\xi(X_i), i = 1, \ldots, n-1, Y_n \leqslant p_\xi(X_n), X_n \leqslant x) =$$
$$= \sum_{n=1}^{\infty} q^{n-1}\, \mathbf{P}(Y_n \leqslant p_\xi(X_n), X_n \leqslant x) = \sum_{n=1}^{\infty} q^{n-1}\, \frac{\int_a^x p_\xi(u)\, du}{C(b-a)} =$$
$$= \sum_{n=1}^{\infty} q^{n-1}\, [p\, F_\xi(x)] = F_\xi(x),$$

что и требовалось доказать. ■

Вопрос 6. Сколько в среднем точек (X_i, Y_i) потребуется «вбросить» в прямоугольник $[a, b] \times [0, C]$ для получения одного значения ξ?

В случае, когда площадь прямоугольника $[a, b] \times [0, C]$ значительно превышает 1, время моделирования можно существенно уменьшить, применяя *модифицированный метод Неймана (расслоенную выборку)*. Он состоит в разбиении $[a, b]$ на отрезки Δ_k, на каждом из которых $p_\xi(x)$ не намного отличается от $C_k = \max\limits_{x \in \Delta_k} p_\xi(x)$ (рис. 11). Тогда $p_\xi(x)$ представляется в виде смеси плотностей $f_k(x)$:

$$p_\xi(x) = \sum_k p_k\, f_k(x), \quad \text{где } 0 < p_k = \int\limits_{\Delta_k} p_\xi(x\, dx,\ f_k(x) = \frac{p_\xi(x)}{p_k}\ I_{\Delta_k}.$$

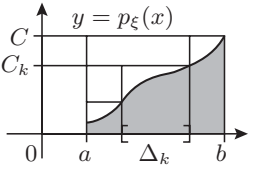


Рис. 11

Отсюда видим, что для моделирования случайной величины ξ надо:

1) разыграть номер отрезка разбиения Δ_k с вероятностями p_k; результат розыгрыша обозначим через k_0,

2) моделировать методом Неймана случайную величину с плотностью $f_{k_0}(x) \leqslant C_{k_0}/p_{k_0}$, бросая случайно точки в прямоугольник $\Delta_{k_0} \times [0, C_{k_0}/p_{k_0}]$ до первого их попадания под график плотности $f_{k_0}(x)$ или, что то же самое, в прямоугольник $\Delta_{k_0} \times [0, C_{k_0}]$ до первого попадания под график функции $p_\xi(x)$.

Из ответа на вопрос 6 и свойства 1 условного математического ожидания (П7) следует, что среднее число бросаний будет равно

$$\sum_k \left[\frac{C_k}{p_k}\ |\Delta_k|\right] p_k = \sum_k C_k\, |\Delta_k|,$$

т. е. сумме площадей прямоугольников $\Delta_k \times [0, C_k]$, которая при измельчении разбиения стремится к 1, так как $p_\xi(x)$ — плотность.

Пример 3. Плотность показательной с параметром $\lambda = 1$ случайной величины τ представляется в виде смеси ($q = e^{-1}, p = 1 - q$):

$$p_\tau(x) = e^{-x} I_{[0, +\infty)} = \sum_{k=0}^{\infty} e^{-k}\, e^{-(x-k)} I_{[k, k+1)} = \sum_{k=0}^{\infty} \left(q^k p\right)\ p_{\{\tau\}}(x-k)$$

(см. утверждение 3 гл. 4). Этот пример демонстрирует возможность моделирования модифицированным методом Неймана случайной величины, плотность которой имеет неограниченный носитель.*)

*) *Носитель* — множество, на котором плотность положительна.

Замечание 2. Рассмотренный в § 3 гл. 4 способ моделирования последовательности $\{\tau\}$ без вычисления логарифмов, является обобщением метода Неймана. Вместо бросания точек в прямоугольник $[a,b]\times[0,C]$ производится выбор случайных точек $\boldsymbol{\eta}_i = (\eta_{i1},\eta_{i2},\ldots)$ из бесконечномерного единичного куба $I^\infty = [0,1]\times[0,1]\times\ldots$. Момент первого попадания под график плотности $p_\xi(x)$, т. е. в множество

$$\{(x,y)\colon a\leqslant x\leqslant b, 0\leqslant y\leqslant p_\xi(x)\},$$

заменяется на момент попадания в множество

$$A=\{\boldsymbol{y}=(y_1,y_2,\ldots)\colon y_1>y_2>\ldots>y_{K-1}<y_K, K \text{ четно}\}\subset I^\infty,$$

имеющее бесконечномерный объем $p=1-e^{-1}$.

§ 4. ПРИМЕР ИЗ ТЕОРИИ ИГР

Добрый пример лучше ста слов.

Представьте, что вам предложили принять участие в следующей простой игре (см. [72, с. 54]). Одновременно со своим противником вы называете одну из двух цифр — «1» либо «2» (поднимаете один или два пальца). Если сумма названных цифр — четное число, то вы выигрываете, а если нечетное, то проигрываете эту сумму. Платежная матрица игры приведена на рис. 12. Стоит ли играть на таких условиях?

	☝	✌
☝	+2	−3
✌	−3	+4

Рис. 12

Давайте разберем, к каким результатам приводят разные стратегии в этой игре.

Прежде всего, заметим, что если бы вы сумели точно предсказать следующую цифру противника, то смогли бы выиграть, назвав такую же. Аналогичной возможностью играть в противофазе обладает и противник. Поэтому обоим игрокам надо применять случайные стратегии.

Однако, совершенно случайно (например, подбрасывая монету) называть «1» или «2» игрокам также нет резона, так как в этом случае их средний выигрыш равен $+2-3-3+4=0$ (говорят, что игра имеет «нулевую сумму»).

Вопрос 7.
Допустим, что противник все же решил называть «1» и «2» равновероятно, невзирая на ваше поведение. Как вам следует играть в таком случае?

Давайте сыграем! Ниже приведена последовательность из «1» и «2», названных противником без учета вашего поведения. Закройте эти цифры листком бумаги и, постепенно сдвигая его вправо, попытайтесь угадывать следующую цифру. Запишите свои выигрыши и проигрыши в соответствии в платежной матрицей, изображенной на рис. 12, и подсчитайте итог.

2 1 2 2 1 2 1 1 1 2 1 1 2 1 1 2 2 1 1 2

Удалось победить? На самом деле эта игра выгодна для противника: используя приведенную ниже стратегию, он сможет при достаточно большом числе партий вас разорить!

Действительно, если вы называете «1» с вероятностью p_1 и «2» с вероятностью $1-p_1$, а противник — с вероятностями p_2

и $1-p_2$ соответственно, то *средний выигрыш за партию* $Z(p_1,p_2) = +2p_1p_2 - 3p_1(1-p_2) - 3(1-p_1)p_2 + 4(1-p_1)(1-p_2) = 12p_1p_2 - 7p_1 - 7p_2 + 4$.

Поверхность $z = Z(p_1,p_2)$ является гиперболическим параболоидом («седлом»), в сечении которого плоскостью $p_2 = 7/12$ получается прямая, параллельная оси абсцисс: $Z(p_1,7/12) = -1/12$ (рис. 13). Таким образом, если противник называет «1» с вероятностью $7/12 \approx 0{,}583$ и «2» с вероятностью $5/12 \approx 0{,}417$, то независимо от значения p_1 вы будете в среднем проигрывать 1 за каждые 12 партий.

Приведенная выше последовательность из «1» и «2» была получена с помощью последнего столбца таблицы случайных чисел Т1 в соответствии с оптимальной стратегией противника: если очередное число в столбце меньше или равно 58, то записывалась «1», иначе — «2».

Однако, чтобы в серии независимых игр выиграть определенную небольшую сумму S (скажем, $S = 20$) с вероятностью близкой к 1, противнику потребуется вовсе не $12S$ $(= 240)$ партий, а намного больше (см. задачу 5).

Вопрос 8.
Допустим, что вам известно значение p_2 и то, что противник не станет его менять, как бы вы не играли. Какой стратегии следует придерживаться в такой ситуации?

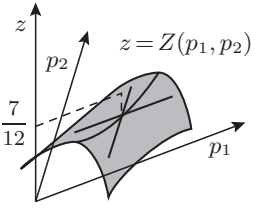


Рис. 13

Медленно, но верно.

ЗАДАЧИ

Счастье в том, чтобы без помех упражнять свои способности, каковы бы они ни были.
Аристотель, «Политика», IV, 11

1. Для равномерной порядковой статистики $\eta_{(k)}$ сравните точку максимума плотности (так называемую *моду распределения*) и $\mathbf{M}\eta_{(k)}$.

2. Чему равно математическое ожидание случайной величины $\widetilde{\xi}$ с функцией распределения $\widetilde{F}(x)$ из примера 1?

3. Пусть случайные величины $X \sim \Gamma(r,\lambda)$ и $Y \sim \Gamma(s,\lambda)$ независимы. Проверьте, что тогда $X/(X+Y)$ имеет бета-распределение с параметрами r и s.
Указание. Примените формулу преобразования плотности случайного вектора из П8.

4. Выборка $\eta_1,\ldots,\eta_n$ взята из равномерного распределения на $[0,1]$. Используйте леммы 1 и 3 гл. 4 для доказательства того, что при $n \to \infty$ имеет место сходимость $n\eta_{(k)} \xrightarrow{d} T \sim \Gamma(k,1)$, где $\eta_{(k)}$ — k-я порядковая статистика.
Указание. Используйте свойства сходимости (П5) и закон больших чисел (П6).

5. Допустим, что в примере из теории игр противник играет по оптимальной стратегии ($p_2 = 7/12$), а вы равновероятно называете «1» и «2» ($p_1 = 1/2$). Сколько примерно партий потребуется противнику для выигрыша $S = 20$ с вероятностью 0,975?
Указание. Для оценки вероятности используйте центральную предельную теорему (П6) и таблицу Т2.

РЕШЕНИЯ ЗАДАЧ

1. Дифференцируя правую часть формулы (2), нетрудно установить, что плотность случайной величины $\eta_{(k)}$ имеет максимум в точке $m_k = (k-1)/(n-1)$. По формуле (3) находим $\mathbf{M}\eta_{(k)} = k/(n+1)$ (выбранные наудачу n точек разбивают отрезок $[0,1]$ на $n+1$ интервалов в среднем одинаковой длины). Плотность случайной величины $\eta_{(k)}$ симметрична плотности величины $\eta_{(n+1-k)}$ относительно прямой $x = 1/2$. При $k < (n+1)/2$ мода $m_k < \mathbf{M}\eta_{(k)}$, распределение «скошено» влево (правый «хвост» тяжелее). При $k > (n+1)/2$ — наоборот.

2. $\mathbf{M}\widetilde{\xi} = \int x\,\widetilde{F}(dx) = \frac{1}{2}\int x\,F_\xi(dx) + \frac{1}{2}\int x\,F_\eta(dx) = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}$.

3. Используем тот же прием, что и при решении задачи 6 гл. 4:
а) дополним случайную величину $X/(X+Y)$ до двумерного случайного вектора;
б) найдем его плотность по формуле преобразования (см. П8);
в) проинтегрируем плотность по последней координате.

Пусть $u = x/(x+y), v = x+y$. Обратное преобразование задается формулами $x = uv$, $y = v - uv$, $0 < u < 1$, $v > 0$. Отсюда $\partial x/\partial u = v$, $\partial x/\partial v = u$, $\partial y/\partial u = -v$, $\partial y/\partial v = 1-u$ и $|J| = v$.

$$p_{(U,V)}(u,v) = \frac{v}{\Gamma(r)\Gamma(s)}\,(uv)^{r-1}(v-uv)^{s-1}e^{-v},\ 0<u<1, v>0.$$

$$p_U(u) = \frac{u^{r-1}(1-u)^{s-1}}{\Gamma(r)\Gamma(s)}\int\limits_0^\infty v^{r+s-1}e^{-v}dv = \frac{u^{r-1}(1-u)^{s-1}}{B(r,s)}\,.$$

4. Пусть $(\tau_1,\ldots,\tau_{n+1})$ — показательная выборка с параметром $\lambda = 1$, $S_k = \tau_1 + \ldots + \tau_k$. По лемме 1 гл. 4 $S_k \sim \Gamma(k,1)$. В силу леммы 3 гл. 4 $\eta_{(k)} \sim S_k/S_{n+1}$. Согласно закону больших чисел (П6) $S_{n+1}/(n+1) \xrightarrow{\mathsf{P}} \mathbf{M}\tau_1 = 1$. Из непрерывности функции $\varphi(x) = 1/x$ при $x > 0$ и представления

$$n\,\frac{S_k}{S_{n+1}} = \frac{n}{n+1}\left(\frac{1}{S_{n+1}/(n+1)}\right)S_k,$$

используя свойства сходимости (П5), получаем $\Gamma(k,1)$ в качестве предельного распределения для случайной величины $n\eta_{(k)}$.

5. Приведем решение, следуя [79, с. 107]. При $p_1 = 1/2$ и $p_2 = 7/12$ размер выигрыша в одной партии X_i имеет не зависящее от i распределение: он принимает значения $-3, 2, 4$ с вероятностями $1/2, 7/24, 5/24$ соответственно. При этом $\mu = \mathbf{M}X_1 = -1/12$, $\sigma^2 = \mathbf{D}X_1 = \mathbf{M}X_1^2 - (\mathbf{M}X_1)^2 \approx \mathbf{M}X_1^2 = (9\cdot12 + 4\cdot7 + 16\cdot5)/24 = 9$. Поскольку $|\mu| = 1/12$ значительно меньше, чем $\sigma \approx 3$, в одной партии преобладает случайный разброс, а не снос. Однако, при продолжительной игре систематический отрицательный снос

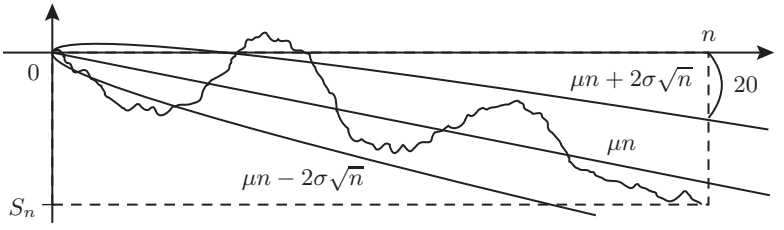


Рис. 14

$\mu = -1/12$ в соответствии с усиленным законом больших чисел неминуемо (почти наверное) приводит к уходу траектории блуждания на $-\infty$ (рис. 14).

На большом пути и малая ноша тяжела.

Насколько быстро происходит этот уход?

Обозначим через $S_n = X_1 + \ldots + X_n$ размер выигрыша за n партий. Тогда в силу центральной предельной теоремы (П6) при достаточно больших n

$$\mathsf{P}(S_n \leqslant -20) = \mathsf{P}\left(\frac{S_n - \mu n}{\sigma\sqrt{n}} \leqslant \frac{-20 - \mu n}{\sigma\sqrt{n}}\right) \approx \Phi\left(\frac{-20 + n/12}{3\sqrt{n}}\right),$$

где $\Phi(x)$ — функция распределения закона $\mathcal{N}(0,1)$. Из таблицы Т2 находим, что $0{,}975 \approx \Phi(1{,}96) \approx \Phi(2)$. Отсюда получаем квадратное уравнение для определения требуемого числа партий: $n/12 - 6\sqrt{n} - 20 = 0$. Оно имеет положительный корень $n_0 \approx 5600$. Таким образом, пройдет немало времени, пока проигрышная тенденция перевесит случайные колебания S_n.

Игра не стоит свеч.

ОТВЕТЫ НА ВОПРОСЫ

Ответа не хочу, я знаю ваш ответ.

Софья в «Горе от ума» А. С. Грибоедова

1. Для центрированных и нормированных сумм бернуллиевских случайных величин в силу центральной предельной теоремы (П6) предельным законом является $\mathcal{N}(0,1)$.

2. Так как при $n \to \infty$, $p \to 0$, $np \to \lambda$ биномиальное распределение стремится к пуассоновскому, то можно предположить, что и дисперсии сходятся: $\lim\limits_{n\to\infty} np(1-p) = \lambda$. Однако, в общем случае из сходимости по распределению (П5) не следует сходимость дисперсий, поэтому вычислим $\mathsf{D}N$ непосредственно:

$$\mathsf{M}[N(N-1)] = \sum_{k=2}^{\infty} k(k-1)\,p(k,\lambda) = \lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!}\, e^{-\lambda} = \lambda^2,$$

$$\mathsf{D}N = \mathsf{M}[N(N-1)] + \mathsf{M}N - (\mathsf{M}N)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

3. Учитывая непрерывность $F(x)$, воспользуемся методом обратной функции (см. § 1 гл. 4):

$$F_{\xi_{(k)}}(x) = \mathsf{P}\left(\xi_{(k)} \leqslant x\right) = \mathsf{P}\left(F(\xi_{(k)}) \leqslant F(x)\right) =$$

$$= \mathsf{P}\left(\eta_{(k)} \leqslant F(x)\right) = \sum_{i=k}^{n} C_n^i F(x)^i (1 - F(x))^{n-i}.$$

4. Можно. Например, $\widetilde{F}(x) = \frac{1}{2}\,\widetilde{F}(x) + \frac{1}{2}\,\widetilde{F}(x)$. С другой стороны, разложение на дискретную и непрерывную составляющие единственно (см. [12, с. 53]).

5. Степенной ряд $p_\xi(x) = \sum\limits_{k=0}^{\infty} a_k x^k$ сходится на $[0,1]$. Поэтому можно почленно интегрировать на $(0,1)$. Будучи функцией распределения,

$$F_\xi(x) = \sum_{k=0}^{\infty} \frac{a_k}{k+1}\, x^{k+1} \to 1 \quad \text{при } x \to 1.$$

Доказательство законности суммирования (*тауберовой теоремы*) приведено, например, в [33, с. 57].

Так как коэффициенты этого ряда неотрицательны, то его можно суммировать при $x = 1$.

6. Согласно задаче 4 гл. 1 в среднем потребуется $\mathbf{M}\nu + 1 = q/p + 1 = 1/p = C\,(b-a)$ точек.

7. Всегда называя «2», вы будете выигрывать больше $(+4)$, чем проигрывать (-3).

8. Надо максимизировать $Z(p_1,p_2) = p_1(12p_2 - 7) - 7p_2 + 4$ по $p_1 \in [0,1]$ при фиксированном значении p_2. Очевидно, что для $p_2 < 7/12$ максимум достигается при $p_1 = 0$, а для $p_2 > 7/12$ — при $p_1 = 1$.

Часть II
ОЦЕНИВАНИЕ ПАРАМЕТРОВ

Эта часть книги рассчитана в основном на студентов, изучающих математическую статистику. Поэтому к большинству утверждений и теорем приведены доказательства, что может оказаться полезным при подготовке к экзамену. Рассмотрено много примеров, на которых проясняется смысл важнейших понятий статистики. Тем читателям, кто интересуется в первую очередь методами обработки прикладных данных, можно сразу после просмотра §§ 1–2 гл. 6 перейти к части III.

Глава 6

СРАВНЕНИЕ ОЦЕНОК

Если у тебя спрошено будет: что полезнее, солнце или месяц? — ответствуй: месяц. Ибо солнце светит днем, когда и без того светло, а месяц — ночью.

Но, с другой стороны: солнце лучше тем, что светит и греет, а месяц только светит, и то лишь в лунную ночь!

Козьма Прутков

Анализируемые методами математической статистики данные обычно рассматриваются как реализация выборки из некоторого распределения, известного с точностью до параметра (или нескольких параметров). При таком подходе для определения распределения, наиболее подходящего для описания данных, достаточно уметь оценивать значение параметра по реализации. В этой главе будет рассказано, как сравнивать различные оценки *по точности*.

§ 1. СТАТИСТИЧЕСКАЯ МОДЕЛЬ

Эксперимент. Пусть θ — некоторое *неизвестное* положительное число. Ниже приведены (с точностью до 0,1) координаты x_i десяти точек, взятых наудачу из отрезка $[0, \theta]$.

3,5 3,2 25,6 8,8 11,6 26,6 18,2 0,4 12,3 30,1

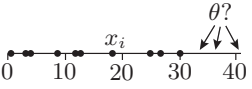


Рис. 1

Они были получены по формуле $x_i = \theta y_i$, $i = 1, \ldots, 10$, где y_i — псевдослучайные числа (см. гл. 2). Попробуйте угадать значение параметра θ с помощью рис. 1, на котором изображены эти точки.

Вопрос 1.
Может ли это значение равняться
а) 28, б) 100?

С формальной точки зрения в данном эксперименте мы имеем дело со следующей моделью: набор x_i — это реализация независимых и *равномерно распределенных* на отрезке $[0, \theta]$ случайных величин X_i с функцией распределения

$$F_\theta(x) = \begin{cases} 0, & \text{если } x \leqslant 0, \\ x/\theta, & \text{если } 0 < x < \theta, \\ 1, & \text{если } x \geqslant 1 \end{cases}$$

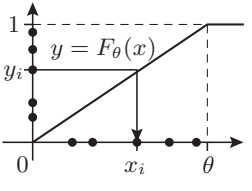


Рис. 2

(рис. 2). Здесь $\theta \in \Theta = (0, +\infty)$ — неизвестный параметр масштаба.

Статистическая модель. В общем случае задается семейство функций распределения $\{F_\theta(x), \theta \in \Theta\}$, где Θ — множество возможных значений параметра; данные $x_1, \ldots, x_n$ рассматриваются как реализация выборки $X_1, \ldots, X_n$, элементы которой имеют функцию распределения $F_{\theta_0}(x)$ при некотором неизвестном значе-

нии $\theta_0 \in \Theta$. Задача состоит в том, чтобы оценить (восстановить) θ_0 по выборке $x_1, \ldots, x_n$, по возможности, *наиболее точно*.

Те, кто знаком с методом обратной функции из § 1 гл. 4, могут представлять себе задачу так: кто-то задумал θ_0, а затем получил реализацию по формуле $x_i = F_{\theta_0}^{-1}(y_i)$, где y_i — псевдослучайные числа (рис. 3). Как «угадать» задуманное значение, основываясь на наблюдениях $x_1, \ldots, x_n$?

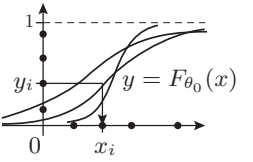


Рис. 3

Будем оценивать θ_0 при помощи некоторых функций $\widehat{\theta}$ от n переменных $x_1, \ldots, x_n$.*)

Для приведенных выше данных эксперимента в качестве оценок неизвестного параметра масштаба можно использовать, скажем, $\widehat{\theta}_1 = x_{(n)} = \max\{x_1, \ldots, x_n\}$ и $\widehat{\theta}_2 = 2\,(x_1 + \ldots + x_n)/n$. Интуитивно понятно, что при увеличении n каждая из оценок будет приближаться именно к тому значению θ, с которым моделировалась выборка. Но какая из них точнее? Каким образом вообще можно сравнивать оценки? Прежде чем дать ответы на эти вопросы, познакомимся с важнейшими свойствами оценок — несмещенностью и состоятельностью.

§ 2. НЕСМЕЩЕННОСТЬ И СОСТОЯТЕЛЬНОСТЬ

Определение. Оценка $\widehat{\theta}(x_1, \ldots, x_n)$ параметра θ называется *несмещенной*, если $\mathbf{M}_\theta\, \widehat{\theta}(X_1, \ldots, X_n) = \theta$ для всех $\theta \in \Theta$.

Здесь индекс θ у $\mathbf{M}_\theta$ означает, что имеется в виду математическое ожидание случайной величины $\widehat{\theta}(X_1, \ldots, X_n)$, где X_i распределены с функцией распределения $F_\theta(x)$. В дальнейшем этот индекс будет опускаться, чтобы формулы не выглядели слишком громоздко.

Замечание. Важно, чтобы условие несмещенности выполнялось для всех $\theta \in \Theta$. *Тривиальный контрпример*: оценка $\widehat{\theta}(x_1, \ldots, x_n) \equiv \equiv 1$, идеальная при $\theta = 1$, при других значениях θ имеет *смещение* $b(\theta) = \mathbf{M}\widehat{\theta} - \theta = 1 - \theta$.

b: bias (*англ.*) — смещение.

Иногда представляет интерес получение оценки не для самого параметра θ, а для некоторой заданной функции $\varphi(\theta)$.

Пример 1. Для выборочного контроля из партии готовой продукции отобраны n приборов. Пусть $X_1, \ldots, X_n$ — их времена работы до поломки. Допустим, что X_i одинаково показательно распределены с неизвестным параметром θ: $F_\theta(x) = 1 - e^{-\theta x}$, $x > 0$. Требуется оценить *среднее время до поломки прибора*

$$\varphi(\theta) = \mathbf{M}X_1 = \theta \int\limits_0^\infty x\, e^{-\theta x} dx = \frac{1}{\theta} \int\limits_0^\infty y\, e^{-y} dy = \frac{1}{\theta}\ .$$

Вопрос 2. Будут ли несмещенными определенные выше оценки $\widehat{\theta}_1$ и $\widehat{\theta}_2$? (Посмотрите решения задач 2 и 3 из гл. 1.)

По свойствам математического ожидания (П2) *выборочное среднее* $\overline{X}$ будет несмещенной оценкой для функции $\varphi(\theta)$: $\mathbf{M}\overline{X} = \varphi(\theta)$.

*) Предполагается, что функции являются борелевскими (см. П2). В частности, годятся любые непрерывные функции $\widehat{\theta}(x_1, \ldots, x_n)$.

Замечание. Если в примере 1 попытаться оценить сам параметр θ при помощи $\widehat{\theta} = 1/\overline{X}$, то получим смещенную оценку. Это следует из строгой выпуклости функции $\varphi(x) = 1/x$ при $x > 0$ и неравенства Иенсена (см. П4). Несмещенная оценка для θ приведена в задаче 6.

Следующий пример показывает, что не всякую функцию φ в заданной статистической модели можно несмещенно оценить.

Пример 2. Пусть элементы выборки X_i имеют распределение Бернулли с неизвестной вероятностью «успеха» $\theta \in \Theta = (0, 1)$:

$$\mathbf{P}(X_i = 1) = \theta, \ \mathbf{P}(X_i = 0) = 1 - \theta.$$

В этой модели при $n = 1$ нельзя несмещенно оценить $\varphi(\theta) = 1/\theta$. Действительно, условие несмещенности имеет вид

$$\widehat{\varphi}(0)\,(1 - \theta) + \widehat{\varphi}(1)\,\theta = 1/\theta.$$

При $\theta \to 0$ линейная функция в левой части стремится к $\widehat{\varphi}(0)$, а гипербола в правой — к бесконечности.

Пример 3. Рассмотрим выборку из какого-либо распределения с двумя параметрами μ и σ, где $\mu = \mathbf{M}X_1$ и $\sigma^2 = \mathbf{D}X_1$ (скажем, нормального закона $\mathcal{N}(\mu, \sigma^2)$ из § 2 гл. 3). По свойствам математического ожидания (П2) выборочное среднее $\overline{X}$ несмещенно оценивает параметр μ. В качестве оценки для неизвестной дисперсии $\varphi(\sigma) = \sigma^2$ можно взять *выборочную дисперсию*

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n}\sum_{i=1}^{n}X_i^2 - \overline{X}^2. \tag{1}$$

Вопрос 3.
Как доказать последнее равенство, не проводя вычислений, а опираясь только на формулу (4) гл. 1 и теорему о замене переменных (П2)?

Однако, оценка S^2 *имеет смещение*. Действительно, так как случайные величины X_i независимы и одинаково распределены, то, применяя свойства математического ожидания (П2) на основе формулы (1), получаем:

$$\mathbf{M}S^2 = \mathbf{M}\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2 - \frac{1}{n^2}\sum_{i,j=1}^{n}X_iX_j\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{M}X_i^2 - \frac{1}{n^2}\sum_{i=1}^{n}\mathbf{M}X_i^2 -$$
$$- \frac{1}{n^2}\sum_{i\neq j}^{n}\mathbf{M}X_i\,\mathbf{M}X_j = \mathbf{M}X_1^2 - \frac{1}{n}\,\mathbf{M}X_1^2 - \frac{n-1}{n}\,(\mathbf{M}X_1)^2 = \frac{n-1}{n}\,\mathbf{D}X_1.$$

Чтобы устранить смещение, достаточно домножить S^2 на $n/(n-1)$.

В *нормальной модели* $\mathcal{N}(\mu, \sigma^2)$ можно несмещенно оценить само *стандартное отклонение* σ с помощью оценки $\widehat{\sigma} = c_n\sqrt{nS^2/(n-1)}$

(см. [15, с. 29]), где $c_n = \sqrt{\frac{n-1}{2}}\,\Gamma\left(\frac{n-1}{2}\right)\Big/\Gamma\left(\frac{n}{2}\right)$, $\Gamma(x)$ — гамма-функция Эйлера, определенная ранее в § 4 гл. 3. Отметим, что с ростом n коэффициент c_n довольно быстро убывает к 1:

n	2	3	4	5	10	20	50
c_n	1,253	1,128	1,085	1,064	1,028	1,012	1,005

Само по себе свойство несмещенности *не достаточно* для того, чтобы оценка хорошо приближала неизвестный параметр. Например, первый элемент X_1 выборки из закона Бернулли служит несмещенной оценкой для θ: $\mathbf{M}X_1 = 0\cdot(1-\theta) + 1\cdot\theta = \theta$. Однако, его возможные значения 0 и 1 даже не принадлежат $\Theta = (0,1)$. Необходимо, чтобы погрешность приближения стремилась к нулю с увеличением размера выборки. Это свойство в математической статистике называется *состоятельностью*.

Определение. Оценка $\widehat{\theta}(x_1,\ldots,x_n)$ параметра θ называется *состоятельной*, если для всех $\theta\in\Theta$ последовательность

$$\widehat{\theta}_n = \widehat{\theta}(X_1,\ldots,X_n) \xrightarrow{\mathsf{P}} \theta \quad \text{при } n\to\infty.$$

Здесь $\xrightarrow{\mathsf{P}}$ обозначает *сходимость по вероятности* (см. П5):

для любого $\varepsilon>0$ $\quad \mathbf{P}(|\widehat{\theta}_n - \theta| > \varepsilon) \to 0 \quad$ при $n\to\infty$.

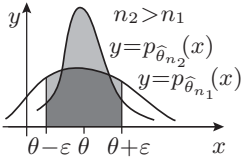


Рис. 4

Состоятельность оценки (а точнее — последовательности оценок $\{\widehat{\theta}_n\}$) означает концентрацию вероятностной массы около истинного значения параметра с ростом размера выборки n (рис. 4).

Как установить, будет ли данная оценка состоятельной? Обычно оказывается полезным один из следующих **трех способов**.

1) Иногда удается доказать состоятельность, непосредственно вычисляя функцию распределения оценки (задачи 1 и 2).

2) Другой способ проверки состоит в использовании закона больших чисел (П6) и свойства сходимости 3 из П5. (Так, оценка $\widehat{\theta} = 1/\overline{X}$ из примера 1 будет состоятельной ввиду непрерывности функции $\varphi(x) = 1/x$ при $x>0$.)

3) Часто установить состоятельность помогает

Лемма. Если смещение $b_n(\theta) = \mathbf{M}\widehat{\theta}_n - \theta$ и дисперсия $\mathbf{D}\widehat{\theta}_n$ стремятся к нулю при $n\to\infty$, то оценка $\widehat{\theta}$ состоятельна.

Доказательство. По неравенству Чебышева (П4)

$$\mathbf{P}(|\widehat{\theta}_n - \theta| > \varepsilon) \leqslant \frac{\mathbf{M}(\widehat{\theta}_n - \theta)^2}{\varepsilon^2}.$$

Но

$$\mathbf{M}(\widehat{\theta}_n - \theta)^2 = \mathbf{M}(\widehat{\theta}_n - \mathbf{M}\widehat{\theta}_n + \mathbf{M}\widehat{\theta}_n - \theta)^2 = \mathbf{M}\left[(\widehat{\theta}_n - \mathbf{M}\widehat{\theta}_n) + b_n(\theta)\right]^2 =$$
$$= \mathbf{D}\widehat{\theta}_n + 2\,b_n(\theta)\,\mathbf{M}(\widehat{\theta}_n - \mathbf{M}\widehat{\theta}_n) + b_n^2(\theta) = \mathbf{D}\widehat{\theta}_n + b_n^2(\theta). \qquad \blacksquare$$

§ 3. ФУНКЦИИ РИСКА

Кто не рискует, тот не пьет шампанского.

Проблема. Как измерить точность оценки?

Пусть $\rho(u) \geqslant 0$ обозначает *функцию штрафа (потерь)* в том смысле, что мы платим штраф $\rho(\widehat{\theta} - \theta)$ за отклонение оценки $\widehat{\theta}(x_1, \ldots, x_n)$ от истинного значения параметра θ. Обычно $\rho(0) = 0$ и $\rho(u)$ возрастает с ростом $|u|$.

Если используется функция штрафа $\rho(u) = |u|$, то риск называют *абсолютным*, а если $\rho(u) = u^2$ — *квадратичным*.

Определение. *Функцией риска* оценки $\widehat{\theta}$ называется

$$R_{\widehat{\theta}}(\theta) = \mathbf{M}\rho(\widehat{\theta}(X_1, \ldots, X_n) - \theta),$$

т. е. средняя величина потерь при оценивании θ с помощью $\widehat{\theta}$.

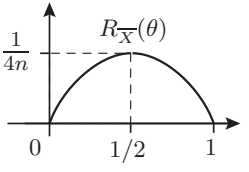


Рис. 5

Пример 4. Вычислим квадратичный риск частоты $\overline{X}$ для схемы Бернулли из примера 2. В силу несмещенности $\overline{X}$ и свойств дисперсии $R_{\overline{X}}(\theta) = \mathbf{M}(\overline{X} - \theta)^2 = \mathbf{M}(\overline{X} - \mathbf{M}\overline{X})^2 = \mathbf{D}\overline{X} = n^{-2}\mathbf{D}(X_1 + \ldots + X_n) = n^{-1}\mathbf{D}X_1 = \theta(1-\theta)/n$ (дисперсия случайной величины X_1 была найдена в § 2 гл. 1). График функции $R_{\overline{X}}(\theta)$ (верхняя часть параболы) для $\theta \in (0, 1)$ приведен на рис. 5.

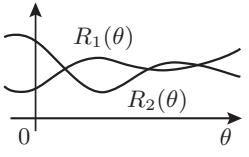


Рис. 6

Как же сравнивать оценки? Можно считать ту оценку лучшей, у которой риск меньше. Но риск — это функция от θ. Каким образом выбрать «наименьшую» из двух функций, скажем, таких, как на рис. 6? (*Что больше, синус или косинус?*)

Рассмотрим **три подхода** к этой проблеме.

1) *Минимаксный* (осторожный) подход заключается в сравнении функций по их наибольшему значению на множестве Θ (см. рис. 7): выбирается та оценка, у которой при самом неблагоприятном значении θ риск меньше. Таким образом, при этом подходе выбор оценки диктуется желанием избежать крупного штрафа, если θ окажется вблизи точки максимума функции риска.

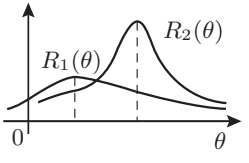


Рис. 7

Однако, возможна ситуация, когда минимаксный подход противоречит здравому смыслу (рис. 8). В подобных случаях более разумным представляется

2) *Байесовский* (интегральный) подход: сравниваются два интеграла, $I_1 = \int R_1(\theta)\, dQ$ и $I_2 = \int R_2(\theta)\, dQ$, где Q — некоторая мера на множестве Θ. В частности, когда θ — скалярный параметр и Q — равномерная мера на Θ, лучшей при таком подходе считается оценка, у которой меньше площадь под графиком функции риска.

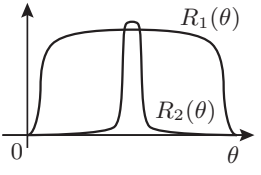


Рис. 8

Т. Байес (1702–1761), английский математик.

Байесовскими в математической статистике называются методы, при которых априорная информация о параметре, если таковая имеется, формализуется в виде некоторого распределения Q (не обязательно вероятностного) на параметрическом множестве Θ (см. [38, с. 168]). При этом, если Q — вероятностная мера (см. П1), на сам параметр θ можно смотреть, как на случайную величину. В некоторых ситуациях предположение о случайности параметра θ

выглядит весьма естественно: «природа» как бы разыгрывает значение θ в соответствии с распределением Q перед моделированием очередной выборки.

На практике можно оценить функцию распределения меры Q по частоте появления θ в ранее проведенных экспериментах. Иногда из соображений равной возможности всех значений θ априори полагают, что Q — это равномерная мера на множестве Θ.

3) *Ограничение множества оценок*: для некоторых статистических моделей существуют оценки, обладающие *равномерно минимальным риском в заданном классе оценок* (рис. 9).

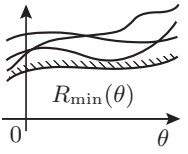


Рис. 9

Так, в нормальной модели из примера 3 оценки $\widehat{\mu} = \overline{X}$ и $\widehat{\sigma^2} = nS^2/(n-1)$ имеют равномерно минимальную дисперсию среди *всех несмещенных оценок с конечной дисперсией* (см. [50, с. 83]). То же самое справедливо (см. [50, с. 75]) для частоты $\overline{X}$ как оценки неизвестной вероятности «успеха» в схеме Бернулли и для несмещенной оценки $\widehat{\theta}_3 = \dfrac{n+1}{n}\max\{X_1,\ldots,X_n\} = \dfrac{n+1}{n}\,\widehat{\theta}_1$ параметра масштаба θ в модели равномерного распределения на отрезке $[0,\theta]$ из § 1.

Для последней оценки согласно задаче 3 гл. 1 имеем

$$\mathbf{D}\widehat{\theta}_3 = \left(\frac{n+1}{n}\right)^2 \mathbf{D}\widehat{\theta}_1 = \frac{(n+1)^2}{n^2}\ \frac{n}{(n+1)^2(n+2)}\ \theta^2 = \frac{1}{n(n+2)}\ \theta^2. \qquad (2)$$

Отметим, что за устранение смещения максимума $\widehat{\theta}_1$ пришлось заплатить увеличением дисперсии в $\left(\dfrac{n+1}{n}\right)^2$ раз. Ввиду задачи 2 гл. 1 дисперсия $\mathbf{D}\widehat{\theta}_2 = 4\,\mathbf{D}\overline{X} = \theta^2/(3n)$. Отсюда находим, что отношение $\mathbf{D}\widehat{\theta}_3/\mathbf{D}\widehat{\theta}_2 = 3/(n+2) \leqslant 1$ и стремится к 0 при $n \to \infty$.

Сравним точность этих оценок для данных эксперимента. На самом деле выборка была получена умножением на $\theta_0 = 35$ десяти псевдослучайных чисел из первой строки таблицы Т1. Легко вычислить, что $\widehat{\theta}_2(x_1,\ldots,x_n) = 28{,}1$ и $\widehat{\theta}_3(x_1,\ldots,x_n) = 33{,}1$. Значит, оценка $\widehat{\theta}_3$ в данном случае точнее, а значение $\widehat{\theta}_2$ оказалось даже меньше, чем $x_{10} = \max\{x_1,\ldots,x_n\} = 30{,}1$.

В заключение, обсудим **проблему выбора функции штрафа**. Во многих статистических моделях существуют несмещенные оценки с равномерно минимальным риском для любой *выпуклой* (П4) штрафной функции (см. [50, с. 79]).

С другой стороны, реальные (например, финансовые) потери всегда ограничены. Но ни одна ограниченная на $(-\infty,+\infty)$ функция $\rho(u) \neq \text{const}$ не может быть выпуклой. К сожалению, для ограниченных штрафных функций (скажем, для $\rho(u) = \text{const}\cdot I_{\{|u|>\delta\}}$), как правило, не существует несмещенных оценок не только с равномерно, но и с локально минимальным риском (см. [50, с. 81]).

К счастью, для выборок большого размера ситуация упрощается. Для гладкой функции потерь ее разложение в ряд Тейлора в нуле

дает

$$\rho(\theta' - \theta) = a + b\,(\theta' - \theta) + c\,(\theta' - \theta)^2 + \varepsilon,$$

где остаток ε пренебрежимо мал при достаточно малом $|\theta' - \theta|$. Условия $\rho(0) = 0$ и $\rho(u) \geqslant 0$ влекут, соответственно, равенства $a = 0$ и $b = 0$. Следовательно, $\rho(\theta' - \theta) = c\,(\theta' - \theta)^2 + \varepsilon$. Таким образом, для состоятельных оценок $\widehat{\theta}$ минимизация риска $\mathsf{M}\rho(\widehat{\theta} - \theta)$ при больших n, по существу, *равносильна минимизации квадратичного риска* $\mathsf{M}(\widehat{\theta} - \theta)^2$ (т. е. не зависит от конкретного вида функции ρ).

Однако, оценки, наилучшие при квадратичной функции потерь, часто бывают слишком чувствительными к выделяющимся значениям элементов выборки (так называемым «выбросам»). Эту проблему мы рассмотрим подробнее в главе 8 при обсуждении устойчивости оценок к «выбросам».

Взирая на солнце, прищурь глаза свои, и ты смело разглядишь в нем пятна.

Козьма Прутков

§ 4. МИНИМАКСНАЯ ОЦЕНКА В СХЕМЕ БЕРНУЛЛИ

Как было отмечено выше, выборочное среднее $\overline{X}$ в схеме Бернулли — несмещенная оценка вероятности «успеха» θ, обладающая равномерно минимальной дисперсией. Однако *минимаксной* (имеющей наименьший максимум риска) *для квадратичного штрафа* является (см. [50, с. 228]) *оценка Ходжеса—Лемана*

$$\tilde{\theta} = \overline{X} + \frac{1}{1 + \sqrt{n}} \left(\frac{1}{2} - \overline{X} \right). \tag{3}$$

Давайте проведем **эксперимент** по сравнению точности оценок $\overline{X}$ и $\tilde{\theta}$.

1) Задумайте вероятность «успеха» $\theta_0 \in (0, 1)$.

2) Смоделируйте выборку размера $n = 9$ из распределения Бернулли с помощью таблицы Т1 (см. вопрос 2 гл. 2).

3) Вычислите значения оценок и определите, какая из них оказалась ближе к θ_0.

Например, пусть $\theta_0 = 0{,}17$. По первой строке в таблице Т1 получаем реализацию выборки: 1, 1, 0, 0, 0, 0, 0, 1, 0 (если число в таблице меньше 17, то пишем «1», иначе — «0»). Для таких x_i находим, что $\overline{x} = \frac{1}{3} \approx 0{,}333$ и $\tilde{\theta} = \frac{3}{8} = 0{,}375$. Видим, что в данном случае частота оказалась точнее. А как у вас?

Если проводить этот эксперимент многократно (каждый раз загадывая новое значение θ_0), то примерно в половине случаев оценка $\tilde{\theta}$ будет ближе к θ_0, чем $\overline{x}$. Но согласно изложенной ниже «теории» такое, казалось бы, должно происходить намного чаще!

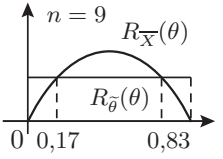


Рис. 10

Не всякой сказке верь.

Действительно, из рис. 10, на котором приведены графики квадратичных рисков при $n = 9$ оценок $\overline{X}$ и $\tilde{\theta}$ (см. задачу 4), следует, что доля тех θ, при которых $R_{\tilde{\theta}}(\theta) < R_{\overline{X}}(\theta)$, равна примерно $0{,}83 - 0{,}17 = 0{,}66$. Почему же на практике оценка $\tilde{\theta}$ обычно выигрывает лишь в 50% случаев?

Объяснение. Дело в том, что, сравнивая $\overline{X}$ и $\tilde{\theta}$, мы смотрим, значение какой из них оказалось ближе к θ_0, а не платим квадратичный штраф за погрешность оценивания параметра θ_0.

Другими словами, средний выигрыш в пари при ставке C на оценку $\tilde{\theta}$ равен $Cp(\theta) + (-C)[1 - p(\theta)] = 2C\,[p(\theta) - 1/2]$, где $p(\theta) = \mathsf{P}(|\tilde{\theta} - \theta| < |\overline{X} - \theta|)$. Игра выгодна, когда $p(\theta) > 1/2$.

Функция $p(\theta)$ симметрична относительно $1/2$. Нетрудно вывести, что при $\theta \leqslant 1/2$

$$p(\theta) = 1 - \mathsf{P}\left(\frac{\theta - d_n}{1 - 2d_n} \leqslant \overline{X} \leqslant \frac{1}{2}\right), \quad \text{где } d_n = \frac{1}{4\,(1 + \sqrt{n})}\,.$$

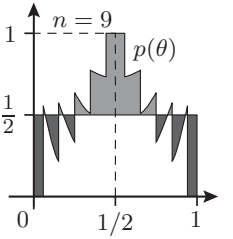


Рис. 11

Ее можно вычислить, используя то, что случайная величина $n\overline{X}$ имеет биномиальное распределение (см. утверждение 1 гл. 5).

Предполагая равномерность выбора значения θ_0 из $[0, 1]$ (байесовский подход), заметим, что при $n = 9$ площади над и под уровнем $1/2$ графика $p(\theta)$ на рис. 11 отличаются всего на 0,008.

Похожая ситуация возникает при оценке параметра сдвига μ в нормальной модели $X_i \sim \mathcal{N}(\mu, \sigma^2)$ с *известным* параметром масштаба σ. Эта модель может использоваться для измерений, точность которых заранее известна. В ней выборочное среднее $\overline{X}$ будет не только несмещенной оценкой с равномерно минимальной дисперсией, но и минимаксной оценкой для квадратичного риска. Тем не менее, существует такая оценка $\tilde{\mu}$, что $\mathsf{P}(|\tilde{\mu} - \mu| < |\overline{X} - \mu|) > 1/2$ *при всех* μ. Эта оценка выглядит так:

$$\tilde{\mu} = \overline{X} - \frac{\sigma}{2\sqrt{n}}\ \operatorname{sign}(\overline{X})\ \min\left\{\sqrt{n}\,|\overline{X}|/\sigma,\ \Phi(-\sqrt{n}\,|\overline{X}|/\sigma)\right\},$$

где $\Phi(x)$ — функция распределения закона $\mathcal{N}(0, 1)$ (см. [15, с. 14]).

На хорошее всегда найдется лучшее.

ЗАДАЧИ

1. Для случайных величин X_i $i = 1, \ldots, n$, взятых наудачу из отрезка $[0, \theta]$, докажите состоятельность оценки $\widehat{\theta}_1 = X_{(n)} = \max\{X_1, \ldots, X_n\}$

а) непосредственно из определения,

б) применяя лемму из § 2.

2. Для статистической модели из задачи 1 проверьте, что оценка $\widehat{\theta}_4 = (n + 1)\,X_{(1)}$, где $X_{(1)} = \min\{X_1, \ldots, X_n\}$,

а) не имеет смещения,

б) *не является состоятельной.*

3. Придумайте какую-нибудь несмещенную и состоятельную оценку для параметра θ в модели сдвига показательного распределения (см. рис. 12)

$$F_\theta(x) = \begin{cases} 1 - e^{-(x-\theta)} & \text{при } x > \theta, \\ 0 & \text{при } x \leqslant \theta. \end{cases}$$

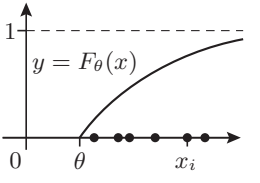


Рис. 12

4. Вычислите смещение, дисперсию и квадратичный риск минимаксной оценки Ходжеса—Лемана в схеме Бернулли (см. формулу (3)).

5. Для схемы Бернулли при $n = 3$ нарисуйте график *абсолютного* риска $\overline{X}$. Будет ли эта функция

а) непрерывной,

б) гладкой?

6. Для показательной выборки $X_1, \ldots, X_n$ из примера 1 положим $S_n = X_1 + \ldots + X_n$. Установите несмещенность оценок

а) $\widehat{\theta} = (n-1)/S_n$ для параметра θ при $n > 1$,

б) $\widehat{\varphi} = (1 - t/S_n)^{n-1} I_{\{S_n > t\}}$ для *функции надежности* $\varphi(\theta) = \mathbf{P}(X_1 > t) = e^{-\theta t}$, где $t > 0$.

Указание. Примените лемму 1 гл. 4 и теорему о замене переменных (П2).

7*. Для модели из задачи 1 сравните дисперсии несмещенных оценок $\widehat{\theta}_3 = \dfrac{n+1}{n}\, X_{(n)}$ и $\widehat{\theta}_5 = X_{(1)} + X_{(n)}$.

Указание. Найдите функцию распределения $F(x,y)$ и плотность $p(x,y)$ вектора $(X_{(1)}, X_{(n)})$ и вычислите математическое ожидание $\mathbf{M}X_{(1)}X_{(n)} = \int\int xy\, p(x,y)\, dxdy$.

РЕШЕНИЯ ЗАДАЧ

1. **а)** Ввиду задачи 3 гл. 1 $F_{X_{(n)}}(x) = (x/\theta)^n$ при $0 \leqslant x \leqslant \theta$. Поскольку $\mathbf{P}(X_{(n)} \leqslant \theta) = 1$, для любого ε из $(0, \theta)$ имеем

$$\mathbf{P}(|\widehat{\theta}_1 - \theta| > \varepsilon) = \mathbf{P}(X_{(n)} \leqslant \theta - \varepsilon) = (1 - \varepsilon/\theta)^n \to 0$$

при $n \to \infty$.

б) Из той же задачи $\mathbf{M}\widehat{\theta}_1 = \dfrac{n\theta}{n+1} \to \theta$ и $\mathbf{D}\widehat{\theta}_1 = \dfrac{n\theta^2}{(n+1)^2(n+2)} \to 0$.

2. **а)** Очевидно, что случайные величины $X'_i = \theta - X_i$, $i = 1, \ldots, n$, также образуют выборку из равномерного распределения на отрезке $[0, \theta]$. Поэтому случайная величина $X'_{(1)} = \min\{X'_1, \ldots, X'_n\} = \theta - X_{(n)}$ распределена так же, как $X_{(1)}$. Отсюда с учетом решения предыдущей задачи получаем $\mathbf{M}X_{(1)} = \theta - \mathbf{M}X_{(n)} = \theta - n\theta/(n+1) = \theta/(n+1)$.

б) Используя независимость случайных величин X_i, находим, что при $n \to \infty$

$$\mathbf{P}(\widehat{\theta}_4 > \theta + \varepsilon) = \prod_{i=1}^{n} \mathbf{P}\left(X_i > \frac{\theta + \varepsilon}{n+1}\right) = \left(1 - \frac{\theta + \varepsilon}{\theta(n+1)}\right)^n \to e^{-(\theta+\varepsilon)/\theta} > 0.$$

3. Рисунок 12 подсказывает взять $X_{(1)} = \min\{X_1, \ldots, X_n\}$ в качестве состоятельной оценки параметра сдвига. Поскольку

$\mathbf{P}(X_{(1)} > \theta) = 1$, эта оценка имеет смещение. Вычислим его. Для $\Delta = X_{(1)} - \theta$ в силу независимости величин X_i находим:

$$\mathbf{P}(\Delta > x) = \mathbf{P}(X_1 > x + \theta) \cdot \ldots \cdot \mathbf{P}(X_n > x + \theta) = e^{-nx}.$$

Иными словами, случайная величина Δ имеет показательное распределение с параметром n. С учетом ответа на вопрос 3 гл. 4 получаем $\mathbf{M}\Delta = 1/n$ и $\mathbf{D}\Delta = 1/n^2$. Следовательно, $\widehat{\theta} = X_{(1)} - 1/n$ — несмещенная и состоятельная (в силу леммы из § 2) оценка.

Другим решением задачи может служить несмещенная оценка $\overline{X} - 1$ (случайные величины $X_i - \theta$ показательно распределены с параметром 1). Ее состоятельность вытекает из закона больших чисел (П6) и свойств сходимости (П5).

4. Минимаксную оценку $\tilde{\theta}$ можно записать так:

$$\tilde{\theta} = (1 - \varepsilon_n)\,\overline{X} + (1/2)\,\varepsilon_n, \text{ где } \varepsilon_n = 1/(1 + \sqrt{n}). \tag{4}$$

Поскольку $\mathbf{M}\overline{X} = \theta$ и $\mathbf{D}\overline{X} = \theta(1 - \theta)/n$, по свойствам из П2 смещение $b(\theta) = \mathbf{M}\tilde{\theta} - \theta = (1/2 - \theta)\,\varepsilon_n$, $\mathbf{D}\tilde{\theta} = (1 - \varepsilon_n)^2\,\mathbf{D}\overline{X} = (1 - \varepsilon_n)^2\,\theta(1 - \theta)/n$. Таким образом, $\tilde{\theta}$ — смещенная, однако состоятельная (в силу леммы из § 2) оценка.

Квадратичный риск $\tilde{\theta}$ легко найти по формуле, полученной при доказательстве леммы: $R_{\tilde{\theta}}(\theta) = \mathbf{D}\tilde{\theta} + b^2(\theta) = 1/[4(1 + \sqrt{n})^2]$.

Замечание. Так как $R_{\overline{X}}(1/2) = 1/(4n)$, то $R_{\tilde{\theta}}(\theta) < R_{\overline{X}}(\theta)$ в некоторой окрестности точки $1/2$ при любом n (см. рис. 10). Поэтому для достаточно близких к $1/2$ значений параметра θ оценка $\tilde{\theta}$ предпочтительнее. Это и понятно: ввиду формулы (4) она, являясь «взвешенным средним» $\overline{X}$ и $1/2$, подправляет оценку $\overline{X}$, притягивая ее к $1/2$.

5. Приведем решение для произвольного n (см. [15, с. 31]). Абсолютный риск частоты $\overline{X}$ в схеме Бернулли есть

$$R_{\overline{X}}(\theta) = \mathbf{M}\left|\overline{X} - \theta\right| = \sum_{i=0}^{n} \left|\frac{i}{n} - \theta\right| C_n^i \theta^i (1 - \theta)^{n-i}.$$

Пусть $d_i(\theta) = (\theta - \frac{i}{n})\, C_n^i\, \theta^i (1 - \theta)^{n-i}$. Так как $\mathbf{M}\overline{X} = \theta$, то $\sum\limits_{i=0}^{n} d_i(\theta) = 0$. С учетом этого для $\theta \in \left[\dfrac{k-1}{n}, \dfrac{k}{n}\right]$, $k = 1,2,\ldots,n$, абсолютный риск представляется в следующем виде:

$$R_{\overline{X}}(\theta) = \sum_{i=0}^{k-1} d_i(\theta) - \sum_{i=k}^{n} d_i(\theta) = 2\sum_{i=0}^{k-1} d_i(\theta) - \sum_{i=0}^{n} d_i(\theta) = 2\sum_{i=0}^{k-1} d_i(\theta).$$

Аналогично выводу формулы (2) из § 2 гл. 5 устанавливается, что правая часть равна $2C_{n-1}^{k-1}\,\theta^k(1-\theta)^{n-k+1}$ (проверьте!). В случае $n=3$ имеем

$$R_{\overline{X}}(\theta)=\begin{cases}2\,\theta(1-\theta)^3 & \text{при } 0\leqslant\theta\leqslant 1/3,\\ 4\,\theta^2(1-\theta)^2 & \text{при } 1/3\leqslant\theta\leqslant 2/3,\\ 2\,\theta^3(1-\theta) & \text{при } 2/3\leqslant\theta\leqslant 1.\end{cases}$$

График этой функции приведен на рис. 13. На отрезке $\left[0,\frac{1}{3}\right]$ риск имеет локальный максимум при $\theta=\frac{1}{4}$. Части графика на концах отрезков $\left[\frac{k-1}{3},\,\frac{k}{3}\right]$ $(k=1,2,3)$ не стыкуются гладко.

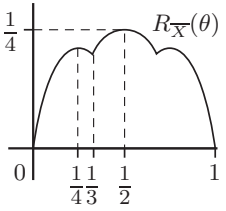


Рис. 13

6. **а)** В силу леммы 1 гл. 4 случайная величина $S_n\sim\Gamma(n,\theta)$ с плотностью $p_{S_n}(s)=\theta^n s^{n-1}e^{-\theta s}/\Gamma(n)$ при $s>0$. Далее, по теореме о замене переменных из П2, определению и основному свойству гамма-функции (см. § 4 гл. 3) имеем при $n>1$

$$\mathbf{M}\widehat{\theta}=\frac{(n-1)\,\theta^n}{\Gamma(n)}\int\limits_0^\infty s^{n-2}e^{-\theta s}ds=\frac{\theta(n-1)}{\Gamma(n)}\;\Gamma(n-1)=\theta.$$

б) Аналогично, сделав замену $x=\theta(s-t)$, получим

$$\mathbf{M}\widehat{\varphi}=\int\limits_t^\infty(1-t/s)^{n-1}p_S(s)\,ds=\frac{\theta^n}{\Gamma(n)}\int\limits_t^\infty(s-t)^{n-1}e^{-\theta s}ds=e^{-\theta t}.$$

7. Так как θ — параметр масштаба для распределений оценок $\widehat{\theta}_3$ и $\widehat{\theta}_5$, то достаточно найти отношение $\mathbf{D}\widehat{\theta}_5/\mathbf{D}\widehat{\theta}_3$ при $\theta=1$. Вычислим совместное распределение случайных величин $X_{(1)}$ и $X_{(n)}$:

$$F(x,y)=\mathbf{P}(X_{(1)}\leqslant x,X_{(n)}\leqslant y)=\mathbf{P}(X_{(n)}\leqslant y)-\mathbf{P}(X_{(1)}>x,X_{(n)}\leqslant y)=$$
$$=\begin{cases}y^n-(y-x)^n, & \text{если } 0<x<y<1,\\ y^n, & \text{если } 0<x\geqslant y<1,\end{cases}$$

поскольку $\mathbf{P}(X_{(1)}>x,X_{(n)}\leqslant y)=\mathbf{P}(x<X_i\leqslant y,\,i=1,\ldots,n)$, а эта вероятность равна $(y-x)^n$, если $x<y$, и равна 0 в противном случае.

Плотность $p(x,y)=\dfrac{\partial^2F(x,y)}{\partial x\partial y}=n(n-1)(y-x)^{n-2}I_{\{0<x<y<1\}}$,

$\mathbf{M}X_{(1)}X_{(n)}=n(n-1)\int\limits_0^1 x\int\limits_x^1 y(y-x)^{n-2}dy\,dx$. С помощью замены $z=y-x$ легко получить, что внутренний интеграл равен

$(1-x)^n/\,n+x\,(1-x)^{n-1}/\,(n-1)$.

Согласно определению бета-функции и формуле (10) гл. 3, выражающей ее через гамма-функцию, находим:

$$\mathbf{M}X_{(1)}X_{(n)} = (n-1)\,B(2,n+1) + n\,B(3,n) =$$
$$= [(n-1)\,\Gamma(2)\,\Gamma(n+1) + n\,\Gamma(3)\,\Gamma(n)]\,/\,\Gamma(n+3).$$

Поскольку $\Gamma(n) = (n-1)!$, окончательно получаем, что

$$\mathbf{M}X_{(1)}X_{(n)} = [(n-1)1!\,n! + n\,2!\,(n-1)!\,]\,/\,(n+2)! = 1/(n+2).$$

С учетом свойств дисперсии (П2) и решений задач 1–2

$$\mathbf{D}\widehat{\theta}_5 = \mathbf{D}X_{(1)} + \mathbf{D}X_{(n)} + 2\,(\mathbf{M}X_{(1)}X_{(n)} - \mathbf{M}X_{(1)}\,\mathbf{M}X_{(n)}) =$$
$$= 2\,\mathbf{D}X_{(n)} + 2\left(\frac{1}{n+2} - \frac{1}{n+1}\,\frac{n}{n+1}\right) = 2\left(1+\frac{1}{n}\right)\mathbf{D}X_{(n)}.$$

Отсюда и из формулы (2) имеем $1 \leqslant \mathbf{D}\widehat{\theta}_5/\mathbf{D}\widehat{\theta}_3 = \dfrac{2}{1+1/n} \to 2$ при $n \to \infty$.

ОТВЕТЫ НА ВОПРОСЫ

1. **а)** Значение θ не может равняться 28, так как $x_{10} = 30{,}1$.
б) Хотя значение 100 для θ теоретически возможно, но интуитивно представляется крайне маловероятным (вероятность получить наблюдаемое расположение точек при $\theta = 100$ не превосходит $(\max x_i/\theta)^n = 0{,}301^{10} \approx 6{,}1 \cdot 10^{-6}$).

2. Оценка $\widehat{\theta}_1$ всегда недооценивает θ и поэтому смещена. Несмещенность оценки $\widehat{\theta}_2$, очевидно, вытекает из симметрии распределения величины X_1 относительно $1/2$ (откуда $\mathbf{M}X_1 = \theta/2$) и свойств математического ожидания (П2).

3. Для дискретной случайной величины ξ, принимающей значения x_i $(i = 1,\ldots,n)$ с одной и той же вероятностью $1/n$, $\mathbf{M}\xi = \overline{x}$. В силу теоремы о замене переменных (П2)

$$\mathbf{M}\xi^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2, \quad \mathbf{D}\xi = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2.$$

Таким образом, при любых x_i проверяемое равенство — частный случай формулы (4) гл. 1. Из пушки по воробьям.

Глава 7

АСИМПТОТИЧЕСКАЯ НОРМАЛЬНОСТЬ

Все верят в универсальность нормального распределения: физики верят потому, что думают, что математики доказали его логическую необходимость, а математики верят, так как считают, что физики проверили это лабораторными экспериментами.

А. Пуанкаре

§ 1. РАСПРЕДЕЛЕНИЕ КОШИ

В [11, с. 178] приведено описание следующего эксперимента по *локализации источника излучения*. В некоторой точке трехмерного пространства с неизвестными координатами (a, b, c) находится источник γ-излучения. Регистрируются координаты (x_i, y_i), $i = 1, \ldots, n$, точек пересечения траекторий γ-квантов с поверхностью детекторной плоскости $z = 0$. Требуется оценить параметры a и b по этим данным, предполагая, что направления траекторий γ-квантов случайны, т. е. равномерно распределены на сфере с центром в точке (a, b, c) (рис. 1).

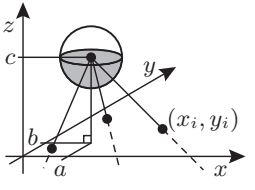


Рис. 1

Какую оценку можно было бы предложить для (a, b)? Первое, что приходит в голову, — это $(\overline{X}, \overline{Y})$. Ясно, что точки пересечения траекторий с плоскостью $z = 0$ располагаются гуще непосредственно под источником излучения. В подобных случаях прибегают к усреднению данных, чтобы, по возможности, устранить разброс измерений (предполагается, что при этом происходит взаимная компенсация отклонений в разные стороны).

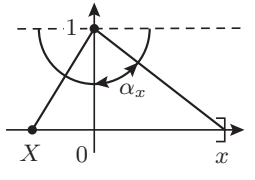


Рис. 2

Однако, в данном случае усреднение *совершенно бесполезно*. Для объяснения, почему это так, рассмотрим одномерный аналог эксперимента (двумерная модель разбирается в задаче 7): из точки $(0, 1)$ выходит случайный луч (рис. 2), направление которого равномерно распределено на нижней полуокружности с центром $(0, 1)$. Случайная величина X — координата пересечения этого луча с осью абсцисс. Какая плотность $p(x)$ у этой величины?

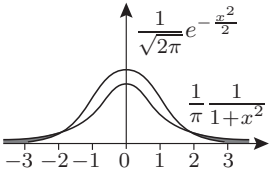


Рис. 3

Решение. Понятно, что плотность — четная функция. Вычислим ее для $x \geqslant 0$. Найдем сначала функцию распределения $F(x) = \mathbf{P}(X \leqslant x)$ (см. рис. 2):

$$F(x) = \mathbf{P}(X \leqslant 0) + \mathbf{P}(0 < X \leqslant x) = \frac{1}{2} + \frac{\alpha_x}{\pi} = \frac{1}{2} + \frac{1}{\pi} \operatorname{arctg} x.$$

Отсюда $p(x) = F'(x) = 1/[\pi\,(1 + x^2)]$. Это — *плотность Коши*. На первый взгляд она похожа на плотность стандартного нормального закона $\mathcal{N}(0, 1)$ (рис. 3). Однако, они различаются по скорости убывания к нулю при $x \to \infty$ вероятностей $\mathbf{P}(X \leqslant -x)$ и $\mathbf{P}(X \geqslant x)$ (так

называемых «*хвостов распределения*»). У закона Коши «хвосты» намного «тяжелее».

Ленивой лошади и хвост в тягость.

Чем опасны «тяжелые хвосты»? Тем, что случайная величина с таким распределением с довольно существенной вероятностью может принимать большие по абсолютной величине значения. Поэтому в реализации выборки большого размера из такого закона обязательно появятся одно или несколько наблюдений, которые сильно отличаются от остальных (их называют «*выбросами*»). В этом случае при оценивании «центра» распределения при помощи выборочного среднего $\overline{X}$ произойдет резкое смещение оценки в сторону наибольшего «выброса» (см. задачу 6).

Куда один баран, туда и все стадо.

Из-за слишком «тяжелых хвостов» у закона Коши не существует даже математического ожидания (см. замечание в § 2 гл. 1). Если бы оно существовало, то по усиленному закону больших чисел (П6) среднее арифметическое сходилось бы к $\mathbf{M}X_1$ с вероятностью 1 при $n \to \infty$. А что происходит с $\overline{X}$ для выборки из распределения Коши? Чтобы выяснить это, используем характеристические функции (П9):

$$\psi_{X_1}(t) = \mathbf{M}e^{itX_1} = \int\limits_{-\infty}^{\infty} (\cos tx + i \sin tx)\, p(x)\, dx = \\ = \frac{1}{\pi} \int\limits_{-\infty}^{\infty} \frac{\cos tx}{1+x^2}\, dx.$$

Этот интеграл, зависящий от параметра t, можно явно вычислить с помощью *теоремы о вычетах* (см. [73, с. 239]). Ответ таков: $\psi_{X_1}(t) = e^{-|t|}$. (Другой способ — применение к $e^{-|t|}$ обратного преобразования Фурье из П9.) Отсюда, используя свойства 2 и 3 характеристических функций из П9 (при $a = 0$, $b = 1/n$), находим:

$$\psi_{n\overline{X}}(t) = \prod_{i=1}^{n} \psi_{X_i}(t) = e^{-n|t|}, \quad \psi_{\overline{X}}(t) = e^{-n|t/n|} = e^{-|t|}.$$

Таким образом, характеристическая функция среднего $\overline{X}$ совпадает с характеристической функцией величины X_1. Так как характеристическая функция однозначно определяет функцию распределения (см. [90, с. 301]), то $\overline{X}$ также имеет распределение Коши при любом n. Поэтому наблюдаемое значение $\overline{X}$ будет отклоняться от 0 ничуть не меньше значений самих X_i.

Как же, все-таки, *состоятельно* оценить θ в *модели сдвига* $F(x-\theta)$, когда F — функция распределения закона Коши? Подходящей оказывается, например, оценка, определяемая в следующем параграфе.

§ 2. ВЫБОРОЧНАЯ МЕДИАНА

Рассмотрим ранее встречавшийся в § 4 гл. 4 вариационный ряд $X_{(1)} \leqslant X_{(2)} \leqslant \ldots \leqslant X_{(n)}$, состоящий из упорядоченных по возрастанию элементов выборки $(X_1, \ldots, X_n)$.

Определение. *Выборочной медианой* называется оценка

$$MED = \begin{cases} X_{(k+1)}, & \text{если } n = 2k+1, \\ (X_{(k)} + X_{(k+1)})/2, & \text{если } n = 2k \end{cases}$$

Рис. 4

(рис. 4 при $n = 5$).

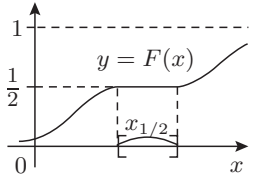


Рис. 5

Выборочная медиана MED служит оценкой для *теоретической медианы* $x_{1/2}$, которая определяется как решение уравнения $F(x) = 1/2$, где $F(x)$ — функция распределения элементов выборки. Для непрерывной функции $F(x)$ решение всегда существует, но может быть не единственным (рис. 5).

Подобно математическому ожиданию (см. § 2 гл. 1), медиана $x_{1/2}$ является характеристикой, показывающей, где располагается «центр» распределения:

$$\mathbf{P}(X_i \leqslant x_{1/2}) = \mathbf{P}(X_i \geqslant x_{1/2}) = \frac{1}{2}$$

(см. также задачу 4).

Пример 1. ***Модель радиоактивного распада*** ([68, с. 5])**.** Как известно, радий (Ra) с течением времени превращается в радон (Rn). В момент распада атом радия излучает α-частицу — ядро атома гелия (He), и происходит переход $Ra \to Rn$. Допустим, что время τ до распада отдельного атома Ra не зависит от состояния других атомов и имеет показательное распределение: $p_t = \mathbf{P}(\tau > t) = e^{-\lambda t}$. Если имеется всего n атомов радия (в одном грамме насчитывается приблизительно 10^{22} атомов), то среднее число остающихся через время t атомов есть $n(t) = np_t = ne^{-\lambda t}$. Определяемая из равенства $n(T) = n/2$ величина T (*период полураспада*) не зависит от исходного количества Ra: $T = \ln 2/\lambda$ (для радия $T \approx 1600$ лет). На языке теории вероятностей T — медиана показательного распределения.

Какими свойствами обладает MED как оценка для $x_{1/2}$?

Теорема 1. Пусть элементы выборки имеют плотность $p(x)$, причем $p(x_{1/2}) > 0$. Тогда при $n \to \infty$

$$\sqrt{n}\,(MED - x_{1/2}) \xrightarrow{d} \xi \sim \mathcal{N}\left(0,\, \frac{1}{4\,p^2(x_{1/2})}\right).$$

Эта сходимость вытекает (см. [50, с. 314]) из теоремы 2, доказываемой в § 3.

Контрпример. Плотность $\tilde{p}(x)$ на рис. 6 симметрична и равна 0 при $|x| \leqslant a$. При $n \to \infty$ MED для нечетных n будет принимать

бесконечное количество раз значения как из интервала $(-\infty, -a)$, так и из интервала $(a, +\infty)$. Поэтому она не может сходиться ни к какой из точек отрезка $[-a,a]$.

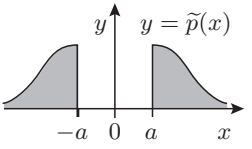


Рис. 6

При выполнении условий теоремы 1 выборочная медиана будет состоятельной оценкой для $x_{1/2}$.

Более того, из теоремы 1 следует, что точность оценки *MED* при больших n имеет порядок малости $1/\sqrt{n}$. Действительно, умножая $(MED - x_{1/2})$ на $\sqrt{n}$ («коэффициент увеличения микроскопа»), мы получаем нечто «практически ограниченное» (с вероятностью 0,997 по «правилу трех сигм» из § 2 гл. 3).

Вопрос 1. Почему *MED* состоятельна? (См. свойства сходимости из П5.)

Задача 5 дает пример состоятельной оценки с иной асимптотикой погрешности.

Возвращаясь к оценке параметра сдвига распределения Коши, видим, что $x_{1/2} = \theta$ (это вытекает из симметрии плотности случайной величины X_i относительно θ), причем $p(x_{1/2}) = 1/\pi > 0$. Следовательно, MED — состоятельная оценка для параметра сдвига. Другие (более точные) оценки θ в этой модели будут приведены в двух следующих главах.

Замечание. Для оценивания координат a и b в эксперименте по локализации источника излучения можно было бы поставить вторую детекторную плоскость, если излучение достаточно сильное, чтобы пройти сквозь первую, или установить над плоскостью непроницаемый экран с «окошком» и использовать «сверхэффективность» (т. е. точность порядка $1/n$) крайних членов усеченной выборки (см. эксперимент по сравнению $2\overline{X}$ и $X_{(n)}$ для выборки из равномерного распределения на $[0, \theta]$ в § 1 гл. 6).

Тем не менее, рассмотренный пример поучителен тем, что такая «естественная» оценка центра симметрии распределения и «сгущения» наблюдаемых значений элементов выборки, как $\overline{X}$, оказывается несостоятельной, и поэтому требуются более сложные оценки.

§ 3. ВЫБОРОЧНЫЕ КВАНТИЛИ

Понятие теоретической медианы можно обобщить.

Определение. Пусть $\alpha \in (0, 1)$. Для непрерывной функции распределения F *теоретической* α*-квантилью* x_α называется решение уравнения $F(x) = \alpha$ (рис. 7).

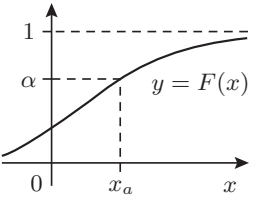


Рис. 7

Так же, как и в случае медианы ($\alpha = 1/2$), это решение может быть не единственным.

Оценить x_α можно с помощью порядковой статистики $X_{([\alpha n]+1)}$, где $[\,\cdot\,]$ обозначает целую часть числа. Эту оценку называют *выбо-*

рочной α-квантилью. Ее состоятельность вытекает из следующей теоремы.

Теорема 2. Пусть элементы выборки имеют плотность $p(x)$, причем $p(x_\alpha) > 0$ для заданного $\alpha \in (0,1)$. Тогда

$$\sqrt{n}\left(X_{([\alpha n]+1)} - x_\alpha\right) \xrightarrow{d} \xi \sim \mathcal{N}\left(0, \alpha(1-\alpha)/p^2(x_\alpha)\right) \quad \text{при } n \to \infty.$$

Для доказательства теоремы 2 потребуется

Лемма 1. Пусть $\varphi(\theta)$ — дифференцируемая функция, причем $\varphi'(\theta) \neq 0$. Если $\sqrt{n}\,(\widehat{\theta}_n - \theta) \xrightarrow{d} \xi \sim \mathcal{N}(0, \sigma^2)$ при $n \to \infty$, то

$$\sqrt{n}\,[\varphi(\widehat{\theta}_n) - \varphi(\theta)] \xrightarrow{d} \varphi'(\theta)\,\xi \sim \mathcal{N}\left(0, \sigma^2[\varphi'(\theta)]^2\right).$$

Пояснение. При $n \to \infty$ распределение оценки $\widehat{\theta}_n$ приближенно нормально и концентрируется около θ. Отображение φ в малой окрестности θ практически является линейным растяжением с коэффициентом $\varphi'(\theta)$ угла наклона касательной к графику $y = \varphi(x)$ в точке θ (рис. 8). Так как распределение оценки $\widehat{\theta}_n$ в основном сосредоточено на расстоянии порядка $1/\sqrt{n}$ от точки θ, то нормальность сохраняется (см. П9), а дисперсия умножается на коэффициент $[\varphi'(\theta)]^2$.

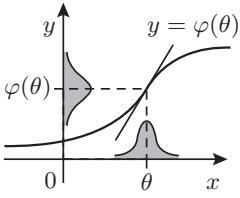


Рис. 8

Доказательство леммы 1. Разложим $\varphi(x)$ в точке θ по формуле Тейлора: $\varphi(\widehat{\theta}_n) - \varphi(\theta) = (\widehat{\theta}_n - \theta)\,[\varphi'(\theta) + \zeta_n]$, где для любого $\varepsilon > 0$ величина $|\zeta_n| < \varepsilon$ при $|\widehat{\theta}_n - \theta| < \delta(\varepsilon)$. Отсюда $\mathbf{P}(|\zeta_n| < \varepsilon) \geqslant \mathbf{P}(|\widehat{\theta}_n - \theta| < \delta) \to 1$ при $n \to \infty$, так как $\widehat{\theta}_n \xrightarrow{\mathsf{P}} \theta$ (см. вопрос 1). Поэтому $\zeta_n \xrightarrow{\mathsf{P}} 0$. С учетом свойства сходимости 1 из П5

$$\sqrt{n}\,[\varphi(\widehat{\theta}_n) - \varphi(\theta)] = [\sqrt{n}\,(\widehat{\theta}_n - \theta)]\,(\varphi'(\theta) + \zeta_n) \xrightarrow{d} \xi \cdot \varphi'(\theta). \qquad \blacksquare$$

Доказательство теоремы 2. Докажем ее сначала для выборки $(\eta_1, \eta_2, \ldots, \eta_n)$ из равномерного распределения на $[0,1]$. По лемме 3 гл. 4 порядковая статистика $\eta_{(k)} \sim S_k/S_{n+1}$, где $S_k = \tau_1 + \ldots + \tau_k$, τ_i — независимые показательные случайные величины с параметром $\lambda = 1$, $i = 1, \ldots, n+1$. Поэтому $\sqrt{n}\,(\eta_{(k)} - \alpha) \sim \sqrt{n}\,(S_k/S_{n+1} - \alpha)$. Проведем следующие простые преобразования:

$$\sqrt{n}\,(S_k/S_{n+1} - \alpha) = \sqrt{n}\,[(1-\alpha)\,S_k - \alpha\,(S_{n+1} - S_k)]/S_{n+1} =$$
$$= [n/S_{n+1}] \cdot [b_n Y_n - c_n Z_n + d_n],$$

где

$$b_n = (1-\alpha)\sqrt{k/n},\ c_n = \alpha\sqrt{(n+1-k)/n},\ d_n = (k - \alpha n - \alpha)/\sqrt{n},$$
$$Y_n = (S_k - k)/\sqrt{k},\ Z_n = [(S_{n+1} - S_k) - (n+1-k)]/\sqrt{n+1-k}\,.$$

В силу закона больших чисел (П6) $S_{n+1}/(n+1) \xrightarrow{\mathsf{P}} \mathbf{M}\tau_1 = 1$ при $n \to \infty$. Применяя свойства сходимости 1 и 3 из П5 для непрерывной при $x > 0$ функции $\varphi(x) = 1/x$, получаем, что $n/S_{n+1} \xrightarrow{\mathsf{P}} 1$.

Заметим, что случайные величины $\xi_n = b_n Y_n$ и $\zeta_n = c_n Z_n$ независимы, так как являются функциями от независимых векторов $(\tau_1,\ldots,\tau_k)$ и $(\tau_{k+1},\ldots,\tau_{n+1})$ соответственно (см. лемму из § 3 гл. 1). Поэтому характеристическая функция вектора (ξ_n,ζ_n) (см. П9) в силу свойства 5 математического ожидания из П2 имеет следующий вид:

$$\psi_{(\xi_n,\,\zeta_n)}(t_1,t_2) = \mathbf{M}e^{i(t_1\xi_n+t_2\zeta_n)} = \mathbf{M}\left[e^{i\,t_1\xi_n}\cdot e^{i\,t_2\zeta_n}\right] = \psi_{\xi_n}(t_1)\cdot\psi_{\zeta_n}(t_2).$$

Положим $k = k_n = [\alpha n]+1$. Тогда $b_n \to (1-\alpha)\sqrt{\alpha}$, $c_n \to \alpha\sqrt{1-\alpha}$. Согласно центральной предельной теореме (П6) распределение обеих случайных величин Y_n и Z_n стремится к $\mathcal{N}(0,1)$. Отсюда с учетом свойства сходимости 1 из П5 получаем, что

$$\xi_n \xrightarrow{d} \xi \sim \mathcal{N}(0,(1-\alpha)^2\alpha) \quad \text{и} \quad \zeta_n \xrightarrow{d} \zeta \sim \mathcal{N}(0,\alpha^2(1-\alpha)).$$

Применяя теорему непрерывности характеристической функции (П9), из приведенного выше представления для $\psi_{(\xi_n,\,\zeta_n)}(t_1,t_2)$ выводим, что $(\xi_n,\zeta_n) \xrightarrow{d} (\xi,\zeta)$, где (ξ,ζ) — нормальный вектор с независимыми компонентами. Свойство 3 из П5 (для непрерывной функции $\varphi(x,y) = x-y$) обеспечивает сходимость $\xi_n - \zeta_n \xrightarrow{d} \xi - \zeta$. Здесь предельная величина $\xi-\zeta$, являясь линейной комбинацией компонент нормального вектора, также имеет нормальное распределение, причем $\mathbf{M}(\xi-\zeta) = 0$ и $\mathbf{D}(\xi-\zeta) = \mathbf{D}\xi + \mathbf{D}\zeta = \alpha(1-\alpha)$ ввиду независимости случайных величин ξ и ζ (П2).

Чтобы установить сходимость распределения $\sqrt{n}\,(\eta_{([\alpha n]+1)} - \alpha)$ к $\mathcal{N}(0,\alpha(1-\alpha))$, остается заметить, что $d_n \to 0$ при $k = [\alpha n]+1$, и воспользоваться свойством 1 из П5.

Для выборки $(X_1,\ldots,X_n)$ из закона F с плотностью $p(x)$ в силу метода обратной функции (см. § 1 гл. 4) порядковая статистика $X_{(k)}$ распределена как $F^{-1}(\eta_{(k)})$. Производная обратной функции

$$\frac{d}{d\alpha}\,F^{-1}(\alpha) = \frac{1}{p(F^{-1}(\alpha))} = \frac{1}{p(x_\alpha)}\,.$$

Применение леммы 1 завершает доказательство теоремы 2. ■

§ 4. ОТНОСИТЕЛЬНАЯ ЭФФЕКТИВНОСТЬ

В качестве оценок мы использовали различные функции от выборки: $X_{(n)} = \max\{X_1,\ldots,X_n\}$, $\overline{X}$, *MED*.

Определение. *Статистикой* T будем называть произвольную *борелевскую* (см. П2) функцию от выборки $(X_1,\ldots,X_n)$.

Любые непрерывные функции n переменных являются борелевскими. Функции, полученные в результате арифметических операций над борелевскими, а также их суперпозиций и предельного перехода, снова будут борелевскими.

Определение. Статистика $T = T_n$ называется *асимптотически нормальной*, если найдутся такие числовые последовательности a_n и $b_n > 0$, что

$$(T_n - a_n)/b_n \xrightarrow{d} Z \sim \mathcal{N}(0,1) \quad \text{при } n\to\infty.$$

Пример 2. ***Гипотеза случайности*** ([32, с. 133])**.** Такой гипотезой называют предположение о том, что данные $(x_1,\ldots,x_n)$ — это реализация выборки, т. е. случайного вектора $(X_1,\ldots,X_n)$ с независимыми и одинаково распределенными компонентами. Допустим, что X_i имеют непрерывную функцию распределения. Тогда из соображений симметрии все $n!$ вариантов расположений X_i относительно друг друга равновероятны. Одной из статистик, измеряющих степень «беспорядка», является R_n — *общее количество инверсий* в выборке: говорят, что X_i и X_j образуют *инверсию*, если $i<j$, но $X_i>X_j$. Крайние случаи, когда $X_1<\ldots<X_n$ $(R_n=0)$ и $X_1>\ldots>X_n$ $(R_n=(n-1)+(n-2)+\ldots+1=n(n-1)/2)$ естественно рассматривать как свидетельства «полного отсутствия беспорядка». Слишком малые или слишком близкие к числу $n(n-1)/2$ значения статистики R_n служат основанием для того, чтобы отвергнуть гипотезу случайности.

Известно, что статистика R_n асимптотически нормальна:

$$(R_n-\mathbf{M}R_n)\Big/\sqrt{\mathbf{D}R_n}\xrightarrow{d}Z\sim\mathcal{N}(0,1)\quad\text{при } n\to\infty,$$

где $\mathbf{M}R_n=n(n-1)/4$, $\mathbf{D}R_n=n(n-1)(2n+5)/72$ (см. [81, с. 271]). Это позволяет при достаточно больших n проверить гипотезу случайности, например, по «правилу трех сигм» (см. § 2 гл. 3).

Для асимптотически нормальных оценок параметров в *регулярных* (см. § 3 гл. 9) статистических моделях типичным порядком малости коэффициента b_n является $1/\sqrt{n}$. Условие асимптотической нормальности для них представляется в следующем виде:

$$\sqrt{n}\,(\widehat{\theta}_n-\theta)\xrightarrow{d}\xi\sim\mathcal{N}(0,\sigma^2(\theta))\quad\text{при } n\to\infty.$$

Определение. Величина $\sigma^2(\theta)>0$ называется *асимптотической дисперсией* асимптотически нормальной оценки $\widehat{\theta}_n$.

Например, если $0<\mathbf{D}X_1<\infty$, то согласно центральной предельной теореме (П6)

$$\sqrt{n}\,(\overline{X}-\mathbf{M}X_1)\xrightarrow{d}\xi\sim\mathcal{N}(0,\mathbf{D}X_1)\quad\text{при } n\to\infty,\tag{1}$$

т. е. асимптотической дисперсией выборочного среднего $\overline{X}$ (как оценки для $\mathbf{M}X_1$) служит $\mathbf{D}X_1$. Теоремы 1 и 2 дают еще два примера асимптотически нормальных оценок.

Замечание. Вообще говоря, асимптотическая дисперсия $\sigma^2(\theta)$ может не совпадать с пределом при $n\to\infty$ последовательности $c_n=\mathbf{D}(\sqrt{n}\,(\widehat{\theta}_n-\theta))=n\,\mathbf{D}\widehat{\theta}_n$ по той причине, что из сходимости распределений не следует сходимость моментов (см. П5).

Так, условия теоремы 2 выполняются для выборки с функцией распределения $F(x)=1-(1/\ln x)$ при $x>e$ (и плотностью $p(x)=(x\ln^2 x)^{-1}I_{\{x>e\}}$) (это распределение встречалось ранее

в § 2 гл. 4). Легко видеть, что $\mathbf{M}X_1 = \infty$. Покажем, что также бесконечны и математические ожидания порядковых статистик $X_{(k)}$ для всех $k = 1, \ldots, n$.

Действительно, метод обратной функции показывает, что $X_{(k)}$ распределена так же, как $F^{-1}(\eta_{(k)})$. Согласно формуле (2) гл. 5 плотностью $\eta_{(k)}$ является $p_{\eta_{(k)}}(x) = nC_{n-1}^{k-1}\, x^{k-1}(1-x)^{n-k} I_{\{0<x<1\}}$. Применяя формулу преобразования с якобианом $J = F'(x) = p(x)$ из П8, находим, что

$$p_{X_{(k)}}(x) = nC_{n-1}^{k-1}\, F(x)^{k-1}(1-F(x))^{n-k} p(x). \qquad (2)$$

Наконец, заметим, что при любом k (проверьте второе равенство!)

$$\mathbf{M}X_{(k)} \geqslant \mathbf{M}X_{(1)} = \int\limits_{-\infty}^{\infty} x\, p_{X_{(1)}}(x)\, dx = n \int\limits_{e}^{\infty} \left(\frac{1}{\ln x}\right)^{n+1} dx = \infty.$$

Вопрос 2. При каких размерах выборки будет конечна дисперсия *MED* для распределения Коши? (Используйте то, что для закона Коши $x(1-F(x)) \to 1/\pi$ при $x \to +\infty$.)

Асимптотическая дисперсия $\sigma^2(\theta)$ характеризует точность асимптотически нормальной оценки, вычисленной по большой выборке.

Определение. *Относительной асимптотической эффективностью* асимптотически нормальной оценки $\widehat{\theta}_1$ по отношению к асимптотически нормальной оценке $\widehat{\theta}_2$ называется величина $e_{\widehat{\theta}_1,\widehat{\theta}_2} = \sigma_2^2/\sigma_1^2$.

Почему относительная эффективность $e_{\widehat{\theta}_1,\widehat{\theta}_2}$ определяется как σ_2^2/σ_1^2, а не как σ_2/σ_1 или σ_2^4/σ_1^4? Пусть требуется оценить параметр θ с заданной точностью δ, причем за каждое наблюдение x_i мы должны заплатить цену C. Тогда размеры выборок n_1 и n_2, обеспечивающие заданную точность для оценок $\widehat{\theta}_1$ и $\widehat{\theta}_2$ соответственно, определяются из соотношения $\delta = \sigma_1/\sqrt{n_1} = \sigma_2/\sqrt{n_2}$. Таким образом,

$$e_{\widehat{\theta}_1,\widehat{\theta}_2} = \sigma_2^2/\sigma_1^2 = n_2/n_1 = (n_2 C)/(n_1 C),$$

т. е. относительная эффективность $e_{\widehat{\theta}_1,\widehat{\theta}_2}$ представляет собой *отношение затрат* при использовании оценки $\widehat{\theta}_2$ к затратам при использовании оценки $\widehat{\theta}_1$.

Примеры вычисления $e_{\widehat{\theta}_1,\widehat{\theta}_2}$ для некоторых распределений приведены в задачах 1–3 и ряде задач следующей главы.

§ 5. УСТОЙЧИВЫЕ ЗАКОНЫ

На все свои законы есть.
Фамусов в «Горе от ума» А. С. Грибоедова

В определении асимптотической нормальности участвуют константы: центрирующие a_n и масштабирующие $b_n > 0$. В связи с этим возникает следующий вопрос: нельзя ли подобрать другие a'_n и $b'_n > 0$ такие, что $(T_n - a'_n)/b'_n$ сходилось бы к невырожденному*) закону, отличному от нормального? Следующая лемма дает отрицательный ответ.

*) Распределение случайной величины ξ *вырождено*, если $\mathbf{P}(\xi = \text{const}) = 1$.

Лемма 2. Пусть $(T_n - a_n)/b_n \xrightarrow{d} \xi$ и $(T_n - a'_n)/b'_n \xrightarrow{d} \xi'$, причем обе случайные величины ξ и ξ' имеют невырожденное распределение. Тогда существуют такие константы a и $b > 0$, что $b'_n/b_n \to b$, $(a'_n - a_n)/b_n \to a$ и $\xi' \sim a + b\xi$ (т. е. невырожденный предельный закон определяется *однозначно* с точностью до преобразований сдвига и растяжения).

Доказательство приведено в [90, с. 371].

В § 2 гл. 4 в качестве предельных законов для порядковых статистик $(X_{(n)} - a_n)/b_n$, где $X_{(n)} = \max\{X_1, \ldots, X_n\}$, возникали так называемые *распределения экстремальных значений*. Оказывается, и для $S_n = X_1 + \ldots + X_n$ (сумм независимых и одинаково распределенных случайных величин) можно полностью описать класс предельных законов для статистик $(S_n - a_n)/b_n$. (Условия сходимости приведены, например, в [82, с. 643].) Такие законы (и только они) обладают свойством устойчивости.

Определение. (Невырожденное) распределение F *устойчиво*, если для любых a_1, $b_1 > 0$ и a_2, $b_2 > 0$ найдутся a и $b > 0$ такие, что $F(a_1 + b_1 x) * F(a_2 + b_2 x) = F(a + bx)$.

Здесь $*$ означает свертку функций распределения (см. П3).

Другими словами, при сложении независимых случайных величин, имеющих устойчивое распределение, получается снова тот же закон, но, вообще говоря, с другими параметрами сдвига и масштаба. Нормальный закон устойчив: если $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ и $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ независимы, то $X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

П. Леви (1886–1971), французский математик.

А. Я. Хинчин (1894–1959), советский математик.

Согласно теореме Леви — Хинчина характеристические функции устойчивых законов (с точностью до сдвига и масштаба) допускают следующее представление:

$$\psi(t) = \exp\{-|t|^\alpha (1 + i\beta\, G(t,\alpha) \operatorname{sign} t)\}, \tag{3}$$

где α и β — постоянные, $0 < \alpha \leqslant 2$, $-1 \leqslant \beta \leqslant 1$,

$$G(t,\alpha) = \begin{cases} \dfrac{2}{\pi} \ln |t|, & \text{если } \alpha = 1, \\ \operatorname{tg} \dfrac{\pi}{2}\alpha, & \text{если } \alpha \neq 1. \end{cases}$$

Соответствующие распределения имеют непрерывные плотности, которые вычисляются по формуле обратного преобразования Фурье из П9. Явный вид плотностей известен для нормального закона ($\alpha = 2$, $\beta = 0$), распределения Коши ($\alpha = 1$, $\beta = 0$) и законов с $\alpha = 1/2$, $\beta = \pm 1$:

$$p^+(x) = \frac{1}{\sqrt{2\pi}}\, x^{-3/2} e^{-1/(2x)} I_{\{x>0\}}, \quad p^-(x) = p^+(-x).$$

Легко проверить, что функция распределения $F^+(x)$ закона с плотностью $p^+(x)$ удовлетворяет соотношению $F^+(x) = 2\,(1 - \Phi(1/\sqrt{x}))$

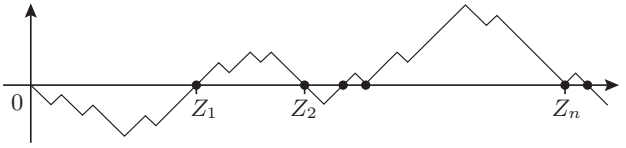


Рис. 9

при $x > 0$, где $\Phi(x)$ — функция распределения случайной величины $Z \sim \mathcal{N}(0,1)$. Иначе говоря, случайная величина $1/Z^2$ имеет функцию распределения $F^+(x)$.

Закон $F^+(x)$ возникает в качестве предельного в задаче о частоте возвращений в 0 *симметричного случайного блуждания* $S_k = X_1 + \ldots + X_k$, где X_i независимы, $\mathbf{P}(X_i = -1) = \mathbf{P}(X_i = 1) = 1/2$, $i = 1,2,\ldots$. Обозначим через Z_n время до n-го возвращения в 0 (рис. 9).

Тогда Z_n — сумма времен между последовательными возвращениями. Интуитивно понятно, что эти времена независимы и одинаково распределены. Поэтому предельный закон для Z_n должен быть устойчивым. В [82, с. 492] доказано, что $\mathbf{P}(Z_n/n^2 \leqslant x) \to F^+(x)$ при $n \to \infty$. Иными словами, число возвращений в 0 растет не пропорционально количеству шагов n, а как $\sqrt{n}$. Это связано с тем, что $\mathbf{M}Z_1 = \infty$, и закон больших чисел (П6) не применим к сумме величин с таким распределением.

Когда ж пространствуешь, воротишься домой...

Чацкий в «Горе от ума» А. С. Грибоедова

Вопрос 3.
а) Какую характеристическую функцию имеет разность двух независимых случайных величин с функцией распределения $F^+(x)$?
б) Будет ли соответствующий закон устойчивым?
в) Как ведет себя $\overline{X}$ для такого закона при возрастании размера выборки n?
(Используйте свойства характеристической функции из П9.)

Пример 3. ***Гравитационное поле звезд.*** Представим, что в шаре радиуса r с центром в начале координат расположены наудачу n звезд единичной массы. Обозначим через $X_1, \ldots, X_n$ x-компоненты гравитационных сил, создаваемых в центре шара отдельными звездами. Положим $S_n = X_1 + \ldots + X_n$. Тогда при таком стремлении $r \to \infty$ и $n \to \infty$, что $\frac{4}{3}\pi r^3/n \to \lambda$, распределение случайной величины S_n стремится (с точностью до масштаба) к устойчивому закону с $\alpha = 3/2$, $\beta = 0$ (так называемому *распределению Хольцмарка*) (см. [82, с. 252]).

ЗАДАЧИ

1. Вычислите $e_{MED,\overline{X}}$ для выборки из закона $\mathcal{N}(\theta, 1)$.

2. Пусть случайные величины X и Y независимы и показательно распределены с $\lambda = 1$.

а) Докажите, что разность $X - Y$ имеет распределение с плотностью $p(x) = \frac{1}{2}\, e^{-|x|}$ (*закон Лапласа*).

б) Вычислите $\mathbf{D}(X - Y)$, используя свойства дисперсии из П2.

в) Для выборки из сдвинутого на θ закона Лапласа найдите $e_{MED,\overline{X}}$.

Первое — это понять правило, второе — научиться его применять. Первое достигается разумом и сразу, второе — опытом и постепенно.

Артур Шопенгауэр, «Афоризмы житейской мудрости»

3*. Элементы выборки распределены согласно закону $F(x-\theta)$, где

$$F(x) = (1-\varepsilon)\,\Phi(x) + \varepsilon\,\Phi(x/3),$$

$0 \leqslant \varepsilon \leqslant 1$, $\Phi(x)$ — функция распределения $\mathcal{N}(0,1)$. Другими словами, $F(x)$ — это *смесь* (см. § 2 гл. 5) с весами $1-\varepsilon$ и ε законов $\mathcal{N}(0,1)$ и $\mathcal{N}(0,9)$ соответственно. Определите приближенно значение ε, при котором *MED* становится эффективней $\overline{X}$.

Указание. Линеаризуйте функцию $e_{MED,\overline{X}}(\varepsilon)$ в окрестности 0.

4. **а)** Найдите значение a, при котором достигается минимум функции $g(a) = \mathbf{M}|\xi - a|$ для случайной величины ξ с плотностью $p(x)$ (ср. с задачей 1 гл. 1).

б) Определим в общем случае *медиану* m распределения случайной величины ξ как любое число, удовлетворяющее неравенствам $\mathbf{P}(\xi \leqslant m) \geqslant 1/2$ и $\mathbf{P}(\xi \geqslant m) \geqslant 1/2$. Пусть $\mathbf{M}|\xi| < \infty$. Докажите, что функция $g(a) = \mathbf{M}|\xi - a|$ имеет минимум при $a = m$.

Указание. Установите, что $g(a) - g(m) \geqslant 0$, если $a \geqslant m$.

5*. Выборка размера n получена из распределения

$$F_\theta(x) = 1 - (1 - x/\theta)^\alpha \text{ при } 0 \leqslant x \leqslant \theta,$$

где $\alpha > 0$ — известный параметр. Каков порядок малости величины $\theta - X_{(n)}$ при $n \to \infty$? Сколько (с точностью до порядка) наблюдений потребуется, чтобы оценить $\theta = 1$ с погрешностью 0,1 при $\alpha = 5$?

6. Докажите, что для максимума выборки из закона Коши $\mathbf{P}(\pi X_{(n)}/n \leqslant x) \to e^{-1/x} I_{\{x>0\}}$ при $n \to \infty$ (т. е. предельным является распределение экстремальных значений II-го типа с $\alpha = 1$, появившееся в § 2 гл. 4).

Указание. Примените правило Лопиталя.

7*. Пусть в эксперименте по локализации источника излучения из § 1 $a = 0$, $b = 0$, $c = 1$. Будут ли независимыми

а) полярные координаты R и Φ точки (X_1, Y_1),

б) сами X_1 и Y_1?

Указание. Перейдите к полярным координатам (П8) и используйте формулу площади поверхности вращения.

РЕШЕНИЯ ЗАДАЧ

1. Медиана $x_{1/2} = \theta$ в силу симметрии закона $\mathcal{N}(\theta,1)$, причем $p(\theta) = 1/\sqrt{2\pi}$. Теорема 1 дает $\sigma^2_{MED} = 1/[4\,p^2(\theta)] = \pi/2$. Согласно формуле (1) имеем $\sigma^2_{\overline{X}} = \mathbf{D}X_1 = 1$. Отсюда $e_{MED,\overline{X}} = 2/\pi \approx 0{,}64$. Таким образом, для нормального закона оценка $\overline{X}$ эффективней *MED* примерно на 36%.

2. **а)** Так как $X - Y \sim Y - X = -(X-Y)$, в силу следствия из П8 $p_{X-Y}(x) = p_{Y-X}(x) = p_{X-Y}(-x)$, т. е. эта плотность — четная

функция. Вычислим ее при $x \geqslant 0$. Ввиду того, что $p_{-Y}(y) = p_Y(-y) = e^y I_{\{y\leqslant 0\}}$, используя формулу свертки (П3), запишем

$$p_{X-Y}(x) = \int\limits_{-\infty}^{\infty} e^{-(x-y)} I_{\{y\leqslant x\}}\, e^y I_{\{y\leqslant 0\}}\, dy = e^{-x} \int\limits_{-\infty}^{0} e^{2y} dy = \frac{1}{2}\, e^{-x}.$$

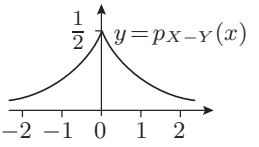


Рис. 10

График плотности закона Лапласа приведен на рис. 10.

б) Из ответа на вопрос 3 гл. 4 имеем $\mathbf{D}X = 1/\lambda^2 = 1$. Согласно лемме из § 3 гл. 1 случайные величины X и $(-Y)$ независимы. Применяя свойства дисперсии 2 и 1 из П2, находим: $\mathbf{D}(X-Y) = \mathbf{D}X + \mathbf{D}(-Y) = \mathbf{D}X + \mathbf{D}Y = 2$.

в) В силу симметрии $x_{1/2} = \theta$, причем $p(\theta) = 1/2$. Теорема 1 дает $\sigma^2_{MED} = 1/[4\,p^2(\theta)] = 1$. С учетом пункта б) из (1) получаем, что $\sigma^2_{\overline{X}} = e_{MED,\overline{X}} = 2$. Таким образом, выборочная медиана MED вдвое эффективней выборочного среднего $\overline{X}$ как оценка параметра сдвига закона Лапласа.

3. $\mathbf{M}X_1 = x_{1/2} = \theta$ ввиду симметрии функции распределения $F(x)$, причем $p(\theta) = \dfrac{1}{\sqrt{2\pi}}\,(1 - \varepsilon + \varepsilon/3)$. Дисперсия случайной величины X_1 — смесь вторых моментов: $\mathbf{D}X_1 = \int x^2\, dF(x) = (1-\varepsilon)\int x^2\, d\Phi(x) + \varepsilon \int x^2\, d\Phi\Big(\dfrac{x}{3}\Big) = 1 - \varepsilon + 9\varepsilon = 1 + 8\varepsilon$. Отсюда в силу теоремы 1 и формулы (1) имеем:

$$e(\varepsilon) \equiv e_{MED,\overline{X}} = \frac{2}{\pi}\,(1+8\varepsilon)\left(1 - \frac{2}{3}\,\varepsilon\right)^2 = \frac{2}{\pi}\left(1 + \frac{20}{3}\,\varepsilon - \frac{92}{9}\,\varepsilon^2 + \frac{32}{9}\,\varepsilon^3\right).$$

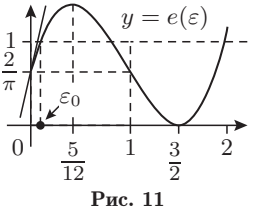


Рис. 11

График этого многочлена приведен на рис. 11. Ближний к 0 корень ε_0 уравнения $e(\varepsilon) = 1$ можно приближенно найти из равенства $\dfrac{2}{\pi}\left(1 + \dfrac{20}{3}\,\varepsilon\right) = 1$ (функция заменяется на касательную в нуле). Получаем $\varepsilon_0 \approx 0{,}1$. Отметим также, что преимущество выборочной медианы MED максимально при $\varepsilon = 5/12$ и составляет около 44%.

4. **а)** По формуле замены переменных из П2 запишем функцию $g(a) = \mathbf{M}|\xi - a|$ в виде

$$g(a) = \int\limits_{-\infty}^{a} (a-x)\, dF(x) + \int\limits_{a}^{+\infty} (x-a)\, dF(x) = 2a\,F(a) - 2\int\limits_{-\infty}^{a} x\,p(x)\,dx - a + \mathbf{M}\xi.$$

Дифференцируя правую часть по a, приходим к равенству $g'(a) = 2F(a) - 1$. Поэтому $g(a)$ имеет минимум при $x_{1/2}$.

б) Так как $\mathbf{M}|\xi - a| \leqslant \mathbf{M}|\xi| + |a| < \infty$, то функция $g(a)$ определена при всех a. Используем для записи приращения $\Delta = g(a) - g(m)$ формулу замены переменных из П2:

$$\Delta = (a-m)F(m) + \int_m^a (a+m-2x)\,dF(x) - \int_a^{+\infty} (a-m)\,dF(x).$$

Добавляя и вычитая интеграл $\int_m^a (a-m)\,dF(x)$, получаем:

$$\Delta = (a-m)\,[2F(m)-1] + 2\int_m^a (a-x)\,dF(x).$$

Оба слагаемых в правой части неотрицательны: первое — в силу определения медианы m, второе — из неравенства $a \geqslant m$ и неотрицательности интегрируемой функции на области интегрирования. Случай $a \leqslant m$ рассматривается аналогично.

5. Найдем δ_n, убывающие к нулю при $n \to \infty$, из условия, чтобы вероятность $\mathbf{P}(\theta - X_{(n)} \leqslant \delta_n)$ сходилась к пределу, отличному от 0 и 1. Из задачи 3 гл. 1 имеем $\mathbf{P}(X_{(n)} \leqslant x) = [F_\theta(x)]^n$. Следовательно,

$$\mathbf{P}(X_{(n)} \geqslant \theta - \delta_n) = 1 - \left[1 - \left(1 - \frac{\theta - \delta_n}{\theta}\right)^{\alpha}\right]^n = 1 - \left[1 - \left(\frac{\delta_n}{\theta}\right)^{\alpha}\right]^n.$$

Поэтому величина $(\delta_n/\theta)^\alpha$ должна убывать со скоростью $1/n$. Это влечет для δ_n порядок малости $n^{-1/\alpha}$. В частности, для оценивания θ с точностью $\delta_n = 0{,}1$ при $\alpha = 5$ потребуется примерно 10^5 наблюдений.

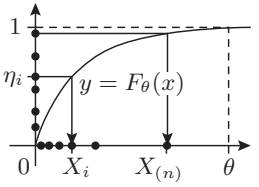


Рис. 12

Причина столь большого значения необходимого размера выборки кроется в гладкости при $\alpha > 1$ функции распределения $F_\theta(x)$ в точке $x = \theta$. В соответствии с методом обратной функции (см. § 1 гл. 4) для близости $X_{(n)}$ к θ нужна «сверхблизость» к 1 одной из координат η_i точек, взятых наудачу из отрезка $[0, 1]$ (рис. 12).

Хотя $X_{(n)}$ в этой модели и *не является* асимптотически нормальной оценкой (это вытекает из теоремы 1 гл. 4 и леммы 2 текущей главы), она, по крайней мере, очевидно, состоятельна при любом $\alpha > 0$.

На безрыбьи и рак — рыба.

6. Применим для вычисления $\lim\limits_{x \to +\infty} x[1 - F(x)]$ правило Лопиталя (см. [45, с. 284]):

Г. Лопиталь (1661–1704), французский математик.

$$\lim_{x\to+\infty} \frac{1-F(x)}{1/x} = \lim_{x\to+\infty} \frac{-p(x)}{-1/x^2} = \lim_{x\to+\infty} \frac{1/[\pi(1+x^2)]}{1/x^2} = \frac{1}{\pi}\,.$$

Чтобы установить для $x > 0$ сходимость $\mathbf{P}(\pi X_{(n)}/n \leqslant x) \to$ $\to e^{-1/x}$ при $n \to \infty$, достаточно сослаться на теорему 1 гл. 4.

Вывод:. максимум $X_{(n)}$ выборки из закона Коши имеет порядок n при $n \to \infty$.

7. Без ограничения общности можно считать, что сфера имеет радиус 1. Вычислим $F_{R,\Phi}(r,\varphi) = \mathbf{P}(R \leqslant r, \Phi \leqslant \varphi)$. В силу симметрии она, очевидно, равна $\dfrac{\varphi}{2\pi}\, S_h/S_1$. Здесь S_h — площадь поверхности шарового сегмента («шапочки»), отсекаемого плоскостью $z = h$, где $h = h_r$ определяется из пропорции $(1-h) : 1 = 1 : \sqrt{1+r^2}$ (рис. 13). S_h можно вычислить по формуле площади поверхности вращения, образованной дугой функции $f(x) = \sqrt{1-x^2}$ (проверьте второе равенство!):

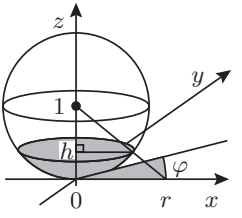


Рис. 13

$$S_h = 2\pi \int\limits_{1-h}^{1} f(x)\sqrt{1+[f'(x)]^2}\,dx = 2\pi h. \qquad (4)$$

Поясним происхождение формулы для вычисления площади поверхности вращения (подробнее см. [45, с. 652]). Разобьем отрезок $[a,b]$, на котором задана вращаемая дуга $y = f(x)$, на части длины Δx_i (рис. 14). При малых Δx_i общая площадь поверхности вращения приближенно равна сумме площадей поверхностей усеченных конусов, получаемых вращением вокруг оси абсцисс хорд длины

$$\Delta l_i = \sqrt{(\Delta x_i)^2 + (\Delta y_i)^2}.$$

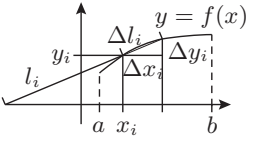


Рис. 14

Для неусеченного конуса с образующей длины l_i и радиусом основания y_i (см. рис. 14) площадь поверхности $S_i = \pi y_i l_i$. Действительно, эту поверхность можно развернуть в сектор круга радиуса l_i и длиной дуги $2\pi y_i$ (рис. 15). Тогда S_i находится из пропорции

$$2\pi y_i : 2\pi l_i = S_i : \pi l_i^2.$$

Следовательно, для усеченного конуса, образованного хордой длины l_i, площадь равна

$$\pi\,[(y_i + \Delta y_i)(l_i + \Delta l_i) - y_i l_i] = \pi\,[y_i \Delta l_i + l_i \Delta y_i + \Delta y_i \Delta l_i]. \qquad (5)$$

Подобие прямоугольных треугольников на рис. 14 влечет пропорцию

$$\Delta y_i : \Delta l_i = y_i : l_i \iff \Delta y_i = (y_i/l_i)\,\Delta l_i.$$

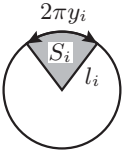


Рис. 15

Подставив выражение для Δy_i в (5), получим

$$2\pi y_i\,\Delta l_i + (y_i/l_i)\,(\Delta l_i)^2 = 2\pi y_i\,\Delta l_i + o(\Delta l_i).$$

Отсюда видим, что главная часть интегральной суммы есть

$$2\pi \sum y_i\,\Delta l_i = 2\pi \sum y_i \sqrt{1 + (\Delta y_i/\Delta x_i)^2}\,\Delta x_i.$$

При измельчении разбиения она стремится к $2\pi \int\limits_a^b y\sqrt{1+(y')^2}\,dx$.

Вопрос 4.
Как вычислить объем тела вращения?

Из (4) следует, что $F_{R,\Phi}(r,\varphi) = \dfrac{\varphi}{2\pi}\, h = \dfrac{\varphi}{2\pi}\left(1 - \dfrac{1}{\sqrt{1+r^2}}\right)$. Поэтому R и Φ независимы. Плотность $p_{R,\Phi}(r,\varphi) = \dfrac{1}{2\pi}\, r(1+r^2)^{-3/2}$ получается при замене координат на полярные из функции $p_{X_1,Y_1}(x,y) = \dfrac{1}{2\pi}\,(1+x^2+y^2)^{-3/2}$ (*двумерная плотность Коши*). Интегрированием по y (см. П8) находим маргинальную плотность

$$p_{X_1}(x) = \frac{1}{2\pi}\int\limits_{-\infty}^{+\infty} \frac{dy}{(1+x^2+y^2)^{3/2}} = \frac{1}{2\pi}\,\frac{1}{1+x^2}\,\frac{y}{\sqrt{1+x^2+y^2}}\,\Bigg|_{-\infty}^{+\infty} = \frac{1}{\pi}\,\frac{1}{1+x^2}.$$

Таким образом, случайные величины X_1 и Y_1 имеют распределение Коши, но они зависимы, так как $p_{X_1,Y_1}(x,y) \neq p_{X_1}(x)\, p_{Y_1}(y)$.

ОТВЕТЫ НА ВОПРОСЫ

1. Представим $(MED - x_{1/2})$ в виде $\beta_n \xi_n$, где $\beta_n = \dfrac{1}{\sqrt{n}} \xrightarrow{\text{п. н.}} 0$ и $\xi_n = \sqrt{n}\,(MED - x_{1/2}) \xrightarrow{d} \xi \sim \mathcal{N}\left(0, 1/[4\,p^2(x_{1/2})]\right)$ при $n \to \infty$. Тогда $MED \xrightarrow{\mathsf{P}} x_{1/2}$ в силу свойств сходимости 1 и 2 из П5.

2. Дисперсия порядковой статистики $X_{(k)}$ конечна тогда и только тогда, когда $\mathbf{M}X_{(k)}^2 < \infty$. Так как для закона Коши $x^2\, p(x)\, F^{k-1}(x) \to 1$ при $x \to +\infty$, то с учетом формулы (2) заключаем, что $\int x^2\, p_{X_{(k)}}(x)\, dx$ сходится и расходится одновременно с интегралом $\int [1 - F(x)]^{n-k} dx$. Используя указание, выводим отсюда, что $\mathbf{M}X_{(k)}^2 < \infty$ при $3 \leqslant k \leqslant n-2$. Следовательно, дисперсия выборочной медианы MED конечна для $n \geqslant 5$.

3. **а)** По формуле (3) $\psi^+(t) = \exp\{-\sqrt{|t|}\,(1 + i\,\mathrm{sign}\,t)\}$. Согласно свойствам характеристической функции (П9) разность таких независимых случайных величин имеет характеристическую функцию $\psi^+(t)\,\psi^+(-t) = e^{-2\sqrt{|t|}}$.

б) С точностью до масштабного коэффициента эта функция совпадает с характеристической функцией устойчивого закона при $\alpha = 1/2$ и $\beta = 0$.

в) Очевидно, характеристическая функция суммы $X_1 + \ldots + X_n$ равна $e^{-2n\sqrt{|t|}}$, откуда $\overline{X}$ имеет характеристическую функцию $e^{-2\sqrt{|nt|}}$, т. е. распределение $\overline{X}$ растянуто в n раз по сравнению с распределением X_1.

Бывает, что усердие превозмогает и рассудок.

Козьма Прутков

4. Разрезая тело вращения на слои толщины Δx_i, получаем интегральную сумму $\sum \pi y_i^2\, \Delta x_i$, которая сходится к $\pi \int\limits_a^b y^2 dx$.

Глава 8

СИММЕТРИЧНЫЕ РАСПРЕДЕЛЕНИЯ

§ 1. КЛАССИФИКАЦИЯ МЕТОДОВ СТАТИСТИКИ

Симметрия является той идеей, посредством которой человек на протяжении веков пытался постичь и создать порядок, красоту и совершенство.

Г. Вейль

Нам нравится смотреть на проявление симметрии в природе, на идеально симметричные сферы планет или Солнца, на симметричные кристаллы, на снежинки, наконец, на цветы, которые почти симметричны.

Р. Фейнман

Согласно одному из подходов к классификации статистических методов, их можно условно разделить на **три класса** (см. [84, с.21]): параметрические, робастные и непараметрические.

Первые, как правило, обладают максимальной эффективностью в рамках заданной модели $F_\theta(x)$, $\theta \in \Theta \subseteq \mathbb{R}^m$, т. е. на некоторой *$m$-параметрической кривой* в пространстве всех функций распределения (рис. 1а).

Так (см. пример 4 гл. 9), $\overline{X}$ — эффективная оценка параметра сдвига μ закона $\mathcal{N}(\mu, \sigma^2)$. Здесь $m = 2$, $\boldsymbol{\theta} = (\mu, \sigma)$, $\Theta = \mathbb{R} \times (0{,}\infty)$. Однако, она весьма чувствительна к утяжелению «хвостов» распределения, приводящему к появлению в реализации выборки выделяющихся наблюдений («выбросов»).

Robust (*англ.*) — крепкий, надежный, устойчивый.

Оценки, которые обладают высокой эффективностью для заданной параметрической модели, и, кроме того, не боятся небольших отклонений от нее, т. е. достаточно точны в некоторой окрестности m-параметрической кривой (рис. 1б), называются *робастными* (см. § 4).

Наконец, *непараметрические* методы успешно работают на целом классе законов распределения (рис. 1в), скажем, на множестве Ω_c всех *непрерывных* функций распределения. В этой главе мы обсудим поведение ряда оценок параметра сдвига на классе Ω_s *симметричных гладких распределений*.

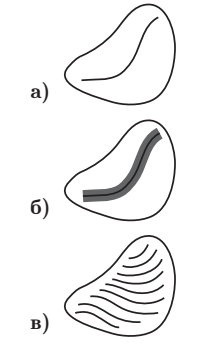


Рис. 1

Определение. Функция распределения F *принадлежит классу* Ω_s, если существует такое c: $0 < c \leqslant +\infty$, что $F(-c) = 0$, $F(c) = 1$ и $F(x)$ на $(-c, c)$ имеет четную, непрерывную и положительную плотность $p(x)$ (рис. 2).

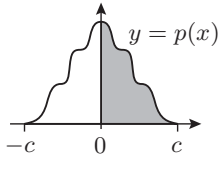


Рис. 2

Обратим внимание на то, что распределение с плотностью $\tilde{p}(x)$, график которой приведен на рис. 6 гл. 7, *не входит* в класс Ω_s, так как носитель*) этой плотности не является интервалом.

*) *Носитель* — множество, на котором плотность положительна.

Хотя условие симметричности распределения может показаться искусственным и редко выполняющимся в точности на практике, бывают задачи, где оно возникает довольно естественно.

Пример 1. ***Контроль и обработка.*** Пусть имеются два ряда наблюдений: $X_1, \ldots, X_n$ (так называемая «контрольная» выборка) и $Y_1, \ldots, Y_n$ («обработка»). Это могут быть, скажем, размеры растений на двух грядках, на второй из которых применялся определенный вид удобрений, а на первой — нет. Нас интересует, есть ли эффект, т. е. значимое увеличение размера растений, от применения удобрения.

Рассмотрим следующую статистическую модель: $X_i = \mu + \varepsilon_i$, $Y_i = \mu + \theta + \varepsilon_i'$, где μ — средний размер растений, θ — увеличение размера за счет удобрения, ε_i и ε_i' — случайные величины, включающие в себя влияние неучтенных факторов на размер конкретного растения. Допустим, что $\mathbf{M}\varepsilon_i = \mathbf{M}\varepsilon_i' = 0$, случайные величины $\{\varepsilon_i,\ \varepsilon_i',\ i = 1, \ldots, n\}$ независимы и одинаково распределены с непрерывной и ограниченной на своем носителе (a, b) $(-\infty \leqslant a < b \leqslant +\infty)$ плотностью.

Образуем новые случайные величины $Z_i = Y_i - X_i = \theta + \delta_i$, где $\delta_i = \varepsilon_i' - \varepsilon_i$. Зная плотность распределения $p_{\varepsilon_1}(x)$, можно записать плотность $p_{\delta_1}(x)$ по формуле свертки (П3), откуда вытекают ее *четность*, непрерывность и ограниченность (убедитесь!).

Вопрос 1. Какой носитель имеет плотность $p_{\delta_1}(x)$?

Следовательно, получаем, что распределение $F_{\delta_1}(x)$ принадлежит классу Ω_s.

Вопрос 2. Останется ли верным это утверждение, если отказаться от предположения, что носителем $p_{\varepsilon_1}(x)$ является (a, b)?

Таким образом, мы приходим к *модели сдвига* симметричного распределения. Как проверить гипотезу $H\colon \theta > 0$ и как оценить параметр θ, если гипотеза H подтвердилась? Проверке гипотез посвящена часть III этой книги. Сейчас же мы познакомимся с двумя оценками для параметра сдвига в распределении $F(x) \in \Omega_s$ и обсудим их поведение с точки зрения эффективности и робастности.

§ 2. УСЕЧЕННОЕ СРЕДНЕЕ

На соревнованиях по некоторым видам спорта (например, по прыжкам в воду, гимнастике) при учете оценок, выставленных судьями, наименьшая и наибольшая отбрасываются, а остальные усредняются.

Определение. Пусть $0 < \alpha < 1/2$, $k = [\alpha n]$, где $[\,\cdot\,]$ — *целая часть* числа, а n — объем выборки. *Усеченным средним* порядка α называется

$$\overline{X}_\alpha = \frac{1}{n - 2k}\left(X_{(k+1)} + \ldots + X_{(n-k)}\right),$$

где $X_{(1)} \leqslant X_{(2)} \leqslant \ldots \leqslant X_{(n)}$ — вариационный ряд (см. § 4 гл. 4).

Предельные случаи $\alpha = 0$ и $\alpha = 1/2$ соответствуют оценкам $\overline{X}$ и *MED* (рис. 3).

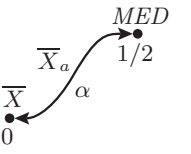


Рис. 3

Оценка $\overline{X}$ достаточно эффективна (в смысле точности) на распределениях, близких к нормальному, но слишком чувствительна к «выбросам». С другой стороны, на нормальном законе *MED* проигрывает $\overline{X}$ в эффективности 36% (см. задачу 1 гл. 7), но весьма устойчива: даже если «выбросами» окажутся почти половина x_i, она не почувствует их присутствие (не сместится в их сторону). Изменяя значение α от 0 до 1/2, будем получать *компромиссные* оценки $\overline{X}_\alpha$. Наибольшая доля «выбросов» в выборке, которую игнорирует усеченное среднее $\overline{X}_\alpha$ (так называемая *толерантность*, см. § 4), определяется выбором α. А как ведет себя асимптотическая дисперсия усеченного среднего $\overline{X}_\alpha$ (см. § 4 гл. 7) при изменении α? Ответ дает следующая теорема об асимптотической нормальности $\overline{X}_\alpha$ на Ω_s.

Теорема 1. Пусть элементы выборки X_i распределены согласно закону $F(x-\theta)$, где $F \in \Omega_s$. Тогда для $0 < \alpha < 1/2$ имеем

$$\sqrt{n}\,(\overline{X}_\alpha - \theta) \xrightarrow{d} \xi \sim \mathcal{N}\left(0, \sigma_\alpha^2\right) \quad \text{при } n \to \infty,$$

где

$$\sigma_\alpha^2 = \frac{2}{(1-2\alpha)^2}\left[\int_0^{x_{1-\alpha}} t^2\, p(t)\, dt + \alpha\, x_{1-\alpha}^2\right].$$

Здесь $p(t)$ — плотность, отвечающая функции распределения F, $x_{1-\alpha}$ — (единственное) решение уравнения $F(x) = 1-\alpha$, т. е. $(1-\alpha)$-квантиль распределения F (см. § 3 гл. 7).

Таблица, полученная на основе приведенной выше формулы для σ_α^2, демонстрирует, как уменьшается с ростом α асимптотическая относительная эффективность $e_{\overline{X}_\alpha, \overline{X}}$ для *нормальной* модели сдвига ($F(x) = \Phi(x)$ — функция распределения закона $\mathcal{N}(0,1)$):

α	0	1/20	1/8	1/4	3/8	1/2
$e_{\overline{X}_\alpha, \overline{X}}$	1,00	0,99	0,94	0,84	0,74	0,64

Вопрос 3.
Чему равно точное значение эффективности в последней графе этой таблицы?
а) Догадайтесь.
б) Вычислите.

В частности, при $\alpha = 1/8$ (при защите от 12,5%-го «загрязнения» выборки) потеря эффективности составляет всего 6%!

Варианты поведения асимптотической дисперсии σ_α^2 на некоторых других симметричных распределениях рассмотрены в задаче 2. А что можно сказать об относительной эффективности оценок $\overline{X}_\alpha$ и $\overline{X}$ на всем классе Ω_s? Оказывается, верна следующая теорема.

Теорема 2. Для всех распределений $F \in \Omega_s$ справедливы неравенства $(1-2\alpha)^2 \leqslant e_{\overline{X}_\alpha, \overline{X}}(F) \leqslant \infty$.

Доказательство. Первое неравенство немедленно вытекает из того, что

$$\frac{1}{2}\,\mathbf{D}X_1 = \int\limits_0^{x_{1-\alpha}} t^2 p(t)\,dt + \int\limits_{x_{1-\alpha}}^{\infty} t^2 p(t)\,dt \geqslant \int\limits_0^{x_{1-\alpha}} t^2\,p(t)\,dt + x_{1-\alpha}^2 \int\limits_{x_{1-\alpha}}^{\infty} p(t)\,dt =$$

$$= \int\limits_0^{x_{1-\alpha}} t^2 p(t)dt + \alpha x_{1-\alpha}^2 = \frac{1}{2}\,\sigma_\alpha^2 (1-2\alpha)^2.$$

Бесконечная верхняя граница достигается на распределениях с бесконечным вторым моментом (например, на законе Коши). ■

Несложно установить, что нижняя граница является точной, рассматривая последовательность распределений, которые могут быть произвольными внутри $[x_\alpha, x_{1-\alpha}]$, но лежащая вне этого отрезка вероятностная масса которых все более концентрируется около его концевых точек.

Приведем таблицу нескольких значений нижней границы:

α	0	1/20	1/8	1/4	3/8	1/2
$(1-2\alpha)^2$	1,00	0,81	0,56	0,25	0,06	0,00

Из нее видно, что, скажем, при $\alpha = 1/8$ потеря эффективности может составить 44% (сравните с 6% на нормальном законе). Это слишком много. Таким образом, на всем классе Ω_s усеченное среднее $\overline{X}_\alpha$ не обеспечивает удовлетворительный компромисс между точностью и робастностью. Следующая оценка справляется с этой задачей существенно лучше.

§ 3. МЕДИАНА СРЕДНИХ УОЛША

По выборке $X_1, \ldots, X_n$ построим $N = n(n+1)/2$ новых случайных величин $Y_k = \frac{1}{2}\,(X_i + X_j)$, $i \leqslant j$ (их называют *средними Уолша*). Рассмотрим статистику $W = MED\{Y_1, \ldots, Y_N\}$.

Вопрос 4.
а) Зависимы ли величины $Y_1, \ldots, Y_N$?
б) Верно ли, что $W = MED\{Z_1, \ldots, Z_N\}$, где $Z_k = \frac{1}{2}\,(X_{(i)} + X_{(j)})$, $i \leqslant j$?

Теорема 3. Если X_i имеют функцию распределения $F(x-\theta)$, где $F \in \Omega_s$, то

$$\sqrt{n}\,(W-\theta) \xrightarrow{d} \xi \sim \mathcal{N}\left(0,\ \sigma_F^2\right) \text{ при } n \to \infty,$$

где $\sigma_F^2 = 1/E(F)$, $E(F) = 12\left(\int p^2(t)\,dt\right)^2$, $p(t)$ — плотность, отвечающая функции распределения F.

Вопрос 5.
А все-таки, чему равно точное значение $e_{W,\overline{X}}$?

Отсюда нетрудно подсчитать, что для *нормального закона* $e_{W,\overline{X}} \approx 0{,}955$, т. е. потеря эффективности всего 4,5%.

Более того, на всем классе Ω_s верна следующая оценка снизу.

Теорема 4. Для всех распределений $F \in \Omega_s$ справедливо неравенство $e_{W,\overline{X}}(F) \geqslant 108/125 \approx 0{,}864$.

Таким образом, используя W вместо $\overline{X}$ для оценки параметра сдвига симметричного распределения, мы в самом худшем случае потеряем только 14% эффективности! (Этот случай реализуется (см. [86, с. 86]) при плотности $p(x) = \dfrac{3\sqrt{5}}{100}\,(5 - x^2)\,I_{\{|x| \leqslant \sqrt{5}\}}$.)

С другой стороны, $e_{W,\overline{X}}$ может быть сколь угодно велика (бесконечна, если $\mathbf{D}X_1 = \infty$).

При сохранении высокой эффективности медиана средних Уолша оказывается достаточно робастной оценкой: она «не боится» даже того, что доля «выбросов» в реализации выборки достигнет 29% (задача 4).

В § 4 определяется одна из характеристик устойчивости оценок — *асимптотическая толерантность*, и приводится пример, показывающий, как резко может уменьшаться точность неробастных оценок даже при крайне малом возмущении модели.

§ 4. РОБАСТНОСТЬ

В реальных данных доля «выбросов» (выделяющихся значений) обычно составляет от 1% до 10%. Это происходит из-за большого числа неучтенных факторов (в медицине, психологии), сбоев оборудования, скажем, скачков напряжения в электросети (в экспериментальной физике), ошибках при вводе с клавиатуры чисел в компьютер и т. д. Даже в астрономических таблицах встречается до 0,1% ошибок.

Казалось бы, можно придерживаться такой стратегии борьбы с «выбросами»: найти их и исключить, а затем применить эффективные параметрические методы для анализа оставшихся данных. Конечно, среди точек на прямой «выброс» хорошо заметен. Но реальные данные, как правило, многомерные. На рис. 4 приведена двумерная выборка, где обведенная кружком точка (очевидный «выброс») не выделяется среди остальных ни по координате X, ни по координате Y. Однако, если попытаться формально подогнать прямую под это «облако» точек, например, с помощью метода наименьших квадратов (см. § 1 гл. 21), то ее угловой коэффициент будет существенно искажен под влиянием «выброса».

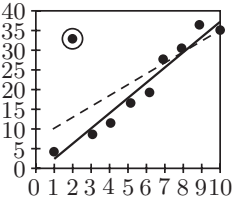


Рис. 4

Возможна ситуация, когда даже проецирование многомерных данных на всевозможные двумерные плоскости не позволит выявить выделяющиеся наблюдения.

Вопрос 6. Как может выглядеть такое «облако» в $\mathbb{R}^3$?

Так что исключение «многомерного выброса» (или группы «выбросов») — весьма непростая задача. По-видимому, лучше использовать робастные методы, которые за счет небольшой, как

правило, потери в точности по сравнению с параметрическими процедурами автоматически уменьшают влияние выделяющихся наблюдений и не допускают существенного смещения оценок параметров модели.*)

Пусть во всяком деле лучшим советником будет умеренность. Хорошо обрабатывать землю необходимо, а превосходно — убыточно.

Плиний Старший

Замечание. Анализ многомерных данных часто сопряжен с большим числом подвохов и сложностей, возникающих из-за так называемого «проклятия размерности». При вычислении объема k-мерного шара в § 4 гл. 3 мы уже встречались с тем, что наша трехмерная интуиция не помогает предвидеть результат с должной точностью.

Параметрические методы, как правило, *крайне чувствительны* (уже при $k \approx 7$) к возмущению модели (см. пример 2 в гл. 16). По-видимому, чтобы не ввести себя в заблуждение, следует *совместно* анализировать не более четырех—пяти столбцов таблицы данных.

Даже и для одномерного случая, как показывает следующий пример, точность некоторых оценок может резко уменьшаться при незначительном утяжелении «хвостов» закона распределения элементов выборки.

Пример 2. ***Смесь нормальных законов*** (Дж. Тьюки, 1960, см. [89, с. 10])**.** Пусть X_i имеют функцию распределения

$$F_{\mu,\sigma,\varepsilon}(x) = (1-\varepsilon)\,\Phi\left(\frac{x-\mu}{\sigma}\right) + \varepsilon\,\Phi\left(\frac{x-\mu}{3\sigma}\right),$$

где $\Phi(x)$ — функция распределения $\mathcal{N}(0,1)$, а параметры $\mu \in \mathbb{R}$, $\sigma > 0$ и $0 \leqslant \varepsilon \leqslant 1$ неизвестны. Данная модель — смесь законов $\mathcal{N}(\mu,\sigma^2)$ и $\mathcal{N}(\mu,9\sigma^2)$ с весами $1-\varepsilon$ и ε соответственно (сравните с задачей 3 гл. 7). Все наблюдения имеют общее среднее μ, а разброс у некоторых из них (в количестве $\approx \varepsilon n$) в 3 раза больше, чем у остальных.

Рассмотрим задачу оценивания разброса. Сравним следующие две оценки: *среднее абсолютное отклонение* R_n и *среднее квадратичное отклонение* S_n: $R_n = \frac{1}{n}\sum\limits_{i=1}^{n}\left|X_i - \overline{X}\right|$, $S_n^2 = \frac{1}{n}\sum\limits_{i=1}^{n}\left(X_i - \overline{X}\right)^2$.

Следует учесть, что R_n и S_n оценивают разные характеристики разброса. Скажем, если $\varepsilon = 0$ (наблюдения имеют в точности нормальное распределение), то $S_n \xrightarrow{\mathsf{P}} \sigma$ (см. пример 3 гл. 6), в то время как $R_n \xrightarrow{\mathsf{P}} \sigma\sqrt{2/\pi} \approx 0{,}798\,\sigma$. Кроме того, эти пределы зависят от ε (например, $S_n \xrightarrow{\mathsf{P}} \sigma\sqrt{1+8\varepsilon}$). Поэтому необходимо уточнить, как проводить сравнение эффективности этих оценок. Возьмем в качестве меры относительной точности оценок при больших n

*) Для данных на рис. 4 можно применить, скажем, робастный метод Тейла из § 1 гл. 21.

предел отношения (не зависящих от масштаба) *коэффициентов вариации*:

$$\tilde{e}_{R_n,S_n}(\varepsilon) = \lim_{n\to\infty} \frac{\mathbf{D}S_n/(\mathbf{M}S_n)^2}{\mathbf{D}R_n/(\mathbf{M}R_n)^2} = \frac{[3\,(1+80\varepsilon)/(1+8\varepsilon)^2 - 1]/4}{\pi\,(1+8\varepsilon)/[2\,(1+2\varepsilon)^2] - 1}\,.$$

Приведем таблицу некоторых значений этой функции из [89, с. 11]:

ε	0	0,002	0,01	0,05	0,1	0,5	1
$\tilde{e}_{R_n,S_n}(\varepsilon)$	0,88	1,02	1,44	2,04	1,90	1,02	0,88

Как видно, функция очень быстро возрастает: 12%-ное преимущество S_n при $\varepsilon = 0$ исчезает уже при $\varepsilon = 0{,}002$ (достаточно всего двух «плохих» наблюдений на 1000 для того, чтобы оценка R_n стала эффективнее). Если же доля наблюдений из распределения с чуть более «тяжелыми хвостами» составит 5%, то среднее абсолютное отклонение окажется точнее более чем в 2 раза!

Этот пример показывает, что малое возмущение моделей может приводить к качественному изменению статистических выводов, в данном случае — выводов о сравнительной эффективности R_n и S_n.

За малым погнался — большое потерял.

Робастные оценки, как правило, осуществляют компромисс между точностью и защищенностью от «выбросов». Если асимптотическая дисперсия — это характеристика точности асимптотически нормальных оценок, то каким образом можно измерить защищенность?

Одной из простых мер робастности является *асимптотическая толерантность* (см. [86, с. 31]). Содержательно она выражает ту *наибольшую долю «выбросов»* в выборке, которую «выдерживает» статистика, не смещаясь вслед за «выбросами» на $-\infty$ или $+\infty$. Дадим формальное

Tolerantia (*лат.*) — терпение.

Определение. Пусть для оценки $\widehat{\theta}(x_{(1)},\ldots,x_{(n)})$ найдется такое целое число k, $0 \leqslant k < n$, что

а) если $x_{(k+2)},\ldots,x_{(n)}$ фиксированы, а $x_{(k+1)} \to -\infty$, то $\widehat{\theta}(x_{(1)},\ldots,x_{(n)}) \to -\infty$;

б) если $x_{(1)},\ldots,x_{(n-k-1)}$ фиксированы, а $x_{(n-k)} \to +\infty$, то $\widehat{\theta}(x_{(1)},\ldots,x_{(n)}) \to +\infty$.

Обозначим через k_n^* наименьшее такое k (тем самым $\widehat{\theta}$ допускает *по крайней мере* k_n^* выделяющихся наблюдений). *Асимптотической толерантностью* оценки $\widehat{\theta}$ называется предел $\tau_{\widehat{\theta}} = \lim\limits_{n\to\infty} k_n^*/n$ (если, конечно, этот предел существует).*)

Очевидно, что $\tau_{\overline{X}} = 0$, $\tau_{\overline{X}_\alpha} = \alpha$ и $\tau_{MED} = 1/2$. Примеры вычисления толерантности других оценок см. в задачах 3 и 4.

*) Отметим, что толерантность основывается на поведении оценки $\widehat{\theta}$ как функции от n переменных и (в отличие от относительной эффективности) *не связана с распределением* элементов выборки $X_1,\ldots,X_n$.

ЗАДАЧИ

Главное, делайте все с увлечением, это страшно украшает жизнь.
Л. Д. Ландау

1. Сравните W как оценку параметра сдвига распределения Лапласа (см. задачу 2 гл. 7) с оценками $\overline{X}$ и *MED*.

2*. Постройте график асимптотической дисперсии σ_α^2 (см. теорему 1) для модели сдвига закона Лапласа. Будет ли эта функция монотонной? Найдите пределы при $\alpha \to 0$ и $\alpha \to 1/2$.

Ф. Гальтон (1822–1911), английский психолог и антрополог.

3. Вычислите асимптотическую толерантность *оценки Гальтона*:

$$\widehat{\theta} = MED\left\{\frac{1}{2}\,(X_{(i)} + X_{(n-i+1)}),\ i = 1, \ldots, \left[\frac{n+1}{2}\right]\right\},$$

где $[\,\cdot\,]$ обозначает *целую часть* числа.

4. Найдите точное значение τ_W.

5. Пусть величины X_i равномерно распределены на $[\theta - \frac{1}{2}, \theta + \frac{1}{2}]$.

а) Смоделируйте с помощью таблицы Т1 выборку размера $n = 5$ для $\theta = \frac{1}{2}$ и сравните точность оценок W и $\widehat{\theta} = \frac{1}{2}\,(X_{(1)} + X_{(n)})$ на этих данных.

б) Какая из этих оценок предпочтительнее при больших n?
Указание. Используйте результат задачи 3 гл. 1 и формулу (6) гл. 3.

6*. Рассмотрим модель сдвига $F(x - \theta)$, где F имеет четную плотность $p(x)$, причем $p(0) \geqslant p(x)$ при всех x. Докажите, что среди всех таких распределений наименьшую асимптотическую эффективность $e_{MED,\overline{X}}(F) = 1/3$ имеет равномерное распределение.

РЕШЕНИЯ ЗАДАЧ

Усердный в службе не должен бояться своего незнанья; ибо каждое новое дело он прочтет.
Козьма Прутков

1. Вычислим σ_F^2 (см. теорему 3) для закона Лапласа:

$$I_F = \frac{1}{4}\int\limits_{-\infty}^{+\infty} e^{-2|t|}\,dt = \frac{1}{4}\int\limits_{0}^{+\infty} e^{-u}\,du = \frac{1}{4}.$$

Отсюда $E(F) = 12\,I_F^2 = 3/4$ и $\sigma_F^2 = 4/3$. В свою очередь, из решения задачи 2 гл. 7 имеем $\sigma_{\overline{X}}^2 = 2$ и $\sigma_{MED}^2 = 1$, т. е. оценка W точнее, чем $\overline{X}$, но проигрывает *MED* примерно 33%.

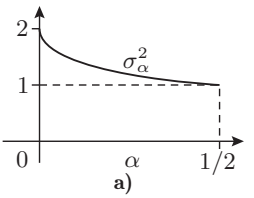
2. Интегрируя плотность Лапласа, находим функцию распределения: $F(x) = 1 - \frac{1}{2}\,e^{-x}$ при $x > 0$. Следовательно, для $\alpha \leqslant \frac{1}{2}$ квантиль $x_{1-\alpha} = -\ln 2\alpha$. Дважды интегрируя по частям, нетрудно получить, что

$$\frac{1}{2}\int\limits_{0}^{-\ln 2\alpha} t^2\,e^{-t}\,dt = 1 - 2\alpha + 2\alpha\ln 2\alpha - \alpha\,(\ln 2\alpha)^2.$$

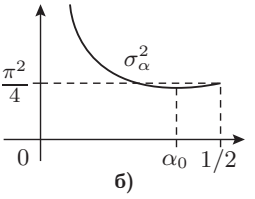
Из теоремы 1 находим асимптотическую дисперсию оценки $\overline{X}_\alpha$:

$$\sigma_\alpha^2 = \frac{2}{(1-2\alpha)^2}\,(1-2\alpha+2\alpha\ln 2\alpha).$$

Функция σ_α^2 монотонно убывает от 2 до 1: для закона Лапласа *MED* оказывается эффективней, чем любая из оценок $\overline{X}_\alpha$ (рис. 5а).



а)



б)

Рис. 5

Любопытно, что у закона Коши, несмотря на более «тяжелые хвосты», оптимальной долей усечения будет не 1/2, а $\alpha_0 \approx 0{,}38$ (рис. 5б). При этом выигрыш в эффективности по сравнению с *MED* составляет около 8%. (В примерах 11 и 12 гл. 9 появятся еще более точные оценки для параметра сдвига распределения Коши.)

3. Всего имеется $\left[\frac{n+1}{2}\right]$ пар вида $(X_{(i)},X_{(n-i+1)})$ (рис. 6а). Если при некотором $k+1$ значение статистики $X_{(k+1)}$ равно $-\infty$, то и у всех пар с $i \leqslant k+1$ первая координата — тоже $-\infty$. Медиана полусумм $\frac{1}{2}\,(X_{(i)} + X_{(n-i+1)})$, $i = 1,\ldots,\left[\frac{n+1}{2}\right]$, «уходит» в $-\infty$ при $k+1 \geqslant \frac{1}{2}\left[\frac{n+1}{2}\right]$. Деля на n и переходя к пределу, находим, что толерантность оценки Гальтона равна $\frac{1}{4}$.

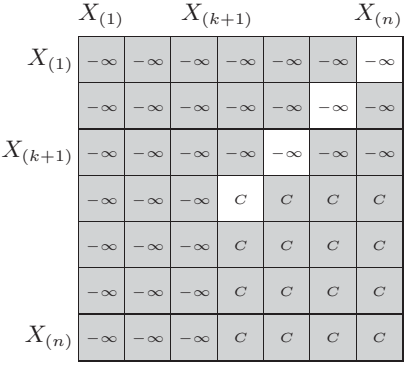
Эксперимент можно считать удавшимся, если нужно отбросить не более 50% сделанных измерений, чтобы достичь соответствия с теорией. (*Следствие из законов Мейерса.*)

4. Аналогично (рис. 6б), чтобы W не обращалась в $-\infty$, необходимо, чтобы $(n-k-1)(n-k)/2 \geqslant n(n+1)/4$ полусумм порядковых статистик были конечны. Положив $k \sim \tau_W n$, получаем отсюда, что $(1-\tau_W)^2 = 1/2$ или $\tau_W = 1 - \sqrt{2}/2 \approx 0{,}293$.

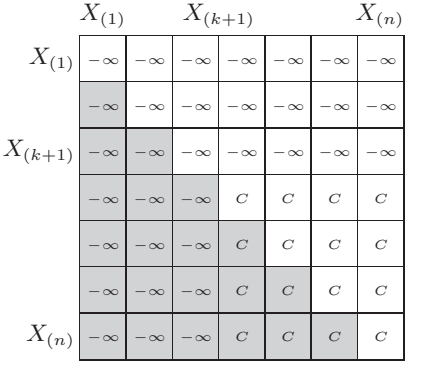
5. а) Возьмем, скажем, из 3-й строки таблицы Т1 первые 5 псевдослучайных чисел: $x_1 = 0{,}08$; $x_2 = 0{,}42$; $x_3 = 0{,}26$; $x_4 = 0{,}89$; $x_5 = 0{,}53$. Для этих данных $W = 0{,}395$ и $\widehat{\theta} = 0{,}485$, т. е. значение оценки $\widehat{\theta}$ оказывается ближе к 1/2.

б) Равномерное распределение на $\left[-\frac{1}{2},\frac{1}{2}\right]$ принадлежит классу Ω_s. В силу теоремы 3 точность оценки W имеет порядок $1/\sqrt{n}$.

С другой стороны, из задачи 3 гл. 1 порядок малости дисперсии $\mathbf{D}X_{(n)}$ равен $1/n^2$. Случайные величины $\theta - X_{(1)}$



а)



б)

Рис. 6

и $X_{(n)} - \theta$ ввиду симметрии одинаково распределены. Поэтому $\mathbf{D}X_{(1)} = \mathbf{D}X_{(n)}$. В соответствии с формулой (6) гл. 3 стандартное отклонение $\sqrt{\mathbf{D}\widehat{\theta}}$ оценивается сверху величиной порядка $1/n$. Таким образом, при больших n для модели сдвига равномерного распределения оценка $\widehat{\theta} = \frac{1}{2}(X_{(1)} + X_{(n)})$ значительно точнее оценки W.

И в ком не сыщешь пятен?

Чацкий в «Горе от ума» А. С. Грибоедова

Замечание. Распределение величины $n\left(X_{(1)} - \theta + \frac{1}{2}\right)$ согласно задаче 4 гл. 5 стремится при $n \to \infty$ к показательному закону с $\lambda = 1$. В силу симметрии величина $n(\theta + \frac{1}{2} - X_{(n)})$ имеет тот же предельный закон. Можно доказать, что случайные величины $X_{(1)}$ и $X_{(n)}$ асимптотически независимы (см. задачу 7 гл. 6). Отсюда и из задачи 2 гл. 7 получаем, что имеет место сходимость $n\left(\frac{1}{2}(X_{(1)} + X_{(n)}) - \theta\right) \xrightarrow{d} \xi$, где ξ распределена по закону Лапласа.

6. По теореме 1 гл. 7 и формуле (1) гл. 7 относительная эффективность $e_{MED,\overline{X}} = 4\,p^2(0)\,\mathbf{D}X_1$. Причем она не зависит от масштаба. (Действительно, если $Y = cX$, то согласно следствию из П8 плотность $p_Y(x) = |c|^{-1}p_X(x/c)$, откуда $p_Y^2(0) = c^{-2}p_X^2(0)$, а согласно свойствам дисперсии (П2) $\mathbf{D}Y = c^2\mathbf{D}X$.) Поэтому можно считать, что $p(0) = 1$. Таким образом, в силу четности плотности $p(x)$, задача 6 сводится к следующей:

минимизировать по f функционал

$$\frac{1}{2}\,\mathbf{D}X_1 = \int\limits_0^\infty x^2 f(x)\,dx$$

при выполнении условий

$$0 \leqslant f(x) \leqslant f(0) = 1, \quad \int\limits_0^\infty f(x)\,dx = \frac{1}{2}.$$

Утверждение. Минимум функционала достигается на $f^*(x) = I_{[0,\,\frac{1}{2}]}$.

Доказательство. Оценим снизу приращение функционала:

$$\int\limits_0^\infty x^2(f(x) - f^*(x))\,dx = \int\limits_0^{1/2} x^2(f(x) - 1)\,dx + \int\limits_{1/2}^\infty x^2 f(x)\,dx \geqslant$$

$$\geqslant \frac{1}{4}\left(\int\limits_0^{1/2}(f(x) - 1)\,dx + \int\limits_{1/2}^\infty f(x)\,dx\right) = \frac{1}{4}\left(\int\limits_0^\infty f(x)\,dx - \frac{1}{2}\right) = 0. \quad \blacksquare$$

Замечание. Эту задачу можно обобщить (см. [50, с. 321]): для распределений $F \in \Omega_s$, плотность которых имеет максимум в нуле,

$$e_{\overline{X}_\alpha, \overline{X}}(F) \geqslant \frac{1}{1+4\alpha}$$

(сравните с теоремой 2). Минимум достигается при равномерном распределении на отрезке $\left[-\frac{1}{2}, \frac{1}{2}\right]$.

ОТВЕТЫ НА ВОПРОСЫ

1. Носителем плотности $p_{\delta_1}(x)$ является интервал $(-c, c)$, где $c = b - a$.

2. Нет. Скажем, для плотности с носителем $(0,1) \cup (8,9)$ носителем $p_{\delta_1}(x)$ служит множество $(-9, -7) \cup (-1, 1) \cup (7, 9)$.

3. **а)** Точное значение равно $2/\pi$ (т. е. асимптотической эффективности *MED* относительно $\overline{X}$).

б) Выражая асимптотическую дисперсию σ_α^2 через квантиль $x = x_{1-\alpha}$ и дважды применяя правило Лопиталя, находим:

$$\lim_{\alpha \to 1/2} \sigma_\alpha^2 = 2 \cdot \lim_{x \to 0} \frac{\int\limits_0^x t^2\, p(t)\, dt + (1 - F(x))x^2}{[2F(x) - 1]^2} = \frac{1}{4\, p^2(0)}\,.$$

4. **а)** Вообще говоря, зависимы, поскольку увеличение X_1 приводит к росту и $X_1 + X_2$, и $X_1 + X_3$.

б) Верно, так как в обоих случаях пробегаются всевозможные значения $\frac{1}{2}(X_i + X_j)$, где $1 \leqslant i \leqslant j \leqslant n$.

5. $I_F = \frac{1}{2\pi} \int e^{-t^2} dt = \frac{1}{2\sqrt{\pi}}$, откуда $e_{W,\overline{X}} = 12\, I_F^2 = 3/\pi$.

6. Например, «облако» сферической формы с «выбросом» (или группой «выбросов») вблизи центра сферы.

Глава 9

МЕТОДЫ ПОЛУЧЕНИЯ ОЦЕНОК

Следует поставить перед собой цель изыскать способ решения всех задач одним и притом простым методом.

Ж. Даламбер

В этой главе рассматриваются несколько простых и универсальных методов получения оценок параметров статистических моделей, в том числе — метод моментов и метод максимального правдоподобия. Прежде всего, познакомимся с графическим анализом на вероятностной бумаге.

§ 1. ВЕРОЯТНОСТНАЯ БУМАГА

Занимаясь той или иной научной проблемой, лучше исходить из ее индивидуальных особенностей, чем полагаться на общие методы.

Д. Курант, Г. Роббинс

Построим по выборке $X_1, \ldots, X_n$ случайную ступенчатую функцию $\widehat{F}_n(x)$, возрастающую скачками величины $\frac{1}{n}$ в точках $X_{(i)}$ (рис. 1). Она называется *эмпирической функцией распределения*. Чтобы за-

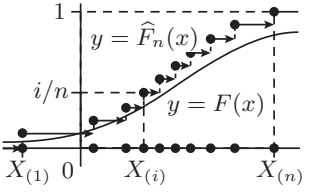


Рис. 1

дать значения в точках разрывов, формально определим ее так, чтобы она была непрерывна справа:

$$\widehat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_{(i)} \leqslant x\}} = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i \leqslant x\}}.$$

Вопрос 1.
а) Какое распределение имеет случайная величина $I_{\{X_i \leqslant x\}}$?
б) Как называется последовательность $I_{\{X_1 \leqslant x\}}, \ldots, I_{\{X_n \leqslant x\}}$?
в) Что такое $\widehat{F}_n(x)$ по отношению к этой последовательности?
г) К чему сходится $\widehat{F}_n(x)$ при $n \to \infty$ для фиксированного x?

Проблема. Пусть элементы выборки $X_1, \ldots, X_n$ имеют функцию распределения $F((x-\mu)/\sigma)$, где F известна, а параметры сдвига μ и масштаба $\sigma > 0$ — нет. Как их оценить?

Из ответа на вопрос 1 вытекает, что эмпирическая функция распределения $\widehat{F}_n(x)$ служит естественным приближением к теоретической функции распределения $F((x-\mu)/\sigma)$. Среди функций этого двухпараметрического семейства следовало бы выбрать такую функцию $F((x-\widehat{\mu})/\widehat{\sigma})$, чтобы она «меньше всего» отличалась от $\widehat{F}_n(x)$, и взять соответствующие $\widehat{\mu}$ и $\widehat{\sigma}$ в качестве искомых оценок. Однако, в общем случае из-за нелинейности F это сделать затруднительно. Идея метода оценивания, приведенного ниже,

состоит в «распрямлении» графика $F((x-\mu)/\sigma)$ и последующей подгонки прямой, сглаживающей соответствующее «облако» точек плоскости.

Для простоты допустим, что F непрерывна и строго монотонна. Тогда для нее определена обратная функция F^{-1}. Посмотрим, во что переходит график функции $y = F((x-\mu)/\sigma)$ при преобразовании $(x, y) \to (x, F^{-1}(y))$:

$$(x, F((x-\mu)/\sigma)) \to (x, F^{-1}(F((x-\mu)/\sigma))) = (x, (x-\mu)/\sigma).$$

Значит, график переходит в прямую $y = (x-\mu)/\sigma$.

Отсюда вытекает следующий способ оценивания μ и σ: преобразуем график эмпирической функции распределения $y = \widehat{F}_n(x)$ в $y = F^{-1}(\widehat{F}_n(x))$ и подберем «на глаз» наиболее тесно прилегающую к нему прямую $y = (x-\widehat{\mu})/\widehat{\sigma}$. При этом оценка $\widehat{\mu}$ — это координата точки пересечения с осью абсцисс, а $\widehat{\sigma}$ — котангенс угла наклона построенной прямой.

Если функция $y = F^{-1}(\widehat{F}_n(x))$ слишком сильно отличается от линейной, то предположение о том, что выборка взята из совокупности с функцией распределения $F((x-\mu)/\sigma)$, скорее всего не выполняется.

Для реализации этого способа получения оценок нет необходимости строить целиком график $y = F^{-1}(\widehat{F}_n(x))$. Достаточно отметить только точки $(x_{(i)}, F^{-1}(i/n))$, отвечающие скачкам функции $\widehat{F}_n(x)$, и подогнать прямую к этому «облаку» точек (это можно осуществить с помощью метода наименьших квадратов или других регрессионных (сглаживающих) методов из гл. 21).

Чтобы избежать неудобства, связанного с построением точки $(x_{(n)}, F^{-1}(1))$, когда случайная величина X_1 не ограничена сверху, обычно используют точки $(x_{(i)}, F^{-1}((i-0{,}5)/n))$, $i = 1, \ldots, n$.

Пример 1. Моделируем нормальную выборку с помощью таблицы нормальных случайных чисел из [10, с. 371]. Взяв первые $n = 10$ чисел z_i из 3-й строки этой таблицы, преобразуем их в реализацию выборки из распределения $\mathcal{N}(\mu, \sigma^2)$ по формуле $x_i = \mu + \sigma z_i$ для $\mu = 1$ и $\sigma = 2$. Получим следующие значения:

3,97 0,29 −0,27 2,39 2,85 3,75 2,57 −0,93 −0,71 −2,73.

По таблице обратной функции $\Phi^{-1}(y)$ к функции распределения $\mathcal{N}(0,1)$ (см. Т2 или [10, с. 136]) вычислим $\Phi^{-1}((i-0{,}5)/n)$, $i = 1, \ldots, n$:

−1,65 −1,04 −0,68 −0,39 −0,13 0,13 0,39 0,68 1,04 1,65.

Точки $\bigl(x_{(i)}, \Phi^{-1}((i-0{,}5)/n)\bigr)$ и подогнанная к ним «на глаз» прямая приведены на рис. 2. Отсюда находим, что $\mu \approx 1{,}2$ и $\sigma \approx 1{,}8$.

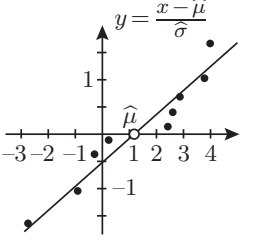


Рис. 2

Как видно, графический анализ позволяет в данном случае получить довольно точные оценки параметров, несмотря на малый размер выборки.

Будто — тяп-ляп, да и корабль.

Для сравнения: выборочное среднее $\overline{X}$ и стандартное отклонение S, где $S^2 = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2$, которые в этой модели являются оценками максимального правдоподобия (см. § 4 и задачу 2), имеют значения 1,12 и 2,16 соответственно.

Графический анализ удобно проводить на так называемой *вероятностной бумаге*. Для ее изготовления строится неравномерная шкала на оси ординат на основе преобразования $y' = F^{-1}(y)$. Шкала на оси абсцисс остается прежней. В новых шкалах непосредственно наносятся точки $(x_{(i)}, (i-0{,}5)/n)$.

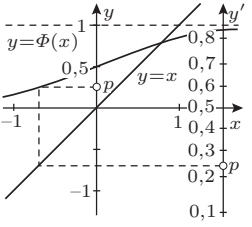


Рис. 3

Оцифровка новой оси ординат для нормального закона показана на рис. 3.

Следует отметить, что рассмотренный метод применяется исключительно к модели сдвига-масштаба и моделям, сводимым к ней при помощи некоторых преобразований. (Так, *логнормальная модель* с плотностью $p_{\mu,\sigma}(x) = \dfrac{1}{x\sigma\sqrt{2\pi}} \exp\Big\{-\dfrac{1}{2\sigma^2}(\ln x - \mu)^2\Big\} I_{\{x>0\}}$ при логарифмировании $X_i' = \ln X_i$ сводится к $\mathcal{N}(\mu, \sigma^2)$.)

Излагаемые далее методы получения оценок можно использовать для более широкого класса статистических моделей.

§ 2. МЕТОД МОМЕНТОВ

Моментом k-го порядка случайной величины X называется величина $\alpha_k = \mathbf{M}X^k$. Моменты существуют не всегда. Например, у закона Коши математическое ожидание α_1 не определено (см. § 2 гл. 1).

А. М. Ляпунов (1857–1918), русский математик.

Из *неравенства Ляпунова*

$$\big(\mathbf{M}|X|^k\big)^{1/k} \leqslant \big(\mathbf{M}|X|^l\big)^{1/l} \quad \text{при } k \leqslant l$$

и свойства $|\mathbf{M}\xi| \leqslant \mathbf{M}|\xi|$ следует, что конечность $\mathbf{M}|X|^m$ гарантирует существование всех моментов α_k для $k = 1, \ldots, m$.

Вопрос 2. Как вывести неравенство Ляпунова из неравенства Иенсена (П4)?

Положим $A_k = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i^k$. Если момент α_k существует, то в силу закона больших чисел (П6) $A_k \xrightarrow{\mathsf{P}} \alpha_k$ при $n \to \infty$. Поэтому для реализации $x_1, \ldots, x_n$ выборки достаточно большого размера можно утверждать, что $a_k = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i^k \approx \alpha_k$, т. е. эмпирические моменты k-го порядка a_k близки к теоретическим моментам α_k. На этом соображении основывается так называемый

Метод моментов

Допустим, что распределение элементов выборки зависит от m неизвестных параметров $\theta_1, \ldots, \theta_m$, где вектор $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$ принадлежит некоторой области Θ в $\mathbb{R}^m$. Пусть $\mathbf{M}|X|^m < \infty$ для всех $\boldsymbol{\theta} \in \Theta$. Тогда существуют все $\alpha_k = \alpha_k(\boldsymbol{\theta})$, $k = 1, \ldots, m$, и можно

записать систему из m (вообще говоря, нелинейных) уравнений

$$\alpha_k(\boldsymbol{\theta}) = y_k, \quad k = 1, \ldots, m. \tag{1}$$

Предположим, что левая часть системы задает взаимно однозначное отображение $g\colon \Theta \to B$, где B — некоторая область в $\mathbb{R}^m$, и что обратное отображение $g^{-1}\colon B \to \Theta$ непрерывно. Другими словами, для всех $(y_1, \ldots, y_m)$ из B система (1) имеет единственное решение, которое непрерывно зависит от правой части. Компоненты решения $\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_1, \ldots, \widehat{\theta}_m)$ при $y_k = A_k$ называются *оценками метода моментов.*[*)]

Пример 2. Рассмотрим модель сдвига показательного закона, в которой плотностью распределения величин X_i служит функция $p_\theta(x) = e^{-(x-\theta)} I_{\{x \geqslant \theta\}}$ (рис. 4). Здесь

$$\alpha_1(\theta) = \mathbf{M}X_1 = \int\limits_\theta^\infty x\, e^{-(x-\theta)}\, dx = \int\limits_0^\infty (y+\theta)\, e^{-y}\, dy = 1 + \theta.$$

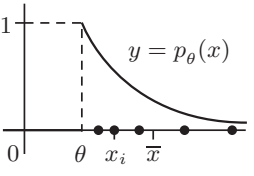


Рис. 4

Из уравнения $1 + \theta = A_1 = \overline{X}$, находим по методу моментов оценку $\widehat{\theta} = \overline{X} - 1$.

Пример 3. Пусть величины X_i имеют гамма-распределение с двумя неизвестными параметрами θ_1 и θ_2: соответствующая плотность имеет вид

$$p_{\boldsymbol{\theta}}(x) = \theta_1^{\theta_2} x^{\theta_2 - 1} e^{-\theta_1 x} I_{\{x>0\}} \,/\, \Gamma(\theta_2).$$

Согласно формуле (2) гл. 4 находим $\alpha_1 = \theta_2/\theta_1$ и $\alpha_2 = \theta_2(\theta_2+1)/\theta_1^2$. Решив систему (1), получаем в качестве оценок метода моментов $\widehat{\theta}_1 = A_1/(A_2 - A_1^2) = \overline{X}/S^2$ и $\widehat{\theta}_2 = A_1^2/(A_2 - A_1^2) = \overline{X}^{\,2}/S^2$.

Какими **статистическими свойствами** обладают оценки, полученные методом моментов?

Их *состоятельность* вытекает из непрерывности определенного выше отображения g^{-1} и свойства сходимости 3 из П5.

Для гладких отображений g^{-1} такие оценки будут также *асимптотически нормальными.* Это следует из того что, во-первых, в силу центральной предельной теоремы (П6) имеет место сходимость

$$\sqrt{n}\,(A_k - \alpha_k) \xrightarrow{d} \xi \sim \mathcal{N}(0, \alpha_{2k} - \alpha_k^2),$$

если $\alpha_{2k} < \infty$. А во-вторых, справедливо обобщение на многомерный случай леммы 1 гл. 7 (см. [11, с. 33]).

Однако обычно асимптотическая дисперсия (см. § 4 гл. 7) оценок, полученных по методу моментов, довольно велика. Поэтому в §§ 4–6 будут рассмотрены оценки с *наименьшей* возможной асимптотической дисперсией для так называемых *регулярных* (см. § 3) статистических моделей.

[*)] Величины $\widehat{\theta}_k$ случайны, так как A_k — функции от $X_1, \ldots, X_n$.

§ 3. ИНФОРМАЦИОННОЕ НЕРАВЕНСТВО

Пусть $f(x,\theta)$ обозначает плотность распределения случайной величины X_1. Для дискретных моделей используем это же обозначение для $\mathsf{P}(X_1 = x)$. Допустим, что выполняются следующие **условия регулярности**:

R1) *параметрическое множество Θ — открытый интервал на прямой (возможно, бесконечный)*;

R2) *носитель распределения $A = \{x\colon f(x,\theta) > 0\}$ не зависит от параметра θ*;

R3) *при любых $x \in A$ и $\theta \in \Theta$ производная $\dfrac{\partial}{\partial\theta} f(x,\theta)$ существует и конечна*;

R4) *для случайной величины $U_1 = \dfrac{\partial}{\partial\theta} \ln f(X_1,\theta)$ при всех $\theta \in \Theta$ справедливы тождество $\mathsf{M}U_1 \equiv 0$ и неравенство $0 < I_1(\theta) = \mathsf{D}U_1 < \infty$.*

Заметим, что условие $\mathsf{M}U_1 \equiv 0$ верно для тех статистических моделей, где производная по θ правой части тождества $1 = \int\limits_A f(x,\theta)\,dx$ может быть вычислена дифференцированием под знаком интеграла:

$$0 = \int\limits_A \frac{\partial}{\partial\theta} f(x,\theta)\,dx = \int\limits_A \frac{\frac{\partial}{\partial\theta} f(x,\theta)}{f(x,\theta)}\ f(x,\theta)\,dx = \mathsf{M}\,\frac{\partial}{\partial\theta} \ln f(X_1,\theta) = \mathsf{M}U_1.$$

Контрпримером может служить равномерное распределение на отрезке $[0,\theta]$, где $\theta \in (0,+\infty)$. Носитель $A = [0,\theta]$ зависит от θ, и

$$0 = \frac{\partial}{\partial\theta}\ 1 = \frac{\partial}{\partial\theta} \left(\int\limits_0^\theta \frac{1}{\theta}\ dx \right) \neq \int\limits_0^\theta \frac{\partial}{\partial\theta} \left(\frac{1}{\theta} \right) dx = -\,\frac{1}{\theta^2} \int\limits_0^\theta dx = -\,\frac{1}{\theta}\,.$$

Для выборки $\boldsymbol{X} = (X_1,\ldots,X_n)$ совместная плотность распределения (или вероятность $\mathsf{P}(X_1 = x_1,\ldots,X_n = x_n)$) в силу независимости компонент распадается в произведение: $f(\boldsymbol{x},\theta) = \prod\limits_{i=1}^n f(x_i,\theta)$, где $\boldsymbol{x} = (x_1,\ldots,x_n)$. Рассмотрим случайную величину

$$U_n = \frac{\partial}{\partial\theta} \ln f(\boldsymbol{X},\theta) = \sum_{i=1}^n \frac{\partial}{\partial\theta} \ln f(X_i,\theta).$$

Тогда применение свойств из П2 дает $\mathsf{M}U_n = 0$ и $\mathsf{D}U_n = nI_1(\theta)$.

Р. Фишер (1890–1962), английский статистик.

Определение. *Информацией Фишера* для случайного вектора $\boldsymbol{\xi}$ с плотностью (в дискретном случае — с совместной вероятностью компонент) $f(\boldsymbol{x},\theta)$ называется величина

$$I_{\boldsymbol{\xi}}(\theta) = \mathsf{M} \left[\frac{\partial}{\partial\theta} \ln f(\boldsymbol{\xi},\theta) \right]^2.$$

Отметим, что содержащаяся в выборке информация $I_{\boldsymbol{X}}(\theta) \equiv \equiv I_n(\theta) = \mathbf{D}U_n = nI_1(\theta)$ пропорциональна размеру выборки. Она интересна тем, что участвует в следующем ограничении снизу на дисперсии оценок в регулярных моделях.

Информационное неравенство (Рао — Крамер). Допустим, что выполнены условия R1–R4, $\widehat{\theta}$ — любая оценка с $\mathbf{M}\widehat{\theta}^{\,2} < \infty$, для которой производная по θ от функции

$$a(\theta) = \mathbf{M}\widehat{\theta} = \int\limits_{\mathbb{R}^n} \widehat{\theta}(\boldsymbol{x})\, f(\boldsymbol{x},\theta)\, d\boldsymbol{x}, \quad \text{где } \boldsymbol{x} = (x_1, \ldots, x_n),$$

существует и может быть получена дифференцированием под знаком интеграла. Тогда

$$\mathbf{D}\widehat{\theta} \geqslant \frac{[a'(\theta)]^2}{I_n(\theta)} = \frac{[a'(\theta)]^2}{nI_1(\theta)}\,.$$

В частности, для оценок, имеющих смещение $b(\theta)$ (т. е. для $a(\theta) = \theta + b(\theta)$), нижней границей служит $[1 + b'(\theta)]^2/I_n(\theta)$, а для несмещенных оценок — $1/I_n(\theta)$.

Доказательство. Дифференцируем под знаком интеграла:

$$a'(\theta) = \int\limits_{\mathbb{R}^n} \widehat{\theta}(\boldsymbol{x})\, \frac{\partial}{\partial\theta} f(\boldsymbol{x},\theta)\, d\boldsymbol{x} = \int\limits_{\mathbb{R}^n} \widehat{\theta}(\boldsymbol{x}) \left[\frac{\partial}{\partial\theta} \ln f(\boldsymbol{x},\theta)\right] f(\boldsymbol{x},\theta)\, d\boldsymbol{x}.$$

Справа стоит $\mathbf{M}\,(\widehat{\theta}\, U_n)$. Отсюда, поскольку $\mathbf{M}U_n = n\,\mathbf{M}U_1 = 0$, получаем представление $a'(\theta) = \mathbf{M}[(\widehat{\theta} - \mathbf{M}\widehat{\theta})\, U_n]$. Применяя теперь неравенство Коши — Буняковского (П4), оценим $[a'(\theta)]^2$ сверху:

$$[a'(\theta)]^2 \leqslant \left[\mathbf{M}(\widehat{\theta} - \mathbf{M}\widehat{\theta})^2\right] \left[\mathbf{M}U_n^2\right] = (\mathbf{D}\widehat{\theta})\, I_n(\theta),$$

что и требовалось установить. ■

Несмещенные оценки, на которых достигается нижняя граница $1/I_n(\theta)$, называются *эффективными*.

Пример 4. Пусть случайные величины X_i имеют нормальное распределение $\mathcal{N}(\theta, \sigma^2)$, где параметр масштаба σ известен, а параметр сдвига θ — нет. Здесь

$$U_1 = \frac{\partial}{\partial\theta} \ln \left[\frac{1}{\sigma\sqrt{2\pi}}\; e^{-\frac{1}{2}(X_1-\theta)^2/\sigma^2}\right] = \frac{\partial}{\partial\theta} \left[-\frac{(X_1-\theta)^2}{2\sigma^2}\right] = \frac{1}{\sigma^2}\,(X_1 - \theta).$$

Отсюда $I_1(\theta) = \mathbf{D}U_1 = \sigma^{-4}\mathbf{D}(X_1 - \theta) = \sigma^{-2}$. Поэтому оценка $\overline{X}$ с дисперсией σ^2/n является эффективной в этой модели.

Информационное неравенство показывает, что для регулярных моделей погрешность оценки $\sqrt{\mathbf{D}\widehat{\theta}}$ не может убывать быстрее, чем $C/\sqrt{n}$. *Контрпримером* служит $X_{(n)}$ — максимум выборки

$X_1,\ldots,X_n$ из равномерного распределения на $[0,\theta]$, который оценивает θ с точностью порядка $1/n$ (см. задачу 3 гл. 1). Подобные оценки называются *сверхэффективными*.

Замечание. Другие условия регулярности, обеспечивающие выполнение неравенства Рао—Крамера, приведены в [11, с. 150].

§ 4. МЕТОД МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ

Метод получил распространение после появления в 1912 г. статьи Р. Фишера, где было доказано, что получаемые этим методом оценки являются *асимптотически наиболее точными* при выполнении условий регулярности из приведенной ниже теоремы 1.

Для знакомства с методом предположим для простоты, что элементы выборки X_i имеют дискретное распределение: $f(x,\boldsymbol{\theta}) = \mathbf{P}(X_1 = x)$ (здесь $\boldsymbol{\theta} = (\theta_1,\ldots,\theta_m)$ — вектор неизвестных параметров модели). Тогда совместная вероятность выборки $f(\boldsymbol{x},\boldsymbol{\theta}) = f(x_1,\boldsymbol{\theta})\cdot\ldots\cdot f(x_n,\boldsymbol{\theta})$ зависит от $n+m$ аргументов (здесь $\boldsymbol{x} = (x_1,\ldots,x_n)$). Рассматриваемая как функция от $\theta_1,\ldots,\theta_m$ при фиксированных значениях элементов выборки $x_1,\ldots,x_n$, она называется *функцией правдоподобия* и обычно обозначается через $L(\boldsymbol{\theta})$. Величину $L(\boldsymbol{\theta})$ можно считать мерой правдоподобия значения $\boldsymbol{\theta}$ при заданной реализации $\boldsymbol{x}$.

Likelihood (*англ.*) — правдоподобие.

Представляется разумным в качестве оценок параметров $\theta_1,\ldots,\theta_m$ взять наиболее правдоподобные значения $\tilde\theta_1,\ldots,\tilde\theta_m$, которые получаются при максимизации функции $L(\boldsymbol{\theta})$ (рис. 5). Такие оценки называются *оценками максимального правдоподобия* (ОМП).

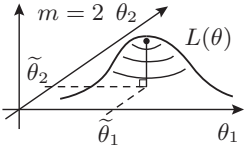


Рис. 5

Часто проще искать точку максимума функции $\ln L(\boldsymbol{\theta})$, которая совпадает с $\tilde{\boldsymbol{\theta}} = (\tilde\theta_1,\ldots,\tilde\theta_m)$ в силу монотонности логарифма.

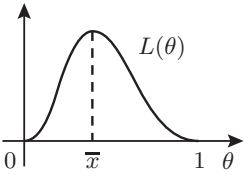


Рис. 6

Пример 5. Для схемы Бернулли $X_1,\ldots,X_n$ с вероятностью «успеха» θ имеем: $f(x,\theta) = \mathbf{P}(X_1 = x) = \theta^x(1-\theta)^{1-x}$, где x принимает значения 0 или 1. Поэтому функция правдоподобия $L(\theta) = \theta^{s_n}(1-\theta)^{n-s_n}$, где $s_n = x_1+\ldots+x_n$, представляет собой многочлен n-й степени (рис. 6 при $s_n > 1$ и $n - s_n > 1$). Найдем точку максимума $\ln L(\theta) = s_n\ln\theta + (n-s_n)\ln(1-\theta)$. Дифференцируя по θ, получаем уравнение $s_n/\theta - (n-s_n)/(1-\theta) = 0$, откуда $\tilde\theta = s_n/n = \overline{x}$. Таким образом, ОМП в схеме Бернулли — это частота «успехов» в реализации $x_1,\ldots,x_n$.

Как и в § 3, в случае непрерывных моделей будем использовать обозначение $f(x,\theta)$ для плотности распределения случайной величины X_1.

Пример 6. Рассмотрим модель сдвига показательного закона с плотностью $f(x,\theta) = e^{-(x-\theta)}I_{\{x\geqslant\theta\}}$. В этом случае функция

правдоподобия равна

$$L(\theta) = \prod_{i=1}^{n} f(x_i,\theta) = e^{-(x_1+\ldots+x_n)} e^{n\theta} I_{\{x_{(1)} \geqslant \theta\}}.$$

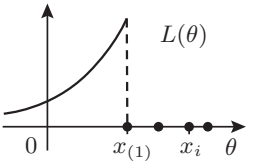


Рис. 7

Отсюда (см. рис. 7) получаем в качестве ОМП $\tilde{\theta} = x_{(1)}$, которая отлична от оценки метода моментов $\widehat{\theta} = \overline{x} - 1$, найденной ранее для этой модели в примере 2. Заметим также, что здесь $L(\theta)$ не является гладкой функцией, и поэтому ОМП нельзя вычислять, приравнивая нулю производную функции правдоподобия.

В случае, когда $L(\boldsymbol{\theta})$ гладко зависит от $\theta_1,\ldots,\theta_m$, оценки максимального правдоподобия являются компонентами решения (вообще говоря, нелинейной) системы уравнений:

$$\frac{\partial}{\partial\theta_j} \ln L(\boldsymbol{\theta}) = \sum_{i=1}^{n} \frac{\partial}{\partial\theta_j} \ln f(x_i,\boldsymbol{\theta}) = 0, \quad j = 1,\ldots,m. \tag{2}$$

Иногда это решение можно найти явно (см. задачу 2), но чаще приходится вычислять его приближенно с помощью итерационных методов (например, метода Ньютона из § 5).

Пример 7. Для закона Вейбулла—Гнеденко, используемого в моделях, описывающих прочность материалов (см. § 2 гл. 4), с функцией распределения $F_{\alpha,\beta}(x) = 1 - \exp\{-x^\alpha/\beta\}$ при $x > 0$ (параметры α и β предполагаются положительными) система (2) после некоторых упрощений приводится к виду

$$\frac{1}{\alpha} = \frac{1}{n}\sum_{i=1}^{n} \ln x_i - \sum_{i=1}^{n} x_i^\alpha \ln x_i \Big/ \sum_{i=1}^{n} x_i^\alpha, \quad \beta = \frac{1}{n}\sum_{i=1}^{n} x_i^\alpha.$$

В [50, с. 387] доказано, что первое уравнение системы (следовательно, и вся система) при любых $x_1,\ldots,x_n$ (не равных одновременно 1) имеет единственное решение.

Как показывают два следующих примера, функция правдоподобия с вероятностью 1 может быть не ограничена сверху, т. е. глобальный максимум $L(\boldsymbol{\theta})$ на множестве Θ может не достигаться.

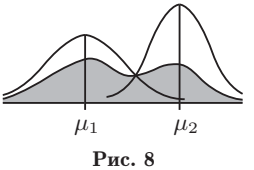


Рис. 8

Пример 8 [50, с. 391]**.** Рассмотрим (рис. 8) смесь плотностей (см. § 2 гл. 5) $\dfrac{1-\varepsilon}{\sigma_1}\, p\left(\dfrac{x-\mu_1}{\sigma_1}\right) + \dfrac{\varepsilon}{\sigma_2}\, p\left(\dfrac{x-\mu_2}{\sigma_2}\right)$, где $\varepsilon \in (0,1)$, $\sigma_1 > 0$, $\sigma_2 > 0$, $p(x)$ — любая положительная при всех x плотность, (скажем, стандартная нормальная $p(x) = \dfrac{1}{\sqrt{2\pi}}\, e^{-x^2/2}$). Тогда плотность выборки

$$\prod_{i=1}^{n} \left\{ \frac{1-\varepsilon}{\sigma_1}\, p\left(\frac{x_i-\mu_1}{\sigma_1}\right) + \frac{\varepsilon}{\sigma_2}\, p\left(\frac{x_i-\mu_2}{\sigma_2}\right) \right\}$$

представляет собой сумму положительных членов, один из которых равен

$$\frac{1-\varepsilon}{\sigma_1}\,p\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{\varepsilon}{\sigma_2}\right)^{n-1}\prod_{i=2}^{n}p\left(\frac{x_i-\mu_2}{\sigma_2}\right).$$

Когда $\mu_1 = x_1$ и $\sigma_1 \to 0$, этот член стремится к бесконечности при любых фиксированных значениях ε, μ_2, σ_2, $x_2,\ldots,x_n$. Стало быть, $L(\boldsymbol{\theta})$ неограничена и *глобальной* ОМП не существует. Однако отметим, что для смеси нормальных плотностей выполняются условия теоремы 1, приведенной ниже, и поэтому при достаточно большом n функция правдоподобия будет иметь (с вероятностью 1) *локальный* максимум в точке, которая находится как решение системы (2).

Пример 9 [61, с. 71]**.** Допустим, что у модели сдвига-масштаба $\frac{1}{\sigma}\,p\left(\frac{x-\mu}{\sigma}\right)$, плотность $p(x)$ такова, что $x^{1+\varepsilon}p(x) \to \infty$ при $x \to \infty$ для любого $\varepsilon > 0$ (например, годится $p(x) = \frac{1}{2}(1+|x|)^{-1}[1+\ln(1+|x|)]^{-2}$). Тогда для $n > 1$ глобальной ОМП не существует.

Доказательство. Как всегда, $x_{(1)} \leqslant x_{(2)} \leqslant \ldots \leqslant x_{(n)}$ — вариационный ряд, построенный по реализации $\boldsymbol{x} = (x_1,\ldots,x_n)$. Возьмем $c \neq 0$ такое, что $p(c) > 0$. Положим $\mu = x_{(1)} - c\sigma$. Тогда

$$f(\boldsymbol{x},\mu,\sigma) =$$
$$= \sigma^{-n}\prod_{i=1}^{n}p((x_i-\mu)/\sigma) = p(c)\,\sigma^{-n}\prod_{i=2}^{n}p(c+(x_{(i)}-x_{(1)})/\sigma) =$$
$$= p(c)\prod_{i=2}^{n}\left\{\frac{[c+(x_{(i)}-x_{(1)})/\sigma]^{n/(n-1)}\,p(c+(x_{(i)}-x_{(1)})/\sigma)}{(c\sigma+x_{(i)}-x_{(1)})^{n/(n-1)}}\right\}.$$

Для $n > 1$ множитель, заключенный в фигурные скобки, стремится к бесконечности при $\sigma \to 0$, независимо от того, равно ли $x_{(i)}$ величине $x_{(1)}$ или нет. Следовательно, $f(\boldsymbol{x},\mu,\sigma) \to \infty$.

В [61, с. 72] доказано *обратное утверждение*: если плотность $p(x)$ ограничена и непрерывна, а функция $|x|^{1+\varepsilon}p(x)$ ограничена при некотором $\varepsilon > 0$, то с вероятностью 1 глобальная ОМП $(\tilde{\mu}_n,\tilde{\sigma}_n)$ существует при достаточно большом n и $(\tilde{\mu}_n,\tilde{\sigma}_n) \xrightarrow{\text{п. н.}} (\mu,\sigma)$ при $n \to \infty$. ■

Для заданной статистической модели координаты точки максимума функции правдоподобия $L(\boldsymbol{\theta})$ зависят от реализации $x_1,\ldots,x_n$, т. е. $\tilde{\theta}_j = \tilde{\theta}_j(x_1,\ldots,x_n)$, $j = 1,\ldots,m$. Подставив вместо аргументов в эти функции компоненты выборки $(X_1,\ldots,X_n)$, получим случайные величины $\tilde{\theta}_j(X_1,\ldots,X_n)$. Какими свойствами обладают такие оценки?

Для простоты далее в этом параграфе ограничимся случаем скалярного параметра (векторный случай рассматривается, скажем, в [11, с. 237] или [50, с. 379]).

Предположим, что помимо условий R1 – R4 из § 3 выполнены **дополнительные условия регулярности**:

R5) *распределения F_θ различны при разных $\theta \in \Theta$;*

R6) *плотность $f(x,\theta)$ при каждом $x \in A$ трижды непрерывно дифференцируема по θ;*

R7) *$\int f(x,\theta)\,dx$ можно дважды дифференцировать по параметру θ под знаком интеграла;*

R8) *существует функция $h(x)$ такая, что при всех $x \in A$ $\left|\dfrac{\partial^3}{\partial\theta^3}\ln f(x,\theta)\right| \leqslant h(x)$ для всех $\theta \in \Theta$ и $\mathbf{M}h(X_1) < \infty$.*

Теорема 1. При выполнении условий R1–R8 для достаточно большого n существует с вероятностью 1 решение $\tilde\theta = \tilde\theta_n$ уравнения правдоподобия

$$\sum_{i=1}^{n} \frac{\partial}{\partial\theta}\ln f(X_i,\theta) = 0, \tag{3}$$

дающее *сильно состоятельную* оценку: $\tilde\theta_n \xrightarrow{\text{п. н.}} \theta$ при $n \to \infty$, причем $\sqrt{n}\,(\tilde\theta_n - \theta) \xrightarrow{d} \xi \sim \mathcal{N}(0, 1/I_1(\theta))$, где $I_1(\theta)$ — информация Фишера случайной величины X_1 (см. § 3).

Доказательство этой теоремы можно найти в учебнике [69, с. 205].

Теорема 2. Пусть выполнены условия R1–R8 и $\widehat\theta$ — некоторая асимптотически нормальная оценка:

$$\sqrt{n}\,(\widehat\theta_n - \theta) \xrightarrow{d} \zeta \sim \mathcal{N}\left(0, \sigma^2(\theta)\right) \quad \text{при } n \to \infty,$$

где асимптотическая дисперсия $\sigma^2(\theta)$ непрерывно зависит от θ. Тогда

$$\sigma^2(\theta) \geqslant 1/I_1(\theta) \quad \text{при всех } \theta \in \Theta. \tag{4}$$

На основании неравенства (4) заключаем, что ОМП (при выполнении условий R1 – R8) имеет наименьшую возможную асимптотическую дисперсию, равную $1/I_1(\theta)$. Это свойство называется *асимптотической эффективностью*.

Ле Кам в 1953 г. доказал, что для разрывных $\sigma^2(\theta)$ неравенство (4) может нарушаться не более, чем на множестве лебеговой меры нуль. Соответствующий контрпример приведен в задаче 7.

§ 5. МЕТОД НЬЮТОНА И ОДНОШАГОВЫЕ ОЦЕНКИ

Для численного решения нелинейного уравнения $\varphi(x) = 0$ можно использовать *метод Ньютона (метод касательных)*, который состоит в следующем. Прежде всего, задается начальное приближение x_0. Затем вычисляются значения x_{k+1}, $k = 0,1,\ldots$, по формуле

$$x_{k+1} = x_k - \varphi(x_k)/\varphi'(x_k), \tag{5}$$

которая получается в результате проведения касательной к графику $y = \varphi(x)$ в точке $(x_k,\varphi(x_k))$ (рис. 9). Действительно, полагая $y = 0$ в уравнении касательной $y = \varphi(x_k) + \varphi'(x_k)(x - x_k)$, находим, что координата x_{k+1} точки ее пересечения с осью абсцисс удовлетворяет соотношению

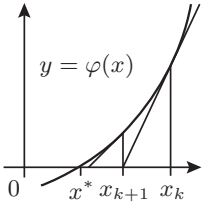


Рис. 9

$$\varphi'(x_k)(x_{k+1} - x_k) = -\varphi(x_k), \tag{6}$$

откуда (при $\varphi'(x_k) \neq 0$) следует итерационная формула (5).

Замечание. Для нелинейной системы уравнений $\boldsymbol{\varphi}(\boldsymbol{x}) = \mathbf{0}$ аналогом уравнения (6) будет линейная система $\boldsymbol{\varphi}'(\boldsymbol{x}_k)(\boldsymbol{x}_{k+1} - \boldsymbol{x}_k) = -\boldsymbol{\varphi}(\boldsymbol{x}_k)$, где $\boldsymbol{\varphi}'(\boldsymbol{x}) = ||\partial\boldsymbol{\varphi}_i(\boldsymbol{x})/\partial\boldsymbol{x}_j||$ — матрица Якоби. Решив линейную систему методом Гаусса (см. [6, с. 137]), найдем очередной шаг $\Delta\boldsymbol{x}_{k+1} = \boldsymbol{x}_{k+1} - \boldsymbol{x}_k$.

К. Ф. Гаусс (1777–1855), немецкий математик.

Оценку скорости сходимости приближений x_k в скалярном случае дает

Теорема 3. Пусть в δ-окрестности корня x^* уравнения $\varphi(x) = 0$ функция φ дважды непрерывно дифференцируема, $0 < \alpha \leqslant |\varphi'(x)|$, $|\varphi''(x)| \leqslant \beta$. Положим $\varepsilon = \min\{\delta/2, \alpha/\beta\}$. Тогда, если $|x_0 - x^*| \leqslant \varepsilon$, то при всех $k \geqslant 0$ выполняются неравенства $|x_{k+1} - x^*| \leqslant |x_{k+1} - x_k|$ и $|x_{k+1} - x^*| \leqslant \delta^{-1}|x_k - x^*|^2$.

Доказательство теоремы 3 приведено в [6, с. 107], обобщение ее на многомерный случай можно найти в [77, с. 192].

Первое неравенство позволяет оценить погрешность текущего приближения через предыдущие.

Второе неравенство показывает, что при выборе начального приближения x_0 из достаточно малой окрестности *простого корня*, т. е. такого, что $\varphi'(x^*) \neq 0$, метод Ньютона сходится *квадратично* (существенно быстрее геометрической прогрессии). Это означает, что на каждой итерации число верных знаков приближения примерно удваивается.

Если x_0 взято достаточно далеко от корня x^*, процесс может расходиться, в частности, могут возникать так называемые «осцилляции» (рис. 10).

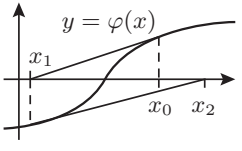


Рис. 10

Приведем пример алгоритма, который является частным случаем метода Ньютона.

Пример 10. Задачу численного извлечения квадратного корня из числа $a > 0$ можно представить как поиск корня уравнения $\varphi(x) = x^2 - a = 0$. Применим метод касательных. Здесь $\varphi'(x) = 2x$ и $x_{k+1} = x_k - (x_k^2 - a)/(2x_k) = \frac{1}{2}(x_k + a/x_k)$, т. е. получаем известный алгоритм усреднения x_k и a/x_k.

Пример 11. Для модели сдвига закона Коши с плотностью $f(x,\theta) = 1/[\pi(1 + (x - \theta)^2)]$ уравнение правдоподобия (3)

имеет вид

$$\sum_{i=1}^{n} \frac{X_i - \theta}{1 + (X_i - \theta)^2} = 0. \qquad (7)$$

Вопрос 3. а) Всегда ли это уравнение имеет хотя бы один корень? б) Какое наибольшее число корней может быть у него?

В этой модели информация Фишера $I_1(\theta) = \frac{4}{\pi} \int\limits_{-\infty}^{+\infty} \frac{x^2}{(1+x^2)^3}\, dx = \frac{1}{2}$. По теореме 1 асимптотическая дисперсия ОМП $\tilde{\theta}$ равна $1/I_1(\theta) = 2$. Следовательно, относительная эффективность $e_{\tilde{\theta},MED} = \pi^2/8 \approx 1{,}234$. Другими словами, в этой модели ОМП $\tilde{\theta}$ асимптотически точнее выборочной медианы *MED* примерно на 23%.

При поиске решения уравнения правдоподобия (3) в качестве начального приближения $\widehat{\theta}_0$ нередко используется значение оценки, полученной по методу моментов, или любой другой легко вычисляемой оценки (желательно, устойчивой к выделяющимся наблюдениям). Оказывается, для регулярных моделей справедлив следующий интересный статистический результат: достаточно сделать *всего один шаг* по методу Ньютона, начиная с любой асимптотически нормальной оценки, чтобы получить асимптотически эффективную оценку $\widehat{\theta}_1$ (т. е. столь же точную, как ОМП). Сформулируем его более строго.

Теорема 4. Пусть выполнены условия R1 – R8 из § 3 и § 4, а оценка $\widehat{\theta}_0$ такова, что $\sqrt{n}\,(\widehat{\theta}_0 - \theta) \xrightarrow{d} \xi \sim \mathcal{N}(0, \sigma^2(\theta))$ при $n \to \infty$. Тогда *одношаговая оценка* $\widehat{\theta}_1 = \widehat{\theta}_0 - \varphi(\widehat{\theta}_0)/\varphi'(\widehat{\theta}_0)$, где $\varphi(\theta) = \dfrac{\partial}{\partial\theta} \ln L(\theta) = \sum\limits_{i=1}^{n} \dfrac{\partial}{\partial\theta} \ln f(x_i,\theta)$, обладает асимптотической дисперсией $1/I_1(\theta)$.

Доказательство теоремы см. в [50, с. 375].

Пример 12. Возьмем для модели сдвига закона Коши из примера 11 в качестве начальной оценки $\widehat{\theta}_0$ выборочную медиану *MED*. Из теоремы 4 и теоремы 1 гл. 7 имеем, что одношаговая оценка $\widehat{\theta}_1$ с

$$\varphi(\theta) = -2 \sum_{i=1}^{n} \frac{X_i - \theta}{1 + (X_i - \theta)^2} \quad \text{и} \quad \varphi'(\theta) = 2 \sum_{i=1}^{n} \frac{1 - (X_i - \theta)^2}{\left(1 + (X_i - \theta)^2\right)^2}$$

является асимптотически эффективной для параметра сдвига θ.

Пример 13 [50, с. 377]**.** Пусть случайные величины X_i имеют плотность $f(x,\theta) = (1-\theta)\, p_1(x) + \theta\, p_2(x)$, где плотности p_1 и p_2 известны, а вес $\theta \in [0,1]$ — нет. Допустим, что у распределений с плотностями $p_1(x)$ и $p_2(x)$ математические ожидания μ_1 и μ_2 различны, а дисперсии — конечны.

Поиск ОМП приводит к уравнению степени $n-1$ относительно θ:

$$\sum_{i=1}^{n} \frac{p_2(X_i) - p_1(X_i)}{(1-\theta)\, p_1(X_i) + \theta\, p_2(X_i)} = 0,$$

Благодарение Всевышнему, что нужное Он сделал нетрудным, а трудное — ненужным.
Г. Сковорода

решить которое при больших n довольно сложно. Применение теоремы 4 позволяет, не решая его, асимптотически эффективно оценить θ. Для этого достаточно использовать в качестве $\widehat{\theta}_0$ оценку метода моментов, которая находится из уравнения $(1-\theta)\mu_1+\theta\mu_2=\overline{X}$: $\widehat{\theta}_0=(\overline{X}-\mu_1)/(\mu_2-\mu_1)$. При сделанных предположениях $\widehat{\theta}_0$ асимптотически нормальна в силу центральной предельной теоремы.

§ 6. МЕТОД СПЕЙСИНГОВ

Познакомимся теперь с оригинальным методом, предложенным Ченом и Амином в 1983 г. (см. [15, с. 90]). Оценки, полученные этим методом, асимптотически эффективны при выполнении условий регулярности и оказываются состоятельными даже для тех моделей, где глобальной ОМП не существует.

Рассмотрим $x_{(1)}\leqslant x_{(2)}\leqslant\ldots\leqslant x_{(n)}$ — вариационный ряд, построенный по реализации выборки из закона с функцией распределения $F(x,\boldsymbol{\theta})$ и плотностью $f(x,\boldsymbol{\theta})$, $\boldsymbol{\theta}\in\Theta$, носителем которой является некоторый интервал (μ_1,μ_2). Параметры μ_1 и μ_2 могут быть неизвестны. В этом случае их следует считать компонентами вектора $\boldsymbol{\theta}$. Положим $x_{(0)}=\mu_1$ и $x_{(n+1)}=\mu_2$. *Спейсингами* называются величины

$$D_i=\int\limits_{x_{(i-1)}}^{x_{(i)}} f(x,\boldsymbol{\theta})\,dx=F(x_{(i)},\boldsymbol{\theta})-F(x_{(i-1)},\boldsymbol{\theta}),\quad i=1,\ldots,n+1.$$

Метод спейсингов рекомендует в качестве оценки векторного параметра $\boldsymbol{\theta}$ взять такую статистику $\breve{\boldsymbol{\theta}}$, которая максимизирует произведение $G(\boldsymbol{\theta})=\prod\limits_{i=1}^{n+1}D_i$ или, что то же самое, максимизирует функцию $H(\boldsymbol{\theta})=\ln G(\boldsymbol{\theta})$.

Мотивировкой метода служит то обстоятельство, что в силу условия $\sum\limits_{i=1}^{n+1}D_i=1$ максимум функции G достигается, когда все D_i одинаковы (задача 8). Выбор значения $\boldsymbol{\theta}$, которое «уравнивает» спейсинги, является разумным ввиду того, что при истинном $\boldsymbol{\theta}$ величины $D_i(X_{(i)},X_{(i-1)},\boldsymbol{\theta})$ представляют собой в силу метода обратной функции из § 1 гл. 4 *одинаково распределенные* спейсинги с равномерным распределением на отрезке $[0,1]$ (см. § 4 гл. 4).

Пример 14. Для модели сдвига показательного закона с плотностью $f(x,\theta)=e^{-(x-\theta)}I_{\{x\geqslant\theta\}}$ из примера 6 легко установить, что

$$H(\theta)=\ln\left(1-e^{\theta-X_{(1)}}\right)+\theta n+\sum_{i=2}^{n+1}\ln\left(e^{-X_{(i-1)}}-e^{-X_{(i)}}\right).$$

Из уравнения $H'(\theta)=0$ находим $\breve{\theta}=X_{(1)}-\ln(1+\frac{1}{n})=$ $=X_{(1)}-\frac{1}{n}+O(n^{-2})$ при $n\to\infty$. Сравнивая ее с ОМП $X_{(1)}$, видим,

что смещение уменьшилось до величины порядка n^{-2} (см. решение задачи 3 гл. 6).

Сопоставим поведение логарифма функции правдоподобия $\ln L(\boldsymbol{\theta}) = \sum\limits_{i=1}^{n} \ln f(X_{(i)},\theta)$ и поведение функции $H(\boldsymbol{\theta}) = \sum\limits_{i=1}^{n+1} \ln D_i$. На основании теоремы Лагранжа о среднем можно записать, что

$$\ln D_i = \ln f(X_{(i)},\theta) + \ln\left(X_{(i)} - X_{(i-1)}\right) + R_i(X_{(i)},X_{(i-1)},\theta).$$

В случае, когда μ_1 и μ_2 известны, остаточный член R_i хоть и зависит от θ, но имеет величину $O(X_{(i)} - X_{(i-1)})$. Так как $X_{(i)} - X_{(i-1)} \xrightarrow{\mathsf{P}} 0$ при $n \to \infty$, вклад остаточного члена в $\ln D_i$ становится пренебрежимо малым, и поведение $\partial H/\partial\theta$ не отличается от $\partial \ln L/\partial\theta$. Это приводит к асимптотической эффективности оценок метода спейсингов.

Если же μ_1 (или μ_2) неизвестен, то вклад R_1 (или R_{n+1}) уже не стремится к нулю при $n \to \infty$. Это приводит к различному поведению $\ln L$ и H: $\ln L$ может быть, скажем, не ограничена сверху, в то время как $H \leqslant 0$, поскольку $0 \leqslant D_i \leqslant 1$.

Пример 15 [15, с. 92]**.** Для выборки из сдвинутого на μ распределения Вейбулла—Гнеденко с функцией распределения $F_{\mu,\alpha}(x) = \left[1 - e^{-(x-\mu)^{\alpha}}\right] I_{\{x>\mu\}}$, где параметры μ и $\alpha > 0$ неизвестны, логарифм функции правдоподобия имеет следующий вид:

$$\ln L(\mu,\alpha) = n \ln\alpha + (\alpha - 1) \sum_{i=1}^{n} \ln(X_{(i)} - \mu) - \sum_{i=1}^{n} (X_{(i)} - \mu)^{\alpha}.$$

При $\alpha < 1$ и $\mu \to X_{(1)}$ эта функция стремится к $+\infty$.

В свою очередь, метод спейсингов позволяет получить *состоятельные* оценки параметров μ и α; правда, решать систему уравнений $\partial H/\partial\mu = \partial H/\partial\alpha = 0$ приходится численно.

Другие примеры применения метода спейсингов встречаются в задачах 4 и 5.

ЗАДАЧИ

1. Вычислите информацию Фишера $I_1(\theta)$ для случайной величины X_1, имеющей распределение Бернулли с неизвестной вероятностью «успеха» θ.
УКАЗАНИЕ. Выразите $U_1 = (\partial/\partial\theta) \ln f(X_1,\theta)$ через X_1.

Числом поболее, ценою подешевле.

Чацкий в «Горе от ума» А. С. Грибоедова

2. Пусть $X_i \sim \mathcal{N}(\mu,\sigma^2)$. Найдите ОМП $\tilde{\mu}$ и $\tilde{\sigma}$. (Почему найденное решение — это точка максимума функции правдоподобия, а не, скажем, седловая точка?)

3. Случайные величины X_i имеют функцию распределения $F\left(\frac{x-\mu}{\sigma}\right)$, где $F(x) = (1-e^{-x})\,I_{\{x\geqslant 0\}}$ (*модель сдвига-масштаба показательного закона*). Найдите ОМП параметров μ и σ.
Указание. Учтите, что плотность разрывна в точке μ.

4. Для равномерного распределения на отрезке $[\mu_1, \mu_2]$ вычислите
а) ОМП,
б) оценки по методу спейсингов.

5. Случайные величины X_i имеют плотность $p(x,\theta) = \frac{1}{2}\,e^{-|x-\theta|}$ (*сдвиг распределения Лапласа*). Как устроено множество, на котором функция правдоподобия максимальна для
а) четного,
б) нечетного размера выборки?
Указание. Сравните с задачей 4 гл. 7.

6*. Для логнормальной модели с плотностью

$$p_{\mu,\sigma}(x) = \frac{1}{x\sigma\sqrt{2\pi}}\,\exp\left\{-\frac{1}{2\sigma^2}\,(\ln x-\mu)^2\right\} I_{\{x>0\}}$$

найдите оценки параметров μ и σ по методу моментов.

7*. Для выборки из закона $\mathcal{N}(\theta,1)$ рассмотрим оценку

$$\widehat{\theta} = \begin{cases} \overline{X}, & \text{если } |\overline{X}| \geqslant a_n, \\ b\overline{X}, & \text{если } |\overline{X}| < a_n, \end{cases}$$

где $|b|<1$, $0<a_n\to 0$, но $a_n\sqrt{n}\to\infty$ при $n\to\infty$. Вычислите асимптотическую дисперсию этой оценки.

8. С помощью метода Лагранжа проверьте, что максимум функции $\prod\limits_{i=1}^{n+1} D_i$ при условиях $\sum\limits_{i=1}^{n+1} D_i = 1,\ 0\leqslant D_i\leqslant 1,\ i=1,\ldots,n+1$, достигается, когда $D_1=\ldots=D_{n+1}=1/(n+1)$.

РЕШЕНИЯ ЗАДАЧ

Принимаясь за дело, соберись с духом.

Козьма Прутков

1. $U_1 = X_1/\theta - (1-X_1)/(1-\theta) = (X_1-\theta)/[\theta(1-\theta)]$. Отсюда

$$I_1(\theta) = \mathbf{M}U_1^2 = [\theta(1-\theta)]^{-2}\mathbf{D}X_1 = [\theta(1-\theta)]^{-1}.$$

Этот результат можно получить и без вычислений: согласно примеру 5, частота $\overline{X}$ является ОМП в этой модели; в силу центральной предельной теоремы $\sqrt{n}\,(\overline{X}-\theta)\xrightarrow{d}\xi\sim\mathcal{N}(0,\theta(1-\theta))$ при $n\to\infty$; наконец, согласно теореме 1, асимптотическая дисперсия $\theta(1-\theta)$ равна $1/I_1(\theta)$.

2. $\ln L(\mu,\sigma) = -n\ln\sqrt{2\pi} - n\ln\sigma - \frac{1}{2\sigma^2}\sum\limits_{i=1}^{n}(X_i-\mu)^2$. Необходимым условием экстремума этой функции является равенство нулю

частных производных:

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{2\sigma^2} \sum_{i=1}^{n} 2\,(X_i - \mu) = \frac{n}{\sigma^2}\,\big(\overline{X} - \mu\big) = 0,$$
$$\frac{\partial \ln L}{\partial \sigma} = -\,\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^{n} (X_i - \mu)^2 = 0.$$

Первое уравнение дает $\tilde{\mu} = \overline{X}$. Подставив $\tilde{\mu}$ во второе уравнение, находим, что $\tilde{\sigma} = S$, где $S^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2$ — *выборочная дисперсия*.

Убедимся, что найденное решение является точкой максимума функции L, а не, скажем, точкой минимума или седловой.

Подробности малейшей не забуду.

Чацкий в «Горе от ума»
А. С. Грибоедова

Для этого вычислим производные второго порядка:

$$\frac{\partial^2 \ln L}{\partial \mu^2} = -\,\frac{n}{\sigma^2}\,, \quad \frac{\partial^2 \ln L}{\partial \sigma^2} = \frac{n}{\sigma^2} - \frac{3}{\sigma^4}\sum_{i=1}^{n}(X_i - \mu)^2,$$
$$\frac{\partial^2 \ln L}{\partial \mu \partial \sigma} = -\,\frac{2n}{\sigma^3}\,(\overline{X} - \mu).$$

Учитывая, что при $n > 1$ с вероятностью 1 статистика $S^2 > 0$, эти производные в точке $(\mu,\sigma) = (\overline{X}, S)$ будут иметь значения $-n/S^2 < 0$, $-2n/S^2 < 0$ и 0 соответственно. Поскольку определитель $(\partial^2 \ln L/\partial\mu^2)(\partial^2 \ln L/\partial\sigma^2) - (\partial^2 \ln L/\partial\mu\partial\sigma)^2 > 0$, то выполняются достаточные условия максимума (см. [46, с. 195]).

3. В данной модели $\boldsymbol{\theta} = (\mu,\sigma) \in \Theta = \mathbb{R} \times (0, +\infty)$,

$$L = \exp\left\{-n\ln\sigma - \frac{1}{\sigma}\sum_{i=1}^{n}(X_i - \mu)\right\} I_{\{X_{(1)} \geqslant \mu\}}.$$

При любом $\sigma > 0$ эта функция возрастает по μ до точки $X_{(1)} = \min\{X_1, \ldots, X_n\}$, а затем обращается в 0. Рассмотрим поведение функции $\ln L$ в сечении $\tilde{\mu} = X_{(1)}$. Приравнивая частную производную нулю, получаем уравнение

$$\frac{\partial \ln L}{\partial \sigma} = -\,\frac{n}{\sigma} + \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - X_{(1)}) = 0,$$

решением которого является статистика $\tilde{\sigma} = \overline{X} - X_{(1)}$. Заметим, что $X_{(1)}$ оценивает сдвиг μ, а $\overline{X}$ служит оценкой для $\mathbf{M}X_1 = \mu + \sigma$ (рис. 11).

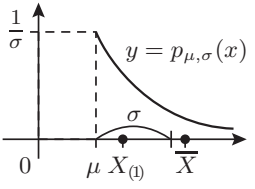


Рис. 11

4. а) Функцию правдоподобия можно записать так:

$$L = (\mu_2 - \mu_1)^{-n} \prod_{i=1}^{n} I_{\{\mu_1 \leqslant X_i \leqslant \mu_2\}} = (\mu_2 - \mu_1)^{-n} I_{\{\mu_1 \leqslant X_{(1)}\}} I_{\{X_{(n)} \leqslant \mu_2\}}.$$

Максимизация L приводит к ОМП $(\tilde{\mu}_1, \tilde{\mu}_2) = (X_{(1)}, X_{(n)})$.

б) В этой модели функция H имеет следующий вид:

$$H(\mu_1,\mu_2) = \ln(X_{(1)} - \mu_1) + \sum_{i=2}^{n} \ln(X_{(i)} - X_{(i-1)}) + \\ + \ln(\mu_2 - \ln X_{(n)}) - (n+1)\ln(\mu_2 - \mu_1).$$

Приравнивая частные производные нулю, получаем систему

$$\frac{\partial H}{\partial \mu_1} = -\frac{1}{X_{(1)} - \mu_1} + \frac{n+1}{\mu_2 - \mu_1} = 0, \quad \frac{\partial H}{\partial \mu_2} = \frac{1}{\mu_2 - X_{(n)}} - \frac{n+1}{\mu_2 - \mu_1} = 0,$$

решение $(\breve{\mu}_1, \breve{\mu}_2)$ которой нетрудно найти:

$$\breve{\mu}_1 = X_{(1)} - \frac{1}{n-1}(X_{(n)} - X_{(1)}), \quad \breve{\mu}_2 = X_{(n)} + \frac{1}{n-1}(X_{(n)} - X_{(1)}).$$

В отличие от ОМП эти оценки не имеют смещения.

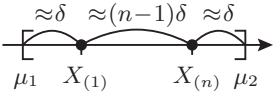


Рис. 12

Действительно, среднее расстояние δ между любыми соседними из n взятых наудачу из $[\mu_1, \mu_2]$ точек, очевидно, равно $(\mu_2 - \mu_1)/(n+1)$. Поэтому математическое ожидание *размаха выборки* $X_{(n)} - X_{(1)}$ есть $(n-1)\delta$ (рис. 12). Следовательно, $\frac{1}{n-1}(X_{(n)} - X_{(1)})$ — это поправка, устраняющая смещение оценок $X_{(1)}$ и $X_{(n)}$.

5. а) Для модели сдвига распределения Лапласа

$$\ln L(\theta) = -n\ln 2 - \sum_{i=1}^{n} |X_i - \theta|.$$

Максимизация этой функции равносильна поиску такого значения θ, при котором сумма расстояний от него до всех элементов выборки X_i была бы минимальной. Пусть $n = 2k$. Покажем, что в качестве решения годится любое θ из $[X_{(k)}, X_{(k+1)}]$. Минимизируем сначала сумму расстояний до $X_{(k)}$ и $X_{(k+1)}$. Она равна $\Delta_1 = X_{(k+1)} - X_{(k)}$ для $\theta \in [X_{(k)}, X_{(k+1)}]$ и больше Δ_1 для $\theta \notin [X_{(k)}, X_{(k+1)}]$. Добавим точки $X_{(k-1)}$ и $X_{(k+2)}$ (рис. 13). Сумма расстояний до них при $\theta \in [X_{(k-1)}, X_{(k+2)}]$ равна $\Delta_2 = X_{(k+2)} - X_{(k-1)}$ и больше Δ_2 вне этого отрезка. Очевидно, что только θ из $[X_{(k)}, X_{(k+1)}] \subseteq [X_{(k-1)}, X_{(k+2)}]$ минимизируют обе суммы сразу. Далее рассуждаем аналогично.

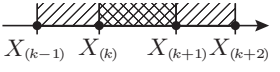


Рис. 13

б) «Склеив» точки $X_{(k)}$ и $X_{(k+1)}$, видим, что в случае выборки нечетного размера ОМП будет выборочная медиана *MED*.

Интересно, что функция H метода спейсингов из § 6 имеет единственную точку максимума $MED = \frac{1}{2}(X_{(k)} + X_{(k+1)})$ и в случае выборки четного размера (убедитесь!).

6. Вычислим сначала момент k-го порядка

$$\alpha_k = \mathbf{M}X_1^k = \frac{1}{\sigma\sqrt{2\pi}} \int_0^{+\infty} x^{k-1} \exp\left\{-\frac{(\ln x - \mu)^2}{2\sigma^2}\right\} dx.$$

Сделаем замену $y = \ln x$ и выделим полный квадрат:

$$\alpha_k = \frac{1}{\sigma\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} \exp\left\{ky - \frac{(y-\mu)^2}{2\sigma^2}\right\} dy =$$

$$= \exp\left\{k\mu + \frac{k^2\sigma^2}{2}\right\} \int\limits_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(y-\mu-\sigma^2 k)^2}{2\sigma^2}\right\} dy.$$

Заметив, что справа стоит интеграл по всей прямой от плотности распределения $\mathcal{N}(\mu+\sigma^2 k, \sigma^2)$, получим $\alpha_k = \exp\{k\mu + k^2\sigma^2/2\}$.

Приравнивая первый и второй теоретические моменты выборочным, получим систему уравнений

$$\exp\{\mu + \sigma^2/2\} = A_1, \quad \exp\{2\mu + 2\sigma^2\} = A_2.$$

Нетрудно убедиться, что она имеет решение

$$\widehat{\mu} = 2\ln A_1 - (\ln A_2)/2, \quad \widehat{\sigma} = (\ln A_2 - 2\ln A_1)^{1/2}.$$

7. В этой модели $\overline{X} \sim \mathcal{N}(\theta, 1/n)$. Поэтому при $n \to \infty$

$$\mathbf{P}(|\overline{X}| \geqslant a_n) = \Phi(\sqrt{n}\,(\theta - a_n)) + \Phi(-\sqrt{n}\,(\theta + a_n)) \to I_{\{\theta \neq 0\}}$$

($\Phi(x)$ — функция распределения закона $\mathcal{N}(0,1)$.) Так как

$$\mathbf{P}(\sqrt{n}\,(\widehat{\theta} - \theta) \leqslant x) = \mathbf{P}(\sqrt{n}\,(\overline{X} - \theta) \leqslant x, |\overline{X}| \geqslant a_n) +$$
$$+ \mathbf{P}(\sqrt{n}\,(b\overline{X} - \theta) \leqslant x, |\overline{X}| < a_n),$$

то при $\theta \neq 0$ и $\theta = 0$ имеем, соответственно,

$$\sqrt{n}\,(\widehat{\theta} - \theta) \xrightarrow{d} \xi \sim \mathcal{N}(0,1) \quad \text{и} \quad \sqrt{n}\,\widehat{\theta} \xrightarrow{d} \xi' \sim \mathcal{N}(0, b^2).$$

При $\theta = 0$ асимптотическая дисперсия $\sigma^2(\theta) = b^2 < 1 = I_1(\theta)$, и неравенство (4) нарушается, несмотря на гладкость модели.

8. Максимизируем $H = \sum \ln D_i$. Для этого составим *функцию Лагранжа* (см. [46, с. 271]): $F = \sum \ln D_i - \lambda\,(\sum D_i - 1)$. Приравнивая нулю частные производные функции F, запишем систему уравнений для поиска экстремальных точек:

$$\frac{\partial F}{\partial D_i} = 1/D_i - \lambda = 0, \ i = 1, \ldots, n+1; \quad \frac{\partial F}{\partial \lambda} = \sum D_i - 1 = 0.$$

Из первых $n+1$ уравнения находим, что $D_i = 1/\lambda$. Подставляя в последнее уравнение, получим $\lambda = n+1$, откуда $D_i = 1/(n+1)$.

ОТВЕТЫ НА ВОПРОСЫ

1. **а)** Случайная величина $I_{\{X_i \leqslant x\}}$ имеет распределение Бернулли с вероятностью «успеха» $F(x)$.

б) Схема Бернулли с вероятностью «успеха» $F(x)$.

в) Для фиксированного x величина $\widehat{F}_n(x)$ есть частота «успехов» — попаданий левее x.

г) По усиленному закону больших чисел (П6) *эмпирическая* функция распределения $\widehat{F}_n(x)$ сходится с вероятностью 1 к *теоретической* функция распределения $F(x)$.

2. Положим $r = l/k > 1$, $\eta = |\xi|^k$. Применяя неравенство Иенсена к функции $\varphi(x) = |x|^r$, находим, что $|\mathsf{M}\eta|^r \leqslant \mathsf{M}|\eta|^r$, т. е. $(\mathsf{M}|\xi|^k)^{l/k} \leqslant \mathsf{M}|\xi|^l$, что и требовалось доказать.

3. **а)** Левая часть уравнения (7) — сумма непрерывных функций, каждая из которых положительна при $\theta < X_{(1)}$ и отрицательна при $\theta > X_{(n)}$. Следовательно, хотя бы один корень всегда есть.
б) Если привести слагаемые к общему знаменателю, то в числителе появится многочлен степени $2n-1$. Поэтому максимальное число решений уравнения (7) равно $2n-1$.

Математика дает наиболее чистое и непосредственное переживание истины; на этом покоится ее ценность для общего образования людей.

Макс Лауэ

Красивый результат для этой модели получил Дж. Ридс в 1980 г. (см. [50, с. 376]): если $2K+1$ есть число корней уравнения (7), то K сходится по распределению к закону Пуассона с параметром $1/\pi$ при $n \to \infty$ (см. § 1 гл. 5).

Удивительно, но для более сложной модели, когда присутствует не только параметр сдвига, но и параметр масштаба, решение $(\tilde{\mu},\tilde{\sigma})$ системы уравнений (2) для закона Коши *всегда единственно*. Это установил Дж. Копас в 1975 г. (см. [50, с. 387]).

Глава 10

ДОСТАТОЧНОСТЬ

§ 1. ДОСТАТОЧНЫЕ СТАТИСТИКИ

По одной капле воды... человек, умеющий мыслить логически, может сделать вывод о возможности существования Атлантического океана или Ниагарского водопада.

А. Конан Дойл, «Этюд в багровых тонах»

Некоторые статистические модели допускают сжатие информации — замену выборки $\boldsymbol{X} = (X_1, \ldots, X_n)$ размера n на статистику $T(\boldsymbol{X})$, которая «эквивалентна» всей выборке в задаче оценивания неизвестного параметра θ.

Рассмотрим для примера схему Бернулли с неизвестной вероятностью «успеха» θ: $\mathsf{P}(X_i = x_i) = \theta^{x_i}(1-\theta)^{1-x_i}$, где $x_i = 0$ или 1. Выборка $\boldsymbol{X}$ имеет совместное распределение

$$f(\boldsymbol{x}, \theta) = \theta^t (1-\theta)^{n-t}, \quad \text{где} \quad \boldsymbol{x} = (x_1, \ldots, x_n), \quad t = T(\boldsymbol{x}) = \sum_{i=1}^{n} x_i.$$

Как было установлено в примере 5 гл. 9, оценкой максимального правдоподобия для схемы Бернулли является частота $\overline{X} = T(\boldsymbol{X})/n$ — функция от T. Найдем условное распределение (см. П7) выборки $\boldsymbol{X}$ при условии $\{T(\boldsymbol{X}) = t\}$, $t = 0, 1, \ldots, n$:

$$\mathsf{P}(\boldsymbol{X} = \boldsymbol{x} \,|\, T(\boldsymbol{X}) = t) = \frac{\theta^t (1-\theta)^{n-t}}{C_n^t \, \theta^t (1-\theta)^{n-t}} = \frac{1}{C_n^t},$$

если точка $\boldsymbol{x}$ такова, что $\sum x_i = t$, иначе вероятность равна 0. Заметим, что это условное распределение *не зависит от* θ.

Определение. Статистика $\boldsymbol{T}(\boldsymbol{X}) = (T_1(\boldsymbol{X}), \ldots, T_m(\boldsymbol{X}))$ в дискретной модели называется *достаточной*, если для всех $\theta \in \Theta$, $\boldsymbol{x} \in \mathbb{R}^n$ и любых возможных значений $\boldsymbol{t} = (t_1, \ldots, t_m)$ условная вероятность $\mathsf{P}(\boldsymbol{X} = \boldsymbol{x} \,|\, \boldsymbol{T}(\boldsymbol{X}) = \boldsymbol{t})$ не зависит от θ.

Это понятие было введено Р. Фишером в 1922 г.

Оказывается, достаточная статистика содержит точно такую же информацию (см. задачу 6) о значении параметра θ, что и вся выборка. Чтобы убедиться в этом, заметим, что моделирование выборки можно разбить на следующие **два этапа**.

1) Розыгрыш значения $\boldsymbol{t}_0$ статистики $\boldsymbol{T}$, имеющей распределение $\mathsf{P}(\boldsymbol{T}(\boldsymbol{X}) = \boldsymbol{t})$. (В схеме Бернулли T распределена по биномиальному закону: $\mathsf{P}(T(\boldsymbol{X}) = t) = C_n^t \, \theta^t (1-\theta)^{n-t}$.)

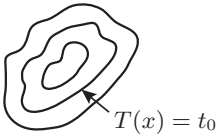


Рис. 1

2) Розыгрыш положения реализации выборки $\boldsymbol{X}$ на множестве $\{\boldsymbol{x}\colon \boldsymbol{T}(\boldsymbol{x}) = \boldsymbol{t}_0\}$ («линии уровня») (рис. 1) в соответствии с условным распределением $\mathsf{P}(\boldsymbol{X} = \boldsymbol{x} \,|\, \boldsymbol{T}(\boldsymbol{X}) = \boldsymbol{t}_0)$. (Для схемы Бернулли на этом этапе надо случайно (равновероятно) расставить t_0 единиц в наборе из нулей и единиц длины n (рис. 2).)

При этом от того, какое значение имеет параметр θ, ввиду достаточности статистики $\boldsymbol{T}$ зависит *только первый этап* — розыгрыш линии уровня.

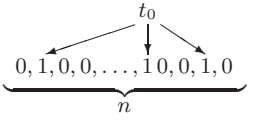


Рис. 2

§ 2. КРИТЕРИЙ ФАКТОРИЗАЦИИ

Как для заданной статистической модели найти достаточную статистику? Ответ на этот вопрос дает

Теорема 1 (критерий факторизации). (Векторная) статистика $\boldsymbol{T}$ в дискретной модели достаточна тогда и только тогда, когда существуют функции g и h такие, что совместная вероятность $f(\boldsymbol{x}, \theta)$ выборки $\boldsymbol{X}$ представляется в виде

$$f(\boldsymbol{x}, \theta) = g(\boldsymbol{T}(\boldsymbol{x}),\theta)\, h(\boldsymbol{x}), \tag{1}$$

Factor (*англ.*) — множитель.

т. е. распадается в произведение двух функций (факторизуется): первая зависит от θ, но от $\boldsymbol{x}$ зависит лишь через $\boldsymbol{T}(\boldsymbol{x})$, а вторая от параметра θ не зависит.

(В частности, для схемы Бернулли можно взять функции $g(t,\theta) = \theta^t(1-\theta)^{n-t}$ и $h(\boldsymbol{x}) \equiv 1$, где $t = T(\boldsymbol{x}) = x_1 + \ldots + x_n$.)

Доказательство [32, с. 55]. Если статистика $\boldsymbol{T}$ достаточна, то при любом $\boldsymbol{x}$ условная вероятность $\mathsf{P}(\boldsymbol{X} = \boldsymbol{x} \,|\, \boldsymbol{T}(\boldsymbol{X}) = \boldsymbol{T}(\boldsymbol{x}))$ не зависит от θ. Возьмем эту вероятность в качестве функции $h(\boldsymbol{x})$. Так как событие $\{\boldsymbol{X} = \boldsymbol{x}\} \subseteq \{\boldsymbol{T}(\boldsymbol{X}) = \boldsymbol{T}(\boldsymbol{x})\}$, то совместная вероятность имеет вид

$$\begin{aligned} f(\boldsymbol{x}, \theta) &= \mathsf{P}(\boldsymbol{X} = \boldsymbol{x}) = \mathsf{P}(\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{T}(\boldsymbol{X}) = \boldsymbol{T}(\boldsymbol{x})) = \\ &= \mathsf{P}(\boldsymbol{T}(\boldsymbol{X}) = \boldsymbol{T}(\boldsymbol{x}))\, \mathsf{P}(\boldsymbol{X} = \boldsymbol{x} \,|\, \boldsymbol{T}(\boldsymbol{X}) = \boldsymbol{T}(\boldsymbol{x})), \end{aligned}$$

где $\mathsf{P}(\boldsymbol{T}(\boldsymbol{X}) = \boldsymbol{T}(\boldsymbol{x}))$ играет роль функции g в представлении (1).

Обратно, пусть имеет место разложение (1). Тогда при любом $\boldsymbol{x}$ таком, что $\boldsymbol{T}(\boldsymbol{x}) = \boldsymbol{t}$, запишем:

$$\begin{aligned} \mathsf{P}(\boldsymbol{X} = \boldsymbol{x} \,|\, \boldsymbol{T}(\boldsymbol{X}) = \boldsymbol{t}) &= \frac{\mathsf{P}(\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{T}(\boldsymbol{X}) = \boldsymbol{t})}{\mathsf{P}(\boldsymbol{T}(\boldsymbol{X}) = \boldsymbol{t})} = \\ &= \frac{\mathsf{P}(\boldsymbol{X} = \boldsymbol{x})}{\mathsf{P}(\boldsymbol{T}(\boldsymbol{X}) = \boldsymbol{t})} = f(\boldsymbol{x}, \theta) \Big/ \sum_{\boldsymbol{x}'\colon \boldsymbol{T}(\boldsymbol{x}') = \boldsymbol{t}} f(\boldsymbol{x}', \theta) = \\ &= g(\boldsymbol{t},\theta)\, h(\boldsymbol{x}) \Big/ \sum_{\boldsymbol{x}'\colon \boldsymbol{T}(\boldsymbol{x}') = \boldsymbol{t}} g(\boldsymbol{t},\theta)\, h(\boldsymbol{x}') = h(\boldsymbol{x}) \Big/ \sum_{\boldsymbol{x}'\colon \boldsymbol{T}(\boldsymbol{x}') = \boldsymbol{t}} h(\boldsymbol{x}'), \end{aligned}$$

т. е. условная вероятность не зависит от θ. Если же $\boldsymbol{x}$ таково, что $\boldsymbol{T}(\boldsymbol{x}) \neq \boldsymbol{t}$, то, очевидно, $\mathsf{P}(\boldsymbol{X} = \boldsymbol{x} \,|\, \boldsymbol{T}(\boldsymbol{X}) = \boldsymbol{t}) = 0$. Таким образом,

при любом $\boldsymbol{x}$ условная вероятность $\mathsf{P}(\boldsymbol{X} = \boldsymbol{x} \,|\, \boldsymbol{T}(\boldsymbol{X}) = \boldsymbol{t})$ не зависит от θ. ■

Замечание. Если $\boldsymbol{T}$ достаточна, то таковой же будет и статистика $\boldsymbol{S} = \varphi(\boldsymbol{T})$, где φ — взаимно однозначная (борелевская) функция (отображение, когда $\boldsymbol{T}$ — векторная статистика). Действительно, в этом случае существует обратная функция (отображение) $\varphi^{-1}\colon \boldsymbol{T} = \varphi^{-1}(\boldsymbol{S})$, и из представления (1) имеем

$$f(\boldsymbol{x},\theta) = g(\varphi^{-1}(\boldsymbol{S}),\theta)\, h(\boldsymbol{x}) = g_1(\boldsymbol{S},\theta)\, h(\boldsymbol{x}).$$

Отсюда в силу теоремы 1 статистика $\boldsymbol{S}$ является достаточной.

Теперь определим понятие достаточности для *непрерывных статистических моделей*. Обозначение $f(\boldsymbol{x},\theta)$ станем использовать для совместной плотности выборки $\boldsymbol{X} = (X_1,\ldots,X_n)$. Тогда для статистики $\boldsymbol{T}(\boldsymbol{x})$, непрерывно зависящей от $\boldsymbol{x}$, верно равенство $\mathsf{P}(\boldsymbol{T}(\boldsymbol{X}) = \boldsymbol{t}) = 0$. Корректное определение условной вероятности $\mathsf{P}(\boldsymbol{X} = \boldsymbol{x} \,|\, \boldsymbol{T}(\boldsymbol{X}) = \boldsymbol{t})$ в этом случае выходит за рамки математического аппарата, используемого в этой книге. Поэтому вместо введения понятия достаточности на языке условных вероятностей,[*)] примем критерий факторизации в качестве определения: будем считать статистику $\boldsymbol{T}$ *достаточной*, если она факторизует плотность $f(\boldsymbol{x},\theta)$ в виде (1).

Пример 1. Для нормальной модели $\mathcal{N}(\mu,\sigma^2)$ с неизвестными параметрами μ и σ (т. е. $\boldsymbol{\theta} = (\mu,\sigma)$) плотностью выборки служит

$$f(\boldsymbol{x},\boldsymbol{\theta}) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2} = \frac{1}{(\sigma\sqrt{2\pi})^n}\, e^{-\frac{1}{2\sigma^2}\left(\sum x_i^2 - 2\mu\sum x_i + n\mu^2\right)}.$$

Из этого представления видно, что статистика $\boldsymbol{T}(\boldsymbol{x}) = (\sum x_i,\, \sum x_i^2)$ является достаточной. Так как компоненты $\boldsymbol{T}$ взаимно однозначно выражаются через $\overline{x} = \frac{1}{n}\sum x_i$ и $s^2 = \frac{1}{n}\sum x_i^2 - (\overline{x})^2$, то вся информация о значениях параметров μ и σ содержится в $\overline{x}$ и s^2.

Пример 2. Для модели сдвига показательного закона совместная плотность $\boldsymbol{X}$ факторизуется так:

$$f(\boldsymbol{x},\theta) = \prod_{i=1}^{n} e^{-(x_i-\theta)} I_{\{x_i \geqslant \theta\}} = e^{n\theta} I_{\{x_{(1)} \geqslant \theta\}} \cdot e^{-\sum x_i}.$$

Таким образом, для оценивания параметра θ достаточно знать значение статистики $X_{(1)} = \min\{X_1,\ldots,X_n\}$.

Вопрос 1. Какая статистика будет достаточной для выборки из равномерного распределения на $[0,\theta]$?

[*)] С таким подходом (и доказательством критерия факторизации для него) можно познакомиться по учебнику [11, с. 431].

§ 3. ЭКСПОНЕНЦИАЛЬНОЕ СЕМЕЙСТВО

Всегда существует тривиальная (векторная) достаточная статистика — вся выборка. Важным является вопрос: для каких статистических моделей существуют достаточные статистики, размерность которых не зависит от длины выборки? Примером класса таких моделей может служить так называемое *экспоненциальное семейство*.

Определение. Модель принадлежит экспоненциальному семейству, если найдутся такие функции $g_0, g_1, \ldots, g_m$ и статистика $\boldsymbol{T} = (T_1, \ldots, T_m)$, что совместная плотность (или вероятность) выборки представляется в следующем виде:

$$f(\boldsymbol{x}, \boldsymbol{\theta}) = \exp\left\{ g_0(\boldsymbol{\theta}) + \sum_{j=1}^{m} g_j(\boldsymbol{\theta})\, T_j(\boldsymbol{x}) \right\} h(\boldsymbol{x}). \tag{2}$$

При этом статистика $\boldsymbol{T}$, очевидно, является достаточной.

Так, нормальная модель из примера 1 удовлетворяет условию (2), если взять

$$g_0(\boldsymbol{\theta}) = -n\left(\ln\sigma + \frac{\mu^2}{2\sigma^2}\right),\ \ g_1(\boldsymbol{\theta}) = \frac{\mu}{\sigma^2},\ \ g_2(\boldsymbol{\theta}) = -\frac{1}{2\sigma^2},\ \ h = (2\pi)^{-n/2}.$$

Некоторые другие распределения из экспоненциального семейства приведены в задаче 2. Отметим, что равномерное распределение на отрезке $[0, \theta]$ и модель сдвига показательного закона из примера 2 в него не входят.

Вопрос 2. Принадлежит ли экспоненциальному семейству модель показательного закона с неизвестным параметром масштаба σ: $F_\sigma(x) = 1 - e^{-x/\sigma}$ при $x \geqslant 0$ (рис. 3)?

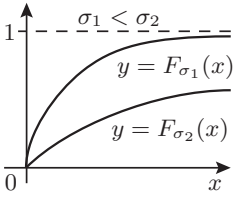


Рис. 3

Замечание. Как доказали Е. Б. Дынкин и Т. С. Фергюсон (см. [50, с. 40]), среди моделей сдвига только нормальный закон и распределение $\mathrm{const} \cdot \ln X$, где случайная величина X имеет гамма-распределение, допускают представление (2) с $m = 1$. В частности, модели сдвига законов Коши и Лапласа не входят в экспоненциальное семейство. Более того: для этих двух законов не существует достаточной статистики, размерность которой меньше длины выборки (см. [50, с. 47]).

В каких еще семействах, помимо экспоненциального, есть достаточные статистики фиксированной размерности? Ответ дает следующая «пессимистическая» теорема из [50, с. 48].

Теорема 2. Предположим, что элементы выборки имеют непрерывную по x плотность $f(x,\theta)$, носитель $A = \{x\colon f(x,\theta) > 0\}$ которой есть интервал, не зависящий от $\theta \in \Theta$. Пусть для совместной плотности выборки существует непрерывная m-мерная достаточная статистика. Тогда

1) если $m = 1$, то найдутся функции g_0, g_1 и h, такие что справедливо равенство (2);

2) если $m > 1$ и $f(x,\theta)$ имеет непрерывную частную производную по x, то существуют функции g_j и h, такие что равенство (2) справедливо с $m' \leqslant m$.

Таким образом, среди гладких семейств распределений с фиксированным носителем *только* экспоненциальное семейство допускает сокращение размерности данных через достаточность. При этом важным условием в теореме 2 является независимость носителя от θ. Если оно не выполняется, то, как показывает пример 2, для любого размера выборки может существовать одномерная достаточная статистика.

Every cloud has a silver lining (*Джон Мильтон*).

≈ Нет худа без добра.

§ 4. УЛУЧШЕНИЕ НЕСМЕЩЕННЫХ ОЦЕНОК

Пусть $\widehat{\theta}$ — несмещенная оценка параметра θ, т. е. $\mathsf{M}\widehat{\theta} = \theta$ для всех $\theta \in \Theta$. Допустим, что в модели существует достаточная статистика $\boldsymbol{T}$. Тогда ее можно использовать для улучшения оценки $\widehat{\theta}$.

Теорема 3 (Колмогоров — Блекуэлл — Рао). Предположим, что дисперсия $\mathsf{D}\widehat{\theta} < \infty$ для всех $\theta \in \Theta$. Рассмотрим оценку $\widehat{\theta}_{\boldsymbol{T}} = \mathsf{M}(\widehat{\theta}(\boldsymbol{X}) \,|\, \boldsymbol{T})$ (см. П7). Тогда $\widehat{\theta}_{\boldsymbol{T}}$ также будет несмещенной оценкой для параметра θ, причем $\mathsf{D}\widehat{\theta}_{\boldsymbol{T}} \leqslant \mathsf{D}\widehat{\theta}$ при всех $\theta \in \Theta$.

Доказательство. Тот факт, что $\widehat{\theta}_{\boldsymbol{T}} = \mathsf{M}(\widehat{\theta}(\boldsymbol{X}) \,|\, \boldsymbol{T})$ является оценкой, т. е. не зависит от θ, следует из достаточности статистики $\boldsymbol{T}$: распределение выборки $\boldsymbol{X}$ при фиксированном значении $\boldsymbol{T}$ от θ не зависит.

Свойство 1 условного математического ожидания (П7) влечет несмещенность оценки $\widehat{\theta}_{\boldsymbol{T}}$: $\mathsf{M}\widehat{\theta}_{\boldsymbol{T}} = \mathsf{M}\mathsf{M}(\widehat{\theta} \,|\, \boldsymbol{T}) = \mathsf{M}\widehat{\theta} = \theta$. Для доказательства неравенства заметим, что

$$\mathsf{D}\widehat{\theta} = \mathsf{M}(\widehat{\theta} - \widehat{\theta}_{\boldsymbol{T}} + \widehat{\theta}_{\boldsymbol{T}} - \theta)^2 =$$
$$= \mathsf{M}(\widehat{\theta} - \widehat{\theta}_{\boldsymbol{T}})^2 + 2\,\mathsf{M}(\widehat{\theta} - \widehat{\theta}_{\boldsymbol{T}})(\widehat{\theta}_{\boldsymbol{T}} - \theta) + \mathsf{D}\widehat{\theta}_{\boldsymbol{T}} \geqslant \mathsf{D}\widehat{\theta}_{\boldsymbol{T}},$$

поскольку первое слагаемое неотрицательно, а второе равно 0. Действительно, так как $\widehat{\theta}_{\boldsymbol{T}}$ — функция от $\boldsymbol{T}$, то согласно свойствам 1 и 3 из П7, имеем:

$$\mathsf{M}(\widehat{\theta} - \widehat{\theta}_{\boldsymbol{T}})(\widehat{\theta}_{\boldsymbol{T}} - \theta) = \mathsf{M}\mathsf{M}\left[(\widehat{\theta} - \widehat{\theta}_{\boldsymbol{T}})(\widehat{\theta}_{\boldsymbol{T}} - \theta) \,|\, \boldsymbol{T}\right] =$$
$$= \mathsf{M}\left[(\widehat{\theta}_{\boldsymbol{T}} - \theta)\mathsf{M}(\widehat{\theta} - \widehat{\theta}_{\boldsymbol{T}} \,|\, \boldsymbol{T})\right] = \mathsf{M}\left[(\widehat{\theta}_{\boldsymbol{T}} - \theta)(\mathsf{M}(\widehat{\theta} \,|\, \boldsymbol{T}) - \widehat{\theta}_{\boldsymbol{T}})\right] = 0. \quad \blacksquare$$

Отметим, что попытка дальнейшего улучшения оценки $\widehat{\theta}_{\boldsymbol{T}}$ при помощи статистики $\boldsymbol{T}$ бесполезна: $\mathsf{M}(\widehat{\theta}_{\boldsymbol{T}} \,|\, \boldsymbol{T}) = \widehat{\theta}_{\boldsymbol{T}}$ по свойству 3 из П7.

Замечание. Можно доказать (см. [15, с.73]), что операция усреднения по достаточной статистике не приводит к увеличению риска

$R_{\widehat{\theta}}(\theta) = \mathbf{M}\,\rho(\widehat{\theta}-\theta)$ (см. § 3 гл. 6) относительно произвольной непрерывной выпуклой функции потерь ρ. Поэтому хорошие оценки, как правило, являются функциями от достаточных статистик.

Следующий пример показывает, что в качестве исходной в теореме 3 может фигурировать даже оценка, не обладающая свойством состоятельности.

Пример 3. Для выборки $\boldsymbol{X} = (X_1,\ldots,X_n)$ из закона Бернулли с неизвестной вероятностью «успеха» θ попробуем улучшить оценку $\widehat{\theta} = X_1$ с помощью достаточной статистики $T = \sum X_i$. Проверим несмещенность $\widehat{\theta}$: $\mathbf{M}X_1 = 1\cdot\mathbf{P}(X_1=1)+0\cdot\mathbf{P}(X_1=0) = \theta$. С учетом П7 выводим, что

$$\mathbf{M}(\widehat{\theta}\,|\,T=t) =$$
$$= 1\cdot\mathbf{P}\left(X_1=1\,\middle|\,\sum_{i=1}^n X_i = t\right) + 0\cdot\mathbf{P}\left(X_1=0\,\middle|\,\sum_{i=1}^n X_i = t\right) =$$
$$= \frac{\mathbf{P}\left(X_1=1,\sum\limits_{i=1}^n X_i = t\right)}{C_n^t\,\theta^t(1-\theta)^{n-t}} = \frac{\mathbf{P}\left(X_1=1,\sum\limits_{i=2}^n X_i = t-1\right)}{C_n^t\,\theta^t(1-\theta)^{n-t}} =$$
$$= \frac{\theta\cdot C_{n-1}^{t-1}\,\theta^{t-1}(1-\theta)^{n-t}}{C_n^t\,\theta^t(1-\theta)^{n-t}} = \frac{C_{n-1}^{t-1}}{C_n^t} = \frac{(n-1)!}{(t-1)!\,(n-t)!}\;\frac{t!\,(n-t)!}{n!} = \frac{t}{n}\,.$$

Отсюда получаем, что $\widehat{\theta}_T = \mathbf{M}(\widehat{\theta}\,|\,T) = T/n = \overline{X}$. Дисперсия при этом уменьшилась в n раз.

§ 5. ШАРИКИ В ЯЩИКАХ

Ниже в задаче 3 встретится частный случай так называемого *полиномиального распределения*, возникающий при случайном размещении шариков по ящикам.

Предположим, что шарики, занумерованные числами от 1 до n, раскладываются по N ящикам (рис. 4). Произвольное размещение описывается набором $\omega = (j_1,j_2,\ldots,j_n)$, где j_i — номер ящика (от 1 до N), куда попал i-й шарик. Всего существует N^n различных наборов. Говоря о «случайном» распределении шаров по ящикам, мы имеем в виду, что все размещения имеют одинаковые вероятности N^{-n}. Вычислим в этой модели вероятности некоторых событий.

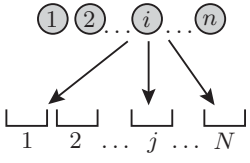


Рис. 4

Пусть l_j — неотрицательные целые числа ($j = 1,\ldots,N$), $l_1+\ldots+l_N = n$. Случайная величина ν_j — количество шариков в j-м ящике. Какова вероятность события $A = \{\nu_1 = l_1,\ldots,\nu_N = l_N\}$?

Событие $\{\nu_1 = l_1\}$ происходит, когда среди номеров j_i оказывается ровно l_1 единиц. Расставить l_1 единиц по n местам можно $C_n^{l_1}$ способами. На оставшихся $n-l_1$ местах надо разместить l_2 двоек $C_{n-l_1}^{l_2}$ способами и т. д. Перемножая все возможности и перекрестно

сокращая факториалы, получаем, что

$$\mathbf{P}(A) = C_n^{l_1} \cdot C_{n-l_1}^{l_2} \cdot \ldots \cdot C_{n-l_1-\ldots-l_{N-1}}^{l_{N-1}} N^{-n} = \frac{n!}{l_1! l_2! \ldots l_N!} N^{-n}. \qquad (3)$$

Подсчитаем теперь вероятность того, что ровно k ящиков из N окажутся пустыми. Для этого потребуется

Теорема 4. Вероятность события B_k, состоящего в том, что произойдут ровно k событий из A_1, $A_2 \ldots, A_N$, вычисляется по формуле

$$\mathbf{P}(B_k) = \sum_{l=k}^{N} (-1)^{l-k} C_l^k S_l,$$

где

$$S_l = \sum_{1 \leqslant j_1 < j_2 < \ldots < j_l \leqslant N} \mathbf{P}(A_{j_1} A_{j_2} \ldots A_{j_l}), \quad S_0 = 1.$$

Доказательство. Проверим, что индикатор I_{B_k} представляется в виде следующей двойной суммы:

$$I_{B_k} = \sum_{l=k}^{N} (-1)^{l-k} C_l^k \left[\sum_{1 \leqslant j_1 < j_2 < \ldots < j_l \leqslant N} I_{A_{j_1} A_{j_2} \ldots A_{j_l}} \right]. \qquad (4)$$

Действительно, пусть элементарное событие ω (см. П1) принадлежит ровно m множествам из A_1, $A_2 \ldots, A_N$ (см. рис. 5 для $N = 3$). Если $m < k$, то $I_{B_k} = 0$ и все члены правой части выражения (4) равны 0. Если $m = k$, то $I_{B_k} = 1$. В правой части равенства (4) член, стоящий в квадратных скобках, равен 1, когда $l = k$, и равен 0 в противном случае. Таким образом, правая часть также есть 1.

Пусть теперь $m > k$. Тогда $I_{B_k} = 0$. Сумма в квадратных скобках при $l > m$ равна 0, а при $l \leqslant m$ равна C_m^l (числу наборов по l событий из m, содержащих ω). При этом вся правая часть превращается в

$$\sum_{l=k}^{m} (-1)^{l-k} C_l^k C_m^l = C_m^k \sum_{j=0}^{m-k} (-1)^j C_{m-k}^j = C_m^k (1-1)^{m-k} = 0.$$

Доказательство теоремы завершается взятием математических ожиданий от обеих частей тождества (4). ■

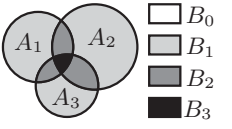


Рис. 5

Следствием теоремы 4 является так называемый принцип **включения—исключения**: вероятность того, что произойдет хотя бы одно из событий A_1, $A_2 \ldots A_N$, равна

$$\mathbf{P}\left(\bigcup_{j=1}^{N} A_j\right) = 1 - \mathbf{P}(B_0) = \sum_{j=1}^{N} (-1)^{j-1} S_j =$$

$$= \sum_j \mathbf{P}(A_j) - \sum_{k<l} \mathbf{P}(A_k A_l) + \ldots + (-1)^{N-1} \mathbf{P}(A_1 A_2 \ldots A_N).$$

Пример 4. ***Задача о совпадениях*** [81, с. 126]**.** Вычислим вероятность наблюдать ровно k совпадений при случайном сопоставлении двух одинаковых колод, состоящих из N различных

карт (или случайной раскладке N писем по конвертам). Множеством элементарных событий (П1) здесь будет пространство перестановок из первых N натуральных чисел. Каждой перестановке $\omega = (i_1,i_2,\ldots,i_N)$, где $i_l \neq i_m$ при $l \neq m$, приписана одна и та же вероятность $1/N!$. Пусть A_m обозначает множество перестановок, оставляющих m на своем месте: $i_m = m$. Тогда $\mathbf{P}(A_{j_1}A_{j_2}\ldots A_{j_l}) = (N-l)!/N!$ при $1 \leqslant j_1 < j_2 < \ldots < j_l \leqslant N$. Таким образом, $S_l = C_N^l(N-l)!/N! = 1/l!$. В силу теоремы 4

$$\mathbf{P}(B_k) = \sum_{l=k}^{N} (-1)^{l-k} C_l^k \frac{1}{l!} = \frac{1}{k!}\left[1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \ldots + \frac{(-1)^{N-k}}{(N-k)!}\right].$$

Следовательно, $\mathbf{P}(B_k) \to e^{-1}/k!$ при $N \to \infty$, т. е. предельным законом для числа совпадений является распределение Пуассона с $\lambda = 1$ (см. § 1 гл. 5). В частности, вероятность $1 - \mathbf{P}(B_0)$ того, что встретится хотя бы одно совпадение, стремится к $1 - e^{-1} \approx 0{,}632$. Точность приближения при небольших N видна из следующей таблицы:

N	2	3	4	5	6
$1 - \mathbf{P}(B_0)$	0,500	0,667	0,625	0,633	0,632

Вернемся к вычислению вероятности p_k того, что *при случайном размещении n шариков по N ящикам окажется ровно k пустых ящиков*. Обозначим через A_j событие $\{j$-й ящик пуст$\}$. Чтобы для заданных l номеров ящиков $1 \leqslant j_1 < j_2 < \ldots < j_l \leqslant N$ произошло событие $A_{j_1}A_{j_2}\ldots A_{j_l}$, все шарики должны попасть в оставшиеся $N-l$ ящиков. Очевидно, вероятность этого равна $(N-l)^n N^{-n} = (1-l/N)^n$. Отсюда, $S_l = C_N^l(1-l/N)^n$. Согласно теореме 4, искомая вероятность задается формулой

$$p_k = \sum_{l=k}^{N} (-1)^{l-k} C_l^k C_N^l \left(1 - \frac{l}{N}\right)^n. \tag{5}$$

Изучим поведение p_k, когда N и n возрастают так, что $Ne^{-n/N} \to \lambda > 0$. Нетрудно проверить, что это условие равносильно соотношению

$$n = N\ln N - (\ln\lambda)N + o(N) \quad \text{при } N \to \infty. \tag{6}$$

Теорема 5. Для $k = 0,1,\ldots$ при выполнении условия (6) вероятность $p_k \to \lambda^k e^{-\lambda}/k!$, т. е. предельным законом для числа пустых ящиков служит распределение Пуассона с параметром λ.

Доказательство. Элементарными выкладками правая часть формулы (5) преобразуется к виду

$$\frac{1}{k!}\sum_{l=k}^{N} \frac{(-1)^{l-k}}{(l-k)!}\left[\left(1-\frac{l}{N}\right)^n N^l \left(1-\frac{l-1}{N}\right)\left(1-\frac{l-2}{N}\right)\cdot\ldots\cdot 1\right]. \tag{7}$$

Так как $1-x \leqslant e^{-x}$ при $0 \leqslant x \leqslant 1$, то $(1-l/N)^n N^l \leqslant \left(Ne^{-n/N}\right)^l$. Таким образом, член в квадратных скобках в формуле (7) оценивается сверху величиной $\left(Ne^{-n/N}\right)^l$. Поскольку $Ne^{-n/N} \to \lambda$, каждое слагаемое в сумме выражения (7) по абсолютной величине не превосходит $(\lambda+1)^l/(l-k)!$ при всех достаточно больших N и n. Сходимость ряда $\sum\limits_{j=0}^{\infty}(\lambda+1)^j/j!$ обеспечивает законность перехода к пределу под знаком суммы.

Проверим, что $\left[(1-x/N)^N e^x\right]^{\ln N} \to 1$ для всех $x>0$. Действительно, раскладывая логарифм по формуле Тейлора, видим, что $\ln N\,[N\ln(1-x/N)+x] = \ln N\left[\left(-x-x^2/(2N)+o(N^{-1})\right)+x\right] = o(1)$.

Отсюда, поскольку при выполнении условия (6) $\ln N - n/N \to \ln\lambda$, имеем

$$\lim_{N\to\infty}(1-l/N)^n N^l = \lim_{N\to\infty}\left[(1-l/N)^N e^l\right]^{n/N} e^{(\ln N - n/N)l} = \lambda^l.$$

Замена индекса $j=l-k$ в формуле (7) завершает доказательство. ■

Пример 5. ***Неразличимые шары*** [81, с. 58]**.** Рассмотрим варианты размещения n неразличимых шариков по N ящикам. На рис. 6 приведены для сравнения всевозможные варианты размещения при $n=2$ и $N=2$ как различимых (занумерованных) шариков (вверху), так и неразличимых (внизу). Для неразличимых шариков каждый вариант описывается вектором $\omega=(x_1,x_2,\ldots,x_N)$, где x_i — число шариков в j-м ящике. Взглянем на него как на последовательность символов, в которой «0» обозначает шарик, а «|» — стенку, разделяющую два соседних ящика (рис. 7). Так как имеется N ящиков, то количество вертикальных черточек на приведенной схеме равно $N-1$. Общее число позиций, занятых либо ноликом, либо чертой равно $n+N-1$. Причем n из них заняты ноликами. Поэтому всего разных вариантов размещения будет C^n_{n+N-1}. Если считать, что все они равновероятны, то каждому ω надо приписать вероятность $1/C^n_{n+N-1}$.

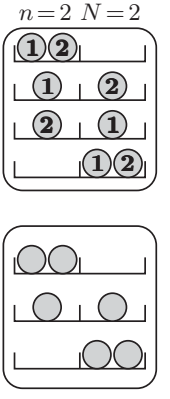


Рис. 6

$$\underbrace{0\,0\,\ldots\,0}_{x_1}\,|\,\underbrace{0\,0\,\ldots\,0}_{x_2}\,|\,\ldots\,|\,\underbrace{0\,0\,\ldots\,0}_{x_N}$$

Рис. 7

Вопрос 3. Какова вероятность, что при размещении n неразличимых шаров по N ящикам все ящики окажутся занятыми?

Приведем небольшой отрывок из [81, с. 60], посвященный применению этой модели в статистической физике.

«Рассмотрим механическую систему, состоящую из n неразличимых частиц. В статистической механике обычно разбивают фазовое пространство на большое число N малых областей или ячеек, так что каждая частица приписывается ровно одной ячейке. В результате состояние всей системы описывается как случайное

размещение n частиц по N ячейкам. На первый взгляд кажется, что (во всяком случае при подходящем выборе n ячеек) все N^n размещений будут равновероятны. Если это так, то физики говорят о *статистике Максвелла — Больцмана* (термин «статистика» используется здесь в смысле, специфическом для физики). Делались многочисленные попытки доказать, что физические частицы ведут себя в соответствии со статистикой Максвелла — Больцмана, однако современная теория, вне сомнения, показала, что эта статистика *не применима ни к каким известным частицам*; ни в одном случае все N^n размещений не являются примерно равновероятными. Были введены две различные вероятностные модели, каждая из которых удовлетворительно описывает поведение некоторого класса частиц.

(...) В *статистике Бозе—Эйнштейна* каждому из размещений приписывается вероятность $1/C^n_{n+N-1}$. В статистической механике показано, что это предположение справедливо для фотонов, атомных ядер и атомов, содержащих четное число элементарных частиц. Для описания других частиц должно быть введено третье возможное распределение вероятностей. *Статистика Ферми — Дирака основана на следующих предположениях:* 1) *в одной ячейке не могут находиться две или более частиц и* 2) *все различные размещения, удовлетворяющие первому условию, имеют одинаковую вероятность*. Для выполнения первого предположения необходимо, чтобы $n \leqslant N$. Тогда размещение полностью описывается указанием того, какие из N ячеек содержат частицу, и, так как существует n частиц, соответствующие ячейки могут быть выбраны C^n_N способами. Следовательно, в *статистике Ферми — Дирака существует* C^n_N *возможных размещений, каждое из которых имеет вероятность* $1/C^n_N$. Эта модель применима к электронам, нейтронам и протонам. Здесь мы имеем поучительный пример невозможности выбора и обоснования вероятностной модели на основе априорных соображений. Действительно, нет оснований говорить, что фотон и протон не подчиняются одним и тем же вероятностным законам».

Приведенные ниже результаты показывают, что размещения неразличимых частиц (*статистика Бозе—Эйнштейна*) имеют **ряд существенных отличий** по сравнению с размещениями различимых частиц (*статистика Максвелла—Больцмана*).

Обозначим через r_k вероятность того, что фиксированный (скажем, первый) ящик содержит ровно k *различимых* шариков. Так как остальные $(n-k)$ шариков надо разместить по $(N-1)$ ящикам, то

$$r_k = C^k_n\,(N-1)^{n-k}N^{-n} = C^k_n\,(1/N)^k(1-1/N)^{n-k}.$$

Это — биномиальное распределение с $p = 1/N$. Как было доказано в § 1 гл. 5, при $N \to \infty$ и $n \to \infty$ так, что среднее число шариков

на ящик n/N стремится к $\lambda > 0$, это распределение сходится к *закону Пуассона* с параметром λ. Рассматривая отношение r_{k+1}/r_k, несложно установить (проверьте!), что максимум вероятностей r_k достигается при $k^* = [(n+1)/N] \approx \lambda$, где $[\,\cdot\,]$ — целая часть числа.

Пусть $\tilde{r}_k$ — вероятность того, что фиксированный ящик содержит ровно k *неразличимых* шариков. Рассуждая так же, как выше, получаем $\tilde{r}_k = C^{n-k}_{(n-k)+(N-1)-1} \,/\, C^n_{n+N-1} = C^{n-k}_{n+N-k-2} \,/\, C^n_{n+N-1}$. При том же предельном переходе, что и выше, для $k = 0,1,\ldots$ имеем:

$$\tilde{r}_k = \frac{N-1}{N} \prod_{i=1}^{k} \frac{n-k+i}{N} \Bigg/ \prod_{i=1}^{k+1} \frac{n+N-k-2+i}{N} \to \lambda^k/(\lambda+1)^{k+1}.$$

Это — *геометрическое распределение* с $p = (\lambda+1)^{-1}$ (см. задачу 4 гл. 1). Поскольку $\tilde{r}_{k+1}/\tilde{r}_k = 1 - (N-2)/(n-k)$ (убедитесь!), то при $N > 2$ вероятности $\tilde{r}_k$ монотонно убывают по k. Следовательно, наиболее вероятным является то, что фиксированный ящик пуст.

В случае неразличимых шаров относительное преобладание размещений с большим количеством пустых ящиков еще заметнее проявляется при сравнении предельного поведения вероятности события B_k *обнаружить ровно k пустых ящиков*.

В случае различимых шаров вероятность p_k этого события задается формулой (5), а ее асимптотика приведена в теореме 5.

Для неразличимых шаров обозначим вероятность события B_k через $\tilde{p}_k$. Пусть A_j — это событие, состоящее в том, что j-й ящик пуст. Для заданных $1 \leqslant j_1 < j_2 < \ldots < j_l \leqslant N$ события A_{j_1}, $A_{j_2}, \ldots,$ A_{j_l} означают, что l ящиков с номерами j_1, $j_2, \ldots,$ j_l пусты. Число таких размещений равно числу способов, которыми n одинаковых шариков могут быть распределены по $N-l$ оставшимся ящикам. Поэтому $\mathbf{P}(A_{j_1} A_{j_2} \ldots A_{j_l}) = C^n_{n+N-l-1}/C^n_{n+N-1}$. В силу теоремы 4

$$\tilde{p}_k = \sum_{l=k}^{N} (-1)^{l-k}\, C^k_l\, C^l_N C^n_{n+N-l-1}/C^n_{n+N-1}. \tag{8}$$

Приведем без доказательства предельную теорему для вероятностей $\tilde{p}_k$.

Теорема 6. Распределение вероятностей $\tilde{p}_k$, заданное формулой (8), сходится к закону Пуассона, если $n, N \to \infty$ так, что $N^2/n \to \lambda > 0$.

В частности, чтобы при $\lambda = 1$ для достаточно большого N с вероятностью $\tilde{p}_0 \approx 1/e \approx 0{,}368$ не осталось ни одного пустого ящика, потребуется случайно бросить N^2 неразличимых шариков. Это отличается по порядку от величины $N \ln N$, которая (согласно теореме 5) понадобится в случае различимых шариков.

ЗАДАЧИ

Достоинство человека измеряется не той истиной, которой он владеет, а тем трудом, который он приложил для ее приобретения.

Г. Лессинг

1. С помощью критерия факторизации найдите достаточную статистику для

а) равномерного распределения на отрезке $[\theta_1, \theta_2]$,

б) модели сдвига-масштаба $F((x-\theta_1)/\theta_2)$ показательного закона с функцией распределения $F(x) = 1 - e^{-x}$ при $x > 0$.

2. Докажите, что экспоненциальному семейству принадлежат

а) биномиальное распределение: $f(x,\theta) = C_m^x\, \theta^x (1-\theta)^{m-x}$, $x = 0, 1, \ldots, m$ (в частности, при $m = 1$ — закон Бернулли),

б) гамма-распределение с неизвестным параметром θ: $f(x,\theta) = \dfrac{\theta^\alpha}{\Gamma(\alpha)}\, x^{\alpha-1} e^{-\theta x} I_{\{x>0\}}$ (при $\alpha = 1$ — показательный закон).

3. Пусть $\boldsymbol{X}$ — выборка из распределения Пуассона с неизвестным параметром $\theta > 0$: $f(x,\theta) = \theta^x e^{-\theta}/x!$, $x = 0,1,2,\ldots$. Найдите

а) распределение суммы $\xi + \eta$, где ξ и η — независимые пуассоновские случайные величины с параметрами λ и μ соответственно,

б) условное распределение $\boldsymbol{X}$ при условии $X_1 + \ldots + X_n = m$.

4*. Пусть величины $X_1, \ldots, X_n$ выбраны случайно (с повторением) из множества $\{1,2,\ldots,N\}$. Значение N неизвестно.

а) Докажите достаточность и найдите распределение статистики $X_{(n)}$.

б) Укажите такую функцию g, чтобы статистика $g(X_{(n)})$ несмещенно оценивала значение N.

5*. Элементы выборки имеют гамма-плотность из задачи 2 с $\alpha = 2$.

а) Проверьте несмещенность оценки $\widehat{\theta} = 1/X_1$.

б) Улучшите ее с помощью достаточной статистики (см. теорему 3).

6*. Рассмотрим $I_{\boldsymbol{\xi}}(\theta) = \mathsf{M}\left[\dfrac{\partial}{\partial\theta} \ln f(\boldsymbol{\xi},\theta)\right]^2$ — информацию Фишера, введенную в § 3 гл. 9. Докажите, что для выборки $\boldsymbol{X}$ из произвольной дискретной модели

а) для любой статистики $\boldsymbol{T}$ верно неравенство $I_{\boldsymbol{T}(\boldsymbol{X})}(\theta) \leqslant I_{\boldsymbol{X}}(\theta)$ при всех θ,

б) для достаточной статистики неравенство в п. а) превращается в равенство.

7*. Пусть $\eta_1, \ldots, \eta_n$ — координаты точек, взятых наудачу из отрезка $[0,1]$. Точки разбивают $[0,1]$ на $(n+1)$ частей, длины которых Δ_j называются *равномерными спейсингами* (см. следствие в § 4 гл. 4). Найдите распределение наибольшего спейсинга $\Delta_{(n+1)}$ при помощи принципа включения—исключения из § 5 и следующего результата Б. де Финетти (см. [82, с. 57]):

$$\mathsf{P}\,(\Delta_1 > x_1, \ldots, \Delta_{n+1} > x_{n+1}) = (1 - x_1 - \ldots - x_{n+1})_+^n$$

для произвольных $x_1 \geqslant 0, \ldots, x_{n+1} \geqslant 0$ (здесь $f_+ = \max\{0, f\}$).

РЕШЕНИЯ ЗАДАЧ

1. а) Плотностью выборки $\boldsymbol{X} = (X_1, \ldots, X_n)$ служит

$$f(\boldsymbol{x},\boldsymbol{\theta}) = \frac{\prod\limits_{i=1}^{n} I_{\{x_i \geqslant \theta_1\}}\, I_{\{x_i \leqslant \theta_2\}}}{(\theta_2 - \theta_1)^n} = \frac{I_{\{x_{(1)} \geqslant \theta_1\}}\, I_{\{x_{(n)} \leqslant \theta_2\}}}{(\theta_2 - \theta_1)^n}\,.$$

Стало быть, $(X_{(1)}, X_{(n)})$ — достаточная статистика.

б) Поскольку плотность величины X_i равна $\theta_2^{-1} e^{-(x-\theta_1)/\theta_2} I_{\{x \geqslant \theta_1\}}$, то

$$f(\boldsymbol{x},\boldsymbol{\theta}) = \prod_{i=1}^{n} \theta_2^{-1} e^{-(x_i-\theta_1)/\theta_2} I_{\{x_i \geqslant \theta_1\}} = \theta_2^{-n} e^{-\left(\sum x_i - n\theta_1\right)/\theta_2} I_{\{x_{(1)} \geqslant \theta_1\}}.$$

Следовательно, статистика $\boldsymbol{T} = (X_{(1)}, \sum X_i)$ является достаточной. Отметим, что вектор оценок максимального правдоподобия в этой модели $(X_{(1)}, \overline{X} - X_{(1)})$ (см. задачу 3 гл. 9) связан со статистикой $\boldsymbol{T}$ взаимно однозначным (линейным) преобразованием.

2. а) Совместная вероятность выборки $\boldsymbol{X} = (X_1, \ldots, X_n)$

$$f(\boldsymbol{x},\theta) = \left(\prod_{i=1}^{n} C_m^{x_i}\right) \theta^{\sum x_i} (1-\theta)^{mn - \sum x_i} =$$

$$= \exp\left\{ mn \ln(1-\theta) + \left(\ln \frac{\theta}{1-\theta}\right) \sum_{i=1}^{n} x_i \right\} \prod_{i=1}^{n} C_m^{x_i}.$$

Видим, что $T = \sum X_i$, $g_0(\theta) = mn \ln(1-\theta)$ и $g_1(\theta) = \ln \dfrac{\theta}{1-\theta}$.

б) Плотность выборки $\boldsymbol{X}$ из гамма-распределения имеет вид

$$f(\boldsymbol{x},\theta) = \exp\left\{ n\alpha \ln\theta - \theta \sum_{i=1}^{n} x_i \right\} \Gamma(\alpha)^{-n} I_{\{x_{(1)} > 0\}} \prod_{i=1}^{n} x_i^{\alpha-1}.$$

Таким образом, $T = \sum X_i$, $g_0(\theta) = n\alpha \ln\theta$ и $g_1(\theta) = -\theta$.

3. а) Используя формулу полной вероятности (П7) и независимость случайных величин ξ и η, получаем, что

$$p_m = \mathbf{P}(\xi + \eta = m) = \sum_{k=0}^{m} \mathbf{P}(\xi + \eta = m, \eta = k) =$$

$$= \sum_{k=0}^{m} \mathbf{P}(\xi = m - k, \eta = k) = \sum_{k=0}^{m} \mathbf{P}(\xi = m - k)\, \mathbf{P}(\eta = k).$$

Тем самым мы вывели *дискретную формулу свертки* из П3. Далее с учетом бинома Ньютона находим, что

$$p_m = \sum_{k=0}^{m} \frac{\lambda^{m-k}}{(m-k)!}\, e^{-\lambda}\, \frac{\mu^k}{k!}\, e^{-\mu} = \frac{(\lambda+\mu)^m}{m!}\, e^{-(\lambda+\mu)},$$

т. е. $\xi + \eta$ имеет распределение Пуассона с параметром $\lambda + \mu$.

Истинное знание самостоятельно.

Л. Н. Толстой

Вопрос 4.
Чему равен предел при $n \to \infty$ последовательности $c_n = e^{-n} \sum\limits_{k=0}^{n} \dfrac{n^k}{k!}$?

(Примените центральную предельную теорему (см. Пб) к пуассоновским случайным величинам.)

б) Согласно пункту а), сумма $X_1 + \ldots + X_n$ распределена по закону Пуассона с параметром $n\theta$. Поэтому

$$\mathsf{P}(\boldsymbol{X} = \boldsymbol{x} \,|\, X_1 + \ldots + X_n = m) = \frac{\mathsf{P}(\boldsymbol{X} = \boldsymbol{x},\, \sum X_i = m)}{\mathsf{P}(\sum X_i = m)} =$$
$$= \frac{\mathsf{P}(X_1 = x_1, \ldots, X_{n-1} = x_{n-1}, X_n = m - x_1 - \ldots - x_{n-1})}{(n\theta)^m e^{-n\theta}/m!} =$$
$$= \left[\prod \frac{\theta^{x_i}}{x_i!}\, e^{-\theta}\right] \Big/ \left[\frac{(n\theta)^m}{m!}\, e^{-n\theta}\right] = \frac{m!}{x_1! x_2! \ldots x_n!}\, n^{-m}.$$

Это — вероятность того, что при случайном размещении m различимых шаров по n ящикам в ящике с номером i $(i = 1, \ldots, n)$ окажется ровно x_i шариков (см. формулу (3) из § 5).

4. а) Так как совместная вероятность выборки равна $f(\boldsymbol{x}, N) = N^{-n} I_{\{x_{(n)} \leqslant N\}}$, то статистика $X_{(n)}$ достаточна в силу критерия факторизации. Ввиду независимости случайных величин $X_1, \ldots, X_n$ имеем $\mathsf{P}(X_{(n)} \leqslant m) = \mathsf{P}(X_1 \leqslant m, \ldots, X_n \leqslant m) = (m/N)^n$. Отсюда

$$\mathsf{P}(X_{(n)} = m) = \mathsf{P}(X_{(n)} \leqslant m) - \mathsf{P}(X_{(n)} \leqslant m - 1) =$$
$$= N^{-n}[m^n - (m-1)^n].$$

б) Запишем условие несмещенности для функции $g(X_{(n)})$:

$$\sum_{m=1}^{N} g(m)\, N^{-n}[m^n - (m-1)^n] = N,$$

т. е. $\sum g(m)\,[m^n - (m-1)^n] = N^{n+1}$, $N = 1,2,\ldots$. Вычитая из суммы до N сумму до $N-1$, находим, что при всех N

$$g(N) = [N^{n+1} - N^n] \,/\, [N^n - (N-1)^n].$$

Нетрудно вывести, что $g(N) \approx \left(1 + \dfrac{1}{n}\right) N$, когда N велико.

5. Проверим несмещенность оценки $\widehat{\theta}$: $\mathsf{M}\widehat{\theta} = \int\limits_0^\infty (1/x)\, \theta^2 x e^{-\theta x}\, dx = \theta$.

Согласно задаче 2 б), статистика $T = \sum X_i$ достаточна. В силу леммы 1 гл. 4 она имеет гамма-распределение с плотностью

$$g_n(t) = \frac{1}{(2n-1)!}\, \theta^{2n} t^{2n-1} e^{-\theta t}\, I_{\{t>0\}}.$$

Плотность вектора $(X_1, \sum\limits_{i=1}^{n} X_i)$ равна плотности вектора $(X_1, \sum\limits_{i=2}^{n} X_i)$ в точке $(x, t-x)$ (см. формулу преобразования плотности из П8):

$$p_{(X_1,T)}(x,t) = p_{X_1}(x)\, g_{n-1}(t-x) = \frac{\theta^{2n} x (t-x)^{2n-3} e^{-\theta t}}{(2n-3)!}\, I_{\{0<x<t\}}.$$

Тогда условная плотность (П7) случайной величины X_1 при условии T есть

$$p_{(X_1|T)}(x,t) = (2n-1)(2n-2)\,x(t-x)^{2n-3}\,t^{-2n+1}I_{\{0<x<t\}}.$$

Наконец, $\mathbf{M}(\widehat{\theta}\,|\,T=t) = \int\limits_0^\infty (1/x)\,p_{(X_1|T)}(x,t)\,dx = (2n-1)/t$, т. е. искомая оценка $\widehat{\theta}_T = \mathbf{M}(\widehat{\theta}\,|\,T) = (2n-1)/(X_1+\ldots+X_n)$.

6. а) Положим $Q_\theta(\boldsymbol{t}) = \mathbf{P}(\boldsymbol{T}(\boldsymbol{X}) = \boldsymbol{t})$ и введем случайные величины $U = \partial \ln f(\boldsymbol{X},\theta)/\partial\theta$ и $V = \partial \ln Q_\theta(\boldsymbol{T}(\boldsymbol{X}))/\partial\theta$. Рассмотрим тождество

$$\mathbf{M}(U-V)^2 = I_{\boldsymbol{X}}(\theta) + I_{\boldsymbol{T}(\boldsymbol{X})}(\theta) - 2\,\mathbf{M}(UV). \tag{9}$$

Распишем $\mathbf{M}(UV)$ из правой части формулы (9):

$$\sum_{\boldsymbol{x}} \frac{f'(\boldsymbol{x},\theta)}{f(\boldsymbol{x},\theta)} \cdot \frac{Q'_\theta(\boldsymbol{T}(\boldsymbol{x}))}{Q_\theta(\boldsymbol{T}(\boldsymbol{x}))}\, f(\boldsymbol{x},\theta) = \sum_{\boldsymbol{t}} \frac{Q'_\theta(\boldsymbol{t})}{Q_\theta(\boldsymbol{t})} \sum_{\boldsymbol{x}:\,\boldsymbol{T}(\boldsymbol{x})=\boldsymbol{t}} f'(\boldsymbol{x},\theta) =$$

$$= \sum_{\boldsymbol{t}} \frac{Q'_\theta(\boldsymbol{t})}{Q_\theta(\boldsymbol{t})}\, \frac{\partial}{\partial\theta} \sum_{\boldsymbol{x}:\,\boldsymbol{T}(\boldsymbol{x})=\boldsymbol{t}} f(\boldsymbol{x},\theta) = \sum_{\boldsymbol{t}} \frac{[Q'_\theta(\boldsymbol{t})]^2}{Q_\theta(\boldsymbol{t})} =$$

$$= \sum_{\boldsymbol{t}} \left[\frac{Q'_\theta(\boldsymbol{t})}{Q_\theta(\boldsymbol{t})}\right]^2 Q_\theta(\boldsymbol{t}) = \mathbf{M}\left[\frac{Q'_\theta(\boldsymbol{T})}{Q_\theta(\boldsymbol{T})}\right]^2 = I_{\boldsymbol{T}(\boldsymbol{X})}(\theta).$$

Итак, правая часть тождества (9) равна $I_{\boldsymbol{X}}(\theta) - I_{\boldsymbol{T}(\boldsymbol{X})}(\theta)$. Так как его левая часть неотрицательна, то неравенство доказано.

б) Для достаточной статистики $\boldsymbol{T}$ просуммируем правую часть формулы (1) по таким $\boldsymbol{x}$, что $\boldsymbol{T}(\boldsymbol{x}) = \boldsymbol{t}$. Получим

$$Q_\theta(\boldsymbol{t}) = g(\boldsymbol{t},\theta) \sum_{\boldsymbol{x}:\,\boldsymbol{T}(\boldsymbol{x})=\boldsymbol{t}} h(\boldsymbol{x}). \tag{10}$$

Подставляя соотношения (1) и (10) в левую часть тождества (9), обратим ее в нуль.

7. Для произвольного $x \geqslant 0$ положим $A_j = \{\Delta_j > x\}$, $j = 1,\ldots,n+1$. Взяв в формуле Б. де Финетти одно из x_j равным x, а все остальные — равными 0, получим, что $\mathbf{P}(A_j) = (1-x)_+^n$. Таким образом, все спейсинги распределены одинаково (так же, как $\Delta_1 = \eta_{(1)}$), но, конечно, зависимы: их совместная плотность, найденная в следствии из § 4 гл. 4, не равна произведению плотностей случайной величины Δ_j.

Ввиду очевидного равенства $\mathbf{P}(\Delta_{(n+1)} \leqslant x) = 1 - \mathbf{P}(\bigcup A_j)$, для применения принципа включения—исключения осталось вычислить вероятности $\mathbf{P}(A_{j_1}A_{j_2}\ldots A_{j_l})$ для любых $1 \leqslant j_1 < j_2 < \ldots < j_l \leqslant n+1$. Положив в формуле Б. де Финетти $x_{j_1},\ldots,x_{j_l}$ равными x, а все остальные — равными 0, находим, что $\mathbf{P}(A_{j_1}A_{j_2}\ldots A_{j_l}) = (1-lx)_+^n$. Группу из l упорядоченных

индексов можно выбрать C_{n+1}^l способами. Окончательно имеем:

$$\rho_n(x) = \mathsf{P}(\Delta_{(n+1)} \leqslant x) = \sum_{l=0}^{n+1} (-1)^l\, C_{n+1}^l\, (1-lx)_+^n.$$

Придадим этому результату другую форму, известную как **теорема о покрытии** (см. [82, с. 43]). Для этого свернем отрезок $[0,1]$ в окружность единичной длины и будем считать точки $0,\eta_1,\ldots,\eta_n$ серединами дуг длины x. Тогда с вероятностью $\rho_n(x)$ дуги покрывают всю окружность.

Вот некоторые значения этой вероятности для $x = 0{,}2$:

n	8	10	15	20	25	30	40	50
$\rho_n(0{,}2)$	0,040	0,134	0,493	0,766	0,903	0,962	0,995	0,999

Моделируйте задачу с помощью таблицы T1. Сколько дуг понадобилось?

Таким образом, вместо 5 дуг, достаточных для регулярного покрытия, в среднем потребуется около 15 случайных дуг, и не менее 30 дуг обеспечивают полное покрытие окружности с вероятностью 0,962. Этот численный пример поясняет причину медленного уменьшения с ростом n погрешности метода Монте-Карло (см. § 3 гл. 3).

ОТВЕТЫ НА ВОПРОСЫ

1. Для выборки из равномерного на отрезке $[0,\theta]$ распределения

$$f(\boldsymbol{x},\theta) = \theta^{-n} \prod_{i=1}^{n} I_{\{x_i \geqslant 0\}} I_{\{x_i \leqslant \theta\}} = \theta^{-n} I_{\{x_{(1)} \geqslant 0\}} I_{\{x_{(n)} \leqslant \theta\}}.$$

Взяв $g(t,\theta) = \theta^{-n} I_{\{t \leqslant \theta\}}$ и $h(\boldsymbol{x}) = I_{\{x_{(1)} \geqslant 0\}}$, видим, что максимум $X_{(n)}$ — достаточная статистика.

2. Плотностью выборки из показательного закона служит

$$f(\boldsymbol{x},\theta) = \exp\left\{-n\ln\theta - \frac{1}{\theta}\sum_{i=1}^{n} x_i\right\} I_{\{x_{(1)} \geqslant 0\}}.$$

Таким образом, $T = \sum X_i$, $g_0(\theta) = -n\ln\theta$ и $g_1(\theta) = -1/\theta$.

3. Положим сразу N шариков по одному в каждый ящик. Оставшиеся $n-N$ шариков разместим $C_{n-N+N-1}^{n-N} = C_{n-1}^{N-1}$ способами. Стало быть, искомая вероятность есть $C_{n-1}^{N-1}/C_{n+N-1}^{n}$.

4. Пусть $X_1,\ldots,X_n$ — выборка из закона Пуассона с $\lambda = 1$, $S_n = X_1 + \ldots + X_n$. Тогда S_n имеет распределение Пуассона с параметром n, $c_n = \mathsf{P}(S_n \leqslant n)$. Поскольку $\mathsf{M}S_n = \mathsf{D}S_n = n$ (см. вопрос 2 гл. 5), то в силу центральной предельной теоремы

$$c_n = \mathsf{P}\left(\frac{S_n - n}{\sqrt{n}} \leqslant 0\right) \to \Phi(0) = 1/2 \quad \text{при } n \to \infty.$$

Глава 11

ДОВЕРИТЕЛЬНЫЕ ИНТЕРВАЛЫ

§ 1. КОЭФФИЦИЕНТ ДОВЕРИЯ

Вместо того, чтобы приближать неизвестный скалярный параметр θ с помощью «точечной» оценки $\widehat{\theta}$, можно локализовать его иначе — указать случайный интервал $(\widehat{\theta}_1, \widehat{\theta}_2)$, который накрывает θ с вероятностью близкой к единице (рис. 1).

Касательно расстояния от Земли до Солнца можно заметить, что каждое новое, более точное определение этой величины не укладывается в доверительный интервал, построенный по старым наблюдениям.

В. Н. Тутубалин, [79, с. 313]

Определение. Пусть $\alpha \in (0,1)$. Две статистики $\widehat{\theta}_1$ и $\widehat{\theta}_2$ определяют границы *доверительного интервала для параметра θ с коэффициентом доверия* $1-\alpha$, если при всех $\theta \in \Theta$ для выборки $\boldsymbol{X} = (X_1, \ldots, X_n)$ из закона распределения $F_\theta(x)$ справедливо неравенство

$$\mathsf{P}\left(\widehat{\theta}_1(\boldsymbol{X}) < \theta < \widehat{\theta}_2(\boldsymbol{X})\right) \geqslant 1-\alpha. \tag{1}$$

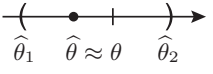


Рис. 1

Часто на практике полагают $\alpha = 0{,}05$. Если вероятность в левой части неравенства (1) стремится к $1-\alpha$ при $n \to \infty$, то интервал называется *асимптотическим*. Как правило, длина доверительного интервала возрастает при увеличении коэффициента доверия $1-\alpha$ и стремится к нулю с ростом размера выборки n.

Пример 1. Для модели сдвига показательного закона с плотностью $p_\theta(x) = e^{-(x-\theta)} I_{\{x \geqslant \theta\}}$ оценкой максимального правдоподобия согласно примеру 6 гл. 9 является $X_{(1)} = \min\{X_1, \ldots, X_n\}$. Поскольку $\theta < X_{(1)}$, можно взять $X_{(1)}$ в качестве $\widehat{\theta}_2$. Попробуем подобрать константу c_α так, чтобы для $\widehat{\theta}_1 = X_{(1)} - c_\alpha$ (см. рис. 2) при всех θ выполнялось тождество

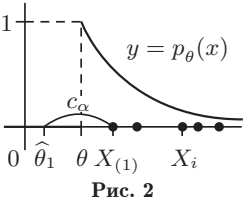


Рис. 2

$$\mathsf{P}\left(X_{(1)} - c_\alpha < \theta < X_{(1)}\right) = \mathsf{P}\left(X_{(1)} - c_\alpha < \theta\right) = 1-\alpha. \tag{2}$$

Используя независимость и показательность величин $X_i - \theta$, перепишем условие (2):

$$\alpha = \mathsf{P}\left(X_{(1)} - \theta \geqslant c_\alpha\right) = \\ = \mathsf{P}\left(X_i - \theta \geqslant c_\alpha, i = 1, \ldots, n\right) = \exp\{-n c_\alpha\}.$$

Откуда находим, что длина интервала $c_\alpha = (-\ln \alpha)/n$. Отметим, что $c_\alpha \to \infty$ при $\alpha \to 0$ и $c_\alpha \to 0$ при $n \to \infty$.

Вопрос 1.
Какое c_α подойдет для построения доверительного интервала с помощью статистики $X_{(1)}$ в модели равномерного распределения на отрезке $[\theta, \theta+1]$?

§ 2. ИНТЕРВАЛЫ В НОРМАЛЬНОЙ МОДЕЛИ

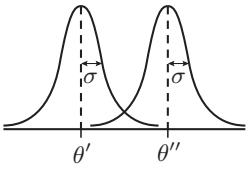


Рис. 3

Пример 2. Допустим, что элементы выборки X_i распределены по закону $\mathcal{N}(\theta, \sigma^2)$, причем параметр масштаба σ известен, а параметр сдвига θ — нет. Эту модель часто применяют к данным, полученным при независимых измерениях некоторой величины θ с помощью прибора (или метода), имеющего известную среднюю погрешность (стандартную ошибку) σ (рис. 3).

Пусть $\Phi(x) = (2\pi)^{-1/2} \int\limits_{-\infty}^{x} e^{-u^2/2} du$ — функция распределения закона $\mathcal{N}(0,1)$. Для $0 < \alpha < 1$ обозначим через x_α так называемую α-квантиль этого закона, т. е. решение уравнения $\Phi(x_\alpha) = \alpha$ (см. § 3 гл. 7). Приведем некоторые значения $x_{1-\alpha/2}$ (см. также таблицу T2):

α	0,05	10^{-2}	10^{-3}	10^{-5}
$x_{1-\alpha/2}$	1,96	2,58	3,29	4,26

Согласно примеру 4 гл. 9, эффективной оценкой для θ служит $\overline{X}$. Известно, что $\overline{X} \sim \mathcal{N}(\theta, \sigma^2/n)$. Тогда $\sqrt{n}\,(\overline{X} - \theta)/\sigma \sim \mathcal{N}(0,1)$. Поэтому в качестве границ интервала с коэффициентом доверия $1-\alpha$ можно взять $\widehat{\theta}_1 = \overline{X} - \sigma x_{1-\alpha/2}/\sqrt{n}$ и $\widehat{\theta}_2 = \overline{X} - \sigma x_{\alpha/2}/\sqrt{n}$:

$$\mathbf{P}(\widehat{\theta}_1 < \theta < \widehat{\theta}_2) = \mathbf{P}(x_{\alpha/2} < \sqrt{n}\,(\overline{X} - \theta)/\sigma < x_{1-\alpha/2}) = 1 - \alpha.$$

В силу четности плотности закона $\mathcal{N}(0,1)$ верно равенство $x_{\alpha/2} = -x_{1-\alpha/2}$. Таким образом, из приведенной выше таблицы видим, что с вероятностью 0,95 истинное значение параметра сдвига θ находится в интервале $\overline{X} \pm 1{,}96\,\sigma/\sqrt{n} \approx \overline{X} \pm 2\sigma/\sqrt{n}$ (**правило двух сигм**).

Замечание 1. В теории измерений обычно предполагают, что наблюдения $X_i = \theta + \varepsilon_i$, где ошибки ε_i — независимые и одинаково распределенные случайные величины (не обязательно нормальные) с $\mathbf{M}\varepsilon_i = 0$ и $\mathbf{D}\varepsilon_i = \sigma^2 < \infty$. В силу центральной предельной теоремы (П6)

$$\sqrt{n}\,(\overline{X} - \theta)/\sigma \xrightarrow{d} Z \sim \mathcal{N}(0,1) \quad \text{при } n \to \infty.$$

Значит, $\overline{X} \pm 1{,}96\,\sigma/\sqrt{n}$ — *асимптотический* 95%-й доверительный интервал. Если значение σ неизвестно, то на практике его заменяют на состоятельную оценку $\widehat{\sigma} = S$, где $S^2 = \dfrac{1}{n}\sum (X_i - \overline{X})^2$, получая *приближенный* 95%-й интервал $\overline{X} \pm 2S/\sqrt{n}$.

Однако нередко на самом деле $X_i = \theta + \Delta + \varepsilon_i$, где Δ — это *систематическая ошибка* измерения. Попытка оценить величину возможного отклонения $\overline{X}$ от θ лишь на основе разброса наблюдений X_i вокруг $\overline{X}$, вообще говоря, может приводить к ошибочным выводам.

В «Философском очерке теории вероятностей» Лаплас, определив отношение массы Юпитера к массе Солнца, предлагает пари, что будущие поколения ученых не изменят найденное им число более чем на 1% (вероятность того, что это произойдет, согласно Лапласу, ничтожно мала). Но современное значение отличается от найденного Лапласом на 2%.

В. Н. Тутубалин, [79, с. 313]

Другими причинами неверных заключений могут быть *зависимость наблюдений* X_i, наличие которой не предполагалось исследователем (подробнее см. § 4 гл. 15), или *нарушение однородности* данных при слишком большом увеличении их количества.

Пример 3. Предположим, что элементы выборки X_i распределены по закону $\mathcal{N}(\mu, \theta^2)$, причем значение параметра сдвига μ известно, а параметра масштаба θ — нет. Такую модель можно использовать для определения средней точности прибора (или метода) путем многократных измерений эталона (рис. 4).

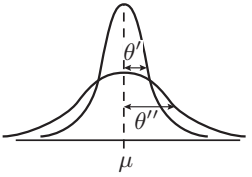


Рис. 4

Чтобы построить доверительный интервал для θ, потребуется следующее

Определение. Пусть случайные величины $Z_1, \ldots, Z_k$ распределены по закону $\mathcal{N}(0,1)$ и независимы. Тогда распределение случайной величины $R_k^2 = Z_1^2 + \ldots + Z_k^2$ называют *распределением хи-квадрат с k степенями свободы* (кратко: $R_k^2 \sim \chi_k^2$).

Отметим, что каждое слагаемое Z_i^2 имеет гамма-распределение (см. § 4 гл. 4) с параметрами $\alpha = \lambda = 1/2$, т. е. $Z_i^2 \sim \Gamma(1/2{,}1/2)$.

Вопрос 2. Как доказать это утверждение?

Отсюда согласно лемме 1 гл. 4 находим, что $R_k^2 \sim \Gamma(k/2{,}1/2)$. Таким образом, плотностью закона хи-квадрат служит функция

$$p_{R_k^2}(x) = c_k\, x^{k/2-1} e^{-x/2} I_{\{x>0\}}, \quad \text{где } c_k = \frac{1}{2^{k/2}\Gamma(k/2)}\,.$$

Здесь $\Gamma(p) = \int\limits_0^\infty x^{p-1}e^{-x}\,dx$ — гамма-функция Эйлера, график которой изображен на рис. 9 гл. 3. Дифференцируя $p_{R_k^2}(x)$, нетрудно убедиться, что при $k > 1$ плотность имеет единственный максимум в точке $k-2$ (рис. 5). Закон χ_1^2 дает пример распределения с неограниченной плотностью.

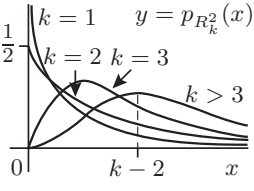


Рис. 5

Отметим, что $\mathbf{M}R_k^2$ расположено правее точки максимума плотности (так называемой *моды распределения*) из-за того, что правый «хвост» плотности «тяжелее» левого, т. е. убывает медленнее.

Дисперсию $\mathbf{D}R_k^2$ легко подсчитать с помощью формулы (2) гл. 4. Она равна $2k$ (проверьте!). Поскольку R_k^2 — это сумма независимых и одинаково распределенных случайных величин Z_i^2, согласно центральной предельной теореме имеет место сходимость $(R_k^2 - k)/\sqrt{2k} \xrightarrow{d} Z \sim \mathcal{N}(0,1)$ при $k \to \infty$. Нормальное приближение является довольно точным уже при $k > 30$. Значения некоторых квантилей закона χ_k^2 для $k \leqslant 30$ приведены в таблице Т3.

Вопрос 3. Чему равно $\mathbf{M}R_k^2$? (Попробуйте догадаться без вычислений.)

Перейдем теперь к построению доверительного интервала для θ по выборке X_i из распределения $\mathcal{N}(\mu, \theta^2)$. Введем $D_n = \sum\limits_{i=1}^n (X_i - \mu)^2$. Заметим, что величина $R_n^2 = D_n/\theta^2 = \sum\limits_{i=1}^n [(X_i-\mu)/\theta]^2$ распределена по закону χ_n^2. Обозначим p-квантиль этого закона через x_p (рис. 6).

Тогда для $\widehat{\theta}_1 = \sqrt{D_n/x_{1-\alpha/2}}$ и $\widehat{\theta}_2 = \sqrt{D_n/x_{\alpha/2}}$ при всех $\theta > 0$ справедливо равенство

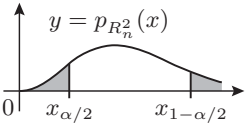


Рис. 6

$$1-\alpha = \mathbf{P}\left(x_{\alpha/2} < R_n^2 < x_{1-\alpha/2}\right) = \mathbf{P}\left(\widehat{\theta}_1 < \theta < \widehat{\theta}_2\right),$$

т. е. $(\widehat{\theta}_1,\widehat{\theta}_2)$ — искомый интервал.

Пример 4. Пусть $X_i \sim \mathcal{N}(\theta_1, \theta_2^2)$, где оба параметра θ_1 и θ_2 неизвестны. Для построения доверительных интервалов понадобится

Определение. Пусть случайные величины Z и R_k^2 независимы и распределены согласно законам $\mathcal{N}(0,1)$ и χ_k^2 соответственно. Тогда распределение случайной величины $T_k = Z\,\big/\sqrt{R_k^2/k}$ называют *распределением Стьюдента с k степенями свободы* или *t-распределением* (кратко: $T_k \sim t_k$).

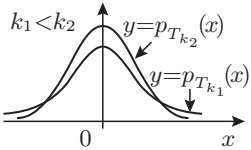


Рис. 7

Вопрос 4.
К какому распределению стремится t_k при $k\to\infty$?
(Используйте закон больших чисел (П6) и свойства сходимости (П5) или же найдите предел последовательности плотностей $p_{T_k}(x)$.)

При каждом k случайная величина T_k имеет четную плотность (см. [38, с. 58])

$$p_{T_k}(x) = d_k\left(1+\frac{x^2}{k}\right)^{-(k+1)/2}, \quad \text{где } d_k = \frac{1}{\sqrt{\pi k}}\,\frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)}. \tag{3}$$

В частности, при $k=1$ получаем закон Коши с плотностью $p_{T_1}(x) = 1/[\pi\,(1+x^2)]$. С ростом k «хвосты» t_k становятся «легче» (рис. 7).

Математическое ожидание случайной величины T_k не существует при $k = 1$ (см. замечание в § 2 гл. 1) и равно 0 при $k > 1$ в силу четности плотности $p_{T_k}(x)$. Дисперсия величины T_k бесконечна при $k = 2$ и равна $k/(k-2)$, если $k > 2$ (см. [50, с. 315]).

Теорема 1. Для нормальной выборки $X_i \sim \mathcal{N}(\theta_1, \theta_2^2)$ выборочное среднее $\overline{X} = \frac{1}{n}\sum X_i$ и выборочная дисперсия $S^2 = \frac{1}{n}\sum (X_i - \overline{X})^2$ независимы,[*)] причем $nS^2/\theta_2^2 \sim \chi_{n-1}^2$, а $\sqrt{n-1}\,(\overline{X}-\theta_1)/S \sim t_{n-1}$.

Доказательство теоремы опирается на следующую лемму.

Лемма 1. Пусть $(n\times n)$-матрица $\boldsymbol{C}$ ортогональна: $\boldsymbol{C}^{-1} = \boldsymbol{C}^T$ (см. П10), случайный вектор $\boldsymbol{Y} = (Y_1,\ldots,Y_n)$ имеет независимые $\mathcal{N}(0,1)$ компоненты. Тогда вектор $\boldsymbol{Z} = \boldsymbol{C}\boldsymbol{Y}$ распределен так же, как и $\boldsymbol{Y}$.

Доказательство леммы. Заметим, что при ортогональном преобразовании $\boldsymbol{z} = \boldsymbol{C}\boldsymbol{y}$ длины векторов в $\mathbb{R}^n$ не меняются:

$$|\boldsymbol{z}|^2 = \boldsymbol{z}^T\boldsymbol{z} = (\boldsymbol{C}\boldsymbol{y})^T\,\boldsymbol{C}\boldsymbol{y} = \left(\boldsymbol{y}^T\boldsymbol{C}^T\right)\boldsymbol{C}\boldsymbol{y} = \boldsymbol{y}^T\left(\boldsymbol{C}^T\boldsymbol{C}\right)\boldsymbol{y} = |\boldsymbol{y}|^2. \tag{4}$$

[*)] Оказывается, что из независимости выборочного среднего $\overline{X}$ и выборочной дисперсии S^2 при $n \geqslant 2$ следует (как доказал Р. Гири в 1936 г.) нормальность распределения элементов выборки $X_1,\ldots,X_n$ (см. [72, с. 198]). Однако, для выполнения условия $\boldsymbol{\rho}(\overline{X},S^2) = 0$ достаточно, чтобы этот *коэффициент корреляции* (П2) существовал и распределение случайных величин X_i было симметричным.

Обратная матрица $\boldsymbol{C}^{-1} = \boldsymbol{C}^T$ также будет ортогональной, причем $|\det \boldsymbol{C}^T| = 1$.

Вопрос 5. Как это доказать с помощью свойств из П10?

Согласно условиям леммы плотность $p_{\boldsymbol{Y}}(\boldsymbol{y}) = (2\pi)^{-n/2} e^{-\frac{1}{2}|\boldsymbol{y}|^2}$. По формуле преобразования плотности (П8) с учетом (4) получаем:

$$p_{\boldsymbol{Z}}(\boldsymbol{z}) = |\det \boldsymbol{C}^T|\, p_{\boldsymbol{Y}}(\boldsymbol{C}^T\boldsymbol{z}) = (2\pi)^{-n/2} e^{-\frac{1}{2}|\boldsymbol{C}^T\boldsymbol{z}|^2} = (2\pi)^{-n/2} e^{-\frac{1}{2}|\boldsymbol{z}|^2}.$$

■

Другое доказательство леммы 1 немедленно вытекает из формулы для преобразования ковариационной матрицы нормального случайного вектора при умножении его на числовую матрицу (см. П9).

Доказательство теоремы. Возьмем в качестве $\boldsymbol{C}$ ортогональную $(n \times n)$-матрицу с последней строкой $(1/\sqrt{n}, \ldots, 1/\sqrt{n})$. Значения c_{ij} при $i < n$ не важны, лишь бы $\boldsymbol{C}$ была ортогональной (можно, скажем, дополнить последнюю строку до ортогонального базиса в $\mathbb{R}^n$ при помощи стандартного алгоритма Грама—Шмидта (см. соотношения (9) гл. 21). Нетрудно проверить, что годятся

$$c_{ij} = \begin{cases} 1/\sqrt{i(i+1)}, & \text{если } j \leqslant i, \\ -i/\sqrt{i(i+1)}, & \text{если } j = i+1, \\ 0, & \text{если } j > i+1. \end{cases}$$

В качестве $\boldsymbol{Y}$ возьмем вектор с компонентами $Y_i = (X_i - \theta_1)/\theta_2$. Рассмотрим $\boldsymbol{Z} = \boldsymbol{C}\boldsymbol{Y}$. При умножении последней строки матрицы $\boldsymbol{C}$ на вектор $\boldsymbol{Y}$ получается соотношение

$$Z_n = \frac{1}{\sqrt{n}}(Y_1 + \ldots + Y_n) = \sqrt{n}\,\overline{Y}. \tag{5}$$

С другой стороны, из формулы (4) следует, что $\sum\limits_{i=1}^{n} Z_i^2 = \sum\limits_{i=1}^{n} Y_i^2$. С учетом формулы (1) гл. 6 из этого равенства и соотношения (5) выводим, что

$$\frac{nS^2}{\theta_2^2} = \sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{\theta_2^2} = \sum_{i=1}^{n} (Y_i - \overline{Y})^2 = \sum_{i=1}^{n} Y_i^2 - n\overline{Y}^2 =$$

$$= \sum_{i=1}^{n} Z_i^2 - Z_n^2 = \sum_{i=1}^{n-1} Z_i^2.$$

В силу леммы 1 имеем $nS^2/\theta_2^2 \sim \chi^2_{n-1}$. Из соотношения (5) получаем, что $\sqrt{n}\,(\overline{X} - \theta_1)/\theta_2 = \sqrt{n}\,\overline{Y} = Z_n \sim \mathcal{N}(0,1)$. Так как $\overline{X}$ — функция от Z_n, а S — от $Z_1, \ldots, Z_{n-1}$, то $\overline{X}$ и S независимы (см. лемму о независимости из § 3 гл. 1). Наконец, случайная величина

$$T_{n-1} = \frac{\sqrt{n-1}\,(\overline{X} - \theta_1)}{S} = \frac{\sqrt{n}\,(\overline{X} - \theta_1)/\theta_2}{\sqrt{(nS^2/\theta_2^2)/(n-1)}} \tag{6}$$

распределена по закону t_{n-1} согласно определению. ■

На основе доказанной теоремы построим доверительные интервалы для неизвестных параметров θ_1 и θ_2 закона $\mathcal{N}(\theta_1, \theta_2^2)$.

Доверительный интервал для параметра сдвига θ_1 выглядит так:

$$\mathbf{P}\left(\overline{X} - \frac{y_{1-\alpha/2}\,S}{\sqrt{n-1}} < \theta_1 < \overline{X} - \frac{y_{\alpha/2}\,S}{\sqrt{n-1}}\right) = \mathbf{P}(y_{\alpha/2} < T_{n-1} < y_{1-\alpha/2}) = 1-\alpha,$$

где y_p — p-квантиль распределения Стьюдента t_{n-1} (см. таблицу Т4). В силу симметрии закона $y_{\alpha/2} = -y_{1-\alpha/2}$.

Доверительный интервал для параметра масштаба θ_2 таков:

$$\mathbf{P}\left(\frac{\sqrt{n}\,S}{\sqrt{z_{1-\alpha/2}}} < \theta_2 < \frac{\sqrt{n}\,S}{\sqrt{z_{\alpha/2}}}\right) = \mathbf{P}\left(z_{\alpha/2} < \frac{nS^2}{\theta_2^2} < z_{1-\alpha/2}\right) = 1-\alpha,$$

где z_p — p-квантиль закона χ^2_{n-1} (см. таблицу Т3).

Замечание 2 [32, с. 95]**.** Было бы неверным считать, что двумерный параметр (θ_1,θ_2) с вероятностью $(1-\alpha)^2$ накрывается случайным прямоугольником

$$\left(\overline{X} - \frac{y_{1-\alpha/2}S}{\sqrt{n-1}}\,,\, \overline{X} - \frac{y_{\alpha/2}S}{\sqrt{n-1}}\right) \times \left(\frac{\sqrt{n}\,S}{\sqrt{z_{1-\alpha/2}}}\,,\, \frac{\sqrt{n}\,S}{\sqrt{z_{\alpha/2}}}\right),$$

так как случайные величины T_{n-1} и S^2, на основании которых строились эти интервалы, зависимы (обе являются функциями от S).

Чтобы построить доверительную область для вектора (θ_1,θ_2), используем независимость $\overline{X}$ и S^2. Обозначим p-квантиль закона $\mathcal{N}(0,1)$ через x_p, а p-квантиль распределения χ^2_{n-1} через z_p. Тогда

$$\mathbf{P}\left(\sqrt{n}\,|\overline{X} - \theta_1|\,/\,\theta_2 < x_{1-\alpha/2},\quad z_{\alpha/2} < nS^2/\theta_2^2 < z_{1-\alpha/2}\right) = (1-\alpha)^2.$$

Соответствующее (случайное) множество точек плоскости представляет собой трапецию, отсекаемую от угла $\theta_2 = \sqrt{n}\,|\overline{X} - \theta_1|/x_{1-\alpha/2}$ двумя параллельными оси абсцисс прямыми $\theta_2 = \sqrt{nS^2/z_{\alpha/2}}$ и $\theta_2 = \sqrt{nS^2/z_{1-\alpha/2}}$ (рис. 8).

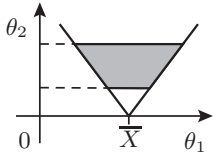


Рис. 8

Замечание 3. Величину T_{n-1}, определяемую формулой (6), можно представить также в виде

$$T_{n-1} = \sqrt{n}\,(\overline{X} - \theta_1)\Big/\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2\right)^{1/2}.$$

При сравнении этого представления с величиной $\sqrt{n}\,(\overline{X} - \theta_1)/\theta_2$, имеющей стандартное нормальное распределение, замечаем, что различие состоит в замене неизвестной дисперсии θ_2^2 на ее несмещенную и состоятельную оценку $\dfrac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2$ (см. пример 3 из § 2 гл. 6).

На то, что при этой подстановке закон $\mathcal{N}(0,1)$ меняется на t_{n-1}, впервые обратил внимание в 1908 г. Уильям Д. Госсет, работавший в то время в Дублине на пивоваренном заводе Гиннеса

(см. [72, с. 122]). Условия контракта не позволяли Госсету публиковать результаты под его собственным именем. Госсет выбрал псевдоним «Student». С тех пор найденное им распределение стало называться **законом Стьюдента**.

Сравним ряд квантилей $x_{1-\alpha/2}$ закона $\mathcal{N}(0,1)$ с соответствующими квантилями $y_{1-\alpha/2}$ распределения t_{n-1} при $n=10$:

α	0,05	10^{-2}	10^{-3}	10^{-5}
$x_{1-\alpha/2}$	1,96	2,58	3,29	4,26
$y_{1-\alpha/2}$	2,26	3,25	4,78	8,83

В частности, видим, что для выборки размера 10 длина интервала для θ_1 с коэффициентом доверия 95% при замене неизвестной дисперсии на ее оценку возрастает примерно на $2{,}26/1{,}96-1 \approx 15\%$.

§ 3. МЕТОДЫ ПОСТРОЕНИЯ ИНТЕРВАЛОВ*)

Метод 1. Использование центральной функции

Рассмотрим выборку $\boldsymbol{X} = (X_1,\ldots,X_n)$. Допустим, что найдется такая функция $g(\boldsymbol{x},\theta)$ (называемая *центральной*), что

1) распределение $g(\boldsymbol{X},\theta)$ не зависит от θ для всех $\theta \in \Theta$;

2) при каждом $\boldsymbol{x} \in \mathbb{R}^n$ функция $g(\boldsymbol{x},\theta)$ непрерывна и строго убывает (возрастает) по θ.

Скажем, для модели сдвига показательного закона из примера 1 можно взять $g(\boldsymbol{X},\theta) = X_{(1)} - \theta \sim \Gamma(1,n)$. Для $X_i \sim \mathcal{N}(\theta_1,\theta_2^2)$ из примера 4 годятся $\sqrt{n-1}\,(\overline{X}-\theta_1)/S \sim t_{n-1}$ для параметра сдвига θ_1 и $nS^2/\theta_2^2 \sim \chi^2_{n-1}$ для параметра масштаба θ_2.

Обозначим p-квантиль распределения $g(\boldsymbol{X},\theta)$ через x_p. Возьмем $0 \leqslant p_1 < p_2 \leqslant 1$ такие, что $p_2 - p_1 = 1-\alpha$. Определим $T_1(\boldsymbol{x})$ и $T_2(\boldsymbol{x})$ как решения относительно θ соответственно уравнений

$$g(\boldsymbol{x},\theta) = x_{p_1} \quad \text{и} \quad g(\boldsymbol{x},\theta) = x_{p_2}. \tag{7}$$

Однозначность их определения гарантируется условием 2. Тогда при всех $\theta \in \Theta$ для $g(\boldsymbol{x},\theta)$, убывающей по θ (рис. 9),

$$\mathsf{P}(T_2(\boldsymbol{X}) < \theta < T_1(\boldsymbol{X})) = \mathsf{P}(x_{p_1} < g(\boldsymbol{X},\theta) < x_{p_2}) = 1-\alpha.$$

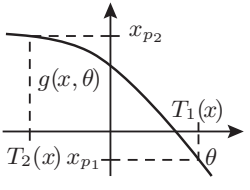


Рис. 9

Когда $p_1 = \alpha/2$, интервал называется *центральным*. При помощи именно этого метода был построен доверительный интервал в примере 1 ($p_1 = 0$), а также центральные интервалы для θ_1 и θ_2 в примере 4.

Нахождение $g(\boldsymbol{x},\theta)$ для конкретной модели — отдельная задача, не всегда имеющая решение. Однако, можно выделить класс моделей, для которых такая функция существует: если функция распределения $F(x,\theta)$ элементов выборки X_i непрерывна и строго монотонна

*) Материал этого параграфа имеет более технический характер.

по θ, то можно взять

$$g(\boldsymbol{X},\theta) = -\sum_{i=1}^{n} \ln F(X_i,\theta). \tag{8}$$

Действительно, ее непрерывность и монотонность очевидны, а в соответствии с методом обратной функции (см. § 1 гл. 4) случайные величины $\eta_i = F(X_i,\theta)$ равномерно распределены на отрезке $[0,1]$. Там же установлено, что каждая из величин $-\ln\eta_i$ имеет показательное распределение с параметром $\lambda = 1$, т. е. $-\ln\eta_i \sim \Gamma(1,1)$. Наконец, в силу леммы 1 гл. 4 находим, что $g(\boldsymbol{X},\theta) = -\sum\limits_{i=1}^{n} \ln\eta_i \sim \Gamma(n,1)$. (Пример применения формулы (8) содержится в задаче 1(а).)

Метод 2. Использование точечной оценки

Предположим, что имеется точечная оценка $T(\boldsymbol{X})$ для параметра θ с функцией распределения $F_T(t,\theta)$, которая непрерывна и строго убывает (возрастает) по θ.

Возьмем $0 \leqslant p_1 < p_2 \leqslant 1$ такие, что $p_2 - p_1 = 1 - \alpha$. Для каждого $\theta \in \Theta$ определим $t_k(\theta)$ как p_k-квантиль распределения $F_T(t,\theta)$, $k = 1,2$. Если функция $F_T(t,\theta)$ убывает по θ, то обе функции $t_k(\theta)$ возрастают (рис. 10).

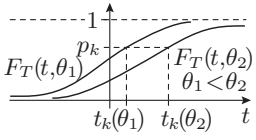


Рис. 10

Обозначим через G_α подмножество $\Theta \times \Theta$ следующего вида: $G_\alpha = \{(\theta,\theta')\colon t_1(\theta) < \theta' < t_2(\theta)\}$ (рис. 11). Определим $\Delta(\theta')$ как сечение G_α при фиксированном θ': $\Delta(\theta') = \{\theta\colon (\theta,\theta') \in G_\alpha\}$. Так как функции $t_1(\theta)$ и $t_2(\theta)$ строго возрастают, то множество $\Delta(\theta')$ является интервалом (возможно, бесконечным). Обозначим его левый и правый концы через $\theta_1(\theta')$ и $\theta_2(\theta')$.

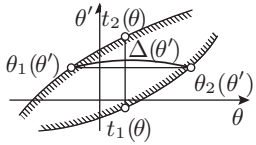


Рис. 11

Рассмотрим случайный интервал $\Delta(T(\boldsymbol{X}))$. Заметим, что событие $\{\theta \in \Delta(T(\boldsymbol{X}))\}$ происходит тогда и только тогда, когда $\{T(\boldsymbol{X}) \in (t_1(\theta), t_2(\theta))\}$ и, значит, при каждом $\theta \in \Theta$ имеет вероятность $1 - \alpha$, что и требуется.

Наглядный смысл метода состоит в том, что сначала строят диаграмму *по вертикали*: для каждой абсциссы θ находят соответствующие квантили $t_1(\theta)$ и $t_2(\theta)$, а затем для наблюдавшейся ординаты $t = T(\boldsymbol{x})$, где $\boldsymbol{x}$ — это реализация выборки, «считывают» *по горизонтали* значения $\theta_1(t)$ и $\theta_2(t)$. Другими словами, концы интервала (θ_1,θ_2) находят как решения относительно θ уравнений

$$F_T(T(\boldsymbol{x}),\theta) = p_k, \quad k = 1,2. \tag{9}$$

Пункт б) задачи 1 дает пример построения доверительного интервала с помощью этого метода.

Аналогичные рассуждения можно провести и для дискретной модели. Отличие состоит в том, что из-за ступенчатости функции распределения $F_T(t,\theta)$ удается, вообще говоря, добиться лишь выполнения неравенства $\mathsf{P}(t_1 < T(\boldsymbol{x}) < t_2) = F_T(t_2-,\theta) - F_T(t_1,\theta) \geqslant 1 - \alpha$. При этом вместо квантилей берут наибольшее t_1 и наименьшее t_2, удовлетворяющие, соответственно, условиям

$$F_T(t_1,\theta) \leqslant p_1 \quad \text{и} \quad F_T(t_2-,\theta) \geqslant p_2.$$

Кривые $\theta' = t_k(\theta)$ также будут ступенчатыми (рис. 12). При «считывании» θ_1 и θ_2 следует взять крайнюю правую точку пересечения горизонтальной прямой с левой кривой и крайнюю левую точку — с правой кривой.

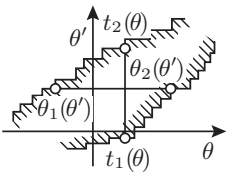


Рис. 12

Вместо уравнений (9) надо решать относительно θ уравнения

$$F_T(t,\theta) = p_1 \quad \text{и} \quad F_T(t-,\theta) = p_2, \quad \text{где } t = T(\boldsymbol{x}). \tag{10}$$

Пример 5. Построим доверительный интервал для неизвестной вероятности «успеха» θ в схеме Бернулли длины n. В качестве точечной оценки возьмем частоту «успехов» $T = \overline{X}$. Случайная величина T принимает значения k/n, $k = 0,1,\ldots,n$, при этом

$$F_T(k/n,\theta) = \mathbf{P}(nT \leqslant k) = \sum_{i=0}^{k} C_n^i \theta^i (1-\theta)^{n-i}. \tag{11}$$

Заметим, что правая часть соотношения (11) равна $1 - F_{\eta_{(k+1)}}(\theta)$, где $\eta_{(k+1)}$ — $(k+1)$-я порядковая статистика выборки размера n из равномерного распределения на отрезке $[0,1]$ (см. утверждение 3 гл. 5). При $k < n$ статистика $\eta_{(k+1)}$ имеет бета-плотность $nC_{n-1}^k x^k (1-x)^{n-k-1} > 0$ для $0 < x < 1$ (см. формулу (2) гл. 5). Следовательно, $F_T(k/n,\theta)$ строго убывает по θ при $k < n$. В соответствии с уравнениями (10), границы θ_1 и θ_2 центрального доверительного интервала находятся из соотношений

$$F_T((k-1)/n,\theta_1) = 1 - \alpha/2 \quad \text{и} \quad F_T(k/n,\theta_2) = \alpha/2. \tag{12}$$

В таблицах [10] они указаны для $1 - \alpha = 0{,}9;\ 0{,}95;\ 0{,}99$.

Пример 6. Пусть X_i распределены по закону Пуассона с неизвестным параметром θ: $\mathbf{P}(X_i = k) = e^{-\theta}\theta^k/k!$, $k = 0,1,\ldots$. Построим доверительный интервал для θ с помощью метода 2. Как и в предыдущем примере, в качестве точечной оценки возьмем $T = \overline{X}$. Ввиду задачи 3 гл. 10 сумма $X_1 + \ldots + X_n$ распределена по закону Пуассона параметром $n\theta$. Поэтому

$$F_T(k/n,\theta) = \mathbf{P}(nT \leqslant k) = e^{-n\theta} \sum_{i=0}^{k} (n\theta)^i / i!\,. \tag{13}$$

В силу формулы (1) гл. 5 правая часть (13) равна $1 - F_{S_{k+1}}(\theta)$, где $S_{k+1} \sim \Gamma(k+1,n)$ с плотностью $n^{k+1}x^k e^{-nx}/k! > 0$ при $x > 0$. Следовательно, $F_T(k/n,\theta)$ строго убывает по θ. Границы центрального доверительного интервала находятся из уравнений (12), где F_T задается формулой (13).

Метод 3. Стабилизация асимптотической дисперсии

Допустим, что известна асимптотически нормальная оценка (см.§ 4 гл. 7): $\sqrt{n}\,(\widehat{\theta}_n - \theta) \xrightarrow{d} \xi \sim \mathcal{N}(0, \sigma^2(\theta))$ при $n \to \infty$. Потребуем, чтобы асимптотическая дисперсия $\sigma^2(\theta)$ была положительна и непрерывна при всех $\theta \in \Theta$.

Построим *асимптотический* интервал для θ с помощью *преобразования, стабилизирующего дисперсию*, основанного на лемме 1

гл. 7. Для этого подберем такую функцию φ, чтобы асимптотическая дисперсия последовательности $\varphi(\widehat{\theta}_n)$ не зависела от неизвестного параметра θ:

$\sigma(\theta)\,\varphi'(\theta) = c,$

т. е.

$$\varphi(\theta) = c\int \frac{d\theta}{\sigma(\theta)}\,. \tag{14}$$

Тогда $\sqrt{n}\,(\varphi(\widehat{\theta}_n) - \varphi(\theta)) \xrightarrow{d} \varphi(\xi) \sim \mathcal{N}(0, c^2)$ при $n \to \infty$. Отсюда, так же, как и в примере 2, получаем:

$$\mathbf{P}\left(\varphi(\widehat{\theta}_n) - \frac{x_{1-\alpha/2}\,c}{\sqrt{n}} < \varphi(\theta) < \varphi(\widehat{\theta}_n) + \frac{x_{1-\alpha/2}\,c}{\sqrt{n}}\right) \to 1-\alpha, \tag{15}$$

где $x_{1-\alpha/2}$ обозначает $(1-\alpha/2)$-квантиль распределения $\mathcal{N}(0,1)$. Ввиду предполагаемой положительности и непрерывности $\sigma(\theta)$, из формулы (14) видим, что функция φ строго монотонна. Тогда из соотношения (15) очевидным образом находим асимптотический доверительный интервал для самого параметра θ.

Метод 4. Подстановка оценки параметра

Пусть выполнены предположения, сформулированные при изложении метода 3. Тогда

$$\frac{\sqrt{n}\,(\widehat{\theta}_n - \theta)}{\sigma(\widehat{\theta}_n)} = \frac{\sqrt{n}\,(\widehat{\theta}_n - \theta)}{\sigma(\theta)} \times \frac{\sigma(\theta)}{\sigma(\widehat{\theta}_n)}\,. \tag{16}$$

Распределение первого сомножителя в этой формуле сходится к $\mathcal{N}(0,1)$. Из асимптотической нормальности оценки $\widehat{\theta}_n$ вытекает ее состоятельность (см. ответ на вопрос 1 гл. 7). Ввиду непрерывности $\sigma(\theta)$ второй сомножитель в формуле (16) стремится к 1 по вероятности (П5). В силу свойства сходимости 1 из П5 произведение имеет в качестве предельного закона $\mathcal{N}(0,1)$, откуда

$$\mathbf{P}\left(\widehat{\theta}_n - x_{1-\alpha/2}\,\sigma(\widehat{\theta}_n)/\sqrt{n} < \theta < \widehat{\theta}_n + x_{1-\alpha/2}\,\sigma(\widehat{\theta}_n)/\sqrt{n}\right) \to 1-\alpha,$$

где $x_{1-\alpha/2}$ — это, как и прежде, $(1-\alpha/2)$-квантиль распределения $\mathcal{N}(0,1)$.

Замечание 4. Используя разные асимптотически нормальные оценки, будем получать различные доверительные интервалы. Чтобы строить интервалы наименьшей длины (при заданном коэффициенте доверия), следует выбирать оценки, имеющие наименьшую возможную асимптотическую дисперсию $\sigma^2(\theta)$ (асимптотически эффективные). Если для модели выполняются условия регулярности, то годятся оценки максимального правдоподобия (см. § 4 гл. 9).

Проиллюстрируем методы 3 и 4 на примере схемы Бернулли.

Для применения метода 3, как и в примере 5, возьмем в качестве оценки неизвестной вероятности «успеха» θ частоту $\overline{X}$. Она, согласно примеру 5 гл. 9, является оценкой максимального правдоподобия с $\sigma^2(\theta) = \theta(1-\theta)$. Условие (14) при $c = 1/2$ приводит к функции $\varphi(\theta) = \frac{1}{2}\int [\theta(1-\theta)]^{-1/2} d\theta = \arcsin\sqrt{\theta}$. При этом в силу формулы (15) интервал с границами

$$\arcsin\sqrt{\overline{X}} \pm \frac{x_{1-\alpha/2}}{2\sqrt{n}}$$

накрывает $\arcsin\sqrt{\theta}$ с вероятностью, стремящейся к $1-\alpha$ при $n \to \infty$. Применив функцию $\varphi^{-1}(x) = \sin^2 x$ к его границам, построим асимптотический доверительный интервал для самого параметра θ.

Метод 4, в свою очередь, приводит к интервалу с границами

$$\overline{X} \pm \frac{x_{1-\alpha/2}}{\sqrt{n}}\ \sqrt{\overline{X}(1-\overline{X})}.$$

Нужно обращать острие ума на самые незначительные и простые вещи и долго останавливаться на них, пока не привыкнешь отчетливо и ясно прозревать в них истину.

Р. Декарт

Р. Декарт (1596—1650), французский математик и философ.

Заметим, что для $\alpha = 5\%$, $n = 100$ и $\overline{x} = 0{,}03$ в качестве левой границы последнего интервала формально получаем отрицательную величину $-0{,}0034$.

Строили мы, строили и, наконец, построили.

Чебурашка

ЗАДАЧИ

1. В модели выбора наудачу n точек с координатами X_i из отрезка $[0,\theta]$ с неизвестным $\theta > 0$ постройте интервал с коэффициентом доверия $1-\alpha$

а) на основе формулы (8),

б) при помощи метода 2 с использованием оценки $X_{(n)} = \max\{X_1,\ldots,X_n\}$.

2. Элементы X_i выборки имеют функцию распределения $F(x-\theta)$. Пусть $F(0) = 0$, а плотность $p(x)$ такова, что $c = p(0) > 0$.

а) Найдите предельный закон для величины $n(X_{(1)} - \theta)$, где $X_{(1)} = \min\{X_1,\ldots,X_n\}$.

б) На его основе постройте асимптотический доверительный интервал для параметра сдвига θ.

3. Рассмотрим модель сдвига из предыдущей задачи, но потребуем теперь, чтобы $F(0) = 1/2$ и F была непрерывной. Найдите коэффициент доверия интервала, образованного парой порядковых статистик (см. § 4 гл. 4) $(X_{(k)},X_{(l)})$, где $k < l$. Вычислите его значение для $k = 2$, $l = 5$, $n = 6$.

4. Докажите теорему 1 при $n = 2$ непосредственно.

5. Для распределения Пуассона из примера 6 постройте асимптотический доверительный интервал методами 3 и 4.

6. Пусть (X,Y) — двумерный нормальный вектор (П9). Неизвестный *коэффициент корреляции* его компонент

$$\rho = \boldsymbol{\rho}(X,Y) = \mathbf{M}(X - \mathbf{M}X)(Y - \mathbf{M}Y) \Big/ \sqrt{\mathbf{D}X\,\mathbf{D}Y}$$

оценивается при помощи *выборочного коэффициента корреляции*

$$\widehat{\rho}_n = \sum (X_i - \overline{X})(Y_i - \overline{Y}) \Big/ \sqrt{\sum (X_i - \overline{X})^2 \sum (Y_i - \overline{Y})^2}\ .$$

Только сокровища ума действительны. Ими можно делиться, ничего не теряя; они даже умножаются, когда ими делятся. Чтобы приобрести такое богатство, надо трудиться.

Демофил

Можно показать (см. [50, с. 391]), что распределение величины $\sqrt{n}\,(\widehat{\rho}_n - \rho)$ сходится при $n \to \infty$ к закону $\mathcal{N}(0, (1-\rho^2)^2)$. Найдите преобразование (впервые предложенное Р. Фишером), стабилизирующее дисперсию $\widehat{\rho}_n$.

РЕШЕНИЯ ЗАДАЧ

Читать следует тогда только, когда иссяк источник собственных мыслей, что нередко случается и с самым умным человеком. Но спугнуть, ради книги, собственную, неокрепшую мысль — это значит совершить преступление против духа.

А. Шопенгауэр

1. а) Функция $F(X_i,\theta) = x/\theta$, $0 \leqslant x \leqslant \theta$, убывает по θ при $\theta > 0$. Согласно формуле (8), $g(\boldsymbol{X},\theta) = n\ln\theta - \sum \ln X_i$. Решая уравнения (7) относительно θ, находим границы интервала

$$T_k(\boldsymbol{X}) = \left(e^{x_{p_k}} \prod_{i=1}^{n} X_i\right)^{1/n}, \quad k = 1,2,$$

где x_{p_k} обозначает p_k-квантиль закона $\Gamma(n,1)$, $0 \leqslant p_1 < p_2 \leqslant 1$ и $p_2 - p_1 = 1 - \alpha$.

б) Функция распределения статистики $X_{(n)}$, равная $(x/\theta)^n$ при $0 \leqslant x \leqslant \theta$, убывает по θ. Поскольку p-квантиль $x_p = p^{1/n}\theta$, множество G_α представляет собой угол между лучами $\theta' = p_1^{1/n}\theta$ и $\theta' = p_2^{1/n}\theta$, где $p_2 - p_1 = 1 - \alpha$ (рис. 13). Сечение $\Delta(X_{(n)})$ является интервалом $\left(X_{(n)}\, p_2^{-1/n},\, X_{(n)}\, p_1^{-1/n}\right)$.

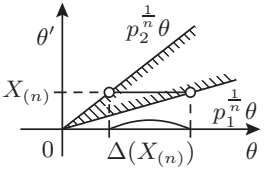


Рис. 13

Вопрос 6.
При каком значении p_2 доверительный интервал из задачи 1б имеет наименьшую длину?

2. а) Используем независимость величин X_i и формулу Тейлора:

$$\begin{aligned}
&\mathbf{P}(n\,(X_{(1)} - \theta) > x) = \mathbf{P}(X_{(1)} > \theta + x/n) = \\
&\quad = \mathbf{P}(X_i - \theta > x/n,\, i = 1,\ldots,n) = [1 - F(x/n)]^n = \\
&\quad = [1 - F(0) - p(0)\,x/n + o(1/n)]^n = \\
&\quad = [1 - p(0)\,x/n + o(1/n)]^n \to \exp\{-p(0)x\} \quad \text{при } n \to \infty.
\end{aligned}$$

б) Осталось, как в примере 1, взять $\widehat{\theta}_2 = X_{(1)}$ и $\widehat{\theta}_1 = X_{(1)} - c_\alpha$, где $c_\alpha = -(\ln\alpha)\,/\,(p(0)n)$.

3. Рассмотрим случайные величины $Y_i = X_i - \theta$, имеющие функцию распределения $F(x)$. Для них $Y_{(i)} = X_{(i)} - \theta$ и

$$\mathbf{P}(X_{(k)} < \theta < X_{(l)}) = \mathbf{P}(Y_{(k)} < 0 < Y_{(l)}).$$

Поскольку $\mathbf{P}(Y_{(k)} < Y_{(l)}) = 1$ при $k < l$, выполняется равенство

$$\mathbf{P}(Y_{(k)} < 0 < Y_{(l)}) = \mathbf{P}(Y_{(k)} < 0) - \mathbf{P}(Y_{(l)} < 0).$$

Согласно ответу на вопрос 3 гл. 5 о распределении случайной величины $Y_{(k)}$,

$$\mathbf{P}(Y_{(k)} < 0) = \sum_{i=k}^{n} C_n^i\, F(0)^i (1 - F(0))^{n-i}.$$

Учитывая условие $F(0) = 1/2$, находим отсюда, что

$$\mathbf{P}(X_{(k)} < \theta < X_{(l)}) = 2^{-n} \sum_{i=k}^{l-1} C_n^i.$$

Для $k = 2$, $l = 5$, $n = 6$ получаем значение коэффициента доверия $2^{-6}(C_6^2 + C_6^3 + C_6^4) = 50/64 \approx 0{,}78$.

Отметим, что в отличие от предыдущей задачи доверительный интервал строится при помощи пары разных статистик, а не за счет «подправления» одной статистики.

4. Для $n = 2$ имеем:

$$\overline{X} = \frac{1}{2}(X_1 + X_2), \quad S^2 = \frac{1}{2}(X_1^2 + X_2^2) - \overline{X}^2 = \frac{1}{4}(X_1 - X_2)^2.$$

Случайные величины $Y_1 = X_1 + X_2$ и $Y_2 = X_1 - X_2$ нормально распределены: $Y_1 \sim \mathcal{N}(2\theta_1, 2\theta_2^2)$, $Y_2 \sim \mathcal{N}(0, 2\theta_2^2)$ (в частности, $\mathbf{M}Y_2 = 0$). Причем ковариация между ними равна

$$\mathbf{M}(Y_1Y_2) - \mathbf{M}Y_1 \cdot \mathbf{M}Y_2 = \mathbf{M}(X_1^2 - X_2^2) = \mathbf{M}X_1^2 - \mathbf{M}X_2^2 = 0.$$

Следовательно (см. П9 или [90, с. 322]), они независимы. Поэтому независимы, как функции от них, $\overline{X}$ и S^2. Наконец, так как $Y_2 \big/ \left(\sqrt{2}\,\theta_2\right) \sim \mathcal{N}(0,1)$, то $2S^2/\theta_2^2 = \left[Y_2 \big/ \left(\sqrt{2}\,\theta_2\right)\right]^2 \sim \chi_1^2$.

5. В силу центральной предельной теоремы (П6) для выборки из закона с конечной дисперсией $\sqrt{n}\,(\overline{X} - \mathbf{M}X_1)$ при $n \to \infty$ сходится по распределению к $\mathcal{N}(0, \mathbf{D}X_1)$. Из ответа на вопрос 2 гл. 5 для закона Пуассона $\mathbf{M}X_1 = \mathbf{D}X_1 = \theta$. По формуле (14) при $c = 1/2$ вычислим $\varphi(\theta) = \frac{1}{2}\int \theta^{-1/2}d\theta = \sqrt{\theta}$. Таким образом, асимптотические доверительные интервалы строятся на основе сходимости $\sqrt{n}\left(\sqrt{\overline{X}} - \sqrt{\theta}\right)$ к закону $\mathcal{N}(0, 1/4)$ (метод 3) и $\sqrt{n}\,(\overline{X} - \theta) \Big/ \sqrt{\overline{X}}$ к закону $\mathcal{N}(0,1)$ (метод 4).

6. Для $c = 1$ найдем

$$\varphi(\rho) = \int \frac{d\rho}{1-\rho^2} = \frac{1}{2}\int \left(\frac{1}{1+\rho} + \frac{1}{1-\rho}\right) d\rho = \frac{1}{2}\ln\frac{1+\rho}{1-\rho} \equiv \operatorname{arcth}\rho,$$

где $\operatorname{arcth} x$ — это функция, обратная к *гиперболическому тангенсу* $\operatorname{th} x = (e^x - e^{-x})/(e^x + e^{-x})$.

При $|\rho|$ близких к 1 и не слишком больших n распределение оценки $\widehat{\rho}_n$ сильно отличается от нормального. Преобразование Р. Фишера $\widehat{z}_n = \operatorname{arcth}\widehat{\rho}_n$ в этом случае существенно повышает точность нормального приближения (на рис. 14, взятом из [13, с. 381], изображены графики плотностей величин $\widehat{\rho}_n$ и $\widehat{z}_n$ для $\rho = 0{,}8$).

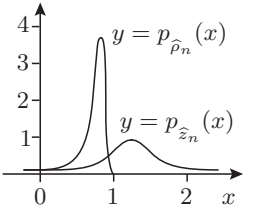


Рис. 14

В [10, с. 51] для статистики $\widehat{z}_n$ приведены следующие асимптотические формулы:

$$\mathbf{M}\widehat{z}_n = \operatorname{arcth}\rho + o(1/n), \quad \mathbf{D}\widehat{z}_n = 1/(n-3) + o(1/n) \text{ при } n \to \infty.$$

Б. Л. Ван дер Варден в [13, с. 41] пишет: «Вообще в математической статистике часто оказывается, что 4 уже является большим числом». В формуле для $\mathbf{D}\widehat{z}_n$, во всяком случае, $n \geqslant 4$.

ОТВЕТЫ НА ВОПРОСЫ

1. Так же, как и в примере 1, запишем:

$$1-\alpha=\mathsf{P}\left(X_{(1)}-c_\alpha<\theta\right)=1-\left[\mathsf{P}\left(X_1-\theta\geqslant c_\alpha\right)\right]^n=1-(1-c_\alpha)^n.$$

Отсюда находим, что длина интервала $c_\alpha = 1-\alpha^{1/n}$.

2. Обозначим через $\Phi(x)$ функцию распределения $\mathcal{N}(0,1)$. Тогда при $x \geqslant 0$

$$\mathsf{P}(Z_1^2 \leqslant x) = \mathsf{P}(-\sqrt{x} \leqslant Z_1 \leqslant \sqrt{x}) = \Phi(\sqrt{x}) - \Phi(-\sqrt{x}).$$

Дифференцируя по x, вычислим плотность случайной величины Z_1^2:

$$p_{Z_1^2}(x)=\frac{1}{2\sqrt{x}}\left(\Phi'(\sqrt{x})+\Phi'(-\sqrt{x})\right)=\frac{1}{2\sqrt{x}}\,\frac{2}{\sqrt{2\pi}}\,e^{-x/2}=\frac{1}{\sqrt{2\pi x}}\,e^{-x/2},$$

которая совпадает с плотностью закона $\Gamma\left(\frac{1}{2},\frac{1}{2}\right)$, поскольку $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$. Этот результат можно получить и с помощью общей формулы преобразования плотности из П8 (убедитесь!), заметив, что $Z_1^2 = |Z_1|^2$, где $\mathsf{P}(|Z_1| \leqslant x) = \Phi(x) - \Phi(-x)$ при $x \geqslant 0$ и, следовательно, $p_{|Z_1|}(x) = 2\,\Phi'(x)$ при $x \geqslant 0$.

3. $\mathsf{M}R_k^2 = \mathsf{M}(Z_1^2 + \ldots + Z_k^2) = k\,\mathsf{M}Z_1^2 = k\,\mathsf{D}Z_1 = k$.

4. В соответствии с законом больших чисел (П6) имеем, что

$$\frac{1}{k}R_k^2 = \frac{1}{k}(Z_1^2 + \ldots + Z_k^2) \xrightarrow{\mathsf{P}} \mathsf{M}Z_1^2 = 1$$

при $k \to \infty$. Функция $1/\sqrt{x}$ непрерывна при $x > 0$. В силу свойств сходимости 3 и 1 из П5 предельным законом для t_k будет $\mathcal{N}(0,1)$.

Другой подход: очевидно, $(1 + x^2/k)^{-(k+1)/2} \to e^{-x^2/2}$ при $k \to \infty$. Используя асимптотику $\Gamma(x) = \sqrt{2\pi}\,x^{x-1/2}e^{-x}(1+o(1))$ при $x \to +\infty$, обобщающую формулу Стирлинга на действительные x (см. [81, с. 84]), нетрудно проверить, что $d_k \to 1/\sqrt{2\pi}$. Наконец, из поточечной сходимости плотностей *к плотности* вытекает сходимость по распределению (см. [90, с. 394]).

5. Во-первых, $(\boldsymbol{C}^T)^{-1} = (\boldsymbol{C}^{-1})^T = (\boldsymbol{C}^T)^T$. Во-вторых, выполняются соотношения:

$$1 = \det \boldsymbol{E} = \det(\boldsymbol{C}^T \cdot \boldsymbol{C}) = \det \boldsymbol{C}^T \cdot \det \boldsymbol{C} = (\det \boldsymbol{C})^2.$$

6. Длина доверительного интервала

$$X_{(n)}\left(p_1^{-1/n} - p_2^{-1/n}\right) = X_{(n)}\,p_2^{-1/n}\left\{[1-(1-\alpha)/p_2]^{-1/n} - 1\right\}$$

убывает по $p_2 \in (1-\alpha, 1]$.

Часть III

ПРОВЕРКА ГИПОТЕЗ

В этой части книги приводятся *основные понятия теории проверки статистических гипотез*: статистика критерия, критическое множество, ошибки I и II рода и др. Рассматриваются методы проверки равномерности, показательности и нормальности распределения элементов выборки. Устанавливается оптимальность критерия Неймана—Пирсона. Обсуждается последовательный анализ Вальда.

Глава 12

КРИТЕРИИ СОГЛАСИЯ

§ 1. СТАТИСТИЧЕСКИЙ КРИТЕРИЙ

Эксперимент. Предположим, что кто-то подбросил 10 раз монетку, и в 8 случаях она упала гербом вверх. Можно ли считать эту монетку симметричной?

Статистическая модель. Используем для описания эксперимента схему Бернулли, т. е. будем считать данные эксперимента реализацией выборки $\boldsymbol{X} = (X_1, \ldots, X_{10})$, где $X_i = 1$ (выпадает герб) с вероятностью θ и $X_i = 0$ (выпадает решка) с вероятностью $1 - \theta$. Как проверить гипотезу H о том, что $\theta = 1/2$?

Правило, позволяющее принять или отвергнуть гипотезу H на основе реализации выборки $x_1, \ldots, x_n$, называется *статистическим критерием*. Обычно критерий задается при помощи *статистики критерия* $T(x_1, \ldots, x_n)$ такой, что для нее типично принимать умеренные значения в случае, когда гипотеза H верна, и большие (малые) значения, когда H не выполняется.

Для приведенного выше эксперимента в качестве статистики T можно взять сумму $x_1 + \ldots + x_n$. Тогда гипотезе $H\colon \theta = 1/2$ противоречат значения, которые близки к 0 или n.

При проверке гипотез с помощью критериев всегда присутствует возможность ошибочно отвергнуть гипотезу H, когда на самом деле она верна. Например, симметричная монета может случайно упасть 10 раз подряд гербом вверх. Но вероятность наблюдать такое событие равна всего лишь $2^{-10} = 1/1024$. Если мы готовы пренебречь возможностью осуществления столь маловероятного события, то появление 10 гербов подряд следует считать основанием для отклонения гипотезы $H\colon \theta = 1/2$.

В общем случае задается малое число α — вероятность, с которой мы можем позволить себе отвергнуть верную гипотезу H (скажем, $\alpha = 0{,}05$). Это число называют *уровнем значимости*. Исходя из предположения, что гипотеза H верна, определяется наименьшее значение $x_{1-\alpha}$, удовлетворяющее условию

$$\mathsf{P}(T(X_1, \ldots, X_n) \geqslant x_{1-\alpha}) \leqslant \alpha. \tag{1}$$

Если функция распределения статистики T непрерывна, то $x_{1-\alpha}$ является, очевидно, ее $(1-\alpha)$-квантилью (см. § 3 гл. 7). Такое $x_{1-\alpha}$ называют *критическим значением*: гипотеза H отвергается, если $t_0 = T(x_1,\ldots,x_n) \geqslant x_{1-\alpha}$ (произошло маловероятное событие), и принимается — в противном случае.

При этом величина $\alpha_0 = \mathbf{P}(T(X_1,\ldots,X_n) \geqslant t_0)$ задает *фактический уровень значимости*. Он равен вероятности того, что статистика T (измеряющая степень отклонения полученной реализации от наиболее типичной) за счет случайности примет значение t_0 или даже больше. Фактический уровень значимости — наибольший уровень, на котором проверяемая гипотеза H принимается (рис. 1).

Если значение T попало в область, имеющую при выполнении гипотезы H высокую вероятность, то можно заключить, что данные *согласуются* с гипотезой H. Отсюда происходит термин «*критерии согласия*».

Проверим для данных эксперимента гипотезу $H\colon \theta = 1/2$ на уровне значимости $\alpha = 0{,}05$ и вычислим α_0. Известно, что сумма $T = x_1 + \ldots + x_n$ имеет биномиальное распределение:

$$\mathbf{P}(T \geqslant k) = \sum_{i=k}^{n} C_n^i\, \theta^i (1-\theta)^{n-i}.$$

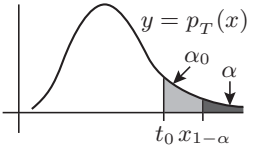


Рис. 1

Для $\theta = 1/2$ правая часть этого выражения при $k = 8$ равна $(45 + 10 + 1)/1024 \approx 0{,}055$ и при $k = 9$ равна $(10 + 1)/1024 \approx 0{,}011$. Поэтому для $\alpha = 0{,}05$ наименьшим $x_{1-\alpha}$, удовлетворяющим условию (1), будет 9. Поскольку полученное в эксперименте значение $t_0 = T(x_1,\ldots,x_n) = 8 < 9$, на заданном уровне значимости гипотеза $H\colon \theta = 1/2$ принимается.

С другой стороны, фактический уровень значимости $\alpha_0 = \mathbf{P}(T \geqslant 8) \approx 0{,}055$, что всего на 0,005 превосходит заданный уровень: уже при $\alpha = 0{,}06$ гипотезу H следует отклонить.

На основе данных эксперимента нельзя уверенно принять или отвергнуть гипотезу H (хотя последнее представляется более правдоподобным). Следовало бы еще несколько раз подбросить монетку, чтобы прийти к более взвешенному заключению.

Вычисление фактического уровня значимости нередко позволяет избегать категоричных (и при этом — ошибочных) выводов, сделанных лишь на основе сравнения t_0 с критическим значением $x_{1-\alpha}$, найденным для формально заданного α.

Вопрос 1.
Чему приближенно равна вероятность наблюдать не менее 60 падений гербом вверх при 100 бросаниях симметричной монеты?

(Воспользуйтесь табл. Т2.)

Не все стриги, что растет.

Козьма Прутков

§ 2. ПРОВЕРКА РАВНОМЕРНОСТИ

Пример 1. ***Орбиты планет и комет*** [72, с. 113]**.** В 1734 г. Французская академия присудила Даниилу Бернулли премию за исследование по орбитам планет, в котором он пытался показать, что схожесть орбит является неслучайной. Если предположить, что Солнце и планеты образовались в результате концентрации вещества первоначального «волчка» (рис. 2), то согласно *закону*

сохранения момента импульса $\boldsymbol{J}$ орбиты планет должны лежать примерно в одной плоскости, что и наблюдается в реальности.

Д. Бернулли (1700–1782), швейцарский математик (племянник Якоба Бернулли (1654–1705), установившего в 1713 г. справедливость закона больших чисел для частоты «успехов» в независимых испытаниях). Д. Бернулли известен своими результатами в области механики жидкостей и газов. В 1778 г. им была опубликована в изданиях Петербургской Академии наук работа «Наиболее вероятное определение по нескольким расходящимся между собой наблюдениям и устанавливаемое отсюда наиболее правдоподобное заключение», где впервые был высказан и использован для оценки неизвестного параметра принцип максимального правдоподобия (см. [19, с. 419]).

В 1812 г. Лаплас исследовал схожую проблему: образовались ли и кометы в общем «волчке» или же они — всего лишь «гости», захваченные притяжением Солнца. В последнем случае углы между нормалями к плоскостям орбит комет и вектором $\boldsymbol{J}$ должны не концентрироваться вблизи нуля, а быть равномерно распределенными на отрезке $[0, \pi/2]$. Проведя статистическую обработку известных к тому времени астрономических данных, Лаплас пришел к выводу, что гипотеза о равномерности не отвергается.

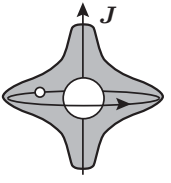


Рис. 2

Каким же образом можно проверить гипотезу равномерности? Рассмотрим несколько разных методов.

Метод 1. Критерий Колмогорова

Статистикой критерия является величина

$$D_n = \sup_{-\infty<x<\infty} |\widehat{F}_n(x) - F(x)|, \tag{2}$$

где $\widehat{F}_n(x) = \dfrac{1}{n}\sum\limits_{i=1}^{n} I_{\{X_i \leqslant x\}}$ — это эмпирическая функция распределения, встречавшаяся ранее в § 1 гл. 9, $F(x)$ — функция распределения элементов выборки (на рис. 3 изображен случай равномерного распределения на $[0,1]$). Для любого фиксированного x_0 согласно усиленному закону больших чисел значение $\widehat{F}_n(x_0)$, равное частоте попаданий X_i левее x_0, при $n \to \infty$ с вероятностью 1 стремится к $F(x_0) = \mathbf{P}(X_1 \leqslant x_0)$. *Теорема Гливенко* утверждает, что для произвольной функции распределения $F(x)$ имеет место сходимость $D_n \xrightarrow{\text{п. н.}} 0$. (Доказательство этой теоремы приведено в [19, с. 206].) Поэтому в случае, когда гипотеза равномерности верна, значение D_n для выборки достаточно большого размера не должно существенно отклоняться от нуля.

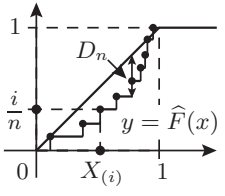


Рис. 3

Как количественно характеризуется значимость отклонения от нуля? В силу центральной предельной теоремы (П6)

$$\sqrt{n}\,(\widehat{F}_n(x_0) - F(x_0)) \xrightarrow{d} \xi \sim \mathcal{N}(0, F(x_0)(1 - F(x_0)).$$

Поэтому в фиксированной точке x_0 величина $|\widehat{F}_n(x_0) - F(x_0)|$ имеет порядок малости $1/\sqrt{n}$. Оказывается, что и величина D_n имеет тот же порядок малости, причем справедлив следующий результат.

Теорема Колмогорова. Если функция распределения элементов выборки $F(x)$ непрерывна, то для $x > 0$

$$\lim_{n\to\infty} \mathbf{P}(\sqrt{n}\, D_n \leqslant x) = K(x) = 1 + 2\sum_{k=1}^{\infty} (-1)^k e^{-2k^2x^2}.$$

Быстрая сходимость к предельному закону позволяет пользоваться этим приближением уже при $n \geqslant 20$.

Замечание 1. Особенностью статистики D_n является то, что закон ее распределения оказывается одним и тем же для всех непрерывных функций F. Он зависит только от размера выборки n. Действительно, полагая в формуле (2) $x = F^{-1}(y) = \sup\{x\colon F(x) = y\}$, $0 \leqslant y \leqslant 1$ (рис. 4), получаем $D_n = \sup\limits_{0\leqslant y\leqslant 1} |\widehat{F}_n(F^{-1}(y)) - y|$.

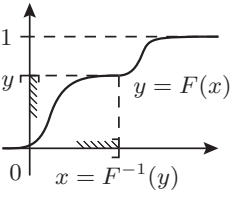


Рис. 4

Согласно методу обратной функции (см. § 1 гл. 4) случайные величины $Y_i = F(X_i)$ образуют выборку из равномерного распределения на отрезке $[0, 1]$. В силу монотонности и непрерывности функции $F(x)$ неравенства $x \leqslant F^{-1}(y)$ и $F(x) \leqslant y$ эквивалентны (см. рис. 4). Поэтому

$$\widehat{F}_n(F^{-1}(y)) = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i\leqslant F^{-1}(y)\}} = \frac{1}{n}\sum_{i=1}^{n} I_{\{Y_i\leqslant y\}}.$$

Правая часть — эмпирическая функция выборки $Y_1, \ldots, Y_n$.

Приведем таблицу некоторых квантилей функции $K(x)$:

α	0,5	0,15	0,1	0,05	0,025	0,01	0,001
$x_{1-\alpha}$	0,83	1,14	1,23	1,36	1,48	1,63	1,95

Таким образом, для заданного уровня значимости α критерий Колмогорова отвергает *гипотезу равномерности*, если для $F(x) = x$ величина $\sqrt{n}\, D_n(x_1, \ldots, x_n) \geqslant x_{1-\alpha}$.

Так как функция распределения $F(x)$ непрерывна и не убывает, а $\widehat{F}_n(x)$ — кусочно-постоянна, то sup в формуле (2) достигается в одной из точек разрыва функции $\widehat{F}_n$. Отсюда получаем простую формулу для вычисления значения $D_n(x_1, \ldots, x_n)$:

$$D_n(x_1, \ldots, x_n) = \max_{1\leqslant i\leqslant n} \left\{ \frac{i}{n} - F(x_{(i)}), F(x_{(i)}) - \frac{i-1}{n} \right\}.$$

Вопрос 2. Как может выглядеть эмпирическая функция распределения $\widehat{F}_n$, для которой $D_n(x_1, \ldots, x_n) \neq \max\limits_{1\leqslant i\leqslant n} \left| \frac{i}{n} - F(x_{(i)}) \right|$?

В задаче 1 критерий применяется для проверки качества таблицы случайных чисел Т1. Задача 3 показывает, что условие непрерывности функции $F(x)$ в теореме Колмогорова необходимо.

Метод 2. Критерий омега-квадрат

Статистика D_n измеряет отклонение эмпирической функции распределения $\widehat{F}_n$ от теоретической функции распределения F в равномерной метрике. Если воспользоваться (взвешенной) квадратичной

метрикой, то получим *статистику критерия омега-квадрат*:

$$\omega_n^2(\psi) = \int\limits_{-\infty}^{\infty} [\widehat{F}_n(x) - F(x)]^2 \, \psi[F(x)] \, dF(x),$$

где $\psi(y)$ — заданная на $[0,1]$ весовая функция. Рассмотрим два варианта: $\psi_1 = 1$ (*критерий Крамéра — Мизеса*), $\psi_2(y) = 1/[y(1-y)]$ (*критерий Андерсона — Дарлинга*).

Первый из них хорошо улавливает расхождение между $\widehat{F}_n$ и F в области «типичных значений» случайной величины с функцией распределения F (часто он оказывается более чувствительным, чем критерий Колмогорова). Второй же, благодаря тому, что $\psi_2(y)$ быстро возрастает при $y \to 0$ и $y \to 1$, способен заметить различие «на хвостах» распределения F, которому придается дополнительный вес.

Так же, как и для статистики D_n, закон распределения величины $\omega_n^2(\psi)$ один и тот же для всех непрерывных функций F.

При выполнении ряда условий относительно ψ можно доказать, что существует $\lim\limits_{n\to\infty} \mathbf{P}(n\omega_n^2(\psi) \leqslant x) = A(x)$, зависящий от ψ. Для ψ_1 и ψ_2 известны разложения в ряды соответствующих законов $A_1(x)$ и $A_2(x)$ (см. [10, с. 83]). Приведем таблицу некоторых квантилей y_p и z_p этих законов ($A_1(y_p) = A_2(z_p) = p$):

α	0,5	0,15	0,1	0,05	0,025	0,01	0,001
$y_{1-\alpha}$	0,12	0,28	0,35	0,46	0,58	0,74	1,17
$z_{1-\alpha}$	0,77	1,62	1,94	2,49	3,08	3,88	5,97

Значения $n\omega_n^2(\psi_1)$ и $n\omega_n^2(\psi_2)$ вычисляются по следующим формулам (первая из них выводится в задаче 4):

$$n\omega_n^2(\psi_1) = \frac{1}{12\,n} + \sum_{i=1}^{n} \left[F(x_{(i)}) - \frac{2i-1}{2n} \right]^2,$$

$$n\omega_n^2(\psi_2) = -n - 2\sum_{i=1}^{n} \left[\frac{2i-1}{2n} \ln F(x_{(i)}) + \left(1 - \frac{2i-1}{2n}\right) \ln(1 - F(x_{(i)})) \right].$$

В гл. 18 появится еще один критерий (так называемый *критерий хи-квадрат*), с помощью которого можно проверять равномерность по сгруппированным данным.

§ 3. ПРОВЕРКА ПОКАЗАТЕЛЬНОСТИ

Многие вещи нам непонятны не потому, что наши понятия слабы; но потому, что сии вещи не входят в круг наших понятий.

Козьма Прутков

Прежде чем познакомиться с методом проверки показательности, введем формально понятие статистической гипотезы.

Напомним, что под статистической моделью в § 1 гл. 6 понималось семейство функций распределения $\{F(x,\theta),\, \theta \in \Theta\}$, где Θ — множество возможных значений параметра. При этом данные

$x_1,\ldots,x_n$ рассматривались как реализация выборки $X_1,\ldots,X_n$, элементы которой имеют функцию распределения $F(x,\theta_0)$ с неизвестным значением $\theta_0 \in \Theta$.

Пусть выделено некоторое подмножество $\Theta_0 \subset \Theta$. Под *статистической гипотезой* H понимается предположение о том, что $\theta_0 \in \Theta_0$. Если множество Θ_0 состоит всего из одной точки, то гипотеза H называется *простой*, иначе — *сложной*. В последнем случае задача заключается в проверке принадлежности закона распределения величин X_i целому классу функций распределения $\{F(x,\theta),\, \theta \in \Theta_0\}$.

Под гипотезой показательности понимается сложная гипотеза, в которой этот класс образуют функции распределения вида $F(x,\theta) = \left(1 - e^{-\theta x}\right) I_{\{x>0\}}$, где $\theta > 0$ (рис. 5). Рассмотрим методы проверки такой гипотезы.

Метод 1. Исключение неизвестного параметра

Согласно лемме 3 гл. 4, вектор $(S_1/S_n,\ldots,S_{n-1}/S_n)$, где $S_k = X_1 + \ldots + X_k$, распределен так же, как вектор порядковых статистик $(\eta_{(1)},\ldots,\eta_{(n-1)})$ для выборки размера $(n-1)$ из равномерного распределения на отрезке $[0,1]$.

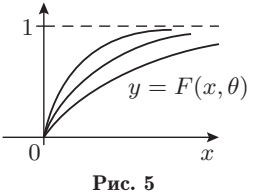


Рис. 5

Так как эмпирическая функция распределения строится по порядковым статистикам, то данное преобразование сводит задачу к проверке равномерности. Однако, за исключение «мешающего» параметра θ приходится платить уменьшением размера выборки на 1.

Пример 2. Следующие данные представляют собой количества летных часов между последовательными отказами установки для кондиционирования воздуха на самолете типа «Боинг-720» [40].

23	261	87	7	120	14	62	47	225	71
246	21	42	20	5	12	120	11	3	14
71	11	14	11	16	90	1	16	52	95

Считая времена между отказами независимыми, проверим гипотезу их показательности. Вычислим $s_k = x_1 + \ldots + x_k$, $k = 1,\ldots,30$:

23	284	371	378	498	512	574	621	846	917
1163	1184	1226	1246	1251	1263	1383	1394	1397	1411
1482	1493	1507	1518	1534	1624	1625	1641	1693	1788

В результате деления на $s_{30} = 1788$, получим ряд значений s_k/s_{30}, $k = 1,\ldots,29$:

0,013	0,159	0,207	0,211	0,279	0,286	0,321	0,347	0,473	0,513
0,650	0,662	0,686	0,697	0,700	0,706	0,773	0,780	0,781	0,789
0,829	0,835	0,843	0,849	0,858	0,908	0,909	0,918	0,947	

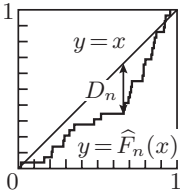


Рис. 6

Построенная по этому ряду эмпирическая функция распределения изображена на рис. 6. Максимальное отклонение $D_n = 0{,}65 - 10/29 \approx 0{,}306$, $\sqrt{29}\,D_n \approx 1{,}65$. Из приведенной выше таблицы квантилей функции Колмогорова находим, что гипотеза показательности отвергается на уровне значимости 1%.

Статистики $n\omega_n^2(\psi_1)$ и $n\omega_n^2(\psi_2)$ критерия омега-квадрат равны, соответственно, 0,627 и 3,036. Из таблицы квантилей $A_1(x)$ и $A_2(x)$ следует, что первая значимо велика на уровне приблизительно 2%, вторая — на уровне 2,5%.

Метод 2. Подстановка оценки параметра

Пусть $\tilde\theta_n$ — оценка максимального правдоподобия для параметра θ (см. § 4 гл. 9). Рассмотрим *модифицированные статистики Колмогорова и Крамéра—Мизеса*:

$$\tilde D_n = \sup_x \left|\widehat{F}_n(x) - F(x,\tilde\theta_n)\right|,$$
$$\tilde\omega_n^2 = \int\limits_{-\infty}^{\infty} \left[\widehat{F}_n(x) - F(x,\tilde\theta_n)\right]^2 dF(x,\tilde\theta_n). \tag{3}$$

Замечание 2 [80, с. 317]**.** Эти статистики, в отличие от их прототипов D_n и $\omega_n^2(\psi_1)$, не обладают свойством «свободы от распределения» элементов выборки, поэтому для каждого параметрического семейства распределений нужны отдельные таблицы. Более того, их распределения могут зависеть и от истинного значения неизвестного параметра (параметров). К счастью, для семейств сдвига-масштаба (к которым относятся, в частности, показательный и нормальный законы) этого последнего осложнения не возникает.

Несложно проверить (см. задачу 3 гл. 9), что оценкой максимального правдоподобия для параметра θ показательного закона является $\tilde\theta_n = 1/\overline{X}$, где $\overline{X} = (X_1 + \ldots + X_n)/n$, которая ранее встречалась в замечании к примеру 1 гл. 6.

Вопрос 3. Почему в случае показательного закона распределение статистики $\tilde D_n$ не зависит от θ?

М. Стефенс (см. [35]) предложил вместо статистик $\sqrt{n}\,\tilde D_n$ и $n\tilde\omega_n^2$ использовать для *показательной модели* их несколько преобразованные варианты $(\sqrt{n} + 0{,}26 + 0{,}5/\sqrt{n})\,(\tilde D_n - 0{,}2/n)$ и $(n + 0{,}16)\,\tilde\omega_n^2$, распределения которых практически не зависят от n, начиная с $n = 5$. Приведем таблицу соответствующих квантилей $\tilde x_{1-\alpha}$ и $\tilde y_{1-\alpha}$ этих распределений (рассчитанную методом Монте-Карло):

α	0,15	0,1	0,05	0,025	0,01
$\tilde x_{1-\alpha}$	0,926	0,990	1,094	1,190	1,308
$\tilde y_{1-\alpha}$	0,149	0,177	0,224	0,273	0,337

Еще один критерий для проверки показательности (*«Новое лучше старого»*) рассматривается в § 1 гл. 13.

§ 4. ПРОВЕРКА НОРМАЛЬНОСТИ

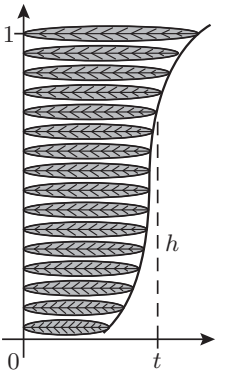


Рис. 7

Л. А. Ж. Кетле (1796–1874), бельгийский социолог.

Б. Л. Ван дер Варден в [13, с. 84] пишет:

> «Я до сих пор живо помню, как однажды, когда я был еще ребенком, мой отец привел меня на край города, где на берегу стояли ивы, и велел мне сорвать наугад сотню ивовых листочков. После отбора листьев с поврежденными кончиками у нас осталось 89 целых листиков. Вернувшись домой, мы расположили их в ряд по росту, как солдат. Затем мой отец через кончики листьев провел кривую и сказал: «Это и есть кривая Кетле. Глядя на нее, ты видишь, что посредственности всегда составляют большинство и лишь немногие поднимаются выше или так и остаются внизу».
>
> Если эту кривую расположить вертикально (рис. 7) и в качестве единицы масштаба на оси ординат выбрать отрезок, длина которого равна высоте всей фигуры, то ордината h, соответствующая абсциссе t, будет, очевидно, представлять собой частоту (или долю) тех ивовых листьев, длина которых меньше t. И так как частота h приближенно равна вероятности, то наша кривая приближенно представляет $p = F(x)$ — функцию распределения длины листьев.»

Как проверить сложную двухпараметрическую *гипотезу нормальности* о том, что выборка была взята из совокупности с функцией распределения $F(x,\mu,\sigma) = \Phi\left(\frac{x-\mu}{\sigma}\right)$ с какими-то неизвестными параметрами μ и $\sigma > 0$? (Здесь, как обычно, $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-u^2/2} du$ — функция распределения стандартного нормального закона.)

Прежде чем применять критерии, полезно посмотреть на данные на вероятностной бумаге (см. § 1 гл. 9): если точки $\left(x_{(i)}, \Phi^{-1}\left(\frac{i-0{,}5}{n}\right)\right)$ не расположены вблизи некоторой прямой, то гипотеза нормальности скорее всего ошибочна.

Чтобы ее формально отвергнуть, можно использовать

Метод 1. Исключение неизвестных параметров

Пусть m — произвольное, но заранее фиксированное целое число от 1 до n. Положим

$$A_m = \frac{1}{n+\sqrt{n}} \sum_{i=1}^{n} X_i + \frac{1}{1+\sqrt{n}} X_m,$$

$$Y_j = \begin{cases} X_j - A_m, & \text{если } j = 1, \ldots, m-1, \\ X_{j+1} - A_m, & \text{если } j = m, \ldots, n-1. \end{cases}$$

Оказывается, случайные величины $Y_1, \ldots, Y_{n-1}$ независимы и одинаково распределены по закону $\mathcal{N}(0, \sigma^2)$ (ввиду нормальности $Y_1, \ldots, Y_{n-1}$ достаточно (см. П9) проверить, что $\mathbf{M}Y_j = 0$, $\mathbf{D}Y_j = \sigma^2$ и $\mathbf{cov}\,(Y_i, Y_j) = 0$ при $i \neq j$).

Переход от выборки $X_1,\ldots,X_n$ к набору случайных величин $Y_1,\ldots,Y_{n-1}$ позволяет избавиться от неизвестного параметра сдвига μ, однако при этом размерность данных уменьшается на 1. (Можно было бы попытаться исключить параметр μ за счет следующего простого преобразования: $X'_i = X_i - \overline{X}$, $i = 1,\ldots,n$. Однако X'_i не будут образовывать выборку.)

Вопрос 4. Зависимы ли а) X'_i и X'_j при $i \neq j$, б) X'_i и $\overline{X}$?

Для исключения оставшегося параметра σ совершим еще одно преобразование:

$$Z_k = Y_k \Big/ \sqrt{B_k}\,, \quad \text{где} \quad B_k = \frac{1}{n-k-1} \sum_{j=k+1}^{n-1} Y_j^2, \quad k = 1,\ldots,n-2.$$

К. Саркади показал (см. [10, с. 57]), что случайные величины $Z_1,\ldots,Z_{n-2}$ также независимы, причем Z_k, очевидно, подчиняется закону Стьюдента t_{n-k-1} (см. § 2 гл. 11).

Обозначим через $F_{n-k-1}(x)$ функцию распределения закона t_{n-k-1}. Если гипотеза нормальности верна, то (согласно методу обратной функции из § 1 гл. 4) случайные величины $F_{n-k-1}(Z_k)$ должны быть независимыми и равномерно распределенными на $[0,1]$. Это проверяется с помощью одного из критериев, рассмотренных в § 2.

Вопрос 5. Применим ли к ним критерий Колмогорова?

Таблицы значений функций F_{n-k-1} для разных степеней свободы приведены в [10, с.174]. Для вычисления ее на компьютере можно численно проинтегрировать плотность, задаваемую формулой (3) гл. 11 (множитель c_{n-k-1} в которой легко определяется из свойства гамма-функции $\Gamma(x+1) = x\Gamma(x)$ и тождества $\Gamma(1/2) = \sqrt{\pi}$), или воспользоваться рекуррентными формулами из [20, с. 22, 51], позволяющими выразить F_{n-k-1} через элементарные функции.

Метод 2. Подстановка оценок параметров

Согласно задаче 2 гл. 9, оценками максимального правдоподобия параметров μ и σ являются соответственно $\overline{X}$ и S, где $\overline{X} = \frac{1}{n}\sum X_i$, $S^2 = \frac{1}{n}\sum (X_i - \overline{X})^2$.

Статистики $\sqrt{n}\,\tilde{D}_n$ и $n\tilde{\omega}_n^2$, определяемые формулой (3) при $F(x,\tilde{\boldsymbol{\theta}}_n) = \Phi((x-\overline{X})/S)$, сходятся при $n \to \infty$ к некоторым предельным законам. М. Стефенс (см. [35]) установил, что для *нормальной модели* распределения (сходящихся к тем же законам) модифицированных статистик $(\sqrt{n} - 0{,}01 + 0{,}85/\sqrt{n})\,\tilde{D}_n$ и $(n+0{,}5)\,\tilde{\omega}_n^2$ практически не зависят от n при $n \geqslant 5$. Приведем таблицу соответствующих квантилей $\tilde{x}_{1-\alpha}$ и $\tilde{y}_{1-\alpha}$ этих распределений:

α	0,15	0,1	0,05	0,025	0,01
$\tilde{x}_{1-\alpha}$	0,775	0,819	0,895	0,955	1,035
$\tilde{y}_{1-\alpha}$	0,091	0,104	0,126	0,148	0,178

Замечание 3. Важно отметить, что предельные распределения статистик $\sqrt{n}\,\tilde{D}_n$ и $n\tilde{\omega}_n^2$ отличаются от $K(x)$ и $A_1(x)$. Дело в том, что при вычислении значения $\tilde{\theta}_n$ используются те же самые $x_1,\ldots,x_n$, что и при построении эмпирической функции распределения. Поэтому $\widehat{F}_n(x)$ и $F(x,\tilde{\theta}_n)$ (в случае, если проверяемая гипотеза верна) оказываются ближе друг к другу, чем $\widehat{F}_n(x)$ и $F(x,\theta_0)$. При этом критическими становятся *существенно меньшие* значения статистик, чем в случае простой гипотезы. Например, сравнение $\tilde{x}_{0,95} = 0{,}895$ с медианой $x_{1/2} = 0{,}83$ и квантилью $x_{0,95} = 1{,}36$ функции распределения $K(x)$ показывает, что значения, типичные для $\sqrt{n}\,D_n$, оказываются *критическими* для статистики $\sqrt{n}\,\tilde{D}_n$.

Метод 3. Центральные выборочные моменты

Простые критерии (см. [13, с. 281]), которые несколько больше, чем критерий Колмогорова, учитывают поведение «хвостов» распределения, основаны на *центральных выборочных моментах* $M_k = \dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i - \overline{X})^k$, $k = 1{,}2,\ldots$.

При помощи величин M_2, M_3 и M_4 вычисляются *выборочные коэффициенты асимметрии* G_1 *и эксцесса* G_2:

$$G_1 = M_3/M_2^{3/2}, \quad G_2 = M_4/M_2^2 - 3.$$

Эти случайные величины можно использовать в качестве оценок для (независящих от сдвига и масштаба) *теоретических коэффициентов асимметрии* $\gamma_1 = \mu_3/\mu_2^{3/2}$ *и эксцесса* $\gamma_2 = \mu_4/\mu_2^2 - 3$, где $\mu_k = \mathbf{M}(X_1 - \mathbf{M}X_1)^k$ — *центральные теоретические моменты.* (Для нормального закона $\gamma_1 = \gamma_2 = 0$.)

При конечных n целесообразно заменить G_1 и G_2 на

$$G_1' = \frac{\sqrt{n(n-1)}}{n-2}\,G_1 \quad \text{и} \quad G_2' = \frac{n-1}{(n-2)(n-3)}\,[(n+1)\,G_2 + 6].$$

Если истинное распределение является *нормальным*, то математические ожидания величин G_1' и G_2' в точности равны нулю, а дисперсии задаются формулами

$$\sigma_1^2 = \frac{6\,n(n-1)}{(n-2)(n+1)(n+3)}, \quad \sigma_2^2 = \frac{24\,n(n-1)^2}{(n-3)(n-2)(n+3)(n+5)}.$$

При этом статистики G_1'/σ_1 и G_2'/σ_2 асимптотически нормальны (см. § 4 гл. 7): распределение каждой из них сходится к закону $\mathcal{N}(0,1)$ при $n \to \infty$. Значимость их отклонения от нуля можно определить по таблице Т2.

Замечание 4 [10, с. 56]**.** Как показал Э. Пирсон (1930 г.), распределение статистики G_1'/σ_1 довольно быстро приближается к $\mathcal{N}(0,1)$, тогда как распределение величины G_2'/σ_2 даже при больших n

оказывается далеким от нормального. Р. Гири (1935 г.) предложил заменить ее на статистику $G_3 = \frac{1}{n}\sum |X_i - \overline{X}|/S$, у которой

$$\mathbf{M}G_3 = \sqrt{\frac{2}{\pi}}\left[1 + \frac{2}{8n-9} + O\left(\frac{1}{n^3}\right)\right],$$
$$\mathbf{D}G_3 = \frac{1}{n}\left[\left(1 - \frac{3}{\pi}\right) - \frac{1}{4\pi n} + O\left(\frac{1}{n^2}\right)\right].$$

Распределение величины $(G_3 - \mathbf{M}G_3)\,/\sqrt{\mathbf{D}G_3}$ удовлетворительно аппроксимируется стандартным нормальным законом при $n \geqslant 50$.

Задача 2 показывает, что, как правило, *нельзя надежно проверить сложную гипотезу по небольшой выборке* (состоящей из нескольких десятков наблюдений): критерии улавливают только очень крупные отклонения, так как за счет варьирования параметра (параметров) обычно удается достаточно хорошо подогнать $F(x,\theta)$ к эмпирической функции распределения $\widehat{F}_n(x)$.

С другой стороны, для выборок большого размера (порядка нескольких сотен наблюдений) трудно гарантировать одинаковость условий при сборе данных (однородность наблюдений).

По-видимому, хороший способ разрешения этой проблемы — использование статистических методов, не предполагающих строгую нормальность наблюдений, для которых требуется лишь непрерывность функции распределения элементов выборки. Именно такие критерии рассматриваются в гл. 14–17.

Проверить нормальность по сгруппированным данным можно также при помощи критерия хи-квадрат (см. гл. 18).

§ 5. ЭНТРОПИЯ

Обозначим через ξ некоторый эксперимент с исходами $A_1, \ldots, A_N$, которые осуществляются с вероятностями $p_1, \ldots, p_N$ соответственно. *Энтропией* этого эксперимента называется величина

$$H = H(\xi) = -\sum_{i=1}^{N} p_i \log_2 p_i, \tag{4}$$

где по непрерывности полагаем $0 \cdot \log_2 0 = 0$. Ясно, что $H \geqslant 0$, причем $H = 0$ тогда и только тогда, когда все вероятности p_i, кроме одной, равны нулю.

Утверждение. Максимум энтропии H, равный $\log_2 N$, достигается при $p_1 = \ldots = p_N = 1/N$.

Доказательство. Поскольку вторая производная функции $\varphi(x) = x \log_2 x$ положительна при $x > 0$, то эта функция выпукла (П4). Записывая неравенство Иенсена (П4) для случайной

величины η, принимающей значения p_i с вероятностями $1/N$, получим неравенство

$$\varphi\left(\frac{1}{N}\sum_{i=1}^{N} p_i\right) = \varphi\left(\frac{1}{N}\right) = -\frac{1}{N}\log_2 N \leqslant \frac{1}{N}\sum_{i=1}^{N}\varphi(p_i) = \frac{1}{N}\sum_{i=1}^{N} p_i \log_2 p_i,$$

которое равносильно доказываемому утверждению. ■

Энтропия может служить количественной характеристикой *меры неопределенности* эксперимента. Если, скажем, $p_1 = 1$, $p_2 = \ldots = p_N = 0$, то с полной уверенностью результатом эксперимента будет осуществление события A_1. Если же $p_1 = \ldots = p_N = 1/N$, то такое распределение обладает максимальной неопределенностью в том смысле, что нельзя отдать предпочтение ни одному из событий A_i.

График значений энтропии $H(\zeta) = -p\log_2 p - (1-p)\log_2(1-p)$ для бернуллиевской случайной величины ζ при разных $0 \leqslant p \leqslant 1$ приведен на рис. 8. Наибольшую неопределенность имеет опыт с равновероятным появлением «успеха» и «неудачи».

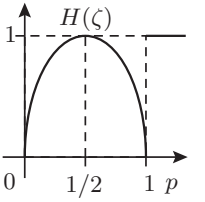


Рис. 8

Ценность понятия энтропии заключается в том, что выражаемая им «степень неопределенности» является именно той характеристикой, которая играет определяющую роль в реальных процессах в природе и технике, связанных с передачей информации.

В конце XIX века психологами было установлено, что среднее время реакции человека на последовательность беспорядочно (равновероятно) чередующихся сигналов N различных типов с увеличением N растет примерно как $H = \log_2 N$. Приведем небольшой отрывок из книги [92] на эту тему.

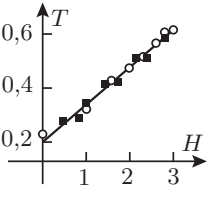


Рис. 9

«На рис. 9, заимствованном из работы американского психолога Р. Хаймана, кружка́ми отмечены данные восьми опытов, состоящих в определении среднего времени, требующегося испытуемому, чтобы указать, какая из N лампочек (где N меняется от 1 до 8) зажглась. Это среднее время определялось из большого числа серий зажиганий, в каждой из которых частоты зажиганий всех лампочек были одинаковыми, причем предварительно испытуемый специально тренировался в подобных опытах. По оси ординат на рис. 9 отложено среднее время реакции (в секундах), по оси абсцисс — величина $\log_2 N$; при этом, как мы видим, все 8 кружков довольно точно укладываются на одну прямую.

Исходя из этих данных, можно было бы предположить, что *среднее время реакции во всех случаях определяется энтропией опыта ξ, состоящего в подаче сигнала.* Из этого предположения следует, что уменьшение степени неопределенности опыта путем замены равновероятных сигналов неравновероятными должно на столько же уменьшить среднее время реакции, на сколько оно уменьшается при уменьшении числа используемых типов сигналов, приводящему к такому же изменению энтропии $H(\xi)$. Это утверждение допускает прямую экспериментальную проверку, полностью его подтверждающую. Так, на том же рис. 9 квадратиками отмечены

результаты восьми опытов (проведенных с теми же испытуемыми, что и раньше), в которых N лампочек (где N равнялось 2, 4, 6 или 8) зажигались с разными относительными частотами $p_1, \ldots, p_N$, причем предварительно испытуемый некоторое время тренировался на сериях зажиганий с такими частотами. Здесь снова по оси ординат откладывалось среднее время реакции T, а по оси абсцисс — энтропия $H(\xi) = -p_1 \log_2 p_1 - \ldots - p_N \log_2 p_N$; при этом оказывается, что квадратики с большой степенью точности укладываются на ту же прямую, что и кружки. Мы видим, таким образом, что энтропия $H(\xi)$ действительно является именно той мерой степени неопределенности исхода опыта, которая решающим образом определяет среднее время, требуемое для определенной реакции на появившийся сигнал.

Причина изменения среднего времени реакции при изменении относительной частоты различных сигналов, очевидно, кроется в том, что испытуемый быстрее реагирует должным образом на более часто повторяющийся (т. е. более привычный для него) сигнал, но зато медленнее реагирует на редкий сигнал, являющийся для него неожиданным. Разумеется, эти факторы носят психологический характер. Тем не менее мы видим, что они могут быть количественно охарактеризованы величиной энтропии $H(\xi)$ опыта ξ».

Понятие энтропии играет также важную роль в статистической механике и в теории кодирования информации. Сам термин был введен Р. Клаузиусом в 1865 г. в качестве меры «хаоса» термодинамической системы (см. [72, с. 138]):

«Проблема обратимости-необратимости — это интересный парадокс классической механики и термодинамики. Суть проблемы заключается в том, что законы классической механики обратимы и поэтому не могут объяснить, почему кусок сахара растворяется в чашке кофе, но мы никогда не наблюдаем обратный процесс. Необратимость нашего мира отражает второй закон термодинамики, впервые сформулированный Л. С. Карно (первый закон термодинамики — это закон сохранения энергии). Спустя сорок лет Р. Клаузиус ввел математическое понятие энтропии, ставшее основным в теории необратимых процессов. (Согласно Клаузиусу слово «энтропия» происходит от греческого $\tau\rho o\pi\eta$, означающего «поворот», «превращение». Клаузиус утверждает, что он добавил «эн», чтобы слово звучало аналогично «энергии».) Используя понятие энтропии, второй закон термодинамики можно сформулировать следующим образом: в изолированной системе энтропия не может уменьшиться, обычно она возрастает. Л. Больцман пытался проверить этот закон с помощью кинематики атомов и молекул. Он показал, что необратимость не противоречит обратимой механике Ньютона: применение последней к большому числу частиц с необходимостью приведет к необратимости, так как системы, состоящие из миллионов молекул, стремятся перейти в состояние, имеющее бóльшую термодинамическую вероятность. Это и есть «основная причина» распада, износа, старения (и, как утверждают некоторые, упадка нравов или цивилизации)».

. . . а понял бы, уединясь,
Вселенной внутреннюю связь,
Постиг все сущее в основе,
И не вдавался в суесловье.

И. В. Гете, «Фауст»

Подробнее о проблеме необратимости можно почитать, скажем, в пятом томе Берклеевского курса физики [67].

Рассмотрим теперь, следуя [69, с. 20], использование понятия энтропии в теории кодирования информации.

При передаче сообщений по каналу связи их необходимо записать в «двоичном коде». Если используется алфавит из N символов (букв), то для кодировки каждого символа потребуется (с точностью до 1) $\log_2 N$ «двоичных» символов 0 и 1. Например, для передачи текста на русском языке, состоящего из букв и пробелов, можно (при объединении «ь» и «ъ») каждый символ закодировать последовательностью из 0 и 1 длины $\log_2 32 = 5$. Для передачи текста из n символов алфавита понадобится код длины $n \log_2 N$.

Рассказывают, что, создавая свой код, Морзе отправился в ближайшую типографию и подсчитал число литер в наборных кассах. Буквам и знакам, для которых литер в этих кассах было припасено больше, он сопоставил более короткие кодовые обозначения

М. Н. Аршинов, А. Е. Садовский, «Коды и математика»

Для больших по объему сообщений можно существенно уменьшить эту величину, используя то, что разные символы алфавита встречаются в тексте с различными частотами (см. таблицу из примера 3 гл. 1). Если $p_1, \ldots, p_N$ — вероятности их появления, то в силу устойчивости частот среди сообщений длины n практически будут встречаться лишь сообщения, в которых каждый i-й символ алфавита будет появляться $\nu_i \approx np_i$ раз. Уточним это утверждение.

Допустим, что каждый символ сообщения появляется независимо от других с соответствующей вероятностью p_i. Для $\delta > 0$ обозначим через $A_{n,\delta}$ множество тех сообщений, у которых $\{|\nu_i - np_i| \leqslant \delta,\ i = 1, \ldots, N\}$. Их станем называть *типичными*, так как в силу закона больших чисел

$$\mathbf{P}\,(A_{n,\delta}) \geqslant 1 - \sum_{i=1}^{N} \mathbf{P}\left(\left|\frac{\nu_i}{n} - p_i\right| > \delta\right) \to 1 \quad \text{при } n \to \infty.$$

Пусть $M_{n,\delta}$ обозначает число «типичных» сообщений. При условии, что все $p_i > 0$ и $0 < \delta < 1$ из *теоремы Макмиллана* (см. [90, с. 64]) следует, что $\dfrac{1}{n} \log_2 M_{n,\delta}$ стремится к энтропии $H = -\sum p_i \log_2 p_i$ при $n \to \infty$. (Обобщение теоремы Макмиллана на стационарные цепи Маркова можно найти в [12, с. 300], см. также задачу 5.) Другими словами, число «типичных» сообщений не превосходит $2^{n(H+\varepsilon)}$, где $\varepsilon > 0$ сколь угодно мало. Каждому такому сообщению можно присвоить порядковый номер, для записи которого потребуется $n(H + \varepsilon)$ «двоичных» символов, и вместо сообщения передавать эту запись. Тем самым, с вероятностью близкой к 1, осуществляется сокращение длины сообщений с *коэффициентом сжатия* $\gamma_1 = H_1/H_0 \leqslant 1$, где $H_0 = \log_2 N$ и $H_1 = H$. Для русского алфавита на основе таблицы из примера 3 гл. 1 имеем $H \approx 4{,}35$, $\gamma_1 \approx 0{,}87$ (см. [92, с. 238]).

Для независимо появляющихся символов *невозможно* предложить способ кодирования (бесконечно большого текста), который давал бы большую экономию, чем γ_1 (см. [92, с. 200]). Однако, символы текста на русском языке, очевидно, зависимы: если оче-

редная буквой является гласной, то следующая вероятнее всего окажется согласной; «ь» не может следовать ни за пробелом, ни за гласной; за буквой «и» после пробела часто следует еще один пробел; после сочетания «тс» естественно ожидать букву «я» и т. п. Эти наблюдения подсказывают разбить текст на блоки длины k и считать эти блоки символами нового алфавита.

Для подсчета частот двухбуквенных и трехбуквенных сочетаний Д. С. Лебедев и В. А. Гармаш использовали отрывок из романа «Война и мир» Л. Н. Толстого, содержащий около 30 000 букв (см. [92, с. 246]). На основе полученных данных были получены оценки соответствующих энтропий: $H_2 \approx 7{,}9$, $H_3 \approx 10{,}9$, что приводит к коэффициентам сжатия $\gamma_2 = H_2/(2H_0) \approx 0{,}79$ и $\gamma_3 = H_3/(3H_0) \approx 0{,}73$. Согласно [92, с. 245] коэффициент сжатия (бесконечно большого) текста не может быть меньше, чем $\gamma_\infty = \lim\limits_{k\to\infty} H_k/(kH_0)$. Лингвист Р. Г. Пиотровский (см. [92, с. 268]) оценил γ_∞ русских литературных текстов как 0,24, а деловых текстов — как 0,17.

К. Шеннон назвал величину $1 - \gamma_\infty$ *избыточностью языка*. Во многих случаях она полезна тем, что позволяет выявлять опечатки и восстанавливать пропуски. (О *кодах Хемминга*, умеющих исправлять подобные ошибки, можно почитать в [91, с. 288] или [92, с. 392].) Последовательность независимых *равновероятных* символов, имеющая энтропию $H = \log_2 N$, несократима. Поскольку сильно сжатый текст похож на нее, практически невозможно восстановить в нем пропущенный или искаженный символ. Это обстоятельство нередко приводит к потере архивированных данных при возникновении дефектов на дискетах.

На практике для кодирования неравновероятно появляющихся символов используют, например, оптимальный *метод Хафмана* (см. [91, с. 276] или [92, с. 206]). Он является самым экономным в следующем смысле: если i-й символ записывается цепочкой из 0 и 1 длины l_i, то код Хафмана имеет наименьшее математическое ожидание $\sum l_i p_i$ длины элементарного кода среди всех кодов, обладающих *свойством префикса*: никакая цепочка не является началом другой, более длинной.

В заключение параграфа приведем утверждение, объединяющее равномерный, показательный и нормальный законы как **распределения с наибольшей энтропией**.

Отыщи всему начало, и ты многое поймешь.

Козьма Прутков

Рассмотрим случайную величину ξ, имеющую плотность распределения $p(x)$ с носителем $A = \{x\colon p(x) > 0\}$. Тогда максимум энтропии $H(\xi) = -\int p(x)\log_2 p(x)\,dx$ при одном из условий

а) $A = (0, 1)$,

б) $A = (0, +\infty)$ и $\mathsf{M}\xi = 1$,

в) $A = (-\infty, +\infty)$, $\mathsf{M}\xi = 0$ и $\mathsf{D}\xi = 1$,

достигается на плотностях $I_{\{0<x<1\}}$, $e^{-x}I_{\{x>0\}}$ и $\dfrac{1}{\sqrt{2\pi}}\,e^{-x^2/2}$ соответственно. Справедливость этого утверждения в случае в) предлагается установить в задаче 6.

Если у тебя есть фонтан, заткни его; дай отдохнуть и фонтану.

Козьма Прутков

ЗАДАЧИ

1. Примените критерий Колмогорова для проверки любого столбца таблицы случайных чисел Т1.

2. Проверьте данные из задачи 1 на показательность. (Если тебе дадут линованную бумагу, пиши поперек. — Хуан Рамон Хименес)

3.* Нарисуйте график плотности предельного закона статистики $\sqrt{n}\,D_n$ для выборки из распределения Бернулли.

4.* Выведите формулу для вычисления $n\omega_n^2(\psi_1)$ из § 2.

5. Установите, что $\lim\limits_{n\to\infty}\dfrac{1}{n}\log_2 M_{n,\delta} = H$ даже при $\delta = 0$.

Указание. Используйте формулу (3) гл. 10 и *формулу Стирлинга*: $\sqrt{2\pi n}\,n^n e^{-n}/n! \to 1$ при $n \to \infty$ (ее доказательство см. в [81, с. 72]).

6.* **а)** Докажите, что для произвольных плотностей $p(x)$ и $q(x)$ из неравенства Иенсена вытекает неравенство

$$\int p(x)\ln p(x)\,dx \geqslant \int p(x)\ln q(x)\,dx. \tag{5}$$

б) С помощью этого неравенства установите справедливость последнего утверждения из § 5 в случае в).

Недостойно многократно опускать сосуд в пустой колодец. Пахарь не понесет зерна на голую скалу. Согласившийся легко примет выгоды, но первым препятствием устрашится. Поэтому испытывайте препятствиями.

Агни-Йога, 264

Если тебе дадут линованную бумагу, пиши поперек.

Хуан Рамон Хименес

РЕШЕНИЯ ЗАДАЧ

1. Выберем для проверки пятый столбец таблицы Т1: его максимум равен всего лишь 74, кроме того, 6 из 20 чисел (30% данных) попали в интервал от 29 до 35 (6% диапазона).

Эмпирическая функция этого столбца (рис. 10, *а*) располагается целиком выше диагонали единичного квадрата и в точке 0,35 отклоняется от диагонали на величину $D_n = 0{,}35$. Значение статистики $\sqrt{n}\,D_n$ равно 1,57, что значимо на уровне $\alpha_0 = 1{,}5\%$. Еще к более категоричным выводам приводит критерий омега-квадрат: $n\omega_n^2(\psi_1) = 0{,}85$ ($\alpha_0 < 0{,}01$) и $n\omega_n^2(\psi_2) = 5{,}1$ ($\alpha_0 < 0{,}01$) (при замене наблюдения 0 на 0,01).

Однако для четвертого столбца картина совершенно иная. Отклонение эмпирической функции распределения от диагонали $D_n = 0{,}17$ (рис. 10, *б*). Отсюда получаем, что $\sqrt{n}\,D_n = 0{,}76$. Эта величина принадлежит области «типичных» значений распределения Колмогорова: $\alpha_0 = 61\%$. Хорошее согласие с гипотезой равномерности подтверждается также критерием омега-

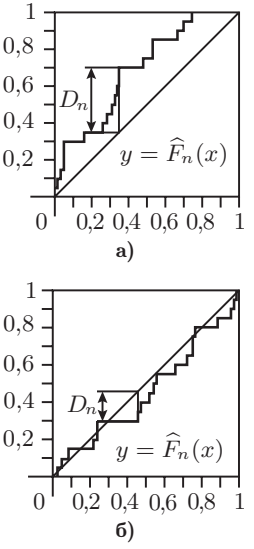
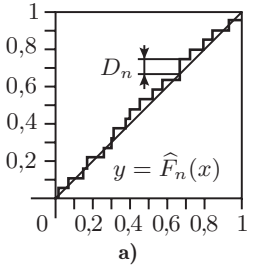


Рис. 10

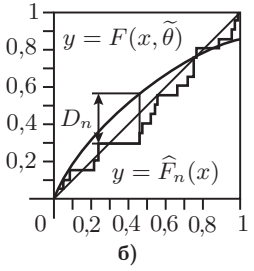
квадрат: $n\omega_n^2(\psi_1) = 0{,}103$ ($\alpha_0 = 57\%$) и $n\omega_n^2(\psi_2) = 0{,}723$ ($\alpha_0 = 54\%$).

Так можно ли пользоваться таблицей T1? С помощью компьютера была проверена вся T1 ($n = 300$). Результаты следующие: $\sqrt{n}\,D_n = 0{,}981$ ($\alpha_0 = 29\%$), $n\omega_n^2(\psi_1) = 0{,}196$ ($\alpha_0 = 37\%$) и $n\omega_n^2(\psi_2) = 1{,}98$ ($\alpha_0 = 9{,}4\%$).

Кроме того, тест на случайность, основанный на количестве инверсий R_n в выборке (см. пример 2 гл. 7) дал в качестве значения статистики $(R_n - \mathbf{M}R_n)/\sqrt{\mathbf{D}R_n}$ величину $-1{,}16$. По таблице T2 находим, что фактический уровень значимости $\alpha_0 = 12{,}3\%$. Таким образом, в целом T1 пригодна для имитации выбора наудачу из $[0, 1]$.



а)



б)

Рис. 11

2. Применим метод исключения неизвестного параметра к данным из четвертого столбца. На рис. 11, *а* по $Y_k = S_k/S_{20}$ ($k = 1, \ldots, 19$), где $S_k = X_1 + \ldots + X_k$, построена эмпирическая функция распределения. Близость к диагонали исключительная! Статистика $\sqrt{n}\,D_n = 0{,}343$. Согласно [10, с. 346] функция Колмогорова в этой точке равна 0,002. Значение статистики попало далеко в область левого «хвоста».

В чем же дело? Может быть, мала выборка? Применим (с помощью компьютера) критерий ко всей таблице T1. Получим $\sqrt{n}\,D_n = 0{,}656$ ($\alpha_0 = 79\%$). Неужели T1 можно использовать в качестве таблицы показательных случайных чисел? Это, конечно же, не так, хотя бы потому, что все $X_i \leqslant 1$.

То, в чем нет загадочности, лишено очарования.

А. Франс

Объяснение. Рассмотрим поведение случайной величины Y_k при $k = [\alpha n]$, где $0 < \alpha < 1$, $[\cdot]$ обозначает целую часть числа. Для любых $X_i > 0$ с $\mathbf{M}X_i < \infty$ в силу свойств сходимости (П5) и закона больших чисел (П6) $Y_k = S_k/S_n = ([\alpha n]/n) \cdot (S_k/k) \cdot (n/S_n) \xrightarrow{\mathsf{P}} \alpha$ при $n \to \infty$. Используя монотонность $\widehat{F}_n(x)$, отсюда так же, как при доказательстве теоремы Гливенко (см. [19, с. 207]), можно вывести, что $D_n \xrightarrow{\mathsf{P}} 0$. Если $\mathbf{D}X_1 < \infty$, то в силу центральной предельной теоремы (П6) и леммы 1 гл. 7 ($\varphi(x) = 1/x$) отклонение $|Y_k - \alpha|$ имеет порядок малости $1/\sqrt{n}$. Поэтому величина $\sqrt{n}\,D_n \approx \text{const}$ и вполне может попасть в область «типичных» значений закона Колмогорова (более аккуратное объяснение приведено в § 3 гл. 26).

Главная цель расчетов — не цифры, а понимание.

Р. В. Хэмминг

Кроме опасности отвергнуть верную гипотезу H из-за случайности (обычно статистика критерия принимает с малой вероятностью любые значения), существует опасность подтвердить H в том случае, когда она ошибочна (подробнее эта проблема обсуждается в следующей главе).

Метод подстановки оценки параметра с поправкой Стефенса для данных из четвертого столбца T1 приводит к $\tilde{\theta} = 1{,}818$,

$\tilde{D}_n = 0{,}28$ и $(\sqrt{n}+0{,}26+0{,}5/\sqrt{n})\,(\tilde{D}_n - 0{,}2/n) = 1{,}308$ (рис. 11, *б*). Это значимо на уровне $\alpha_0 = 1\%$. *)

Не все то волк, что серо.

3. Функция распределения $F_\zeta(x)$ бернуллиевской случайной величины ζ (см. § 1 гл. 1) имеет два скачка: в 0 высоты $q = 1-p$ и в 1 высоты p. Эмпирическая функция выборки $X_1, \ldots, X_n$ из этого закона отличается от нее лишь величиной скачков (рис. 12). Поэтому $D_n = \left|\overline{X} - p\right|$.

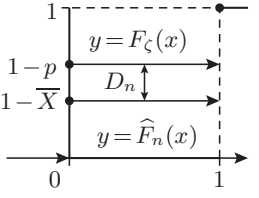


Рис. 12

В силу центральной предельной теоремы при $n \to \infty$ имеет место сходимость $\sqrt{n}\,(\overline{X} - p) \xrightarrow{d} \xi \sim \mathcal{N}(0, pq)$. По свойству сходимости 3 из П5 $\sqrt{n}\,D_n \xrightarrow{d} |\xi|$. Так как случайная величина $|\xi|$ положительна с вероятностью 1, то ее функция распределения и плотность при $x < 0$ равны нулю. При $x \geqslant 0$ в силу симметрии закона $\mathcal{N}(0, pq)$ имеем

$$F_{|\xi|}(x) = \mathbf{P}(-x \leqslant \xi \leqslant x) = F_\xi(x) - F_\xi(-x) = 2F_\xi(x) - 1.$$

Дифференцируя, находим, что при $x \geqslant 0$ плотность распределения величины $|\xi|$ равна удвоенной плотности ξ (рис. 13). (Отметим, что плотность предельного закона зависит от p.)

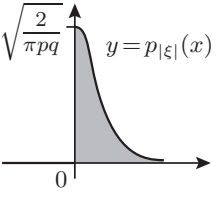


Рис. 13

4. Сделав замену $y = F(x)$, видим, что

$$\omega_n^2(\psi_1) = I_n = \int\limits_0^1 \left(\widehat{F}_n(y) - y\right)^2 dy,$$

где $\widehat{F}_n(y)$ — эмпирическая функция наблюдений $Y_i = F(X_i)$ $(i = 1, \ldots, n)$ из равномерного распределения на отрезке $[0, 1]$ (см. замечание 1). Упорядочим величины Y_i по возрастанию:

$Y_{(1)} \leqslant Y_{(2)} \leqslant \ldots \leqslant Y_{(n)}$.

Положим $Y_{(0)} = 0$, $Y_{(n+1)} = 1$. Разбивая отрезок $[0, 1]$ на части $[Y_{(i)}, Y_{(i+1)}]$, $i = 0, \ldots, n$, представим интеграл I_n в виде

$$I_n = \sum_{i=0}^{n} \int\limits_{Y_{(i)}}^{Y_{(i+1)}} \left(y - \frac{i}{n}\right)^2 dy = \frac{1}{3} \sum_{i=0}^{n-1} \left(Y_{(i+1)} - \frac{i}{n}\right)^3 - \frac{1}{3} \sum_{i=1}^{n} \left(Y_{(i)} - \frac{i}{n}\right)^3.$$

Поменяв индекс суммирования в первой сумме, запишем

$$I_n = \frac{1}{3} \sum_{i=1}^{n} \left[\left(Y_{(i)} - \frac{i-1}{n}\right)^3 - \left(Y_{(i)} - \frac{i}{n}\right)^3 \right].$$

Воспользовавшись тождеством $(a+b)^3 - (a-b)^3 = 2b\,(3a^2 + b^2)$ для $a = Y_{(i)} - \dfrac{2i-1}{2n}$ и $b = \dfrac{1}{2n}$, окончательно получим

$$I_n = \frac{1}{3n} \sum_{i=1}^{n} \left[3\left(Y_{(i)} - \frac{2i-1}{2n}\right)^2 + \frac{1}{4n^2} \right] = \frac{1}{12n^2} + \frac{1}{n} \sum_{i=1}^{n} \left(Y_{(i)} - \frac{2i-1}{2n}\right)^2.$$

*) В примере 2 гл. 13 гипотеза показательности для этих же данных будет отвергнута с помощью еще одного критерия.

5. Согласно формуле (3) гл. 10 $M_{n,0} = \dfrac{n!}{l_1!l_2!\ldots l_N!}$, где $l_i = np_i$. Из формулы Стирлинга имеем $\ln n! = n\ln n - n + o(n)$. Отсюда

$$\frac{1}{n}\ln M_{n,0} = \ln n - 1 - \sum_{i=1}^{N}[p_i \ln np_i - p_i] + o(1) =$$

$$= -\sum_{i=1}^{N} p_i \ln p_i + o(1).$$

Для примера рассмотрим симметричную схему Бернулли ($N = 2$, $p_1 = p_2 = 1/2$). Тогда $M_{2n,0} = C_{2n}^n = (2n)!/(n!)^2$ — количество последовательностей длины $2n$, у которых число нулей совпадает с числом единиц. С помощью формулы Стирлинга легко вывести (проверьте!), что $\mathsf{P}(A_{2n,0}) = 2^{-2n}M_{2n,0} \sim 1/\sqrt{\pi n} \to 0$ при $n \to \infty$, в то время, как $\mathsf{P}(A_{2n,\delta}) \to 1$ при любом $\delta > 0$. Поэтому сообщения из $A_{2n,0}$ составляют пренебрежимо малую часть множества «типичных» сообщений $A_{2n,\delta}$, несмотря на то, что $\lim\limits_{n\to\infty}\dfrac{1}{2n}\log_2 M_{2n,0} = \lim\limits_{n\to\infty}\dfrac{1}{2n}\log_2 M_{2n,\delta} = H = 1$.

6. **а)** Неравенство (5) равносильно утверждению

$$\int_{-\infty}^{\infty} p(x)\ln\frac{q(x)}{p(x)}\,dx = \mathsf{M}\ln\frac{q(\xi)}{p(\xi)} \leqslant 0,$$

где случайная величина ξ имеет плотность $p(x)$. Так как функция $\ln x$ выпукла на $(0,\infty)$, то с учетом неравенства Иенсена (П4) запишем

$$\mathsf{M}\ln\frac{q(\xi)}{p(\xi)} \leqslant \ln\mathsf{M}\,\frac{q(\xi)}{p(\xi)} = \ln\int_{-\infty}^{\infty}\frac{q(x)}{p(x)}\,p(x)\,dx = \ln 1 = 0.$$

б) Возьмем в неравенстве (5) в качестве $q(x)$ плотность закона $\mathcal{N}(0,1)$ и используем то, что $\mathsf{M}\xi^2 = \mathsf{D}\xi + (\mathsf{M}\xi)^2 = 1$:

$$-\int_{-\infty}^{\infty} p(x)\ln p(x)\,dx \leqslant -\int_{-\infty}^{\infty} p(x)\left\{-\ln\sqrt{2\pi} - \frac{x^2}{2}\right\}dx = \ln\sqrt{2\pi e},$$

причем верхняя граница достигается при $p(x) = q(x)$.

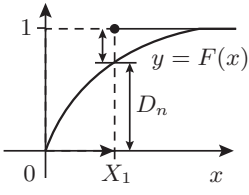


Рис. 14

ОТВЕТЫ НА ВОПРОСЫ

1. По центральной предельной теореме (П6) при $n = 100$

$$\mathsf{P}(S_n \geqslant 60) = \mathsf{P}\left(\frac{S_n - n/2}{\sqrt{n/4}} \geqslant \frac{60 - n/2}{\sqrt{n/4}}\right) \approx 1 - \Phi(2) \approx 0{,}023.$$

2. Например, может выглядеть, как на рис. 14 ($n = 1$).

3. Рассмотрим случайные величины $Y_i = \theta X_i$ с функцией распределения $(1 - e^{-y})\, I_{\{y \geqslant 0\}}$, не зависящей от θ. Выразим через них $\widehat{F}_n(x)$ и $F(x,\tilde{\theta})$:

$$\widehat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} I_{\{X_i \leqslant x\}} = \frac{1}{n} \sum_{i=1}^{n} I_{\{Y_i \leqslant \theta x\}},$$

$$F(x,\tilde{\theta}) = \left(1 - e^{-x/\overline{X}}\right) I_{\{x \geqslant 0\}} = \left(1 - e^{-\theta x/\overline{Y}}\right) I_{\{\theta x \geqslant 0\}}.$$

Когда переменная x пробегает значения от $-\infty$ до $+\infty$, этот же интервал значений пробегает и переменная $y = \theta x$. Поэтому

$$\sup_x \left|\widehat{F}_n(x) - F(x,\tilde{\theta})\right| = \sup_y \left|\frac{1}{n} \sum_{i=1}^{n} I_{\{Y_i \leqslant y\}} - \left(1 - e^{-y/\overline{Y}}\right) I_{\{y \geqslant 0\}}\right|.$$

Распределение правой части равенства от θ не зависит.

4. Так как вектор $(X'_1, \ldots, X'_n)$ получается линейным преобразованием из нормальной выборки $(X_1, \ldots, X_n)$, то он также является нормальным (П9). Некоррелированность компонент такого вектора эквивалентна их независимости (см. [90, с. 322]). Положим $Y_i = X_i - \mu$. Тогда

$$\begin{aligned}\mathbf{cov}\left(X'_i, X'_j\right) &= \mathbf{M}(Y_i - \overline{Y})\,(Y_j - \overline{Y}) = \\ &= -\mathbf{M}(Y_i\,\overline{Y}) - \mathbf{M}(Y_j\,\overline{Y}) + \mathbf{D}\overline{Y}.\end{aligned}$$

Поскольку $\mathbf{M}(Y_i\,\overline{Y}) = \dfrac{1}{n}\,\mathbf{D}Y_i = \dfrac{1}{n}\,\sigma^2$ и $\mathbf{D}\overline{Y} = \dfrac{1}{n}\,\sigma^2$, находим, что $\mathbf{cov}\left(X'_i, X'_j\right) = -\dfrac{1}{n}\,\sigma^2 < 0$. Следовательно, величины X'_i и X'_j зависимы, т. е. набор $X'_1, \ldots, X'_n$ не является выборкой. Более того, все X'_i связаны линейной зависимостью: $X'_1 + \ldots + X'_n = 0$.

В свою очередь,

$$\mathbf{cov}\left(X'_i, \overline{X}\right) = \mathbf{M}\left[(Y_i - \overline{Y})\,\overline{Y}\right] = \mathbf{M}\left(Y_i\,\overline{Y}\right) - \mathbf{D}\overline{Y} = 0.$$

Поэтому $\overline{X}$ не зависит от X'_i, $i = 1, \ldots, n$. Согласно лемме о независимости из § 3 гл. 1, видим, что $\overline{X}$ и выборочная дисперсия $S^2 = \dfrac{1}{n}\sum (X_i - \overline{X})^2$ независимы между собой (см. теорему 1 гл. 11).

5. Нет, так как случайные величины $Z_k \sim t_{n-k-1}$, $k = 1, \ldots, n-2$, имеют разные распределения (меняется число степеней свободы).

Глава 13

АЛЬТЕРНАТИВЫ

Самый отдаленный пункт земного шара к чему-нибудь да близок, а самый близкий от чего-нибудь да отдален.

Козьма Прутков

§ 1. ОШИБКИ I И II РОДА

Пример 1. Рассмотрим модель $X_i \sim \mathcal{N}(\theta, \sigma^2)$, где дисперсия σ^2 известна, а математическое ожидание θ — нет (см. пример 2 гл. 11). Для проверки гипотезы $H_0\colon \theta = \theta_0$ можно применить критерий, основанный на статистике $T(X_1, \ldots, X_n) = \overline{X}$. Если H_0 верна, то $\overline{X} \sim \mathcal{N}(\theta_0, \sigma^2/n)$. Найдем критическое значение t_α из условия $\alpha = \mathsf{P}_{\theta_0}\left(\overline{X} \geqslant t_\alpha\right)$:

$$\alpha = \mathsf{P}_{\theta_0}\left(\frac{\sqrt{n}\,(\overline{X} - \theta_0)}{\sigma} \geqslant \frac{\sqrt{n}\,(t_\alpha - \theta_0)}{\sigma}\right) = 1 - \Phi\left(\frac{\sqrt{n}\,(t_\alpha - \theta_0)}{\sigma}\right),$$

где $\Phi(x)$ — функция распределения закона $\mathcal{N}(0,1)$. Обозначив p-квантиль этого закона (т. е. решение уравнения $\Phi(x) = p$) через x_p, получаем

$$t_\alpha = \theta_0 + \sigma x_{1-\alpha}/\sqrt{n}. \tag{1}$$

Если значение выборочного среднего $\overline{x} \geqslant t_\alpha$, то гипотеза H_0 отвергается. Понятно, что если она верна, то неравенство $\overline{X} \geqslant t_\alpha$ выполняется с вероятностью α. Отвергая в этом случае верную гипотезу H_0, мы совершаем так называемую *ошибку I рода*.

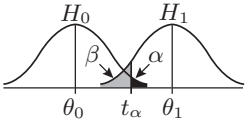


Рис. 1

С другой стороны, может оказаться, что на самом деле верна не гипотеза H_0, а ее *альтернатива* $H_1\colon \theta = \theta_1$, где, скажем, $\theta_1 > \theta_0$. Если при этом случится, что $\overline{x} < t_\alpha$, то мы примем ошибочную гипотезу H_0 вместо H_1. Тем самым мы допустим *ошибку II рода*.

Полезно подчеркнуть, что номинация ошибки как I или II рода зависит от того, какая из возможностей принимается за гипотезу, а какая — за альтернативу. Если поменять их местами, то изменятся и названия ошибок.

Найдем вероятность β ошибки II рода для рассматриваемой модели. Когда верна альтернатива, выборочное среднее $\overline{X}$ распределено по закону $\mathcal{N}(\theta_1, \sigma^2/n)$. Поэтому из равенства (1) имеем

$$\beta = \mathsf{P}_{\theta_1}\left(\overline{X} < t_\alpha\right) = \Phi\left(\frac{\sqrt{n}\,(t_\alpha - \theta_1)}{\sigma}\right) = \Phi\left(x_{1-\alpha} - \frac{\sqrt{n}\,(\theta_1 - \theta_0)}{\sigma}\right). \tag{2}$$

Наглядный смысл вероятностей α и β показывает рис. 1, где приведены графики плотностей распределения среднего $\overline{X}$ при гипотезах H_0 и H_1.

В общем случае для *параметрической* статистической модели $\{F(x,\theta), \theta \in \Theta\}$ в множестве Θ выделяются два непересекающихся

подмножества Θ_0 и Θ_1. Предполагается, что компоненты выборки $\boldsymbol{X} = (X_1, \ldots, X_n)$ имеют функцию распределения $F(x,\theta)$, где θ принадлежит одному из этих подмножеств. Гипотеза H_0 заключается в том, что $\theta \in \Theta_0$, а альтернатива H_1 — в том, что $\theta \in \Theta_1$. Когда множество Θ_0 (Θ_1) состоит из единственной точки, гипотеза H_0 (альтернатива H_1) называется *простой*, иначе — *сложной*.

Чтобы задать критерий уровня α, укажем в $\mathbb{R}^n$ *критическое множество* G_α такое, что $\mathbf{P}_\theta(\boldsymbol{X} \in G_\alpha) \leqslant \alpha$ при всех $\theta \in \Theta_0$.*) Вероятность ошибки II рода такого критерия $\beta = \beta(\theta) = \mathbf{P}_\theta(\boldsymbol{X} \notin G_\alpha)$ при $\theta \in \Theta_1$. *Функцией мощности* критерия называется

$$W(\theta) = 1 - \beta(\theta) = \mathbf{P}_\theta(\boldsymbol{X} \in G_\alpha), \quad \theta \in \Theta_1.$$

Так, если критерий из примера 1 использовать для проверки простой гипотезы $H_0\colon \theta = \theta_0$ против сложной альтернативы $H_1\colon \theta > \theta_0$, то из формулы (2) с учетом симметрии функции распределения $\Phi(x)$ находим (рис. 2)

$$W(\theta) = 1 - \Phi\left(x_{1-\alpha} - \frac{\sqrt{n}\,(\theta - \theta_0)}{\sigma}\right) = \Phi\left(\frac{\sqrt{n}\,(\theta - \theta_0)}{\sigma} - x_{1-\alpha}\right).$$

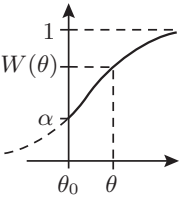


Рис. 2

Когда $W(\theta) \geqslant \alpha$ при всех $\theta \in \Theta_1$, критерий называется *несмещенным*. Несмещенность означает, что попадание в критическое множество G_α (и, следовательно, — отвержение гипотезы H_0) при справедливости любой из альтернатив не менее вероятно, чем попадание в него при выполнении H_0, т. е. правильное отвержение имеет не меньшую вероятность, чем неправильное.

Вопрос 1. Будет ли критерий примера 1 а) несмещенным, б) состоятельным против сложной альтернативы $H_1\colon \theta > \theta_0$?

Если для любого $\theta \in \Theta_1$ функция мощности $W(\theta) \to 1$ при $n \to \infty$, то критерий называется *состоятельным*.

При *непараметрическом* подходе в множестве всех функций распределения выделяются два непересекающихся класса: класс гипотез и класс альтернатив.

Следующий пример содержит критерий для проверки сложной гипотезы показательности распределения элементов выборки, настроенный против непараметрических альтернатив определенного вида.

Пример 2. *«Новое лучше старого» (Холлендер—Прошан),* см. [88, с. 260]**.** Допустим, что времена X_i работы прибора до i-го отказа ($i = 1, \ldots, n$) образуют выборку из некоторого непрерывного закона. Нас интересует гипотеза

$$H_0\colon \mathbf{P}(X_1 \geqslant x + y \,|\, X_1 \geqslant x) = \mathbf{P}(X_1 \geqslant y) \quad \text{для всех } x, y \geqslant 0. \tag{3}$$

(Здесь $\mathbf{P}(A \,|\, B)$ — условная вероятность события A при условии события B (см. П7).) Гипотеза (3) означает, что вероятность отсутствия поломок за дополнительный период времени y при условии,

*) То есть формально статистикой критерия (см. § 1 гл. 12) является индикатор множества G_α.

что прибор уже проработал в течение периода времени x, равна вероятности того, что новый (еще не работавший) прибор прослужит начальный период времени y. Утверждение о том, что это верно для всех $x,y \geqslant 0$ эквивалентно утверждению о том, что работающие приборы любого «возраста» не лучше и не хуже, чем новые.

Гипотеза H_0 равносильна сложной параметрической гипотезе показательности: $\mathbf{P}(X_1 \geqslant x) = e^{-\theta x}$ при некотором $\theta > 0$. Легко убедиться, что показательности достаточно для выполнения гипотезы H_0: при любых $x,y \geqslant 0$

$$\mathbf{P}(X_1 \geqslant x+y \,|\, X_1 \geqslant x) = e^{-\theta\,(x+y)} \big/ e^{-\theta x} = e^{-\theta y} = \mathbf{P}(X_1 \geqslant y).$$

Необходимость доказать сложнее (см. [81, с. 475]).

В качестве альтернатив рассмотрим **два класса законов**.

а) *«Новое лучше старого»*:

$$H_1\colon \mathbf{P}(X_1 \geqslant x+y \,|\, X_1 \geqslant x) \leqslant \mathbf{P}(X_1 \geqslant y) \quad \text{для всех } x,y \geqslant 0, \tag{4}$$

причем хотя бы для некоторых $x,y \geqslant 0$ неравенство (4) строгое.

б) *«Новое хуже старого»*:

$$H_2\colon \mathbf{P}(X_1 \geqslant x+y \,|\, X_1 \geqslant x) \geqslant \mathbf{P}(X_1 \geqslant y) \quad \text{для всех } x,y \geqslant 0, \tag{5}$$

причем хотя бы для некоторых $x,y \geqslant 0$ неравенство (5) строгое.

Ф. Прошан (см. [88, с. 262]) пишет: «Тенденция к удлинению интервалов, если она выявлена, может быть результатом опыта эксплуатации, доводки или смены поврежденных частей, а тенденция к их сокращению — напротив, может быть результатом износа, старения или плохого технического обслуживания».

Упорядочим величины X_i по возрастанию:

$$X_{(1)} \leqslant X_{(2)} \leqslant \ldots \leqslant X_{(n)}.$$

Статистикой критерия Холлендера—Прошана является величина

$$T_n = \sum_{i>j>k} \psi\left(X_{(i)}, X_{(j)} + X_{(k)}\right), \quad \text{где } \psi(a,b) = \begin{cases} 1, & \text{если } a > b, \\ 1/2, & \text{если } a = b, \\ 0, & \text{если } a < b. \end{cases}$$

(Суммирование здесь производится по всем $n(n-1)(n-2)/6$ упорядоченным тройкам (i,j,k), для которых $i > j > k$.)

Поясним, почему T_n можно использовать в качестве статистики критерия для проверки гипотезы H_0 против альтернативы H_1 (или H_2). Положим

$$T'(x,y) = \mathbf{P}(X_1 \geqslant x)\,\mathbf{P}(X_1 \geqslant y) - \mathbf{P}(X_1 \geqslant x+y).$$

Заметим, что $T'(x,y) = 0$ для всех $x,y \geqslant 0$ тогда и только тогда, когда гипотеза H_0 (3) верна. Оказывается, (линейно связанная с T_n) статистика $T^* = 1/4 - 2\,T_n/[n(n-1)(n-2)]$ служит оценкой для параметра $\Delta(F) = \mathbf{M}\,T'(X',Y')$, где X' и Y' независимы и имеют распределение (времени безотказной работы) F. Мы можем рассматривать $T'(x,y)$ как меру отклонения от H_0 в точке (x,y), а $\Delta(F)$ — как среднее значение этого отклонения.

Когда распределение F соответствует H_1 (4) и непрерывно, параметр $\Delta(F)$ положителен. Если выборка берется из такой совокупности, величина T^* возрастает (что эквивалентно убыванию T_n).

В [88, с. 432] приведена таблица критических значений T_n для $n \leqslant 50$. Для достаточно большой выборки можно воспользоваться нормальным приближением: $(T_n - \mathbf{M}T_n)/\sqrt{\mathbf{D}T_n} \xrightarrow{d} \xi \sim \mathcal{N}(0,1)$, где

$$\mathbf{M}T_n = n(n-1)(n-2)/8,$$

$$\mathbf{D}T_n = \frac{3}{2}\, n(n-1)(n-2)\left[\frac{5}{2592}\,(n-3)(n-4) + \frac{7}{432}\,(n-3) + \frac{1}{48}\right].$$

Известно (см. [88, с. 264]), что критерий состоятелен против H_1, H_2.

Применим его для проверки показательности четвертого столбца таблицы Т1 равномерных на $[0,1]$ случайных чисел (см. задачу 2 гл. 12). Для равномерного закона справедлива альтернатива H_1 (проверьте!), поэтому есть надежда отвергнуть H_0.

Прежде, чем использовать критерий «Новое лучше старого», рекомендуется убедиться, что величины X_i образуют случайную выборку из общей совокупности, т. е. проверить их *независимость и одинаковую распределенность*. Одним из способов проверки этого является критерий, основанный на асимптотической нормальности *числа инверсий* R_n в выборке (см. пример 2 гл. 7). Значение статистики $(R_n - \mathbf{M}R_n)/\sqrt{\mathbf{D}R_n}$ для четвертого столбца Т1 равно $-0{,}389$. Ему соответствует фактический уровень значимости (см. § 1 гл. 12) $\alpha_0 = 0{,}697$. Следовательно, гипотеза случайности не отвергается.

Вычисленная на компьютере статистика Холлендера—Прошана T_n приняла значение $-3{,}16$, что значимо мало на уровне $\alpha = 0{,}002$.

§ 2. ОПТИМАЛЬНЫЙ КРИТЕРИЙ НЕЙМАНА—ПИРСОНА

При сравнении двух критериев уровня α, заданных при помощи критических множеств G'_α и G''_α, лучшим будет тот, у которого мощность больше. Если альтернатива сложная, то (как и при сравнении точности оценок (см. § 3 гл. 6)) возникает проблема сравнения двух функций $W'(\theta)$ и $W''(\theta)$ (рис. 3). В случае проверки простой гипотезы H_0 против простой альтернативы H_1 ситуация проще: существует наиболее мощный критерий и можно явно указать его критическое множество G^*_α.

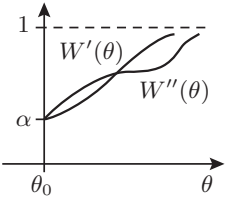


Рис. 3

Прежде чем строго сформулировать и доказать этот результат*), обсудим подробнее проблему выбора критического множества.

Для модели из примера 1 возьмем внутри диапазона «типичных» значений статистики $\overline{X}$ при справедливости гипотезы H_0 маленький отрезок Δ такой, что $\mathbf{P}_{\theta_0}(\overline{X} \in \Delta) = \alpha$ (рис. 4).

Вопрос 2.
Чем плох критерий, задаваемый соответствующим критическим множеством $\{\boldsymbol{x} \in \mathbb{R}^n : \overline{x} \in \Delta\}$?

В свою очередь, если для критерия с критическим множеством $\{\boldsymbol{x} \in \mathbb{R}^n : \overline{x} \geqslant t_\alpha\}$ сдвигать границу t_α вправо для уменьшения

*) Впервые он был получен Ю. Нейманом и Э. Пирсоном в 1933 г.

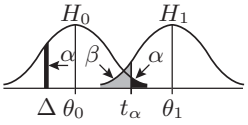


Рис. 4

величины α ошибки I рода, то *дополнение к критическому множеству* $\{\boldsymbol{x} \in \mathbb{R}^n\colon \overline{x} < t_\alpha\}$ будет увеличиваться, и соответствующая величина ошибки II рода β будет возрастать.

Таким образом, не удается добиться того, чтобы α и β были обе сколь угодно малы при фиксированном размере выборки n (*Тришкин кафтан*).

Ничего не доводи до крайности: человек, желающий трапезовать слишком поздно, рискует трапезовать на другой день поутру.

Козьма Прутков

Проблема напоминает *принцип неопределенностей* в квантовой физике, сформулированный в 1927 г. Гейзенбергом: невозможно одновременно сколь угодно точно определить положение и скорость элементарной частицы (см. [14, с. 28]). Для погрешностей измерения координаты Δx и импульса Δp выполняется *соотношение неопределенностей* $\Delta x \cdot \Delta p \geqslant h$, где h обозначает *постоянную Планка* ($h \approx 6{,}626 \cdot 10^{-34}$ Дж $\cdot$ с).

В. Гейзенберг (1901–1976), немецкий физик.

М. Планк (1858–1947), немецкий физик.

Байка. Студент на экзамене по физике, записав формулу для величины кванта энергии света $E = h\nu$, сообщил, что ν — это постоянная Планка. На вопрос «Что же тогда обозначает здесь h?» был немедленно дан ответ: «Высоту этой планки».

Если за ошибку I рода приходится платить цену C_α, а за ошибку II рода — цену C_β, то критическое множество можно постараться выбрать так, чтобы минимизировать «взвешенные» *общие затраты* $\alpha\, C_\alpha + \beta\, C_\beta$ (см. задачу 1).

Не всегда реальное значение ошибок сводится к величине общих затрат. Например, в случае проверки на основе результатов медицинских анализов гипотезы H_0, состоящей в том, что пациент болен, против альтернативы H_1, что он здоров, ошибка I рода приведет к тому, что не будет оказана врачебная помощь больному человеку, а ошибка II рода — к тому, что станут лечить здорового.

В этой ситуации более верным представляется следующий подход: при заданной (достаточно малой) вероятности ошибки I рода α постараться уменьшить вероятность ошибки II рода β насколько возможно за счет подбора критического множества.

Для выборки $\boldsymbol{X} = (X_1, \ldots, X_n)$ и множества $G \in \mathbb{R}^n$ положим $\mathsf{P}_k(G) = \mathsf{P}_{\theta_k}(\boldsymbol{X} \in G)$ для $k = 0$ и 1. Тем самым, гипотеза H_0 и альтернатива H_1 порождают в $\mathbb{R}^n$ меры P_0 и P_1. В этих обозначениях задача сводится к нахождению множества G^* такого, что $\mathsf{P}_0(G^*) \leqslant \alpha$, а $\mathsf{P}_1(G^*)$ была бы как можно больше.

Боюсь,... чтоб множество не накоплялось...

Фамусов в «Горе от ума» А. С. Грибоедова

Пример 3 [38, с. 216]**.** Рассмотрим дискретную модель, в которой при гипотезе H_0 мера P_0 равномерно распределена на конечном множестве из M точек $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M\}$ в $\mathbb{R}^n$, а при альтернативе H_1 j-й точке приписана вероятностная масса $\mathsf{P}_1(\boldsymbol{x}_j) = p_j$, $j = 1, \ldots, M$. Тогда для $\alpha = k/N$ любое подмножество из k точек задает критерий уровня α (рис. 5). Чтобы максимизировать мощность (вероятностный вес) этого подмножества при альтернативе H_1, очевидно, надо упорядочить величины p_j по убыванию и набрать в критическое множество k точек с наибольшими p_j.

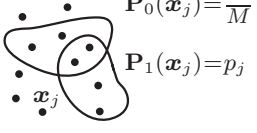


Рис. 5

Немного усложним модель, предположив, что при справедливости гипотезы H_0 точки $\boldsymbol{x}_j$ имеют вероятностную массу $q_j = m_j\delta$. Можно мысленно представить, что j-я точка («молекула») состоит

из m_j частей («атомов») массы δ при гипотезе H_0 и массы p_j/m_j при альтернативе H_1. Оптимальное критическое множество из «атомов» строится так же, как и раньше. Причем, поскольку для «атомов» фиксированной «молекулы» отношение p_j/m_j одно и тоже, можно считать, что при упорядочении они идут подряд. В результате видим, что оптимальное критическое множество «молекул» строится на основе включения в него точек с наибольшим отношением $\dfrac{p_j}{m_j} = \dfrac{p_j}{q_j}\,\delta$, другими словами, — с наибольшим отношением вероятностей (*правдоподобий*) $\mathbf{P}_1(\boldsymbol{x}_j)/\mathbf{P}_0(\boldsymbol{x}_j)$.

Товарищи ученые,
доценты с кандидатами!
Замучились вы с иксами,
запутались в нулях.
Сидите, разлагаете
молекулы на атомы...
В. Высоцкий

Для абсолютно непрерывной модели имеет место то же самое, только отношение вероятностей заменяется на отношение плотностей $p_1(\boldsymbol{x})/p_0(\boldsymbol{x})$ выборки $\boldsymbol{X}$ при H_1 и H_0. Для строгой формулировки теоремы Неймана — Пирсона рассмотрим для $c \geqslant 0$ систему вложенных множеств $G_c = \{\boldsymbol{x} \in \mathbb{R}^n\colon p_1(\boldsymbol{x})/p_0(\boldsymbol{x}) \geqslant c\}$ (рис. 6) и определим функцию $\varphi(c) = \mathbf{P}_0(G_c)$.

Тогда $\varphi(0) = 1$ и $\varphi(c)$ может только убывать с ростом c. Покажем, что $\varphi(c) \leqslant 1/c \to 0$ при $c \to \infty$. Действительно,

$$1 \geqslant \mathbf{P}_1(G_c) = \int\limits_{G_c} p_1(\boldsymbol{x})\,d\boldsymbol{x} \geqslant c \int\limits_{G_c} p_0(\boldsymbol{x})\,d\boldsymbol{x} = c\,\mathbf{P}_0(G_c) = c\,\varphi(c). \tag{6}$$

Рис. 6

Потребуем выполнения **двух условий**:

1) *плотности* $p_0(\boldsymbol{x})$ *и* $p_1(\boldsymbol{x})$ *положительны при всех* $\boldsymbol{x} \in \mathbb{R}^n$;

2) *для заданного уровня значимости* $\alpha \in (0,1)$ *существует* $c = c_\alpha$, *для которого* $\varphi(c_\alpha) = \alpha$ (это всегда выполняется, если функция $\varphi(c)$ непрерывна).

Теорема 1 (Нейман—Пирсон). При сделанных предположениях 1 и 2 наиболее мощный критерий уровня α задается критическим множеством

$$G^* = G_{c_\alpha} = \{\boldsymbol{x} \in \mathbb{R}^n\colon p_1(\boldsymbol{x})/p_0(\boldsymbol{x}) \geqslant c_\alpha\}. \tag{7}$$

Доказательство. Пусть G — критическое множество некоторого критерия уровня α. Согласно определению множества G и условию 2

$$\mathbf{P}_0(G) \leqslant \alpha = \mathbf{P}_0(G^*). \tag{8}$$

Обозначим через $I(\boldsymbol{x})$ и $I^*(\boldsymbol{x})$, соответственно, индикаторы множеств G и G^* и рассмотрим функцию

$$f(\boldsymbol{x}) = (I^*(\boldsymbol{x}) - I(\boldsymbol{x}))\,(p_1(\boldsymbol{x}) - c_\alpha p_0(\boldsymbol{x})). \tag{9}$$

Покажем, что она неотрицательна при всех $\boldsymbol{x} \in \mathbb{R}^n$. В самом деле, при $\boldsymbol{x} \in G^*$ оба сомножителя в формуле (9) неотрицательны: первый — так как $I^*(\boldsymbol{x}) = 1$, а второй — по определению (7). При $\boldsymbol{x} \notin G^*$ первый сомножитель в формуле (9) равен $-I(\boldsymbol{x}) \leqslant 0$,

а ввиду определения (7) и второй сомножитель неположителен. Поэтому

$$0 \leqslant \int f(\boldsymbol{x})\,d\boldsymbol{x} = \int I^*(\boldsymbol{x})\,p_1(\boldsymbol{x})\,d\boldsymbol{x} - \int I(\boldsymbol{x})\,p_1(\boldsymbol{x})\,d\boldsymbol{x} - \\ - c_\alpha\left[\int I^*(\boldsymbol{x})\,p_0(\boldsymbol{x})\,d\boldsymbol{x} - \int I(\boldsymbol{x})\,p_0(\boldsymbol{x})\,d\boldsymbol{x}\right] = \\ = \mathbf{P}_1(G^*) - \mathbf{P}_1(G) - c_\alpha\left[\mathbf{P}_0(G^*) - \mathbf{P}_0(G)\right].$$

Отсюда и из неравенства (8) вытекает, что $\mathbf{P}_1(G^*) \geqslant \mathbf{P}_1(G)$. ■

В учебнике [38, с. 219] доказано, что если мера Лебега множества $\{\boldsymbol{x}\colon p_1(\boldsymbol{x})/p_0(\boldsymbol{x}) = c\}$ равна нулю, то G^* является *единственным* наиболее мощным критическим множеством критерия уровня α с точностью до подмножества $\mathbb{R}^n$ лебеговой меры нуль (контрпример см. в задаче 5).

Покажем, что при выполнении условий 1 и 2 критерий Неймана — Пирсона является *строго несмещенным*.

Теорема 2. Для множества G^*, задаваемого формулой (7), справедливо неравенство $\mathbf{P}_1(G^*) > \alpha$.

Доказательство. Если в формуле (7) константа $c_\alpha > 1$, то из соотношения (6) следует, что

$$\mathbf{P}_1(G^*) \geqslant c_\alpha \mathbf{P}_0(G^*) = c_\alpha\alpha > \alpha.$$

При $c_\alpha \leqslant 1$, поскольку $p_1(\boldsymbol{x}) < c_\alpha p_0(\boldsymbol{x}) \leqslant p_0(\boldsymbol{x})$ при $\boldsymbol{x} \in \overline{G}^*$, имеем

$$\mathbf{P}_1(G^*) = 1 - \int\limits_{\overline{G}^*} p_1(\boldsymbol{x})\,d\boldsymbol{x} > 1 - \int\limits_{\overline{G}^*} p_0(\boldsymbol{x})\,d\boldsymbol{x} = \int\limits_{G^*} p_0(\boldsymbol{x})\,d\boldsymbol{x} = \alpha.$$

Итак, в любом случае $\mathbf{P}_1(G^*) > \alpha$, что и требовалось доказать. ■

Укажем множество G^* для модели примера 1. Ввиду (7) найдем

$$p_1(\boldsymbol{x})/p_0(\boldsymbol{x}) = \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[(x_i-\theta_1)^2 - (x_i-\theta_0)^2\right]\right\} = \\ = \exp\left\{\frac{n}{\sigma^2}(\theta_1-\theta_0)\,\overline{x} - \frac{n}{2\sigma^2}(\theta_1^2-\theta_0^2)\right\}.$$

Неравенство $p_1(\boldsymbol{x})/p_0(\boldsymbol{x}) \geqslant c_\alpha$ эквивалентно неравенству

$$\overline{x} \geqslant \sigma^2 \ln c_\alpha/[n(\theta_1-\theta_0)] + (\theta_1+\theta_0)/2. \tag{10}$$

Отсюда заключаем, что критерий примера 1 является наиболее мощным при t_α, равном правой части неравенства (10). Из формулы (1) можно найти соответствующее значение c_α.

Отметим, что граница t_α в равенстве (1) не зависит от θ_1. Поэтому мощность рассматриваемого критерия максимальна при любом $\theta_1 > \theta_0$. Другими словами, критерий является *равномерно наиболее мощным* против сложной альтернативы $H_1\colon \theta > \theta_0$.

§ 3. ПОСЛЕДОВАТЕЛЬНЫЙ АНАЛИЗ

Приведем небольшой отрывок из [72, с. 120], посвященный истории возникновения последовательного анализа.

> «В классической теории математической статистики предполагается, что элементы выборки (наблюдения) заранее известны. В основе одного из важнейших направлений современной статистики лежит понимание того, что не нужно фиксировать заранее объем выборки, его следует определять в зависимости от результатов более ранних наблюдений. Таким образом, объем выборки случаен. Эта идея последовательного выбора постепенно развивалась в работах Г. Доджа и Г. Ромига (1929 г.), П. Махаланобиса (1940 г.), Г. Хотеллинга (1941 г.) и У. Бертки (1943 г.), но настоящим основателем теории последовательного анализа в математической статистике является А. Вальд (1902–1950 гг.). Его последовательный критерий отношения правдоподобия (1943 г.) стал важным открытием, позволившим (в типичных ситуациях) на 50% уменьшить среднее число наблюдений (при тех же вероятностях ошибок). Неудивительно, что в годы второй мировой войны открытие Вальда было объявлено «секретным». Его основная книга «Последовательный анализ» опубликована лишь в 1947 г. Год спустя Вальд и Дж. Волфовиц доказали, что методы, отличные от последовательного критерия отношения правдоподобия, не дают такого уменьшения числа элементов выборки.»

Рассмотрим последовательный критерий Вальда в случае, когда элементы выборки X_i при гипотезе H_0 имеют известную плотность $p_0(x) > 0$, а при альтернативе H_1 — известную плотность $p_1(x) > 0$. Определим случайные величины $Z_i = \ln\left(p_1(X_i)/p_0(X_i)\right)$. Потребуем, чтобы и при гипотезе H_0, и при альтернативе H_1 были выполнены **два условия**: 1) $\mathbf{M}Z_1 \neq 0$, 2) $0 < \mathbf{D}Z_1 < \infty$.

Положим $S_0 = 0$, $S_k = Z_1 + \ldots + Z_k$, $k = 1,2,\ldots$. Случайная величина S_k представляет собой координату «блуждающей частицы» после k независимых и одинаково распределенных «шагов» Z_i случайного блуждания по прямой. Пусть s_k — это наблюдавшиеся значения величин S_k. На рис. 7 изображена возможная *траектория* (развертка во времени) случайного блуждания — ломаная, соединяющая точки плоскости с координатами (k,s_k).

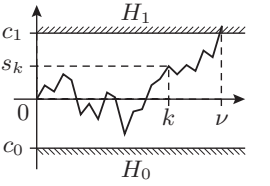


Рис. 7

Последовательный критерий Вальда состоит в следующем. Задаются константы $c_0 < 0$ и $c_1 > 0$, и наблюдения продолжаются до *момента выхода* ν блуждания s_k из интервала (c_0,c_1). Если $s_\nu \leqslant c_0$, то принимается гипотеза H_0; если $s_\nu \geqslant c_1$, то принимается альтернатива H_1 (см. рис. 7).

Может ли блуждание продолжаться как угодно долго?

Теорема 3. При выполнении указанных выше условий 1–2 процедура Вальда с вероятностью 1 заканчивается за конечное число шагов ν, причем моменты $\mathbf{M}\nu^k < \infty$ для всех $k \geqslant 1$.

Доказательство теоремы можно найти в [32, с. 149] или [38, с. 232].

Замечание 1. Первое утверждение теоремы интуитивно понятно вследствие условия 1, приводящего (ввиду закона больших чисел) к систематическому «сносу» блуждания и выходу его из интервала (c_0,c_1) за время порядка $c_1/\mathbf{M}Z_1$ при $\mathbf{M}Z_1 > 0$ и порядка $c_0/\mathbf{M}Z_1$ при $\mathbf{M}Z_1 < 0$.

Как вероятности α и β ошибок I и II рода критерия Вальда связаны с константами c_0 и c_1? Ответ дает следующая теорема.

Теорема 4. Зададим $\alpha' > 0$ и $\beta' > 0$, удовлетворяющие условию $\alpha' + \beta' < 1$. Возьмем

$$c_0 = \ln[\beta'/(1-\alpha')], \quad c_1 = \ln[(1-\beta')/\alpha']. \tag{11}$$

Тогда для вероятностей α и β выполняются неравенства

$$\alpha \leqslant \alpha'/(1-\beta'),\ \beta \leqslant \beta'/(1-\alpha'),\ \alpha+\beta \leqslant \alpha'+\beta'. \tag{12}$$

Замечание 2. Эти неравенства показывают, что каждая из вероятностей α и β может лишь незначительно превысить α' и β', соответственно, когда последние малы. (Например, для $\alpha' = \beta' = 0{,}1$ граница сверху равна $1/9 \approx 0{,}111$.) Кроме того, сумма $\alpha + \beta$ вероятностей ошибок не может превзойти задаваемую величину $\alpha' + \beta'$.

Доказательство. Обозначим через A_n^0 (A_n^1) множество тех результатов наблюдений $(x_1,\ldots,x_n)$, для которых процедура заканчивается на шаге n $(\nu = n)$ принятием гипотезы H_0 (альтернативы H_1). Например,

$$A_n^0 = \{(x_1,\ldots,x_n)\colon c_0 < s_k < c_1, k = 1,\ldots,n-1,\ s_n \leqslant c_0\}. \tag{13}$$

При гипотезе H_0 в силу теоремы 3 справедливо равенство

$$\sum_{n=1}^{\infty} \mathbf{P}_0(A_n^0) + \sum_{n=1}^{\infty} \mathbf{P}_0(A_n^1) = 1. \tag{14}$$

Так как в точках множества A_n^0 согласно определению (13) выполняется неравенство $s_n \leqslant c_0 \iff \prod_{i=1}^{n} p_1(x_i) \leqslant d_0 \prod_{i=1}^{n} p_0(x_i)$, где $d_0 = \exp\{c_0\} = \beta'/(1-\alpha')$, то из соотношения (14) имеем

$$\beta = \sum_{n=1}^{\infty} \mathbf{P}_1(A_n^0) \leqslant d_0 \sum_{n=1}^{\infty} \mathbf{P}_0(A_n^0) = d_0\left(1 - \sum_{n=1}^{\infty} \mathbf{P}_0(A_n^1)\right) = d_0(1-\alpha).$$

Аналогично доказывается, что

$$\alpha \leqslant (1-\beta)/d_1, \quad \text{где } d_1 = \exp\{c_1\} = (1-\beta')/\alpha'.$$

Из этих двух неравенств, соответственно, выводим соотношения

$$\beta \leqslant (1-\alpha)\beta'/(1-\alpha') \leqslant \beta'/(1-\alpha'),$$
$$\alpha \leqslant (1-\beta)\alpha'/(1-\beta') \leqslant \alpha'/(1-\beta').$$

Складывая неравенства $\beta' - \beta'\alpha \geqslant \beta - \alpha'\beta$ и $-\alpha + \alpha\beta' \geqslant -\alpha' + \alpha'\beta$, получаем $\alpha + \beta \leqslant \alpha' + \beta'$. ■

Замечание 3. Сравним критерии Вальда и Неймана — Пирсона. Во-первых, у последнего заранее фиксируется число наблюдений n, и решение отвергнуть гипотезу H_0 принимается в случае попадания конечной точки блуждания s_n в множество $[\ln c_\alpha, \infty)$ (рис. 8).

Во-вторых, для определения c_α по вероятности α ошибки I рода надо знать распределение статистики критерия Неймана—Пирсона при справедливости гипотезы H_0, в то время как для расчета границ c_0 и c_1 критерия Вальда для заданных α и β не возникает проблемы отыскания распределений. Информация о плотностях $p_0(x)$ и $p_1(x)$ требуется только для вычисления математического ожидания числа наблюдений ν до принятия решения. Вычисление опирается на теорему 5.

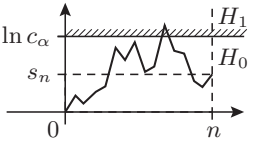


Рис. 8

Теорема 5 (тождество Вальда). При условии конечности величин $\mathbf{M}Z_1$ и $\mathbf{M}\nu$ справедливо тождество $\mathbf{M}S_\nu = \mathbf{M}Z_1 \cdot \mathbf{M}\nu$.

Контрпример. Это тождество справедливо не для всякого случайного момента остановки блуждания. Пусть $\mathbf{P}(Z_1 = -1) = \mathbf{P}(Z_1 = 1) = 1/2$, $\nu' = \min\{n\colon S_n = -1\}$. Тогда $\mathbf{P}(\nu' < \infty) = 1$, но $\mathbf{M}\nu' = \infty$ (см. теорему 2 гл. 14). Поскольку $\mathbf{M}Z_1 = 0$, $\mathbf{M}S_{\nu'} = -1$, слева получаем -1, а справа — неопределенность вида $0 \cdot \infty$.

Доказательство. Положим $Y_n = I_{\{c_0 < S_k < c_1, k=1,\dots,n-1\}} = I_{\{\nu \geqslant n\}}$. Ввиду формулы 3 гл. 1 $\mathbf{M}\nu = \sum\limits_{n=1}^{\infty} \mathbf{P}(\nu \geqslant n) = \sum\limits_{n=1}^{\infty} \mathbf{M}Y_n$. Как функция только от $Z_1, \dots, Z_{n-1}$ случайная величина Y_n не зависит от Z_n. Отсюда

$$\mathbf{M}S_\nu = \sum_{n=1}^{\infty} \mathbf{M}\, S_n I_{\{\nu=n\}} = \sum_{n=1}^{\infty} \sum_{k=1}^{n} \mathbf{M}\, Z_k I_{\{\nu=n\}} = \sum_{k=1}^{\infty} \sum_{n=k}^{\infty} \mathbf{M}\, Z_k I_{\{\nu=n\}} =$$
$$= \sum_{k=1}^{\infty} \mathbf{M}\, Z_k \sum_{n=k}^{\infty} I_{\{\nu=n\}} = \sum_{k=1}^{\infty} \mathbf{M}\, Z_k Y_k = \sum_{k=1}^{\infty} \mathbf{M}Z_k \cdot \mathbf{M}Y_k = \mathbf{M}Z_1 \cdot \mathbf{M}\nu.$$

Здесь первое равенство и перемена порядка суммирования законны благодаря абсолютной суммируемости соответствующих рядов (см. [41, с. 343, 365]):

$$\sum_{k=1}^{\infty} \sum_{n=k}^{\infty} \mathbf{M}\,|Z_k| I_{\{\nu=n\}} = \sum_{k=1}^{\infty} \mathbf{M}\,|Z_k|\, Y_k = \sum_{k=1}^{\infty} \mathbf{M}|Z_k| \cdot \mathbf{M}Y_k =$$
$$= \mathbf{M}|Z_1| \cdot \mathbf{M}\nu < \infty$$

в силу условий теоремы ($\mathbf{M}Z_1$ конечно $\Longleftrightarrow \mathbf{M}|Z_1| < \infty$). ■

Пусть заданы *малые* вероятности α и β ошибок. На основе замечания 2 они примерно равны константам α' и β' из теоремы 4. Так как длина интервала (c_0, c_1) при малых α' и β' будет большой, а $\mathbf{M}Z_1$ конечно, то можно пренебречь величиной «перескока» случайным блужданием границы в момент остановки, т. е. считать, что $S_\nu \approx c_j$, если принимается гипотеза H_j, $j = 0, 1$. Эти рассуждения

приводят к приближенным равенствам

$$\mathbf{M}S_\nu \approx c_0(1-\alpha)+c_1\alpha, \quad \text{если верна гипотеза } H_0,$$
$$\mathbf{M}S_\nu \approx c_0\beta+c_1(1-\beta), \quad \text{если верна альтернатива } H_1.$$

Применяя тождество Вальда, отсюда получаем

Более точные аппроксимации см. в [11, с. 360], [12, с. 223].

$$\mathbf{M}_0\nu \approx \frac{c_0(1-\alpha)+c_1\alpha}{\mathbf{M}_0 Z_1}, \quad \mathbf{M}_1\nu \approx \frac{c_0\beta+c_1(1-\beta)}{\mathbf{M}_1 Z_1}. \tag{15}$$

Выясним, насколько последовательная процедура Вальда *экономичней* критерия Неймана — Пирсона для модели примера 1. Прежде всего, выразим явно Z_i через X_i:

$$Z_i = \ln\frac{p_1(X_i)}{p_0(X_i)} = \frac{(X_i-\theta_0)^2-(X_i-\theta_1)^2}{2\sigma^2} = \frac{\theta_1-\theta_0}{\sigma^2}\left(X_i-\frac{\theta_0+\theta_1}{2}\right).$$

Поэтому если верна H_j (j принимает значения 0 и 1), то

$$\mathbf{M}_j Z_1 = \frac{\theta_1-\theta_0}{\sigma^2}\left(\theta_j-\frac{\theta_0+\theta_1}{2}\right) = (-1)^{j+1}\,\frac{(\theta_1-\theta_0)^2}{2\sigma^2}. \tag{16}$$

Для критерия Неймана — Пирсона (см. задачу 3) с теми же вероятностями ошибок α и β необходимое число наблюдений n^* равно $[\sigma(x_\alpha+x_\beta)/(\theta_1-\theta_0)]^2$ (с точностью до 1), где x_p обозначает p-квантиль закона $\mathcal{N}(0,1)$. Возьмем для простоты $\alpha=\beta$. Из формул (11), (15) и (16) следует, что *экономичность критерия Вальда* равна

$$E(\alpha) = \frac{\mathbf{M}_0\nu}{n^*} = \frac{\mathbf{M}_1\nu}{n^*} \approx \frac{1-2\alpha}{2x_\alpha^2}\,\ln\frac{1-\alpha}{\alpha}. \tag{17}$$

При $\alpha=0{,}05$ по таблице Т2 находим $x_\alpha \approx -1{,}645$. Подставив эти значения в формулу (17), получаем $E(0{,}05)\approx 0{,}49$. Таким образом, в данном случае для принятия решения с помощью последовательного критерия Вальда потребуется в среднем примерно *вдвое меньше наблюдений*, чем для оптимального критерия Неймана — Пирсона с заранее фиксированным размером выборки (см. также задачу 4).

Не во всякой игре тузы выигрывают!

Козьма Прутков

Можно доказать, что критерий Вальда минимизирует средний размер выборки по сравнению с любым другим последовательным критерием, имеющим те же или меньшие вероятности ошибок I и II рода (см. [11, с. 358]).

§ 4. РАЗОРЕНИЕ ИГРОКА

Что ни толкуй Вольтер или Декарт —
Мир для меня — колода карт.
Жизнь — банк: рок мечет, я играю,
И правила игры я к людям применяю.

М. Ю. Лермонтов, «Маскарад»

Представим, что два противника принимают участие в игре, состоящей из большого числа независимых партий. Вероятность выигрыша первого игрока в каждой из партий равна p, а второго — $q=1-p$ (ничьих не бывает). Плата за проигрыш в одной партии равна 1. Начальный капитал первого игрока составляет k,

второго — $(M-k)$. Игра прекращается, когда один из участников проиграет все наличные.

Ход игры можно наглядно представить при помощи поднятой на k единиц вверх траектории случайного блуждания $S_n = Z_1 + Z_2 + \ldots + Z_n$, где «шаги» Z_i независимы, $\mathbf{P}(Z_i = 1) = p$, $\mathbf{P}(Z_i = -1) = q$ (рис. 9). Величина $(S_n + k)$ — капитал первого игрока после n сыгранных партий. Считая суммарный капитал игроков M заданным, найдем r_k — *вероятность разорения первого игрока*.

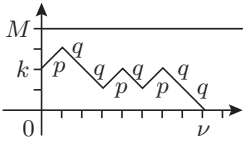


Рис. 9

Перенесем начало координат на рис. 9 в точку $(0,k)$. Обозначим через ν момент окончания игры (разорения одного из игроков). Поскольку $\mathbf{M}Z_1 = p - q$ и $\mathbf{D}Z_1 = 1 - (p-q)^2 < \infty$, при $p \neq 1/2$ выполняются условия теоремы 3. Следовательно, $\mathbf{M}\nu < \infty$. Докажем, что в случае $p = 1/2$ ($\mathbf{M}Z_1 = 0$) $\mathbf{M}\nu$ также конечно.

Приведем, следуя [12, с. 90], доказательство для четного $M = 2m$. Пусть $l = \min\{k, M-k\}$. Тогда с вероятностью $2^{-l} \geqslant 2^{-m}$ игра может закончиться за l партий. Поскольку в течение игры суммарный капитал M не меняется, разбивая блуждание на независимые отрезки длины m, видим, что $\mathbf{P}(\nu > m) \leqslant 1 - 2^{-m}, \ldots, \mathbf{P}(\nu > jm) \leqslant (1-2^{-m})^j$. Но вероятности $\mathbf{P}(\nu > i)$ убывают с ростом i, а при $i = jm$ $(j = 1,2,\ldots)$ мажорируются членами геометрической прогрессии со знаменателем $\gamma = 1 - 2^{-m}$. Поэтому ряд $\sum\limits_{i=0}^{\infty} \mathbf{P}(\nu > i)$ сходится. Согласно формуле (3) гл. 1 его сумма равна $\mathbf{M}\nu$.

Случай нечетного M разберите самостоятельно.

Конечность $\mathbf{M}\nu$ обеспечивает законность всех переходов доказательства теоремы 5 (тождества Вальда) из § 3. Запишем его при $p = 1/2$ для $c_0 = -k$, $c_1 = M - k$, $r_k = \mathbf{P}(S_\nu = c_0)$:

$$S_\nu = c_0 r_k + c_1(1 - r_k) = \mathbf{M}Z_1 \cdot \mathbf{M}\nu = 0 \cdot \mathbf{M}\nu = 0.$$

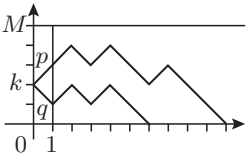


Рис. 10

Отсюда находим $r_k = 1 - k/M$, т. е. шансы на выигрыш *прямо пропорциональны* величине начального капитала.

Получим этот ответ другим способом — *разложением по первому шагу блуждания* (рис. 10). С учетом формулы полной вероятности (П7) интуитивно понятно, что вероятности r_k должны удовлетворять рекуррентным соотношениям

$$r_k = p\, r_{k+1} + q\, r_{k-1}, \quad k = 1, \ldots, M-1, \tag{18}$$

с граничными условиями $r_0 = 1$, $r_M = 0$ (строгий вывод см. в [39, с. 39]). В случае $p = 1/2$ можно переписать соотношение (18) в виде $r_{k+1} - r_k = r_k - r_{k-1}$, из которого следует, что точки плоскости (k, r_k) лежат на одной прямой $r_k = A + Bk$ (рис. 11). Из граничных условий вытекает, что $A = 1$, $B = -1/M$.

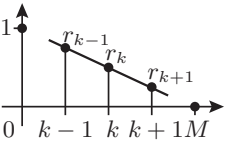


Рис. 11

Прежде чем рассматривать случай $p \neq 1/2$, обсудим аналогию между соотношением (18) и *краевой задачей* для дифференциального уравнения второго порядка

$$y''(x) = f(x,y), \quad y(x_0) = a,\ y(x_1) = b. \tag{19}$$

Я больше всего дорожу аналогиями, моими самыми верными учителями.

И. Кеплер

При малом $h > 0$ производную первого порядка $y'(x)$ можно приближенно заменить на так называемую *конечную разность* $[y(x+h) - y(x)]/h$. Производная второго порядка $y''(x)$ аппроксимируется *второй симметричной разностью*

$$\frac{\dfrac{y(x+h)-y(x)}{h} - \dfrac{y(x)-y(x-h)}{h}}{h} = \frac{y(x+h) - 2y(x) + y(x-h)}{h^2}. \quad (20)$$

При $p = 1/2$ соотношение (18) имеет вид $r_{k+1} - 2r_k + r_{k-1} = 0$, аналогичный виду правой части равенства (20), если положить $x = k$, $y(x) = r_k$ и $h = 1$. Общим решением соответствующего дифференциального уравнения $y'' = 0$ является $y = A + Bx$.

Чтобы найти r_k при $p \neq 1/2$ (задача 6), можно использовать основную идею *метода стрельбы*, применяемого для численного решения краевой задачи. Предположим, что мы умеем численно решать *задачу Коши*

$$y''(x) = f(x,y), \quad y(x_0) = a,\ y'(x_0) = c, \quad (21)$$

например, методом Эйлера из § 6 гл. 2 или Рунге — Кутты (см. [6, с. 439]). Фиксируем a и будем менять c. Обозначим через $g(c)$ значение решения задачи (21) при $x = x_1$ из краевого условия задачи (19). Тогда численное решение краевой задачи (19) сводится к поиску корня c^* уравнения $g(c) = b$. Так как значения $g(c)$ мы умеем вычислять в пробных точках c_n (рис. 12), то можно приближенно найти c^* с помощью деления отрезка пополам или метода Ньютона из § 5 гл. 9. Рисунок 13 объясняет происхождение названия «метод стрельбы».

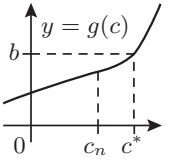


Рис. 12

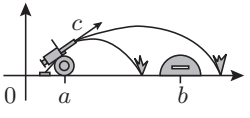


Рис. 13

Приведем ответ задачи 6: $r_k = (\lambda^M - \lambda^k)/(\lambda^M - 1)$, где $\lambda = q/p$. Устремляя M к бесконечности, видим, что предел r_k равен 1, если $p \leqslant 1/2$, и равен λ^k, если $p > 1/2$. Таким образом, искусный игрок может с положительной вероятностью $1 - \lambda^k$ *никогда* не проиграть даже «бесконечно богатому» противнику. (Если он не проиграл из-за случайности сразу, то в дальнейшем его шансы уйти от поражения резко увеличиваются по причине «сноса» вверх траектории блуждания.)

Найдем *среднее время до разорения одного из игроков (среднюю продолжительность игры)* $l_k = \mathbf{M}\nu$, $k = 1, \ldots, M-1$.

Вопрос 3. Какое значение l_k представляется наиболее правдоподобным при $p = 1/2$ для $k = 10$ и $M = 20$ (рис. 14)?

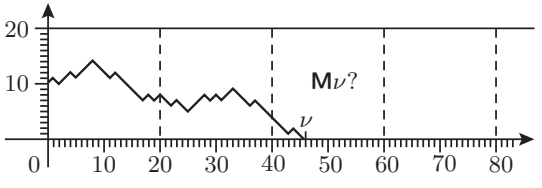


Рис. 14

Те же рассуждения, что и при получении соотношений (18), приводят для l_k к уравнениям

$$l_k = p\, l_{k+1} + q\, l_{k-1} + 1, \quad k = 1, \ldots, M-1, \quad (22)$$

с граничными условиями $l_0 = 0$, $l_M = 0$ (единица в правой части уравнения (22) добавлена потому, что одна партия уже была сыграна). Нетрудно проверить, что при $p = 1/2$ ему удовлетворяют $l_k = k(M - k)$. Следовательно, при больших M и $k \approx \alpha M$ среднее время игры равных по силе противников имеет порядок M^2. Этот результат становится менее удивительным, если принять во внимание центральную предельную теорему (П6), согласно которой траектория симметричного случайного блуждания S_n колеблется в среднем в пределах $\pm\sqrt{n}$.

Иначе дело обстоит при $p \neq 1/2$ (см. замечание 1 из § 3).

Вопрос 4.
Чему равно $\mathbf{M}\nu$ при $p \neq 1/2$? (Используйте для его вычисления тождество Вальда.)

§ 5. ОПТИМАЛЬНАЯ ОСТАНОВКА БЛУЖДАНИЯ

Проведем небольшой вероятностный эксперимент. Пусть в начальный момент частица попадает с равными вероятностями в любую из $M + 1$ точек с целыми координатами отрезка $[0, M]$. Если частица оказывается в одном из концов отрезка, то выигрыш равен 0. В противном случае участник эксперимента должен принять решение: остаться в данной точке и взять приз, равный высоте столбика над этой точкой (рис. 15 для $M = 7$), или же подбросить симметричную монетку и переместить частицу на 1 влево, если выпадет герб, и на 1 вправо, если выпадет решка. После этого снова можно или взять приз, или сделать еще один «случайный шаг» и т. д.

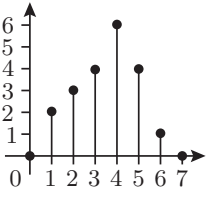


Рис. 15

Если в результате перемещения частица попадет в концевую точку отрезка $[0, M]$, то блуждание принудительно останавливается, и выигрыш оказывается равным 0. Проблема состоит в выборе стратегии, приносящей участнику эксперимента максимальный средний выигрыш. Ответьте на вопрос 5.

Предположим, что в начальный момент частица оказалась в точке с абсциссой 6 на рис. 15. Ответьте на вопрос 6.

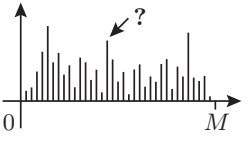


Рис. 16

Давайте проведем моделирование. Прежде всего, разыграем начальное положение частицы. Для этого откройте какую-нибудь книгу «случайным образом» и обратите внимание, скажем, на вторую[*)] цифру справа номера правой страницы. Если эта цифра окажется больше 7, то открывайте книгу до тех пор, пока она не попадет в множество {0,1, . . . ,7}. Возьмите ее в качестве начальной координаты частицы на рис. 15. Если частица оказалась в 0 или 7, то выигрыш равен 0, и моделирование закончено. В противном случае надо принять решение: взять приз или начать блуждание. Если выбрано второе, то снова надо открыть «случайным образом» книгу и переместить частицу на 1 влево или вправо, в соответствии с тем, оказалась ли вторая справа цифра номера меньше 5 или нет. После перемещения надо или взять приз, или продолжить

Вопрос 5.
Стоит или нет останавливаться, если частица находится в точке, указанной стрелкой на рис. 16?

Вопрос 6.
Какая из двух стратегий в данной ситуации лучше: сразу взять приз или сделать ровно один «случайный шаг» и взять приз?

[*)] Первые справа — всегда нечетные, а третьи — слишком медленно меняются.

случайное блуждание и т. д. Если при этом частица попадет в 0 или 7, то моделирование прекращается с нулевым выигрышем. (Повторите моделирование 4 раза, каждый раз разыгрывая заново начальное положение частицы, и подсчитайте общий выигрыш.)

Давайте рассмотрим **некоторые стратегии**.

Стратегия «Никогда»

Если никогда не брать приз, то в соответствии с задачей о разорении игрока из § 4, частица с вероятностью 1 рано или поздно попадет в один из концов отрезка, где выигрыш равен нулю.

Стратегия «Сразу»

Очевидно, что средний выигрыш при немедленной остановке в начальном положении составляет

$$(0+2+3+4+6+4+1+0)/8 = 20/8.$$

Стратегия «Все или ничего»

При использовании такой стратегии блуждание продолжается до момента попадания или в точку 4 с максимальным выигрышем, равным 6, или до «поглощения» на одном из концов отрезка с нулевым выигрышем. Для подсчета среднего выигрыша воспользуемся свойством 1 условного математического ожидания (П7), в котором роль условия будет играть начальное положение.

Рассмотрим, например, случай, когда в начальный момент частица оказалась в точке 5. Тогда блуждание до попадания либо в 4, либо в 7 представляет собой задачу о разорении игрока при $p = 1/2$, имеющего начальный капитал $k = 7 - 5 = 2$, в то время как суммарный капитал двух игроков $M = 7 - 4 = 3$. В § 4 было доказано, что вероятность разорения игрока $r_k = 1 - k/M$. Отсюда находим *условный* средний выигрыш $0 \cdot \frac{1}{3} + 6 \cdot \frac{2}{3} = 4$.

Вопрос 7.
Чему равно среднее время до остановки при использовании стратегии «Все или ничего»?

Аналогично, перебирая все возможные начальные положения от 0 до 7, легко подсчитываем *безусловный* средний выигрыш:

$$6 \cdot \left(0 + \frac{1}{4} + \frac{1}{2} + \frac{3}{4} + 1 + \frac{2}{3} + \frac{1}{3} + 0\right) / 8 = 21/8.$$

Вопрос 8.
Чему равен максимальный средний выигрыш для столбиков, изображенных на рис. 15?

Наилучшая стратегия

В [24] в качестве оптимальной для данной задачи приведена следующая красивая стратегия. Представим, что призовые столбики — это картонные полоски, приклеенные на лист бумаги. Закрепим кнопками резинку в концевых точках отрезка $[0, M]$, оттянем ее вверх и отпустим (рис. 17). При этом столбики разделятся на два класса: такие, которые будут поддерживать резинку, и такие, над

которыми резинка пройдет сверху, их не касаясь. (Мы наглядно построили то, что формально называется *выпуклой оболочкой* графика функции.) Оптимальная стратегия состоит в том, что нужно останавливаться в тот момент, когда частица попадет в точку, над которой располагается столбик из первого класса (на рис. 17 такие точки обведены кружками).

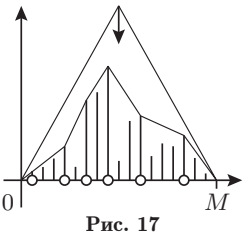


Рис. 17

ЗАДАЧИ

1. Каким следует взять t_α для критерия из примера 1, чтобы минимизировать сумму $\alpha+\beta$?

2. Пусть выборка состоит из единственного наблюдения X_1 с плотностью $p(x-\theta)$. Что представляет собой при разных значениях c_α критическое множество критерия Неймана — Пирсона для проверки гипотезы $H_0\colon \theta=0$ против альтернативы $H_1\colon \theta=1$, если $p(x)=[\pi(1+x^2)]^{-1}$ (плотность закона Коши)?

3. Докажите, что число наблюдений n^*, обеспечивающее заданные вероятности α и β для модели примера 1, равно $[\sigma(x_\alpha+x_\beta)/(\theta_1-\theta_0)]^2$ (с точностью до 1), где x_p обозначает p-квантиль закона $\mathcal{N}(0,1)$.

4. Вычислите предел при $\alpha\to 0$ экономичности критерия Вальда $E(\alpha)$, определяемой формулой (17), с помощью асимптотики $1-\Phi(x)\sim(x\sqrt{2\pi})^{-1}\exp\{-x^2/2\}$ при $x\to+\infty$.

5*. Дайте ответ на вопрос задачи 2 для модели сдвига показательного закона с плотностью $p(x)=e^{-x}I_{\{x\geqslant 0\}}$.

6*. Найдите вероятность r_k разорения игрока при $p\neq 1/2$ методом «стрельбы».

Неусыпный труд препятства преодолевает.

М. В. Ломоносов

РЕШЕНИЯ ЗАДАЧ

1. В силу формулы (2)

$$\alpha+\beta=1-\Phi\left(x_{1-\alpha}\right)+\Phi\left(x_{1-\alpha}-\sqrt{n}\,(\theta_1-\theta_0)/\sigma\right). \tag{23}$$

Положим для краткости $c=\sqrt{n}\,(\theta_1-\theta_0)/\sigma$. Дифференцируя по $x=x_{1-\alpha}$ правую часть равенства (23), получим уравнение

$$\varphi(x)-\varphi(x-c)=0, \tag{24}$$

где $\varphi(x)=\Phi'(x)=\dfrac{1}{\sqrt{2\pi}}\,e^{-x^2/2}$ — плотность закона $\mathcal{N}(0,1)$. Так как $\varphi(x)$ — четная функция, строго убывающая при $x>0$, уравнение (24) имеет единственный корень $x^*=c/2$. Подставив его в соотношение (1), найдем оптимальную границу $t_\alpha=(\theta_0+\theta_1)/2$.

Другим доказательством может служить рис. 18, на котором $\alpha+\beta$ превосходит минимальную сумму на величину γ.

Меньше читайте, меньше учитесь, больше думайте. Учитесь у учителей и в книгах только тому, что вам нужно и хочется узнать.

Л. Н. Толстой

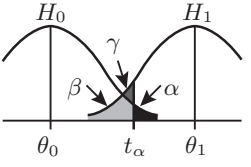


Рис. 18

2. Очевидно, оптимальное критическое множество задается неравенством $p_1(x)/p_0(x)\geqslant c \iff (1+x^2)/[1+(x-1)^2]\geqslant c$.

График функции $y(x) = p_1(x)/p_0(x)$ приведен на рис. 19, где $\varkappa = (\sqrt{5}-1)/2 \approx 0{,}618$ — «золотое сечение», ранее встречавшееся при решении задачи 5 гл. 1.

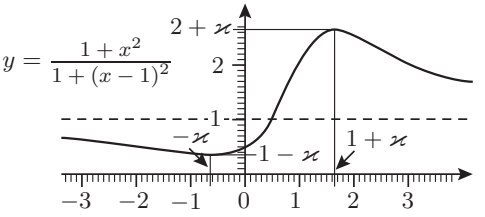


Рис. 19

Изменяя значение c от $+\infty$ до 0, видим, что критическим множеством является

$$\begin{cases} \text{пустое} & \text{при } c > 2+\varkappa, \\ \text{точка} & \text{при } c = 2+\varkappa, \\ \text{отрезок} & \text{при } 1 < c < 2+\varkappa, \\ \text{луч} & \text{при } c = 1, \\ \text{два луча} & \text{при } 1-\varkappa < c < 1, \\ \text{прямая} & \text{при } 0 \leqslant c \leqslant 1-\varkappa. \end{cases}$$

3. В силу четности плотности закона $\mathcal{N}(0,1)$ для квантилей этого распределения верно тождество $x_{1-\alpha} = -x_\alpha$. Поэтому формулу (2) можно переписать в виде

$$\beta = \Phi\left(-x_\alpha - \sqrt{n}\,(\theta_1 - \theta_0)/\sigma\right). \tag{25}$$

Применив обратную функцию Φ^{-1} к обеим частям формулы (25), получим равенство $x_\beta = -x_\alpha - \sqrt{n}\,(\theta_1 - \theta_0)/\sigma$, из которого следует доказываемое утверждение.

4. Выведем сначала саму асимптотику, следуя [81, с. 192]. Установим более сильный результат: для всех $x > 0$ справедливо двойное неравенство

$$(x^{-1} - x^{-3})\,\varphi(x) < 1 - \Phi(x) < x^{-1}\varphi(x). \tag{26}$$

Оно легко доказывается интегрированием по лучу $[x,\infty)$ очевидного неравенства

$$(1 - 3x^{-4})\,\varphi(x) < \varphi(x) < (1 + x^{-2})\,\varphi(x),$$

в котором участвуют производные с обратным знаком членов неравенства (26).

Взяв натуральный логарифм от обеих частей равенства

$$1 - \Phi(x) = x^{-1}\varphi(x)\,(1 + o(1)) = \frac{1}{x\sqrt{2\pi}}\; e^{-x^2/2}\,(1 + o(1)),$$

приходим к асимптотике

$$\ln\left(1 - \Phi(x)\right) = -x^2/2 - \ln x - \ln\sqrt{2\pi} + o(1) \quad \text{при } x \to +\infty.$$

Теперь все готово, чтобы вычислить искомый предел:

$$\lim_{\alpha\to 0} E(\alpha) = \lim_{\alpha\to 0} \frac{-\ln\alpha}{2x_\alpha^2} = \lim_{x\to+\infty} \frac{-\ln(1-\Phi(x))}{2x^2} = \frac{1}{4}.$$

5. Для $x > 0$ отношение плотностей $l(x) = p_1(x)/p_0(x) = \left[e^{-x+1}I_{\{x\geqslant 1\}}\right] / \left[e^{-x}I_{\{x\geqslant 0\}}\right] = e\,I_{\{x\geqslant 1\}}$. В частности, мера Лебега множества $\{x\colon l(x) = e\}$ равна бесконечности, и оптимальное критическое множество G^* не определяется однозначно. Его можно представить в виде $G' \cup G''$: в случае $\alpha \leqslant 1/e$ множество G' пустое, G'' — любое подмножество луча $[1,\infty)$ такое, что $\mathbf{P}_0(G'') = \alpha$ (рис. 20); в случае $\alpha > 1/e$ в качестве G' годится любое подмножество интервала $(0,1)$ такое, что $\mathbf{P}_0(G') = \alpha - 1/e$, $G'' = [1,\infty)$.

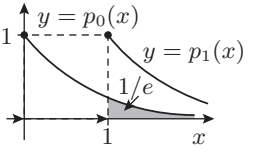


Рис. 20

6. Аналогом производной $y'(x) \approx [y(x+h) - y(x)]/h$ является приращение $\Delta_k = r_{k+1} - r_k$. Запишем соотношения (18) в терминах Δ_k: $p(r_{k+1} - r_k) = q(r_k - r_{k-1}) \iff \Delta_k = \lambda\Delta_{k-1}$, где $\lambda = q/p$ и $k = 0,1,\ldots,M-1$. Начальным условиям задачи Коши $y(x_0) = a$ и $y'(x_0) = c$ соответствуют $r_0 = 1$ и $\Delta_0 = c$. Последовательно выражая Δ_k через предыдущие, находим решение дискретной задачи Коши:

$$r_k(c) = r_0 + \Delta_0 + \ldots + \Delta_{k-1} = 1 + c + c\lambda + \ldots + c\lambda^{k-1} =$$
$$= 1 + c\,(\lambda^k - 1)/(\lambda - 1), \quad \text{если } \lambda \neq 1.$$

В соответствии с методом стрельбы, подберем константу c так, чтобы выполнялось второе краевое условие $r_M = 0$. Получим $c = -(\lambda-1)/(\lambda^M - 1)$, что приводит при $p \neq 1/2$ к ответу:

$$r_k = 1 - (\lambda^k - 1)/(\lambda^M - 1) = (\lambda^M - \lambda^k)/(\lambda^M - 1).$$

ОТВЕТЫ НА ВОПРОСЫ

1. Критерий будет несмещенным из-за возрастания функции $\Phi(x)$ и состоятельным, поскольку при $\theta > \theta_0$

$$W(\theta) = \Phi\left(\sqrt{n}\,(\theta - \theta_0)/\sigma - x_{1-\alpha}\right) \to 1 \quad \text{при } n \to \infty.$$

2. Критерий плох тем, что он имеет вероятность ошибки II рода

$$\beta = \mathbf{P}_{\theta_1}(\overline{X} \notin \Delta) \approx 1.$$

Пример наглядно демонстрирует характерное свойство азартных игр — они обычно длятся намного больше, чем предполагалось.

3. Рисунок 14 подсказывает значение $l_k \approx 50$. На самом деле, как доказано ниже, правильным ответом будет $l_k = k(M-k) = 100$.

В [81, с. 367] выведена явная формула (восходящая к Лагранжу и многократно переоткрывавшаяся после) для вероятности разорения 1-го игрока в n-й партии

$$v_{k,n} = \frac{1}{M}\; p^{(n-k)/2}\, q^{(n+k)/2} \sum_{j=1}^{M-1} \cos^{n-1}\left(\frac{\pi j}{M}\right) \sin\left(\frac{\pi j}{M}\right) \sin\left(\frac{\pi kj}{M}\right).$$

На ее основе была вычислена таблица распределения времени до окончания игры ν при $p = 1/2$, $k = 10$ и $M = 20$:

m	20	40	60	75	100	150	200	300
$\mathsf{P}(\nu \leqslant m)$	0,053	0,235	0,400	0,495	0,634	0,803	0,894	0,969

В частности, медиана распределения $\mu \approx 75$, но (в среднем) в каждом пятом случае ν будет больше, чем 150.

4. $\mathsf{M}\nu = \mathsf{M}S_\nu/\mathsf{M}Z_1 = [-k\,r_k + (M-k)(1-r_k)]\,/(p-q)$, где $r_k = (\lambda^M - \lambda^k)/(\lambda^M - 1)$, $\lambda = q/p$ (см. решение задачи 6).

5. Надо продолжать блуждание, так как и слева, и справа есть более высокие столбики.

6. Вторая стратегия в среднем вдвое выгодней, поскольку $4 \cdot 1/2 + 0 \cdot 1/2 = 2$.

7. Используя формулу $l_k = k(M-k)$, получаем

$$(0 + 1 \cdot 3 + 2 \cdot 2 + 3 \cdot 1 + 0 + 1 \cdot 2 + 2 \cdot 1 + 0)/8 = 14/8 = 1{,}75.$$

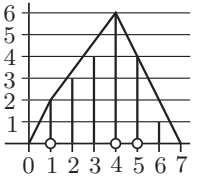


Рис. 21

8. Аналогично подсчету для стратегии «Все или ничего» из рис. 21 находим

$$\Big(0 + 2 + \Big[2 \cdot \frac{2}{3} + 6 \cdot \frac{1}{3}\Big] + \Big[2 \cdot \frac{1}{3} + 6 \cdot \frac{2}{3}\Big] + \\ +6 + 4 + \Big[4 \cdot \frac{1}{2} + 0 \cdot \frac{1}{2}\Big] + 0\Big)/8 = 22/8.$$

Удалось ли вам набрать за 4 моделирования больше 11?

Часть IV
ОДНОРОДНОСТЬ ВЫБОРОК

Нужно различать понимающих и соглашающихся. Понявший Учение не замедлит применить его к жизни. Согласившийся будет кивать головой и превозносить Учение как замечательную мудрость, но не применит эту мудрость в жизни. Согласившихся много, но они, как сухой лес, бесплодны и без тени, только тление ожидает их. Понявших мало, но они, как губка, впитывают драгоценное знание.

Е. И. Рерих, «Путями духа»

В эту часть книги включены наиболее простые и полезные методы статистической обработки данных. Материал и стиль его изложения во многом почерпнут из книги [88] — одного из лучших, по мнению автора, руководств по непараметрической статистике для исследователей и практиков — экономистов, социологов, биологов и специалистов в других областях, использующих статистические методы.

Руководство [88] особенно ценно тем, что около половины его объема занимают таблицы, позволяющие вычислять фактические, а не асимптотические, вероятности ошибок при обработке выборок небольшого размера, которые часто встречаются в прикладных исследованиях.

Один из моих друзей определил практика как человека, ничего не понимающего в теории, а теоретика — как мечтателя, вообще не понимающего ничего.

Л. Больцман

Великая цель образования — это не знания, а действия.

Г. Спенсер

Глава 14

ДВЕ НЕЗАВИСИМЫЕ ВЫБОРКИ

§ 1. АЛЬТЕРНАТИВЫ ОДНОРОДНОСТИ

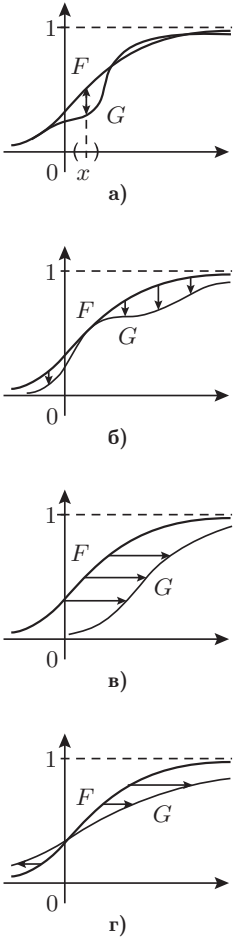


Рис. 1

Данные. Два набора наблюдений $x_1, \ldots, x_n$ и $y_1, \ldots, y_m$ будем рассматривать как реализовавшиеся значения случайных величин $X_1, \ldots, X_n$ и $Y_1, \ldots, Y_m$.

На протяжении всей главы будем считать выполненными

Допущения

Д1. *Случайные величины $X_1, \ldots, X_n$ независимы и имеют общую функцию распределения $F(x)$.*

Д2. *Случайные величины $Y_1, \ldots, Y_m$ независимы и имеют общую функцию распределения $G(x)$.*

Д3. *Обе функции F и G неизвестны, но принадлежат множеству Ω_c всех непрерывных функций распределения.*

Нас будет интересовать

Гипотеза однородности

$$H_0\colon G(x) = F(x) \text{ при всех } x.^{*)}$$

В качестве гипотез, конкурирующих с H_0, выделим следующие **альтернативы** (рис. 1):

а) *неоднородности* $H_1\colon G(x) \neq F(x)$ при некотором x (а в силу непрерывности — и в некоторой окрестности точки x);

б) *доминирования* $H_2\colon G(x) \leqslant F(x)$ при всех x, причем хотя бы для одного x неравенство строгое (говорят, что случайная величина Y_1 *стохастически больше* случайной величины X_1, поскольку $\mathbf{P}(Y_1 \geqslant x) \geqslant \mathbf{P}(X_1 \geqslant x)$ при каждом x);

в) *правого сдвига* $H_3\colon G(x) = F(x-\theta)$, где параметр $\theta > 0$ (эта альтернатива — частный случай предыдущей);

г) *масштаба* $H_4\colon G(x) = F(x/\theta)$, где $0 < \theta \neq 1$.

*) Формально H_0 представляет собой сложную непараметрическую гипотезу (см. § 1 гл. 13): в пространстве $\Omega_c \times \Omega_c$ она задает «диагональ» $\{(F,G)\colon G = F\}$.

Причины, по которым следует рассматривать конкурирующие гипотезы, отличные от H_1, таковы:

1) с практической точки зрения бывает важно уловить отклонения от H_0 только определенного вида (скажем, наличие систематического прироста у y_j по сравнению с x_i);

2) за счет сужения (по сравнению с H_1) множества пар распределений (F,G), составляющих альтернативное подмножество, обычно удается построить более эффективные (чувствительные) критерии, настроенные на обнаружение отклонений от H_0 конкретного вида.

Альтернатива доминирования H_2 встретится в § 3 и § 5. В гл. 15 приведены два полезных критерия, применяемых против альтернативы правого сдвига H_3. Методы анализа альтернативы масштаба H_4 (и ее обобщения, когда присутствует неизвестный «мешающий» параметр сдвига) изложены в гл. 24.

§ 2. ПРАВИЛЬНЫЙ ВЫБОР МОДЕЛИ

При проверке гипотезы однородности двух наборов данных $x_1, \ldots, x_n$ и $y_1, \ldots, y_m$ важно понять, с каким из *двух случаев* мы имеем дело: двумя реализациями независимых между собой выборок или парными повторными наблюдениями.

Примером первого случая может служить определение влияния удобрения на размер растений. Здесь $x_1, \ldots, x_n$ обозначают размеры растений на грядке, где удобрение не применялось, $y_1, \ldots, y_m$ — на соседней грядке, где оно применялось (см. пример 1 гл. 8). В этой ситуации можно предположить *независимость выборок* $X_1, \ldots, X_n$ и $Y_1, \ldots, Y_m$. Формально это выражает допущение

Д4. *Все компоненты случайного вектора* $(X_1, \ldots, X_n, Y_1, \ldots, Y_m)$ *независимы* (см. § 3 гл. 1).

Пример второго случая — исследование эффективности определенного воздействия (лекарства) на величину измеряемого показателя (скажем, артериального давления), где $x_1, \ldots, x_n$ — это значения показателя (у каждого из n наблюдаемых больных) *до воздействия*, а $y_1, \ldots, y_n$ — *после воздействия* ($m = n$). Для каждого фиксированного i ($i = 1, \ldots, n$) числам x_i и y_i в вероятностной модели Д1—Д3 соответствуют случайные величины X_i и Y_i, которые нельзя считать независимыми, так как x_i и y_i относятся к одному и тому же человеку.

Статистические методы, применимые ко второму случаю, рассматриваются в гл. 15. Конечно, их можно использовать и для независимых между собой выборок, отбросив, если $m \neq n$, лишние наблюдения в одной из реализаций (их надо отбирать случайно, скажем, с помощью таблицы Т1). Однако при этом игнорируется важная информация о совместной независимости, что снижает

чувствительность методов по сравнению с критериями, рассматриваемыми в настоящей главе.

В свою очередь, использование приведенных в этой главе критериев для данных, относящихся ко второму случаю, представляет собой *грубую методическую ошибку*, нередко допускаемую неопытными прикладниками, которые пытаются проверить однородность своих наблюдений при помощи первого попавшегося метода.

Не все йогурты одинаково полезны.

Из телерекламы.

Рассмотрим три критерия проверки гипотезы однородности в предположении справедливости допущений Д1—Д4.

§ 3. КРИТЕРИЙ СМИРНОВА

Н. В. Смирнов (1900–1966), русский математик.

Для проверки гипотезы однородности H_0 против альтернативы неоднородности H_1 используется критерий Смирнова , статистикой которого служит величина

$$D_{n,m} = \sup_x \left| \widehat{F}_n(x) - \widehat{G}_m(x) \right|,$$

где $\widehat{F}_n(x) = \dfrac{1}{n} \sum\limits_{i=1}^{n} I_{\{X_i \leqslant x\}}, \quad \widehat{G}_m(x) = \dfrac{1}{m} \sum\limits_{j=1}^{m} I_{\{Y_j \leqslant x\}},$

т. е. $D_{n,m}$ — расстояние в равномерной метрике между эмпирическими функциями выборок (рис. 2).

Слишком большое расстояние противоречит гипотезе H_0. В [10, с. 350] приведена таблица критических значений $D_{n,m}$ для $n, m \leqslant 20$ и уровней значимости 1, 2, 5, 10%.

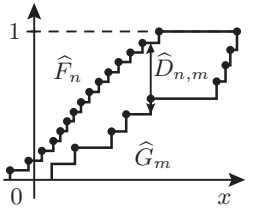


Рис. 2

Для нахождения значения статистики на реализациях $x_1, \ldots, x_n$ и $y_1, \ldots, y_m$ можно либо построить графики функций $\widehat{F}_n$ и $\widehat{G}_m$ и визуально определить их наибольшее расхождение, либо произвести вычисления на компьютере согласно формулам

$$D_{n,m} = \max\{D^+_{n,m}, D^-_{n,m}\},$$

где

$$D^+_{n,m} = \sup_x \left(\widehat{F}_n(x) - \widehat{G}_m(x)\right) = \max_{1 \leqslant i \leqslant n} \left\{ \frac{i}{n} - \widehat{G}_m(X_{(i)}) \right\},$$

$$D^-_{n,m} = \sup_x \left(\widehat{G}_m(x) - \widehat{F}_n(x)\right) = \max_{1 \leqslant j \leqslant m} \left\{ \frac{j}{m} - \widehat{F}_n(Y_{(j)}) \right\}.$$

Здесь $X_{(1)} \leqslant \ldots \leqslant X_{(n)}$ и $Y_{(1)} \leqslant \ldots \leqslant Y_{(m)}$ — упорядоченные по возрастанию элементы каждой из выборок.

Н. В. Смирнов в 1939 г. доказал, что если гипотеза H_0 верна, то при выполнении допущений Д1—Д4 имеет место сходимость

$$\mathbf{P}\left(\sqrt{nm/(n+m)}\, D_{n,m} \leqslant x \right) \to K(x) \text{ при } n, m \to \infty, \tag{1}$$

где $K(x)$ — функция распределения Колмогорова, определенная в § 2 гл. 12 (там же приведена небольшая таблица значений этой функции). Доказательство сходимости (1) при условии

Д5. *Размеры* $n, m \to \infty$ *так, что* $n/(n+m) \to \gamma \in (0,1)$

можно найти в [11, с. 428]. Контрпример в задаче 3 показывает, что условие непрерывности Д3 необходимо. Данное приближение является довольно точным уже при $n, m \geqslant 20$ (см. [32, с. 108]).

Расстояние от пункта A до пункта B равно 1 км. Пусть n — скорость движения из A в B, а m — скорость движения на обратном пути. Тогда $2/(1/n + 1/m) = 2nm/(n + m)$ — средняя скорость. Эту величину называют *средним гармоническим* чисел n и m.

На рис. 3 изображена верхняя половина окружности с центром в точке O, построенная на диаметре PR длины $n + m$, $|PS| = n$, $|SR| = m$. Перпендикуляр ST к диаметру PR пересекает окружность в точке T. При этом $a = |OT| = (n+m)/2$ — *среднее арифметическое* n и m. Из подобия $\triangle PTS$ и $\triangle TRS$ следует, что $b = |ST| = \sqrt{nm}$ — *среднее геометрическое*.

Вопрос 1. Как на этом рисунке построить отрезок длины $c = 2nm/(n+m)$ так, чтобы стало очевидным неравенство $a \geqslant b \geqslant c$?

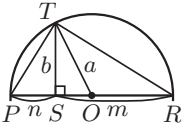


Рис. 3

Величина под корнем в формуле (1) представляет собой половину среднего гармонического n и m. Почему половину? Дело в том, что $\widehat{F}_n - \widehat{G}_m = (\widehat{F}_n - F) + (G - \widehat{G}_m)$, если $G = F$. При сложении независимых случайных величин их дисперсии складываются (П2). Поэтому для фиксированного x дисперсия отклонения $\widehat{F}_n(x) - \widehat{G}_m(x)$ при $m = n$ будет в 2 раза больше дисперсии отклонения $\widehat{F}_n(x) - F(x)$.

Замечание 1. В случае *альтернативы доминирования* (см. § 1) вместо критерия Смирнова надо применять *односторонний критерий*, основанный на следующей предельной теореме для определенной выше статистики $D^+_{n,m}$: при справедливости гипотезы H_0 для любого $x \geqslant 0$ имеет место сходимость

$$\mathsf{P}\left(\sqrt{nm/(n+m)}\, D^+_{n,m} \leqslant x\right) \to 1 - e^{-2x^2} \quad \text{при } n,\, m \to \infty. \tag{2}$$

(Для случая $m = n$ эта сходимость будет установлена в § 6.)

Согласно § 2 гл. 12 для правого «хвоста» распределения Колмогорова справедливо разложение

$$1 - K(x) = 2\left[e^{-2x^2} - e^{-8x^2} + e^{-18x^2} - \dots\right].$$

Второй член заключенного в квадратные скобки ряда представляет собой четвертую степень его первого члена. Пренебрегая им и всеми последующими членами, из сравнения сходимостей (1) и (2) видим, что фактический уровень значимости (см. § 1 гл. 12) данного критерия примерно вдвое меньше, чем у критерия Смирнова.

§ 4. КРИТЕРИЙ РОЗЕНБЛАТТА

Для проверки гипотезы однородности H_0 двух выборок против *альтернативы неоднородности* H_1 (см. § 1) можно воспользоваться

также критерием типа ω^2 из § 2 гл. 12. Статистика этого критерия задается формулой

$$\omega^2_{n,m} = \int\limits_{-\infty}^{\infty} [\widehat{F}_n(x) - \widehat{G}_m(x)]^2 \, d\widehat{H}_{n+m}(x),$$

где $\widehat{H}_{n+m}(x) = \dfrac{n}{n+m}\,\widehat{F}_n(x) + \dfrac{m}{n+m}\,\widehat{G}_m(x)$ представляет собой эмпирическую функцию, построенную по объединенной выборке $(X_1,\ldots,X_n,Y_1,\ldots,Y_m)$.

Согласно [10, с. 86], статистика $\omega^2_{n,m}$ зависит лишь от порядковых номеров (*рангов*) выборочных элементов:

$$\omega^2_{n,m} = \frac{1}{nm}\left[1/6 + \frac{1}{m}\sum_{i=1}^{n}(R_i - i)^2 + \frac{1}{n}\sum_{j=1}^{m}(S_j - j)^2\right] - 2/3,$$

где R_i — ранг $X_{(i)}$, а S_j — ранг $Y_{(j)}$ в объединенном вариационном ряду (см. § 4 гл. 4).

Положим для краткости $Z = Z_{n,m} = \dfrac{nm}{n+m}\,\omega^2_{n,m}$. В 1952 г. М. Розенблатт доказал, что при условии справедливости гипотезы H_0 и выполнении допущений Д1—Д5 имеет место сходимость

$$\mathbf{P}(Z \leqslant x) \to A_1(x), \tag{3}$$

где предельный закон A_1 тот же самый, что встречался в § 2 гл. 12. Математическое ожидание и дисперсия этого закона равны, соответственно, 1/6 и 1/45, в то время как

$$\mathbf{M}Z = \frac{1}{6}\left(1 + \frac{1}{n+m}\right),$$

$$\mathbf{D}Z = \frac{1}{45}\left(1 + \frac{1}{n+m}\right)\left[1 + \frac{1}{n+m} - \frac{3}{4}\left(\frac{1}{n} + \frac{1}{m}\right)\right].$$

Поэтому при вычислении приближенных критических значений рекомендуется вместо Z в формуле (3) использовать статистику $Z^* = (Z - \mathbf{M}Z)/\sqrt{45\,\mathbf{D}Z} + 1/6$.*) Это обеспечивает удовлетворительную точность приближения уже для n, $m \geqslant 7$.

§ 5. КРИТЕРИЙ РАНГОВЫХ СУММ УИЛКОКСОНА

Критерий ранговых сумм Уилкоксона применяется для проверки гипотезы однородности H_0 против *альтернативы доминирования* H_2 (см. § 1), в частности, — против альтернативы правого сдвига H_3.

*) Очевидно, $\mathbf{M}Z^* = 1/6$, $\mathbf{D}Z^* = 1/45$, причем из приведенных выше формул для $\mathbf{M}Z$ и $\mathbf{D}Z$ из (3) с учетом свойств сходимости (П5) следует, что $\mathbf{P}(Z^* \leqslant x) \to A_1(x)$.

Вычислим статистику V *критерия ранговых сумм Уилкоксона.*

1. Обозначим через S_j ранг порядковой статистики $Y_{(j)}$ $(j = 1,\ldots,m)$ в вариационном ряду, построенном по объединенной выборке $(X_1,\ldots,X_n,Y_1,\ldots,Y_m)$ (рис. 4).
2. Положим $V = S_1 + \ldots + S_m$.

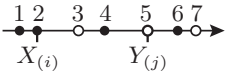


Рис. 4

Критерий, основанный на статистике V, был предложен Ф. Уилкоксоном в 1945 г. для выборок одинакового размера и распространен на случай $m \neq n$ Х. Манном и Д. Уитни в 1947 г.

Суть критерия сводится к следующему: если верна гипотеза H_0, то значения $Y_{(j)}$ должны быть рассеяны по всему вариационному ряду; напротив, достаточно большое значение V указывает на тенденцию преобладания Y_j над X_i, что свидетельствует в пользу справедливости гипотезы H_2. Таким образом, критическая область выбирается в виде луча $\{V > c\}$, где c — некоторая константа.

Малые выборки. Критические значения статистики V для $n, m \leqslant 25$ приведены в таблице [10, с. 357].

Большие выборки. Рассмотрим статистику

$$U = \sum_{i=1}^{n} \sum_{j=1}^{m} I_{\{X_i < Y_j\}}. \tag{4}$$

Вопрос 2. Верно ли, что все слагаемые $I_{\{X_i<Y_j\}}$ в сумме (4) независимы?

При отсутствии совпадений среди X_i и Y_j справедливо равенство (см. задачу 4)

$$U = V - m(m+1)/2, \tag{5}$$

и, следовательно, критерии, основанные на V и U, эквивалентны. Предложенная Уилкоксоном ранговая форма V удобнее для вычислений. С другой стороны, с помощью *считающей формы* U, изученной Манном и Уитни, нетрудно установить (задача 5), что в случае справедливости гипотезы H_0 имеем:

В [86, с. 143] сообщается, что У. Краскел нашел статистику U в работе Г. Дехлера, опубликованной в Германии в 1914 г.

$$\mathbf{M}U = nm/2, \quad \mathbf{D}U = nm(n+m+1)/12. \tag{6}$$

Когда гипотеза H_0 верна и выполнены условия Д1—Д5, имеет место сходимость

$$U^* = (U - \mathbf{M}U)/\sqrt{\mathbf{D}U} \xrightarrow{d} Z \sim \mathcal{N}(0,1). \tag{7}$$

Доказательство этого результата можно найти в [86, с. 145].

Поправка. К сожалению, нормальное приближение (7) не обеспечивает достаточную точность при $n, m \leqslant 50$. Например, при $25 \leqslant n, m \leqslant 50$ (см. [88, с. 87]) в 40% случаев истинные критические точки для статистики V отличаются от точек, полученных на основе сходимости (7), более чем на 1. Существенно точнее следующая аппроксимация, предложенная Р. Иманом в 1976 г. Она использует полусумму нормальной и стьюдентовской квантилей. Положим $N = n + m$. Критическим α-значением статистики

$$\tilde{U}^* = \frac{1}{2}\, U^* \left[1 + \sqrt{(N-2)/\left(N - 1 - (U^*)^2\right)}\right] \tag{8}$$

служит $z_\alpha = (x_{1-\alpha} + y_{1-\alpha})/2$, где $x_{1-\alpha}$ и $y_{1-\alpha}$ обозначают, соответственно, квантили уровня $(1-\alpha)$ закона $\mathcal{N}(0,1)$ и распределения Стьюдента с $(N-2)$ степенями свободы (см. таблицы Т2 и Т4). Таким образом, если наблюдаемое значение статистики $\tilde{U}^*$ окажется больше или равно z_α, то гипотеза H_0 отвергается.

Совпадения. Когда среди $n+m$ наблюдений есть одинаковые, статистику V следует вычислять с учетом средних рангов.*) При подсчете U это соответствует назначению веса $1/2$ нулевой разности $Y_j - X_i$. В приближении (7) надо заменить $\mathbf{D}U$ на

$$\frac{nm}{12}\left[n+m+1-\frac{1}{(n+m)(n+m-1)}\sum_{k=1}^{g} l_k(l_k^2-1)\right], \tag{9}$$

где g — число групп совпадений среди всех $n+m$ наблюдений, l_k — количество элементов в k-й группе. Наблюдения, не совпадающее с другими, рассматриваются как группы размера 1.

Оценка параметра сдвига. Для альтернативы правого сдвига H_3 в качестве оценки параметра θ можно взять

$$\widehat{\theta} = MED\,\{Y_j - X_i,\ 1 \leqslant i \leqslant n, 1 \leqslant j \leqslant m\}. \tag{10}$$

Известно (см. [86, с. 171]), что при выполнении условий Д1—Д5 имеет место сходимость

$$\sqrt{nm/(n+m)}\,(\widehat{\theta}-\theta) \xrightarrow{d} \xi \sim \mathcal{N}\left(0, 1/E(F)\right),$$

$$\text{где } E(F) = 12\left(\int p^2(x)\,dx\right)^2, \tag{11}$$

$p(x)$ — плотность, отвечающая функции распределения F. (Отметим, что величина $E(F)$ ранее встречалась в теореме 3 гл. 8.)

Доверительный интервал. Описание способа построения доверительного интервала для параметра θ в случае малых выборок приведено в [88, с. 96].

При больших n и m приближенный доверительный интервал с коэффициентом доверия $(1-2\alpha)$ образует пара порядковых статистик $(W_{(k_\alpha+1)}, W_{(nm-k_\alpha)})$. Здесь $W_{(1)} \leqslant \ldots \leqslant W_{(nm)}$ — упорядоченные по возрастанию разности $Y_j - X_i$ $(1 \leqslant i \leqslant n,\ 1 \leqslant j \leqslant m)$; k_α — целая часть числа

$$nm/2 - 0{,}5 - x_{1-\alpha}\sqrt{mn(n+m+1)/12},$$

$x_{1-\alpha}$ обозначает $(1-\alpha)$-квантиль распределения $\mathcal{N}(0,1)$; $0{,}5$ — поправка на непрерывность, происхождение которой объясняется в § 2 гл. 15.

*) Пусть, например, наименьшие 4 значения совпадают. Тогда всем им приписывается средний ранг $(1+2+3+4)/4 = 2{,}5$.

Численный пример применения критерия ранговых сумм Уилкоксона—Манна—Уитни содержится в задаче 2.*)

Комментарии

1. Как доказано в [86, с. 167], критерий ранговых сумм состоятелен против альтернативы доминирования H_2, в частности, против альтернативы правого сдвига H_3.

2. Распределение случайной величины $V = S_1 + \ldots + S_m$ можно найти, пользуясь тем, что при справедливости гипотезы H_0 вероятность каждого из C_{n+m}^m возможных сочетаний $S_1, \ldots, S_m$ (соответствующих расстановкам Y_j, $j = 1, \ldots, m$, по $n+m$ местам) одна и та же.

Вопрос 3. Чему равна $\mathbf{P}(V \geqslant 8)$ для $n=3$ и $m=2$?

3. Покажем, как оценка $\widehat{\theta}$, определяемая равенством (10), связана со статистикой U. Ввиду формулы (4) при отсутствии совпадений, U равна числу положительных разностей $Y_j - X_i$. Естественной оценкой параметра θ будет такая величина θ', чтобы наборы $(Y_1' = Y_1 - \theta', \ldots, Y_m' = Y_m - \theta')$ и $(X_1, \ldots, X_n)$ выглядели как выборки из одного и того же закона. Для таких выборок распределение статистики U симметрично относительно среднего $nm/2$. Таким образом, приходим к следующему уравнению относительно θ':

$$\sum_{i=1}^{n}\sum_{j=1}^{m} I_{\{Y_j' - X_i > 0\}} = \sum_{i=1}^{n}\sum_{j=1}^{m} I_{\{Y_j - X_i > \theta'\}} = nm/2.$$

Когда величина θ' становится равной $\widehat{\theta}$ из формулы (10), происходит «перескок» через уровень $nm/2$.

4. Точный доверительный интервал для малых выборок строится с помощью метода 1 из § 3 гл. 11, примененного к

$$g(\boldsymbol{x}, \boldsymbol{y}, \theta) = \sum_{i=1}^{n}\sum_{j=1}^{m} I_{\{y_j - x_i > \theta\}}.$$

Когда известно, что наблюдения имеют *нормальное* распределение (см. § 4 гл. 12), для проверки однородности можно использовать критерии из примера 1.

Пример 1. ***Однородность нормальных выборок.*** Проверим однородность двух *независимых* выборок $(X_1, \ldots, X_n)$ и $(Y_1, \ldots, Y_m)$, где $X_i \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $Y_j \sim \mathcal{N}(\mu_2, \sigma_2^2)$, причем все параметры μ_1, μ_2, σ_1, σ_2 неизвестны. Несмещенными оценками для дисперсий σ_1^2 и σ_2^2 служат

$$S_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \quad \text{и} \quad S_2^2 = \frac{1}{m-1}\sum_{j=1}^{m}(Y_j - \overline{Y})^2$$

(см. пример 3 гл. 6). В силу теоремы 1 гл. 11 $(n-1)\,S_1^2/\sigma_1^2 \sim \chi_{n-1}^2$, $(m-1)\,S_2^2/\sigma_2^2 \sim \chi_{m-1}^2$, причем S_1 не зависит от $\overline{X}$, а ввиду независимости выборок — также и от $\overline{Y}$. Это же верно и для S_2.

*) Обобщение критерия для многомерных данных см. в § 3 гл. 23.

Вопрос 4. Что происходит с законом F_{k_1,k_2} при $k_1,\ k_2 \to \infty$?

Определение. Случайная величина ζ имеет F-распределение (Фишера—Снедекора) с k_1 и k_2 степенями свободы (обозначается $\zeta \sim F_{k_1,k_2}$), если

$$\zeta = \left(\frac{1}{k_1}\,\xi\right) \Big/ \left(\frac{1}{k_2}\,\eta\right), \quad \text{где } \xi \sim \chi^2_{k_1},\ \eta \sim \chi^2_{k_2},\ \xi \text{ и } \eta \text{ независимы.}$$

Критерий Фишера. Если верна гипотеза

$H'\colon \sigma_1 = \sigma_2, \mu_1$ и μ_2 — любые,

то в соответствии с приведенным выше определением статистика S_1^2/S_2^2 распределена по закону $F_{n-1,m-1}$. Ее критические значения можно найти в таблице Т5.

В случае, когда критерий Фишера не отвергает гипотезу H', для проверки однородности остается проверить гипотезу $H''\colon \mu_1 = \mu_2$.

Обозначим неизвестную общую дисперсию через σ^2 . Так как распределение хи-квадрат является частным случаем гамма-распределения ($\chi^2_k \sim \Gamma(k/2,1/2)$), из леммы 1 гл. 4 вытекает, что

$\sigma^{-2}\left[(n-1)\,S_1^2 + (m-1)\,S_2^2\right] \sim \chi^2_{n+m-2}$.

Поскольку математическое ожидание закона χ^2_{n+m-2} равно $n+m-2$, статистика $S^2_{tot} = \left[(n-1)\,S_1^2 + (m-1)\,S_2^2\right]/(n+m-2)$ несмещенно оценивает σ^2 по объединенной выборке.

Total (*англ.*) — общий.

При справедливости гипотезы H'' ввиду независимости выборок имеем: $\overline{X} - \overline{Y} \sim \mathcal{N}(0, (1/n + 1/m)\,\sigma^2)$. При этом $\overline{X} - \overline{Y}$ (функция от $\overline{X}$ и $\overline{Y}$) не зависит от S_{tot} (функции от S_1 и S_2) в силу леммы о независимости из § 3 гл. 1. Отсюда согласно определению закона Стьюдента t_k с k степенями свободы (см. пример 4 гл. 11) имеем:

$$T = \left(\overline{X} - \overline{Y}\right) \Big/ \left(S_{tot}\sqrt{\frac{1}{n} + \frac{1}{m}}\right) = \sqrt{\frac{nm}{n+m}}\,\left(\overline{X} - \overline{Y}\right) \Big/ S_{tot} \sim t_{n+m-2}.$$

Это приводит к так называемому **критерию Стьюдента**, который позволяет проверить гипотезу H''. Критические значения статистики t_{n+m-2} даны в Т4.

Вопрос 5. Какое распределение имеет статистика T^2?

Несмотря на то, что критерий Стьюдента оптимален для нормальных выборок, рассмотренная процедура проверки однородности имеет скорее теоретическое, чем практическое значение. Почему?

Во-первых, это объясняется тем, что критические значения статистики S_1^2/S_2^2 *существенно изменяются* даже при небольших возмущениях модели (см. в гл. 16 задачу 6 и замечание 2 при $k = 2$).*)

Во-вторых, эффективность критерия Стьюдента быстро уменьшается при отклонении от строгой нормальности. (Относительная асимптотическая эффективность двух критериев при альтернативах правого сдвига определена, например, в [86, с. 76].) В частности,

*) Устойчивая ранговая альтернатива критерию Фишера, не предполагающая нормальности наблюдений, описывается в § 2 гл. 24.

эффективность критерия ранговых сумм Уилкоксона—Манна—Уитни по сравнению с критерием Стьюдента равна $E(F)$ (см. формулу (11)).

Рассмотрим для иллюстрации модель Тьюки смеси нормальных законов из примера 2 гл. 8 (при $\mu = 0$ и $\sigma = 1$), у которой функция распределения F выглядит так: $F_\varepsilon(x) = (1-\varepsilon)\,\Phi(x) + \varepsilon\,\Phi(x/3)$, где $\Phi(x)$ — функция распределения $\mathcal{N}(0,1)$, $0 \leqslant \varepsilon \leqslant 1$. Следующая таблица (из [86, с. 85]) показывает изменение эффективности $E(F_\varepsilon)$ в этой модели при небольшом утяжелении «хвостов».

ε	0	0,01	0,03	0,05	0,08	0,10	0,15
$E(F_\varepsilon)$	0,955	1,009	1,108	1,196	1,301	1,373	1,497

В силу теоремы 4 гл. 8 эффективность $E(F) = e_{W,\overline{X}}(F) \geqslant 0{,}864$ при всех $F \in \Omega_s$ и может быть сколь угодно велика.

Отметим также, что у критерия с $m \neq n$ по сравнению с критерием с $m' = n' = (n+m)/2$ эффективность уменьшается в $1/[4\gamma(1-\gamma)] > 1$ раз, $\gamma = n/(n+m)$ (см. [86, с. 171]), поэтому желательно брать выборки одинаковых размеров (если, конечно, есть такая возможность).

§ 6. ПРИНЦИП ОТРАЖЕНИЯ

Материал этого параграфа в основном заимствован из гл. III замечательной книги [81], которую автор настоятельно рекомендует прочитать заинтересовавшемуся читателю. В конце параграфа некоторые из полученных результатов будут использованы для решения двух задач из области проверки однородности выборок.

Рассмотрим случайное блуждание $S_n = \xi_1 + \ldots + \xi_n$, где независимые «шаги» ξ_i принимают значения $+1$ и -1 с одинаковой вероятностью $1/2$. *Траекторией (путем)* блуждания длины n будем называть ломаную, соединяющую точки плоскости с координатами (i, S_i), $i = 1, \ldots, n$. Каждый из 2^n возможных путей имеет одинаковую вероятность 2^{-n}.

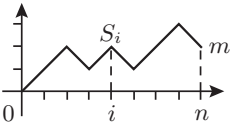


Рис. 5

Обозначим через $N_{n,m}$ количество путей, ведущих из точки $(0,0)$ в точку (n,m) (рис. 5). Пусть для такого пути k — это число шагов вверх ($\xi_i = +1$), l — число шагов вниз ($\xi_i = -1$). Тогда $k + l = n$ и $k - l = m$, откуда $k = (n+m)/2$. Расставить k «плюс единиц» по n местам можно C_n^k способами. Поэтому

$$N_{n,m} = C_n^{(n+m)/2}, \tag{12}$$

где подразумевается, что биномиальный коэффициент равен 0, если $(n+m)/2$ не является целым числом между 0 и n.

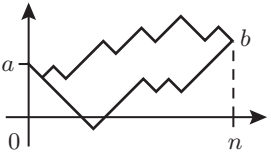


Рис. 6

Пусть a и b — положительные целые числа. Перенесем начальную ординату блуждания из 0 в a и потребуем, чтобы в момент n траектория приходила в точку с координатами (n,b) (рис. 6).

Очевидно, что количество таких путей равно $N_{n,b-a}$. Сколько из них являются *положительными*, т. е. лежат целиком над осью абсцисс? Ответ на этот вопрос получим с помощью следующего утверждения.

Принцип отражения. Число путей, ведущих из $(0,a)$ в (n,b) и касающихся или пересекающих ось абсцисс, совпадает с числом путей, ведущих из $(0,a)$ в $(n,-b)$, которое равно $N_{n,a+b}$.

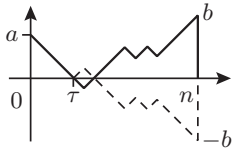


Рис. 7

Доказательство. Обозначим через τ момент первого касания или пересечения траекторией оси абсцисс (рис. 7). Отразим относительно этой оси отрезок пути от $(\tau,0)$ до (n,b). Присоединив к нему отрезок исходного пути от $(0,a)$ до $(\tau,0)$, построим новый путь, ведущий из $(0,a)$ в $(n,-b)$. Очевидно, что по построенному пути исходный восстанавливается однозначно. ■

Следствие. Количество положительных траекторий из $(0,a)$ в (n,b) равно $N_{n,b-a} - N_{n,a+b}$.

Вопрос 6.
В каком месте на дороге надо приземлиться вороне, чтобы общая длина двух перелетов была минимальной?

Геометрическая задача. По всей ширине дороги рассыпано зерно. Слева от дороги на заборе высоты a сидит ворона (рис. 8). Она хочет поклевать зерна, а затем перелететь на забор высоты b (потому, что $b > a$), который находится справа от дороги.

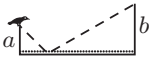


Рис. 8

Частным случаем следствия принципа отражения является следующий результат, полученный У. Уитвортом в 1878 г. и заново Ж. Бертраном в 1887 г. (см. [81, с. 87]).

Задача о баллотировке. Предположим, что на выборах первый кандидат набрал k голосов, а второй кандидат набрал l голосов, причем $k > l$. Тогда вероятность того, что при последовательном подсчете голосов первый кандидат все время был впереди второго, равна $(k-l)/(k+l)$.

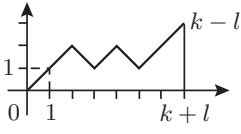


Рис. 9

Доказательство. Траектория, удовлетворяющая условиям теоремы, обязательно проходит через точку (1,1) (рис. 9). Согласно следствию принципа отражения и формуле (12), число положительных путей из (1,1) в $(k+l,k-l)$ равно $(n=k+l-1,\ a=1,\ b=k-l)$

$$N_{n,k-l-1} - N_{n,k-l+1} = C^{k-1}_{k+l-1} - C^{k}_{k+l-1}.$$

Правая часть простыми преобразованиями приводится к виду $N_{k+l,k-l}(k-l)/(k+l)$, что и утверждалось. ■

Среди путей за время $2n$ выделим пути, приходящие в $(2n,0)$:

$$u_{2n} \equiv \mathbf{P}(S_{2n}=0) = N_{2n,0}\cdot 2^{-2n} = C^n_{2n}\,2^{-2n} \quad (u_0=1). \tag{13}$$

Формула Стирлинга[*)] ($n! \sim \sqrt{2\pi n}\,n^n\,e^{-n}$ при $n\to\infty$) позволяет получить следующую асимптотику для u_{2n} (убедитесь!):

$$u_{2n} \sim 1/\sqrt{\pi n} \quad \text{при } n\to\infty. \tag{14}$$

[*)] Простое доказательство этой формулы можно найти в [81, с. 72].

Другими словами, доля среди всех 2^{2n} возможных траекторий путей, приходящих в точку $(2n, 0)$, стремится к нулю со скоростью порядка $1/\sqrt{n}$ (см. решение задачи 5 гл. 12).

Оказывается, что *неотрицательных* путей за время $2n$ ровно столько же, сколько путей, приходящих в $(2n, 0)$.

Теорема 1. При всех значениях n справедливо равенство $\mathbf{P}(S_1 \geqslant 0, \ldots, S_{2n} \geqslant 0) = u_{2n}$.

Доказательство. Э. Нелсон (см. [81, с. 115]) предложил следующее оригинальное преобразование, взаимно однозначно переводящее траектории, приходящие в 0, в неотрицательные.

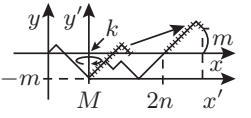


Рис. 10

Обозначим самую левую (если их несколько) точку глобального минимума заданного пути, ведущего в точку $(2n,0)$, через $M = (k,-m)$ (рис. 10). Отразим участок, ведущий из начала координат в точку M, относительно вертикальной прямой $y = k$ и передвинем отраженный участок так, чтобы его начальная точка совпала с точкой $(2n, 0)$. Примем M за начало новой системы координат, в которой построенный путь ведет из начала в точку $(2n, 2m)$, а все его вершины лежат не ниже новой оси абсцисс. ■

Указание. Обратите внимание, что у такого пути обязательно $S_1 = 1$, а также $S_{2n} > 1$, так как $2n$ — четное число.

Вопрос 7. Как вывести, что $\mathbf{P}(S_1>0,\ldots,S_{2n}>0)= \frac{1}{2}u_{2n}$?

Рассмотрим траектории, имеющие единственный глобальный максимум в момент $2n$ (рис. 11).

Траектория блуждания, не возвращающегося в 0 за время $2n$, является либо положительной, либо отрицательной (не считая начальной точки). Из симметрии и вопроса 7 вытекает, что

$$\mathbf{P}(A_{2n}) = \mathbf{P}(S_1 \neq 0, \ldots, S_{2n} \neq 0) = u_{2n}. \tag{15}$$

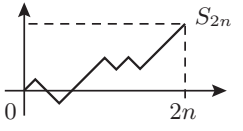


Рис. 11

Введем обозначения: $B_{2n} = \{S_{2n} = 0\}$, $f_{2n} = \mathbf{P}(A_{2n-2}B_{2n})$ — *вероятность вернуться в* 0 *впервые в момент* $2n$. Ввиду (15)

$$f_{2n} = \mathbf{P}(A_{2n-2}) - \mathbf{P}(A_{2n-2}\overline{B}_{2n}) = u_{2n-2} - u_{2n}. \tag{16}$$

Вопрос 8. Чему равна $\mathbf{P}(S_0<S_{2n}, S_1<S_{2n},\ldots, S_{2n-1}<S_{2n})$?

С учетом того, что $u_0 = 1$ и $u_{2n} \to 0$ при $n \to \infty$, из формулы (16) получаем, что *вероятность вернуться в* 0 *когда-нибудь* $f_2 + f_4 + \ldots = 1$. Из соотношений (13) и (16) легко выводится тождество

$$f_{2n} = \frac{1}{2n}\, u_{2n-2}. \tag{17}$$

Асимптотика (14) дает для f_{2n} порядок малости $n^{-3/2}$. Поэтому $\sum\limits_{n=1}^{\infty} 2n\, f_{2n} \sim c \sum\limits_{n=1}^{\infty} \frac{1}{\sqrt{n}} = \infty$, т. е. среднее время до первого возвращения блуждания в 0 бесконечно.

Пусть g_{2n-1} — *вероятность впервые достигнуть уровень* 1 *на* $(2n-1)$*-м шаге.*

Теорема 2. Имеет место равенство $g_{2n-1} = f_{2n}$.

Доказательство. Величина $K_n = 2^{2n-1} g_{2n-1}$ совпадает с числом отрицательных (за исключением крайних точек) путей, ведущих из $(1,-1)$ в $(2n,0)$ (рис. 12). Соединив точку $(1,-1)$ с началом координат, получим траекторию, впервые возвращающуюся в 0 в момент $2n$. Сопоставим исходному пути эту траекторию и симметричную к ней относительно оси абсцисс. Так как исходный путь по этим траекториям определяется однозначно, то $2^{2n} f_{2n} = 2K_n$. ■

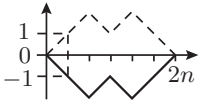


Рис. 12

Из теоремы 2 вытекает, что среднее время до момента первого достижения произвольного уровня $h > 0$ симметричным случайным блужданием бесконечно.

Обозначим через ν_{2n} *момент последнего попадания в* 0 *за время* $2n$ (рис. 13). Найдем распределение этой случайной величины. Положим $\alpha_{2i,2n} = \mathbf{P}(\nu_{2n} = 2i)$, $i = 0,1,\ldots,n$.

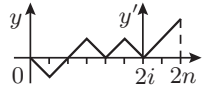


Рис. 13

Теорема 3. Справедливо соотношение $\alpha_{2i,2n} = u_{2i}\, u_{2n-2i}$.

Доказательство. Мы интересуемся путями, у которых $S_{2i} = 0$ и $S_{2i+1} \neq 0, \ldots, S_{2n} \neq 0$. Первые $2i$ вершин можно выбрать $2^{2i} u_{2i}$ различными способами. Взяв точку $(2i,0)$ в качестве нового начала координат, согласно равенству (15) видим, что остальные $(2n-2i)$ вершин можно выбрать $2^{2n-2i} u_{2n-2i}$ способами. Всего получаем $2^{2n} u_{2i}\, u_{2n-2i}$ вариантов. ■

Из теоремы 3 и асимптотики (14) следует, что предельным распределением для случайной величины $\nu_{2n}/(2n)$ является *распределение арксинуса* с функцией распределения $F(x) = (2/\pi) \arcsin \sqrt{x}$ и плотностью $p(x) = 1/(\pi\sqrt{x(1-x)})$, $0 < x < 1$ (график $p(x)$ приведен на рис. 8 в гл. 5). Действительно:

$$\mathbf{P}(\nu_{2n}/(2n) \leqslant x) = \sum_{i<xn} \alpha_{2i,2n} \sim \sum_{i<xn} p\left(\frac{i}{n}\right) \frac{1}{n} \to \int_0^x p(y)\, dy.$$

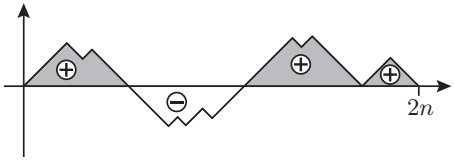


Рис. 14

Оказывается, что время, в течение которого траектория блуждания находилась в полуплоскости $y \geqslant 0$, распределено так же, как случайная величина ν_{2n} (см. § 4 гл. 16). Но если ограничиться только путями, приходящими в момент $2n$ в 0 (рис. 14), то условная вероятность того, что в точности $2i$ $(i = 0,1,\ldots,n)$ их звеньев лежат над осью абсцисс, равна $1/(n+1)$ независимо от i. Это утверждение известно как **теорема о равнораспределенности**.

Доказательство. Рассмотрим отдельно случай $i = n$. Число путей, приходящих в точку $(2n, 0)$, все звенья которых лежат выше оси абсцисс, совпадает с числом положительных путей из $(0, 2)$ в $(2n-1, 1)$. В силу следствия принципа отражения оно равно $C_{2n-1}^{n} - C_{2n-1}^{n+1} = \dfrac{1}{n+1}\, C_{2n}^{n}$. Это доказывает теорему при $i = n$ и, а в силу симметрии — также и при $i = 0$.

При $1 \leqslant i \leqslant n-1$ воспользуемся индукцией. Для случая $n = 1$ теорема очевидна. Предположим, что она верна для всех путей, длина которых меньше $2n$. Пусть первое возвращение в 0 произошло в момент $2r$. Участок пути до $2r$ расположен либо в положительной, либо в отрицательной полуплоскости. В первом случае $1 \leqslant r \leqslant i$ и участок после $2r$ имеет ровно $2i - 2r$ звеньев над осью абсцисс. Согласно предположению индукции и ввиду формулы (17), такой путь может быть выбран

$$\frac{1}{2}\, 2^{2r} f_{2r} \cdot \frac{2^{2n-2r}}{n-r+1}\; u_{2n-2r} = \frac{2^{2n-2}}{r(n-r+1)}\; u_{2r-2}\, u_{2n-2r} \tag{18}$$

различными способами. Во втором случае конечный участок длины $(2n - 2r)$ содержит ровно $2i$ положительных звеньев и, следовательно, $n - r \geqslant i$. Для фиксированного r число путей, удовлетворяющих этим условиям, также определяется величиной (18). Общее количество путей обоих типов получается суммированием слагаемых вида (18) по $1 \leqslant r \leqslant i$ и $1 \leqslant r \leqslant n - i$ соответственно. Во второй сумме заменим индекс r на $j = n - r + 1$. Тогда j меняется от $i + 1$ до n, а слагаемые имеют вид (18) с заменой r на j. Отсюда следует, что число путей, у которых i звеньев лежат в положительной полуплоскости, получается суммированием (18) по $1 \leqslant r \leqslant n$. Так как i не входит в (18), сумма не зависит от i, что и утверждалось. ■

Применим изложенные результаты к некоторым задачам, возникающим при проверке гипотезы однородности H_0 двух независимых выборок $\boldsymbol{X} = (X_1, \ldots, X_n)$ и $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ одинакового размера n. Допустим, что среди всех $2n$ значений обеих выборок нет совпадающих (условие Д3 гарантирует выполнение этого с вероятностью 1). Упорядочим каждую выборку по возрастанию: $X_{(1)} < \ldots < X_{(n)}$ и $Y_{(1)} < \ldots < Y_{(n)}$. Положим L равным числу индексов i, при которых $X_{(i)} < Y_{(i)}$ $(i = 1, \ldots, n)$. Близость этой величины к n указывает на то, что альтернатива доминирования H_2 (см. § 1) предпочтительнее гипотезы однородности H_0. Чтобы вычислить критическую границу, надо найти распределение случайной величины L при условии справедливости гипотезы H_0.

Для этого переведем задачу на язык случайных блужданий. Упорядочим по возрастанию все $2n$ значений обеих выборок в вариационный ряд. Пусть $\xi_k = -1$ или $+1$ в зависимости от того,

элементом выборки $\boldsymbol{X}$ или выборки $\boldsymbol{Y}$ является k-й член построенного ряда ($k = 1, \ldots, 2n$), $S_k = \xi_1 + \ldots + \xi_k$. Полный путь длины $2n$ соединяет начало координат с точкой $(2n,0)$.

Заметим, что событие $X_{(k)} < Y_{(k)}$ происходит тогда и только тогда, когда S_{2k-1} содержит по меньшей мере k «плюс единиц», т. е. когда $S_{2k-1} > 0$. Это влечет неравенство $S_{2k} \geqslant 0$, и поэтому $(2k-1)$-е и $2k$-е звенья траектории лежат выше оси абсцисс. Отсюда следует, что равенство $L = l$ верно в том и только в том случае, когда в точности $2l$ звеньев лежат выше оси абсцисс. По теореме о равнораспределенности вероятность этого события равна $1/(n+1)$ независимо от l.

Статистика L впервые была использована для проверки однородности Ф. Гальтоном при исследовании данных, предоставленных ему Чарльзом Дарвином. Значения l и n были равны 13 и 15 соответственно. Не зная реальных вероятностей, Гальтон отверг гипотезу H_0. Однако в предположении H_0 вероятность того, что L примет значение 13 и более (фактический уровень значимости критерия, определенный в § 1 гл. 12), равна 3/16. Другими словами, в 3 случаях из 16 будет наблюдаться такое же или большее превосходство элементов $\boldsymbol{Y}$ над элементами $\boldsymbol{X}$ при полном совпадении законов распределения элементов выборок. У. Феллер пишет ([81, с. 88]): «Это показывает, что численный анализ может быть полезным дополнением к нашей не совсем надежной интуиции.»

В заключение получим с помощью принципа отражения предельную теорему (2) для статистики одностороннего критерия $D^+_{n,m} = \sup\limits_x (\widehat{F}_n(x) - \widehat{G}_m(x))$ при $m = n$.

Для этого заметим, что для любого $l (l = 1, \ldots, n)$

$$\left\{D^+_{n,n} \geqslant l/n\right\} \Longleftrightarrow \left\{\max_{1 \leqslant k \leqslant 2n} S_k \geqslant l,\ S_{2n} = 0\right\}.$$

В силу принципа отражения (рис. 15), а также формул (12) и (13) имеем

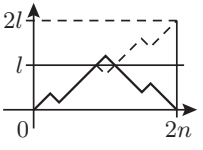


Рис. 15

$$\mathsf{P}\left(D^+_{n,n} \geqslant \frac{l}{n}\right) = \frac{N_{2n,2l}}{N_{2n,0}} = \frac{C^{n-l}_{2n}}{C^n_{2n}}.$$

Множитель $\sqrt{nm/(n+m)}$ в утверждении (2) при $m = n$ равен $\sqrt{n/2}$. Чтобы установить для произвольного $x \geqslant 0$, что при $n \to \infty$

$$\mathsf{P}(\sqrt{n/2}\, D^+_{n,n} \geqslant x) \to e^{-2x^2}, \tag{19}$$

остается применить формулу Стирлинга и разложение логарифма $\ln(1+t) = t - t^2/2 + o(t^2)$ при $t \to 0$ (задача 6).

ЗАДАЧИ

Ум заключается не только в знании, но и в умении применять знание на деле.

Аристотель

1. Пусть $x_1, \ldots, x_{20}$ — реализация выборки из равномерного распределения на отрезке $[0,1]$, построенная по четвертому столбцу таблицы Т1 (см. решение задачи 1 гл. 12); $y_1, \ldots, y_{20}$ — реализация выборки из закона с функцией распределения $G(x) = x^3$

на $[0,1]$ (согласно задаче 3 гл. 1 ее можно моделировать, взяв в качестве y_i наибольшее из трех первых значений в i-й строке табл. Т1, деленное на 100). Проверьте гипотезу однородности с помощью одностороннего критерия из замечания 1.

2. В условиях задачи 1 примените критерий ранговых сумм Уилкоксона—Манна—Уитни.

3. При выполнении условия Д5 найдите предельный закон распределения для статистики критерия Смирнова $\sqrt{nm/(n+m)}\,D_{n,m}$ в случае двух независимых выборок из распределения Бернулли с одинаковой вероятностью «успеха» p.
Указание. См. задачу 3 гл. 12.

4. Выведите формулу (5), связывающую величины U и V при отсутствии совпадений среди всех X_i и Y_j.

5.* Получите выражения (6) для $\mathbf{M}U$ и $\mathbf{D}U$.

6.* Докажите асимптотику (19).

РЕШЕНИЯ ЗАДАЧ

1. На рис. 16 приведены графики эмпирических функций распределения $\widehat{F}_n$ и $\widehat{G}_n$ для реализаций $x_1,\ldots,x_n$ и $y_1,\ldots,y_n$ $(n=20)$. Значение статистики $D^+_{n,n}$ равно 0,4. Отсюда получаем, что $x_0=\sqrt{n/2}\,D^+_{n,n}=0{,}4\sqrt{10}\approx 1{,}265$. Следовательно, фактический уровень значимости $\alpha_0=e^{-2x_0^2}=e^{-3{,}2}\approx 0{,}04$.

Леность всему (дурному) мать: что человек умеет, то позабудет, а чего не умеет, тому не научится.
Владимир Мономах

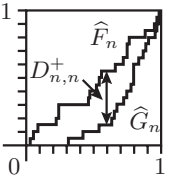


Рис. 16

2. Вычисленная на основе данных задачи 1 статистика V равна 489 (для совпадений были взяты средние ранги). В соответствии с формулой (5) находим $U=V-210=279$. Согласно (6) $\mathbf{M}U=200$. Подсчитанная по формуле (9) с учетом совпадений $\mathbf{D}U\approx 1365{,}2$. Поэтому нормированная статистика U^*, определенная равенством (7), принимает значение 2,138.

Вычислив поправку Имана (8), получим, что $\tilde{U}^*=2{,}192$. Для уровня значимости $\alpha=0{,}025$ по таблицам Т2 и Т4 (для $k=m+n-2=38$) находим критическое значение $z_\alpha=(1{,}96+2{,}024)/2\approx 1{,}992$. Поскольку $2{,}192>1{,}992$, гипотеза однородности H_0 отвергается на уровне 2,5%.

3. Эмпирическая функция распределения $\widehat{F}_n$ выборки $X_1,\ldots,X_n$ из закона Бернулли имеет два скачка: в точке 0 высоты $1-\overline{X}$ и в точке 1 высоты $\overline{X}$. Аналогично устроена функция $\widehat{G}_m$, построенная по выборке $Y_1,\ldots,Y_m$. Поэтому $D_{n,m}=\bigl|\overline{X}-\overline{Y}\bigr|$ (рис. 17). Пусть $q=1-p$. В силу условия Д5, центральной предельной теоремы и свойства 1 сходимости из П5

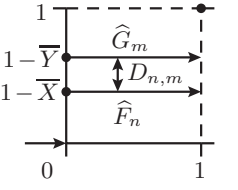


Рис. 17

$$\sqrt{\frac{m}{n+m}}\,\sqrt{n}\,(\overline{X}-p)\xrightarrow{d}\xi\sim\mathcal{N}(0,(1-\gamma)pq),$$

$$\sqrt{\frac{n}{n+m}}\,\sqrt{m}\,(p-\overline{Y})\xrightarrow{d}\eta\sim\mathcal{N}(0,\gamma pq).$$

Следовательно, ввиду независимости $\overline{X}$ и $\overline{Y}$ имеет место сходимость $\sqrt{nm/(m+n)}(\overline{X}-\overline{Y}) \xrightarrow{d} \xi+\eta=\zeta\sim\mathcal{N}(0,pq)$. Согласно свойству 3 сходимости (П5) искомым предельным законом является распределение случайной величины $|\zeta|$ (его плотность изображена на рис. 13 гл. 12).

4. Запишем ранг (т. е. номер в порядке возрастания) S_j статистики $Y_{(j)}$ в вариационном ряду, построенном по объединенной выборке $(X_1,\ldots,X_n,Y_1,\ldots,Y_m)$, в следующем виде:

$$S_j=\sum_{i=1}^{n} I_{\{X_i<Y_{(j)}\}}+\sum_{k=1}^{m} I_{\{Y_k\leqslant Y_{(j)}\}}=\sum_{i=1}^{n} I_{\{X_i<Y_{(j)}\}}+j.$$

Тогда статистика ранговых сумм Уилкоксона равна

$$V=\sum_{j=1}^{m} S_j=U+\sum_{j=1}^{m} j=U+m(m+1)/2.$$

5. Если гипотеза H_0 верна, то случайные величины X_i и Y_j имеют общую непрерывную функцию распределения $F(x)$. Введем величины $\xi_i=F(X_i)$ и $\eta_j=F(Y_j)$. Они независимы в силу независимости X_i и Y_j и равномерно распределены на отрезке $[0,1]$ в соответствии с методом обратной функции (см. § 1 гл. 4). При этом $I_{ij}=I_{\{X_i<Y_j\}}=I_{\{\xi_i<\eta_j\}}$. Отсюда

$$\mathbf{M}I_{ij}=\mathbf{P}(\xi_i<\eta_j)=\iint\limits_{0<x<y<1} dx\,dy=\frac{1}{2}\,.$$

Поэтому $\mathbf{D}I_{ij}=\mathbf{M}I_{ij}^2-(\mathbf{M}I_{ij})^2=\mathbf{M}I_{ij}-\dfrac{1}{4}=\dfrac{1}{2}-\dfrac{1}{4}=\dfrac{1}{4}$.

Поскольку $U=\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} I_{ij}$, для вычисления $\mathbf{D}U$ воспользуемся формулой для дисперсии суммы (см. П2):

$$\mathbf{D}U=\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{n}\sum_{l=1}^{m}\mathbf{cov}\,(I_{ij},I_{kl})\,.$$

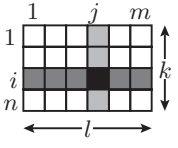


Рис. 18

Для фиксированных i и j разобьем слагаемые в двойной сумме по k и l на четыре группы: $A)\,\{k=i,l=j\}$, $B)\,\{k=i,l\neq j\}$, $C)\,\{k\neq i,l=j\}$, $D)\,\{k\neq i,l\neq j\}$ (рис. 18). В группе A $\mathbf{cov}\,(I_{ij},I_{ij})=\mathbf{D}I_{ij}=\dfrac{1}{4}$. В группе B в силу независимости случайных величин ξ_i, η_j и η_l имеем

$$\mathbf{M}(I_{ij}I_{il})=\mathbf{P}(\xi_i<\eta_j,\xi_i<\eta_l)=\iiint\limits_{\substack{0<x<y<1\\0<x<z<1}} dx\,dy\,dz=\frac{1}{3},$$

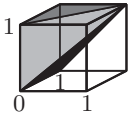


Рис. 19

так как интеграл равен объему пирамиды, основанием которой служит грань единичного куба, лежащая в плоскости $x=0$, а вершиной — точка с координатами (1,1,1) (рис. 19). Следовательно, $\mathbf{cov}\,(I_{ij},I_{il})=\mathbf{M}(I_{ij}I_{il})-\mathbf{M}I_{ij}\cdot\mathbf{M}I_{il}=\dfrac{1}{3}-\dfrac{1}{4}=\dfrac{1}{12}$. Из соображений симметрии ковариация в группе C такая же, как

в B. Наконец, согласно лемме о независимости из § 3 гл. 1 и свойству 5 математического ожидания (П2), ковариация в группе D равна 0. Собирая все вместе, получаем

$$\mathbf{D}U = nm\left(\frac{1}{4} + \frac{m-1}{12} + \frac{n-1}{12}\right) = \frac{nm(n+m+1)}{12}.$$

6. Пусть l равно целой части числа $x\sqrt{2n}$. Чтобы установить асимптотику (19), достаточно вывести, что

$$\ln\left(C_{2n}^{n-l}/C_{2n}^{n}\right) = -l^2/n + o(1) \quad \text{при } n\to\infty,\ l = O(\sqrt{n}).$$

К этому результату можно прийти и без формулы Стирлинга (см. [39, с. 58]), но с ней получается немного короче. Действительно, $\ln n! = n\ln n - n + \frac{1}{2}\ln n + \ln\sqrt{2\pi} + o(1)$. Элементарные выкладки показывают, что

$$-\ln\left(C_{2n}^{n-l}/C_{2n}^{n}\right) = \ln(n+l)! + \ln(n-l)! - 2\ln n! =$$
$$= (n+l)\ln\left(1+\frac{l}{n}\right) + (n-l)\ln\left(1-\frac{l}{n}\right) + \frac{1}{2}\ln\left(1-\frac{l^2}{n^2}\right) + o(1).$$

Разложение в ряд $\ln(1+t)$ при $t\to 0$ с учетом условия $l = O(\sqrt{n})$ позволяет завершить доказательство (проверьте!).

ОТВЕТЫ НА ВОПРОСЫ

1. Опустим перпендикуляр SX на OT (рис. 20). Тогда отрезок TX — искомый. Это вытекает из подобия прямоугольных треугольников $\triangle STX$ и $\triangle OTS$, имеющих общий острый угол:

$$c/b = b/a \iff c = b^2/a = \frac{nm}{(n+m)/2} = \frac{2nm}{n+m}.$$

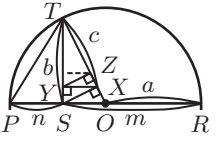


Рис. 20

Неравенство $a \geqslant b \geqslant c$ верно потому, что катет короче гипотенузы.

> Можно строить и другие «средние», продолжающие это неравенство. Для начала опустим перпендикуляр XY на отрезок ST и положим $d = |TY| = c^2/b = 4(nm)^{3/2}(n+m)^{-2}$. Затем построим перпендикуляр YZ к отрезку OT. Получим $e = |TZ| = d^2/c = 8(nm)^2(n+m)^{-3}$ и т. д.

2. Например, $I_{\{X_1<Y_1\}}$ и $I_{\{X_1<Y_2\}}$, являющиеся функциями от X_1, очевидно, зависимы (строгое доказательство вытекает из решения задачи 5). Однако для фиксированной пары индексов (i,j), подавляющая часть индикаторов, а именно $nm - n - m + 1$ (см. рис. 18), не зависят от $I_{\{X_i<Y_j\}}$ в силу леммы о независимости из § 3 гл. 1.

3. Из $C_{n+m}^{m} = C_5^2 = 10$ возможных сочетаний трех «X» и двух «Y» только наборы «$XXXYY$» и «$XXYXY$» имеют сумму рангов $V \geqslant 8$ (9 и 8). Поэтому $\mathbf{P}(V \geqslant 8) = 1/5$.

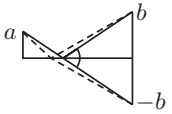


Рис. 21

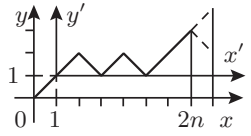


Рис. 22

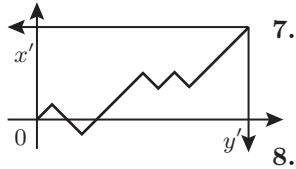


Рис. 23

4. По определению закона χ^2_k (см. пример 3 гл. 11)

$$\zeta = \frac{1}{k_1}\sum_{i=1}^{k_1} Z_i^2 \Big/ \frac{1}{k_2}\sum_{i=k_1+1}^{k_1+k_2} Z_i^2, \tag{20}$$

где все Z_i $(i = 1,\ldots,k_1 + k_2)$ подчинены стандартному нормальному распределению и независимы. Согласно закону больших чисел (см. П6), делимое и делитель в формуле (20) стремятся по вероятности к $\mathbf{M}Z_1^2 = \mathbf{D}Z_1 = 1$. Так как функция $\varphi(x,y) = x/y$ непрерывна на множестве $\{x > 0, y > 0\}$, то по свойству сходимости 3 из П5 распределение случайной величины ζ при $k_1, k_2 \to \infty$ вырождается в 1.

5. По определению закона t_k (см. пример 4 гл. 11)

$$T = \xi \Big/ \sqrt{\frac{1}{n+m-2}\,\eta}\,,$$

где $\xi \sim \mathcal{N}(0,1)$, $\eta \sim \chi^2_{n+m-2}$, ξ и η независимы. Согласно определению F-распределения имеем $T^2 \sim F_{1,n+m-2}$.

6. Из рис. 21 понятно, что вороне следует лететь по траектории, у которой «угол падения равен углу отражения», так как прямая короче ломаной.

7. Перенесем начало координат в точку $(1,1)$ и продолжим путь из концевой точки еще на один шаг вверх или вниз (рис. 22). При этом из одного положительного получим два неотрицательных пути длины $2n$.

8. Изменим систему координат так, как показано на рис. 23. В новой системе траектория будет положительной.

Глава 15

ПАРНЫЕ ПОВТОРНЫЕ НАБЛЮДЕНИЯ

§ 1. УТОЧНЕНИЕ МОДЕЛИ

Да вместе вы зачем?
Нельзя, чтобы случайно.

Фамусов в «Горе от ума»
А. С. Грибоедова

Методы этой главы предназначены для выявления неоднородности реализаций выборок $X_1,\ldots,X_n$ и $Y_1,\ldots,Y_n$ одинакового размера, которые *нельзя считать независимыми* между собой (см. § 2 гл. 14).

Прежде всего, уточним статистическую модель из § 1 гл. 14 применительно к данной ситуации. Вычислим *приращения* $Z_i = Y_i - X_i$, $i = 1,\ldots,n$, и разложим каждое из них на две части: $Z_i = \theta + \varepsilon_i$, где θ — интересующий нас *эффект воздействия* — систематический сдвиг, который мы будем считать положительным, ε_i — *случайная ошибка*, включающая в себя влияние неучтенных факторов на Z_i.

В дополнение к допущениям Д1—Д3 из § 1 гл. 14 предположим, что выполняется условие

Д6. Случайные величины $\varepsilon_1,\ldots,\varepsilon_n$ *независимы и имеют непрерывные (вообще говоря, разные) распределения такие, что*

$$\mathbf{P}(\varepsilon_i \leqslant 0) = \mathbf{P}(\varepsilon_i \geqslant 0) = 1/2, \quad i = 1,\ldots,n.$$

Это означает, что равны нулю медианы функций распределения случайных величин ε_i (см. § 2 гл. 7).

Замечание 1. Предположения Д1—Д3 из § 1 гл. 14 не обеспечивают одинаковой распределенности ε_i. Действительно, пусть случайные величины X_1 и X_2 распределены по стандартному нормальному закону $\mathcal{N}(0,1)$ (см. § 2 гл. 3) и независимы. Положим $Y_1 = X_1 + X_2$ и $Y_2 = X_1 - X_2$. Нетрудно проверить, что Y_1 и Y_2 распределены по закону $\mathcal{N}(0,2)$ и независимы, так как $\mathbf{cov}\,(Y_1, Y_2) = 0$ (см. П9). Но $Z_1 = Y_1 - X_1 = X_2 \sim \mathcal{N}(0,1)$, $Z_2 = Y_2 - X_2 = X_1 - 2X_2 \sim \mathcal{N}(0,5)$. Отсюда $\varepsilon_1 = Z_1 - \theta \sim \mathcal{N}(-\theta,1)$, а $\varepsilon_2 = Z_2 - \theta \sim \mathcal{N}(-\theta,5)$. Кроме того, что ε_1 и ε_2 имеют разные распределения, они еще и зависимы, т. е. нарушается условие Д6: $\mathbf{cov}\,(\varepsilon_1,\varepsilon_2) = \mathbf{cov}\,(X_2, X_1) - 2\,\mathbf{cov}\,(X_2, X_2) = -2 \neq 0$.

Итак, пусть выполнено предположение Д6. Рассмотрим задачу проверки гипотезы $H_0'\colon \theta = 0$ против альтернативы $H_3'\colon \theta > 0$

(штрих указывает на то, что проверяемая гипотеза и сдвиговая альтернатива задаются не для пары законов (F, G) (см. § 1 гл. 14), а для распределений приращений Z_i). Для ее решения используем критерии знаков (§ 2) и знаковых рангов Уилкоксона (§ 3).*)

§ 2. КРИТЕРИЙ ЗНАКОВ

Выполним **следующие шаги**.

1) Зададим уровень значимости (см. § 1 гл. 12) — малую вероятность α ошибочно отвергнуть верную гипотезу H_0'.

2) Положим $U_i = I_{\{Z_i>0\}}$, $i = 1, \ldots, n$.

3) В качестве *статистики критерия знаков* возьмем сумму $S = U_1 + \ldots + U_n$ и подсчитаем ее значение s на реализациях $x_1, \ldots, x_n$ и $y_1, \ldots, y_n$.**)

Малые выборки. При $n \leqslant 15$ вычисляем фактический уровень значимости, определенный в § 1 гл. 12 (см. рис. 1):

$$\alpha_0 = \mathsf{P}_0(S \geqslant s) = 2^{-n} \sum_{i=s}^{n} C_n^i = 2^{-n} \sum_{i=0}^{n-s} C_n^i. \tag{1}$$

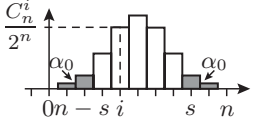


Рис. 1

Если $\alpha_0 \leqslant \alpha$, отвергаем гипотезу H_0', в противном случае — принимаем. В [10, с. 402] приведена таблица биномиальных коэффициентов, облегчающая вычисление α_0.

Большие выборки. Для расчета α_0 при $n > 15$ можно применить нормальную аппроксимацию распределения стандартизованной статистики

$$S^* = (S - \mathsf{M}S) \Big/ \sqrt{\mathsf{D}S} = (S - n/2) \Big/ \sqrt{n/4}\,.$$

Если гипотеза H_0' верна, то в соответствии с центральной предельной теоремой (П6) распределение величины S^* при $n \to \infty$ сходится к стандартному нормальному закону $\mathcal{N}(0,1)$ (см. § 2 гл. 3).

Пусть $x_{1-\alpha}$ — квантиль закона $\mathcal{N}(0,1)$ уровня $1-\alpha$ (см. § 3 гл. 7), s^* — наблюдаемое значение статистики S^*. Если $s^* \geqslant x_{1-\alpha}$, то отвергаем гипотезу H_0', в противном случае — принимаем.

Поправка. Можно значительно улучшить качество приближения дискретного биномиального распределения непрерывным нормальным законом за счет введения *поправки на непрерывность*. Рассмотрим «подправленную» статистику

$$\tilde{S}^* = (S - 0{,}5 - n/2) \Big/ \sqrt{n/4}\,. \tag{2}$$

*) Их обобщения для многомерных данных приведены в § 2 гл. 23.

**) Если значение i-го приращения $z_i = y_i - x_i > 0$, то это отмечают знаком «+», если $z_i < 0$ — знаком «−». Отсюда происходит название критерия.

Как показывает рис. 2, сдвиг *влево* на 0,5 позволяет точнее аппроксимировать сумму площадей прямоугольников площадью под графиком *правого* «хвоста» нормальной плотности.

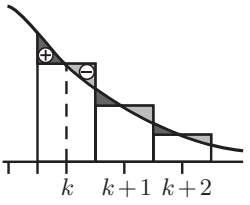


Рис. 2

Совпадения. Если среди значений Z_i встречаются нули, то их надо отбросить и соответственно уменьшить n до числа ненулевых значений Z_i.

Оценка параметра. Когда гипотеза H_0' отвергнута, принимается альтернатива H_3'. В этом случае представляет интерес величина сдвига θ. В качестве ее оценки $\widehat{\theta}$ можно взять *выборочную медиану приращений* $MED\{Z_i,\, i=1,\ldots,n\}$ (см. § 2 гл. 7).

Доверительный интервал. Определим номер k_α как наибольшее число слагаемых, при котором

$$2^{-n}\sum_{i=0}^{k_\alpha} C_n^i \leqslant \alpha. \tag{3}$$

Тогда пара порядковых статистик $(Z_{(k_\alpha+1)},Z_{(n-k_\alpha)})$ (см. § 4 гл. 4) образует доверительный интервал для θ с коэффициентом доверия $1-2\alpha$ (см. § 1 гл. 11). Для нахождения k_α можно также воспользоваться таблицей из [10, с. 353].

При большом n значение k_α с учетом поправки на непрерывность приближенно равно целой части числа

$n/2-0{,}5-x_{1-\alpha}\sqrt{n/4}$, где $x_{1-\alpha}$ — квантиль закона $\mathcal{N}(0,1)$.

Пример 1. ***Времена реакции*** **[80, с. 123].** Числа x_i и y_i в приведенной ниже таблице представляют собой времена реакции i-го испытуемого на световой и звуковой сигналы соответственно, $z_i=y_i-x_i$, $i=1,\ldots,12$.

i	1	2	3	4	5	6	7	8	9	10	11	12
x_i	176	163	152	155	156	178	160	164	169	155	122	144
y_i	168	215	172	200	191	197	183	174	176	155	115	163
z_i	−8	+52	+20	+45	+35	+19	+23	+10	+7	0	−7	+19

Поскольку $z_{10}=0$, отбросим это наблюдение, уменьшив размер выборки до $n=11$. Статистика знаков S имеет значение $s=9$. По формуле (1) находим $\alpha_0=(1+11+55)/2048\approx 0{,}033$. Следовательно, на уровне значимости $\alpha\geqslant 3{,}3\%$ гипотеза $H_0'\colon \theta=0$ отвергается.

Хотя в нашем случае $n<15$, подсчитаем для сравнения значение статистики $\tilde{S}^*$ по формуле (2). Получим 1,809. В таблице Т2 этому значению соответствует уровень значимости 3,5%. Упорядочив z_i по возрастанию, вычисляем оценку параметра сдвига $\widehat{\theta}=MED\{z_1,\ldots,z_9,z_{11},z_{12}\}=19$. Наконец, для $\alpha=0{,}05$ из неравенства (3) находим $k_\alpha=2$, что приводит к интервалу $(7,35)$ с коэффициентом доверия 90%.

Комментарии

1) Если потребовать, чтобы все величины ε_i в допущении Д6 имели одинаковое распределение, у которого нуль — единственная медиана, то в силу закона больших чисел (П6) критерий знаков будет состоятельным против альтернативы $H_3'\colon \theta > 0$.

2) В случае альтернативы $H_3''\colon \theta < 0$, очевидно, достаточно поменять местами выборки $X_1,\ldots,X_n$ и $Y_1,\ldots,Y_n$.

3) Покажем, как оценка *MED* параметра сдвига связана со статистикой S критерия знаков. Интуитивно понятно, что сдвиг разумно оценить такой величиной θ', чтобы набор $Z_i' = Z_i - \theta'$ $(i = 1,\ldots,n)$ выглядел как выборка из распределения с нулевой медианой. Для такой выборки S имеет биномиальное распределение (см. § 1 гл. 5), симметричное относительно своего математического ожидания $n/2$. Эти соображения приводят к следующему уравнению относительно θ':

$$\sum_{i=1}^{n} I_{\{Z_i'>0\}} = \sum_{i=1}^{n} I_{\{Z_i>\theta'\}} = n/2.$$

Когда величина θ' становится равной *MED*, происходит «перескок» через уровень $n/2$.

4) Нетрудно убедиться, что приведенный выше доверительный интервал получается в результате применения метода 1 из § 3 гл. 11 к функции $g(\boldsymbol{z},\theta) = \sum I_{\{z_i>\theta\}}$.

5) Ходжес и Леман показали (см. [88, с. 66]), что при оценивании сдвига с помощью *MED* следует использовать выборку четного размера $n = 2k$, поскольку выборочная медиана для выборки размера $2k+1$ имеет ту же самую точность.

§ 3. КРИТЕРИЙ ЗНАКОВЫХ РАНГОВ УИЛКОКСОНА

Пусть кроме допущения Д6 выполнено условие

Д7. *Случайные величины $\varepsilon_1,\ldots,\varepsilon_n$ имеют одинаковое распределение, симметричное относительно нуля*:

$F_{\varepsilon_1}(-x) = 1 - F_{\varepsilon_1}(x)$ *для всех* x.

Для проверки гипотезы H_0' против альтернативы H_3' (см. § 1) совершим **следующие шаги**.

1) Зададим уровень значимости критерия α (малую вероятность ошибочно отвергнуть верную гипотезу H_0').

2) Вычислим $Z_i = Y_i - X_i$, $i = 1,\ldots,n$, и упорядочим $|Z_1|,\ldots,|Z_n|$ по возрастанию. Пусть R_i обозначает *ранг* (порядковый номер) величины $|Z_i|$.

3) Положим $U_i = I_{\{Z_i>0\}}$, $i = 1,\ldots,n$.

4) В качестве *статистики критерия знаковых рангов* возьмем $T = R_1U_1 + \ldots + R_nU_n$ и подсчитаем ее значение t на реализациях $x_1, \ldots, x_n$ и $y_1, \ldots, y_n$.

Малые выборки. При $n \leqslant 15$ отвергнем гипотезу H_0', если окажется, что $t \geqslant t_\alpha$, где критическое значение t_α берется из таблицы $A.4$ книги [88].

Большие выборки. Для $n > 15$ можно использовать стандартизированную статистику

$$T^* = \frac{T - \mathbf{M}T}{\sqrt{\mathbf{D}T}} = \frac{T - [n(n+1)/4]}{\sqrt{n(n+1)(2n+1)/24}}, \tag{4}$$

распределение которой сходится к $\mathcal{N}(0,1)$ при $n \to \infty$, если справедлива гипотеза H_0' и выполнены условия Д6—Д7 (задачи 4–6).

В случае, когда наблюдаемое значение этой статистики $t^* \geqslant x_{1-\alpha}$, где $x_{1-\alpha}$ — $(1-\alpha)$-квантиль закона $\mathcal{N}(0,1)$, гипотеза H_0' отвергается, иначе — принимается.

Поправка. В 1974 г. Р. Иман предложил следующую аппроксимацию, обеспечивающую значительное снижение относительной ошибки для критических значений. Она использует линейную комбинацию нормальной и стьюдентовской квантилей (см. [88, с. 47]). Положим

$$\tilde{t}^* = \frac{1}{2}\, t^* \left[1 + \sqrt{(n-1)/[n - (t^*)^2]}\right]. \tag{5}$$

С помощью таблиц Т2 и Т4 вычислим $z_\alpha = (x_{1-\alpha} + y_{1-\alpha})/2$, где $x_{1-\alpha}$ и $y_{1-\alpha}$ обозначают соответственно квантили уровня $(1-\alpha)$ закона $\mathcal{N}(0,1)$ и распределения Стьюдента с $(n-1)$ степенями свободы (см. § 2 гл. 11). Если $\tilde{t}^* \geqslant z_\alpha$, то гипотеза H_0' отвергается, иначе — принимается.

Совпадения. Если среди значений Z_i встречаются нули, то их надо отбросить и, соответственно, уменьшить n до числа ненулевых значений Z_i. Если среди ненулевых $|Z_i|$ есть равные, то для вычисления статистики T надо использовать средние ранги. В формуле (4) дисперсию $\mathbf{D}T$ следует заменить на

$$\frac{1}{24}\left[n(n+1)(2n+1) - \frac{1}{2}\sum_{k=1}^{g} l_k(l_k^2 - 1)\right], \tag{6}$$

где g — число групп совпадений, l_k — количество элементов в k-й группе.*)

*) Не совпадающие с другими наблюдения считаются группой размера 1. Если совпадений нет вовсе, то сумма в выражении (6) пропадает.

Оценка параметра. Когда гипотеза H_0' отвергается, в качестве оценки параметра сдвига θ можно взять *медиану средних Уолша* (см. § 3 гл. 8)

$$W = MED\{(Z_i + Z_j)/2,\ 1 \leqslant i \leqslant j \leqslant n\}.$$

Доверительный интервал. Построение доверительного интервала для случая $n \leqslant 15$ описано в [88, с. 55]. При больших n пара порядковых статистик $(V_{(k_\alpha+1)}, V_{(M-k_\alpha)})$ образует приближенный доверительный интервал с коэффициентом доверия $1 - 2\alpha$. Здесь $V_{(1)} \leqslant \ldots \leqslant V_{(M)}$ — упорядоченные по возрастанию *средние Уолша* $(Z_i + Z_j)/2$ при $1 \leqslant i \leqslant j \leqslant n$ и $M = n(n+1)/2$; k_α — это целая часть числа

$$n(n+1)/4 - 0{,}5 - x_{1-\alpha}\sqrt{n(n+1)(2n+1)/24}, \tag{7}$$

где $x_{1-\alpha}$ обозначает, как и ранее, $(1-\alpha)$-квантиль закона $\mathcal{N}(0,1)$, а 0,5 представляет собой поправку на непрерывность (см. § 2).

Проверка симметрии. Прежде чем применять критерий знаковых рангов, следует удостовериться в справедливости допущения Д7. Простой графический метод проверки основан на сходимости выборочных квантилей к теоретическим (см. § 3 гл. 7). Так как для теоретических квантилей z_p симметричного относительно медианы $z_{1/2}$ закона верно равенство $z_{1/2} - z_p = z_{1-p} - z_{1/2}$, то для порядковых статистик $Z_{(i)}$ можно ожидать выполнения соотношений

$$\xi_i = MED - Z_{(i)} \approx \eta_i = Z_{(n+1-i)} - MED, \quad i = 1, \ldots, [n/2],$$

(здесь $[\,\cdot\,]$ обозначает целую часть числа). Поэтому для выборки $Z_1, \ldots, Z_n$ из симметричного относительно медианы распределения точки плоскости (ξ_i, η_i) должны располагаться вблизи диагонали $y = x$.

Замечание 2. Условие *строгой симметрии* относительно медианы является почти столь же нереалистичным, как и предположение, что распределение величин Z_i в точности нормально. Как правило, надежно проверить симметрию можно лишь по выборке из нескольких сотен наблюдений. Асимптотический критерий Гупты для решения этой проблемы приведен в [88, с. 76]. Ссылки на другие критерии см. там же на с. 81.

Предположение о симметрии иногда оказывается справедливым в силу специфики получения наблюдений, приводящей к одинаковым вероятностям отклонения на произвольную величину от медианы как влево, так и вправо.

Симметрия распределения величин Z_i довольно естественно возникает в модели «контроль — обработка» (см. пример 1 гл. 8). Однако подчеркнем еще раз, что в случае совместной независимости выборок $X_1, \ldots, X_n$ и $Y_1, \ldots, Y_n$ для проверки гипотезы

однородности следует использовать не критерии знаков и знаковых рангов Уилкоксона, а методы, изложенные в гл. 14.

Пример 2. Для данных из примера 1 проверим гипотезу $H_0'\colon \theta = 0$ с помощью критерия знаковых рангов. После отбрасывания $Z_{10} = 0$ в выборке останется $n = 11$ наблюдений. Упорядочим их:

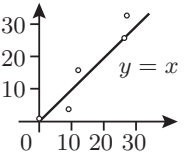


Рис. 3

$Z_{(i)}$	-8	-7	7	10	19	19	20	23	35	45	52

Видим, что $MED = 19$. Для визуальной проверки симметрии построим точки (ξ_i,η_i), определенные выше. Проведем прямую $y = x$ (рис. 3). Хотя выборка слишком мала для уверенного заключения, построенная диаграмма, по-видимому, не противоречит допущению о симметрии распределения случайной величины Z_i.

Упорядочим по возрастанию величи́ны $|Z_i|$ и присвоим средние ранги совпадающим значениям:

$\lvert Z_i\rvert$	7	7	8	10	19	19	20	23	35	45	52
R_i	1,5	1,5	3	4	5,5	5,5	7	8	9	10	11
U_i	0	1	0	1	1	1	1	1	1	1	1

Согласно приведенной таблице статистика критерия знаковых рангов $T = \sum R_i U_i = 61{,}5$. Учитывая, что среди величин $|Z_i|$ есть две группы совпадений, по формуле (6) вычислим дисперсию $\mathbf{D}T = (11 \cdot 12 \cdot 23 - 3 - 3)/24 = 126{,}25$. Отсюда по формуле (4) для нормированной статистики T^* получаем значение $t^* \approx 2{,}54$. Положив $\alpha = 0{,}005$, из таблиц Т2 и Т4 (при $k = n - 1 = 10$) находим $z_\alpha = (2{,}576 + 3{,}169)/2 \approx 2{,}87$. Согласно формуле (5) имеем $\tilde{t}^* = 3{,}14$. Так как $3{,}14 \geqslant 2{,}87$, то гипотеза H_0' отвергается на уровне значимости 0,005.

С помощью компьютера вычисляем значение оценки параметра сдвига $W = MED\{(Z_i + Z_j)/2,\ i \leqslant j\} = 19{,}25$. На основе формулы (7) строим 90%-ный доверительный интервал $\bigl(V_{(15)},V_{(52)}\bigr) = (15/2, 31)$, который несколько у́же интервала $(7, 35)$, полученного ранее при применении критерия знаков к этим же данным.

Комментарии

1) Если в условии Д7 все ε_i имеют одинаковое *симметричное гладкое распределение* (см. § 1 гл. 8), то критерий знаковых рангов будет состоятельным против альтернативы $H_3'\colon \theta > 0$ (см. [86, с. 64]).

2) Покажем, что связь между статистикой T критерия знаковых рангов и медианой средних Уолша W аналогична рассмотренной ранее связи между статистикой S критерия знаков и выборочной медианой MED. Согласно задаче 1 при отсутствии нулевых значений и совпадений среди величин $|Z_i|$ статистика знаковых

рангов T равна $\sum\limits_{i\leqslant j} I_{\{(Z_i+Z_j)/2>0\}}$, т. е. T есть число положительных средних Уолша. Естественной оценкой параметра сдвига θ будет такая величина $\widehat{\theta}$, чтобы набор $Z_i' = Z_i - \widehat{\theta}$ $(i = 1,\ldots,n)$ выглядел как выборка из закона с $\theta = 0$. Но при $\theta = 0$ распределение статистики T симметрично относительно своего среднего $n(n+1)/4$. Тем самым, приходим к соотношению

$$\sum_{i\leqslant j} I_{\{(Z_i'+Z_j')/2>0\}} = \sum_{i\leqslant j} I_{\{(Z_i+Z_j)/2>\widehat{\theta}\}} = n(n+1)/4.$$

Когда величина $\widehat{\theta}$ становится равной медиане средних Уолша, происходит «перескок» через уровень $n(n+1)/4$.

3) Построенный выше доверительный интервал получается в результате применения метода 1 из § 3 гл. 11 к функции

$$g(\boldsymbol{z},\theta) = \sum_{i\leqslant j} I_{\{(z_i+z_j)/2>\theta\}}.$$

4) Сравним критерий знаков с критерием знаковых рангов по их «чувствительности» к обнаружению сдвиговой альтернативы $H_3'\colon \theta > 0$. Их относительная точность при больших n (см. [86, с. 77]) совпадает с относительной асимптотической эффективностью (см. § 4 гл. 7) связанных с критериями оценок: выборочной медианы *MED* и медианы средних Уолша W (доказательство приведено в [86, с. 90]). Согласно теоремам 1 и 3 гл. 8 на классе Ω_s гладких симметричных распределений верно равенство

$$e_{\mathrm{MED},W}(F) = \frac{p^2(0)}{3\left(\int p^2(x)\,dx\right)^2}, \tag{8}$$

где $p(x) = F'(x)$ — плотность закона с функцией распределения $F(x)$. В частности, для нормального закона $e_{MED,W} = 2/3 < 1$ (см. задачу 1 гл. 7 и вопрос 5 гл. 8), а для распределения Лапласа $e_{MED,W} = 4/3 > 1$ (см. задачу 1 гл. 8).

Оказывается, чем «легче» хвосты у распределения F, тем предпочтительнее оценка W по сравнению с *MED*. Для уточнения этого утверждения приведем отрывок из [86, с. 122], в котором обсуждается предложенное У. Ван Цветом в 1970 г. **упорядочение симметричных распределений по весу их хвостов**:

> «Пусть F и G из Ω_s. Будем говорить, что хвосты F «легче» хвостов G (или G имеет хвосты «тяжелее», чем у F), что обозначается $F \preccurlyeq G$, если функция $G^{-1}(F(x))$ выпукла при $x \geqslant 0$.*)
>
> Заметим следующее: 1) $F \preccurlyeq F$, 2) $F \preccurlyeq G$ и $G \preccurlyeq H$ влечет за собой $F \preccurlyeq H$. Следовательно, $\preccurlyeq$ — слабое упорядочение. Если $F \preccurlyeq G$ и $G \preccurlyeq F$, то мы называем F и G *эквивалентными*.
>
> Пусть $F(x) = G(ax)$ при $a > 0$, тогда $G^{-1}(F(x)) = ax$, так что $F \preccurlyeq G$, также $F^{-1}(G(x)) = x/a$ и поэтому $G \preccurlyeq F$. Отсюда мы видим, что распределения, различающиеся лишь по параметрам масштаба, эквивалентны.

*) Определения выпуклости и строгой выпуклости функции приведены в П4.

Так как F и G из Ω_s, то их плотности f и g положительны в нуле. Далее положим $f(0) = g(0)$ (общность при этом не теряется), что достигается преобразованием масштаба: $\tilde{F}(x) = F(x/\sigma)$ с $\sigma = f(0)/g(0)$. Теперь допустим, что $F \preccurlyeq G$, причем они не эквивалентны. Тогда $q(x) = G^{-1}(F(x))$ строго выпукла для некоторого x, а $q'(x) = f(x)/g(G^{-1}(F(x)))$ строго возрастает для некоторых x. Поскольку $q'(0) = f(0)/g(0) = 1$, то $q'(x) > 1$ для некоторых x и, наконец, $G^{-1}(F(x)) > x$. Отсюда $F(x) > G(x)$ и $1 - F(x) < 1 - G(x)$, так что вероятность попадания наблюдения «на хвост» G больше».

Нетрудно проверить, что верно следующее упорядочение:

равномерное $\preccurlyeq$ *нормальное* $\preccurlyeq$ *закон Лапласа* $\preccurlyeq$ *закон Коши*

Можно доказать (см. [86, с. 137]), что для F, $G \in \Omega_s$ отношение $F \preccurlyeq G$ влечет неравенство $e_{MED,W}(F) \leqslant e_{MED,W}(G)$, где $e_{MED,W}$ задается формулой (8).

§ 4. ЗАВИСИМЫЕ НАБЛЮДЕНИЯ

До сих пор мы предполагали, что приращения Z_i независимы (допущение Д6). Для иллюстрации того, что происходит при отказе от этого допущения, рассмотрим следующий пример.

Пример 3. ***Влияние сериальной корреляции*** **[86, с. 38].** Пусть компоненты нормального случайного вектора $(Z_1, \ldots, Z_n)$ (см. П9) таковы, что $Z_i \sim \mathcal{N}(\theta, 1)$, $\mathbf{M}(Z_i Z_j) = \rho$ при $j = i \pm 1$ и 0 — в противном случае, причем коэффициент корреляции $|\rho| \leqslant 1/2$. (Корректность данного определения проверяется в задаче 2.) При $\rho = 0$ получаем независимые Z_i, а при $\rho \neq 0$ зависимы лишь стоящие рядом случайные величины. Эта модель — частный случай m-зависимой последовательности с $m = 1$ (см. П6).

Проверим гипотезу $H_0'\colon \theta = 0$ против альтернативы $H_3'\colon \theta > 0$. При $\rho = 0$ равномерно наиболее мощный критерий Неймана—Пирсона уровня значимости α задается критическим множеством $\{\boldsymbol{z} \in \mathbb{R}^n\colon \sqrt{n}\,\overline{z} \geqslant x_{1-\alpha}\}$, где $x_{1-\alpha}$ — квантиль закона $\mathcal{N}(0,1)$ с функцией распределения $\Phi(x)$ (см. пример 1 гл. 13). Выясним, каков истинный уровень значимости этого критерия, если на самом деле $\rho \neq 0$, т. е. вычислим

$$\alpha_\rho(\overline{Z}) = \mathbf{P}\left(\sqrt{n}\,\overline{Z} \geqslant x_{1-\alpha}\right)$$

при справедливости гипотезы H_0'. Как линейная комбинация компонент нормального вектора, статистика $\sqrt{n}\,\overline{Z}$ распределена нормально с параметрами $\mathbf{M}(\sqrt{n}\,\overline{Z}) = 0$ и $\mathbf{D}(\sqrt{n}\,\overline{Z}) = 1 + 2\rho(n-1)/n$ (последнее равенство верно в силу свойств дисперсии 1 и 3 из П2). Отсюда при больших n имеем

$$\alpha_\rho(\overline{Z}) = 1 - \Phi\left(x_{1-\alpha} \Big/ \sqrt{\mathbf{D}\overline{Z}}\right) \approx 1 - \Phi\left(x_{1-\alpha} \Big/ \sqrt{1+2\rho}\right).$$

Вычислим для критерия знаков аналогичную характеристику

$$\alpha_\rho(S) = \mathsf{P}\left((S - n/2) \Big/ \sqrt{n/4} \geqslant x_{1-\alpha}\right).$$

Здесь $S = \sum U_i$, где $U_i = I_{\{Z_i>0\}}$. При выполнении гипотезы H_0' находим, что

$$\begin{aligned}
&\mathsf{M}S = n\,\mathsf{M}U_1 = n\,\mathsf{P}\,(Z_1 > 0) = n/2,\\
&\mathsf{D}S = n\,\mathsf{D}U_1 + 2(n-1)\,\mathsf{cov}\,(U_1, U_2) =\\
&\qquad = n/4 + 2(n-1)\,[\mathsf{P}(Z_1 > 0, Z_2 > 0) - 1/4]\,,
\end{aligned}$$

поскольку $\mathsf{D}U_1 = \mathsf{M}U_1^2 - (\mathsf{M}U_1)^2 = \mathsf{M}U_1 - 1/4 = 1/4$. Из задачи 3

$$\mathsf{P}(Z_1 > 0, Z_2 > 0) = \frac{1}{4} + \frac{1}{2\pi}\,\arcsin\rho. \tag{9}$$

Подставляя этот результат в предыдущую формулу, получаем

$$\mathsf{D}S = \frac{n}{4} + \frac{n-1}{\pi}\,\arcsin\rho.$$

Так как $\mathsf{M}\,|U_1|^3 = \mathsf{M}U_1 = 1/2 < \infty$, то выполняются условия теоремы Хефдинга и Робинса из П6, согласно которой распределение статистики $(S - \mathsf{M}S)/\sqrt{\mathsf{D}S}$ стремится к $\mathcal{N}(0,1)$. Отсюда при $n \to \infty$

$$\alpha_\rho(S) \approx 1 - \Phi\left(\frac{x_{1-\alpha}\sqrt{n/4}}{\sqrt{\mathsf{D}S}}\right) \approx 1 - \Phi\left(\frac{x_{1-\alpha}}{\sqrt{1 + (4/\pi)\,\arcsin\rho}}\right).$$

Наконец, приведем для сравнения аналогичную характеристику для статистики T критерия знаковых рангов Уилкоксона (см. [86, с. 99]). Из задачи 6 имеем $\mathsf{M}T = n(n+1)/4$, $\mathsf{D}T = n(n+1)(2n+1)/24$,

$$\alpha_\rho(T) = \mathsf{P}\left(\frac{T - \mathsf{M}T}{\sqrt{\mathsf{D}T}} \geqslant x_{1-\alpha}\right) \approx 1 - \Phi\left(\frac{x_{1-\alpha}}{\sqrt{1 + (12/\pi)\,\arcsin(\rho/2)}}\right).$$

В таблице указаны значения характеристик при $\alpha = 5\%$.

ρ	-0,4	-0,3	-0,2	-0,1	0	0,1	0,2	0,3	0,4	0,5
$\alpha_\rho(\overline{Z})$	0,000	0,005	0,017	0,033	0,05	0,067	0,082	0,097	0,109	0,122
$\alpha_\rho(S)$	0,009	0,018	0,028	0,039	0,05	0,061	0,071	0,081	0,092	0,101
$\alpha_\rho(T)$	0,000	0,006	0,018	0,033	0,05	0,067	0,081	0,095	0,107	0,120

Мы видим, что истинные уровни значимости всех трех критериев довольно существенно отличаются от 5%. При положительной корреляции далекие от нуля величины Z_i как бы подтягивают к себе следующие за ними наблюдения, что приводит к увеличению дисперсии статистики критерия по сравнению с номинальной. Так, если $\rho = 0{,}4$, то гипотеза H_0' отвергается примерно в 10% случаев вместо 5%.

Заметим также, что строка для $\alpha_\rho(T)$ почти идентична строке для $\alpha_\rho(\overline{Z})$. Это объясняется тем, что $\arcsin\rho = \rho + \rho^3/6 + \ldots$.

Поэтому

$$1+\frac{12}{\pi}\arcsin(\rho/2)\approx 1+\frac{6}{\pi}\,\rho\approx 1+2\rho.$$

Наличие корреляции для реальных наблюдений — скорее правило, чем исключение. Б. Мандельброт и Дж. Уоллис исследовали около 70 рядов геофизических данных: речной сток, количество атмосферных осадков, частота землетрясений, годовые кольца на деревьях, мощность геологических слоев, а также число солнечных пятен (см. [84, с. 420]). В большинстве случаев была выявлена значимая положительная корреляция.

Законы математики, имеющие какое-либо отношение к реальному миру, ненадежны; а надежные математические законы не имеют отношения к реальному миру.

А. Эйнштейн

Пример 4. ***Колебания уровня воды*** **[90, с. 449].** Рассмотрим (упрощенную) вероятностную модель, описывающую отклонения от среднего значения уровня некоторого водного бассейна (например, Каспийского моря), вызванные испарением с водной поверхности и колебаниями в стоке. Обозначим через H_n уровень в бассейне в n-м году. Запишем для него *уравнение баланса*

$$H_{n+1}=H_n-k\,S(H_n)+T_{n+1},$$

где k — *коэффициент испарения*, $S(H)$ — *площадь водной поверхности* на уровне H, T_{n+1} — *величина стока* в $(n+1)$-м году.

Пусть $Z_n=H_n-\overline{H}$, где *средний уровень* $\overline{H}$ можно считать известным из многолетних наблюдений. Предположим, что $S(H)=S(\overline{H})+c(H-\overline{H})$. (Для гладких $S(H)$ это приближенно верно для не очень больших отклонений $H-\overline{H}$.) Тогда величины Z_n подчиняются соотношениям

$$Z_{n+1}=\alpha Z_n+\varepsilon_{n+1}$$

с $\alpha=1-ck$ и $\varepsilon_{n+1}=T_{n+1}-kS(\overline{H})$. Будем считать случайные величины ε_n независимыми и одинаково нормально распределенными с нулевым математическим ожиданием и дисперсией σ^2. Как установлено в [90, с. 448], указанные соотношения имеют при всех целых n и $|\alpha|<1$ единственное *стационарное* (П6) решение

$$\tilde{Z}_n=\sum_{i=0}^{\infty}\alpha^i\varepsilon_{n-i},\qquad \text{причем } \mathbf{cov}\left(\tilde{Z}_0,\tilde{Z}_n\right)=\sigma^2\alpha^n\,/\left(1-\alpha^2\right).$$

Интересным практическим выводом в данной модели является то, что *оптимальным* (в среднеквадратичном смысле в классе линейных функций) прогнозом на следующий год по результатам наблюдений за предшествующие годы $\ldots, Z_{n-1}, Z_n$ служит просто величина αZ_n (см. [90, с. 489]).

§ 5. КРИТЕРИЙ СЕРИЙ

Рассмотрим один простой метод, позволяющий обнаруживать определенного вида корреляции наблюдений Z_i, который называется *критерием серий* (см. [10, с. 91]).

Пусть h — некоторый заданный уровень. На практике обычно в качестве h берут предварительно вычисленное значение выборочного среднего, выборочной медианы или произвольную константу между минимумом и максимумом наблюдений.

Положим $\zeta_i = I_{\{Z_i \leqslant h\}}$. Если $Z_1, \ldots, Z_n$ — независимые одинаково распределенные случайные величины, то $\zeta_1, \ldots, \zeta_n$ — схема Бернулли (см. § 3 гл. 1) с неизвестной вероятностью «успеха» $p = p(h) = \mathbf{P}(Z_1 \leqslant h)$.

Реализация случайных величин $\zeta_1, \ldots, \zeta_n$ представляет собой последовательность из символов 0 и 1 длины n. *Сериями* назовем цепочки символов одного вида. (Например, 111010 содержит 4 серии.) Обозначим *число серий* через T_n. Тогда

$$T_n = 1 + \delta_1 + \ldots + \delta_{n-1}, \quad \text{где } \delta_i = I_{\{\zeta_{i+1} \neq \zeta_i\}} = |\zeta_{i+1} - \zeta_i|. \tag{10}$$

Здесь δ_i — индикатор перемены символа на $(i+1)$-м месте. Если $\zeta_1, \ldots, \zeta_n$ — схема Бернулли, то δ_i образуют стационарную m-зависимую последовательность (П6) с $m = 1$.

Вопрос 1. Чему в этом случае равны $\mathbf{M}T_n$ и $\mathbf{D}T_n$?

По теореме Хефдинга и Робинса (П6) при $0 < p < 1$ распределение статистики $(T_n - \mathbf{M}T_n)/\sqrt{\mathbf{D}T_n}$ сходится к $\mathcal{N}(0,1)$.

Вопрос 2. Чему примерно равна $\mathbf{P}(T_{100} \geqslant 60)$ при $p = 1/2$?

Поскольку в общем случае вероятность p неизвестна, для проверки независимости Z_i применим критерий, основанный на *условной предельной теореме* для T_n. Пусть $S_n = \zeta_1 + \ldots + \zeta_n$. Для $k = 1,2,\ldots,n-1$; $m = 0,1,\ldots,n$; $l = n - m$ имеем

$$\mathbf{P}(T_n = k \,|\, S_n = m) = \begin{cases} 2\, C_{l-1}^{i-1} C_{m-1}^{i-1} \,/\, C_n^m\,, & k = 2i, \\ \left(C_{l-1}^{i-1} C_{m-1}^{i} + C_{l-1}^{i} C_{m-1}^{i-1}\right) / \, C_n^m, & k = 2i+1. \end{cases}$$

Вопрос 3. Как доказать эту формулу, используя модель размещения неразличимых шариков по ящикам из § 5 гл. 10?

Заметим, что условное распределение $\mathbf{P}(T_n = k \,|\, S_n = m)$, $k = 1,2,\ldots,n$, при фиксированном m *не зависит от неизвестного параметра* p. Для него верен следующий результат.

Условная предельная теорема. *Пусть m, $n \to \infty$ так, что $n/(n+m) \to \gamma \in (0,1)$ (допущение Д5 из гл. 14). Тогда*

$$\mathbf{P}(T_n \leqslant k \,|\, S_n = m) = \Phi((k - \mu_{m,n})/\sigma_{m,n}) + o(1),$$

где

$$\begin{aligned} \mu_{m,n} &= \mathbf{M}(T_n \,|\, S_n = m) = 1 + \frac{2lm}{n}\,, \\ \sigma^2_{m,n} &= \mathbf{D}(T_n \,|\, S_n = m) = \frac{2lm(2lm-n)}{n^2(n-1)}\,. \end{aligned} \tag{11}$$

Доказать формулу для $\mu_{m,n}$ предлагается в задаче 7.

Здесь $\Phi(x)$ — функция распределения закона $\mathcal{N}(0,1)$, $l = n - m$.

Этим приближением можно пользоваться при $n \geqslant 20$. Критические значения для меньших n приведены в [10, с. 354].

Для проверки независимости случайных величин $Z_1,\ldots,Z_n$ против альтернативы их положительной коррелированности H^+ (отрицательной коррелированности H^-), ведущей к относительно малому (большому) числу серий, **надо**:

1) задать уровень h (скажем, равный $\overline{z}$);

2) преобразовать реализацию $z_1,\ldots,z_n$ в последовательность из нулей и единиц $I_{\{z_1\leqslant h\}},\ldots,I_{\{z_n\leqslant h\}}$;

3) подсчитать количество серий k и число единиц m в этой последовательности;

4) вычислить $\mu_{m,n}$ и $\sigma^2_{m,n}$ по формулам (11);

5) найти значение

$$t_n^*=\begin{cases}(k+0{,}5-\mu_{m,n})/\sigma_{m,n} & \text{в случае } H^+,\\ (k-0{,}5-\mu_{m,n})/\sigma_{m,n} & \text{в случае } H^-;\end{cases}$$

6) определить по таблице Т2 приближенный фактический уровень значимости

$$\alpha_0=\begin{cases}1-\Phi(-t_n^*) & \text{в случае } H^+,\\ 1-\Phi(t_n^*) & \text{в случае } H^-;\end{cases}$$

7) отвергнуть гипотезу независимости случайных величин $Z_1,\ldots,Z_n$, если значение α_0 достаточно мало, иначе — принять эту гипотезу.

Пример 5 [10, с. 93]. Ниже указаны результаты проверки правильности прогноза температуры воздуха на сутки вперед в течение 28 последовательных дней. Знаками «−» отмечены дни, когда абсолютная ошибка прогноза была более 2°. В остальных случаях результаты прогноза отмечались знаком «+».

$$+\,+\,+\,+\,+\,+\,+\,+\,+\,+\,+\quad+\,-\,-\,-\,-\,-\,+\,+\,-\,-\,+\quad+\,+\,+\,+\,-\,+\,+$$

Можно ли утверждать, что правильные и неправильные результаты прогноза группируются случайно?

В данном примере $k=7$, $l=20$, $m=8$, $n=28$. По формулам (11) вычисляем $\mu_{m,n}=12{,}429$ и $\sigma^2_{m,n}=4{,}414$. Отсюда находим $t_n^*=-2{,}346$, что согласно таблице Т2 значимо мало на уровне 1%. Таким образом, гипотезу о чисто случайном расположении знаков «+» и «−» следует отвергнуть.

ЗАДАЧИ

1. Докажите, что при отсутствии нулевых значений и совпадений среди величин $|Z_i|$ для статистики знаковых рангов T, определенной в § 3, имеет место представление

$$T=\sum_{i\leqslant j}I_{\{(Z_i+Z_j)/2>0\}}=\sum_{j\leqslant i}I_{\left\{(Z_{(i)}+Z_{(j)})/2>0\right\}}. \qquad (12)$$

Мозг гораздо чаще ржавеет, чем изнашивается.

Кристиан Бови

2* Проверьте, что для всех n матрица $\boldsymbol{\Sigma} = \|\sigma_{ij}\|_{n\times n}$, где

$$\sigma_{ij} = \begin{cases} 1, & \text{если } j = i, \\ \rho, & \text{если } j = i \pm 1, \\ 0 & \text{в противном случае,} \end{cases}$$

положительно определена при $|\rho| \leqslant 1/2$.

Указание. Разложите главный минор порядка n сначала по первой строке, затем — по первому столбцу для получения рекуррентной формулы.

3* Выведите формулу (9) из примера 3.

Указание. Запишите формулу плотности нормального вектора (Z_1,Z_2) (П9) и перейдите к полярным координатам.

4* Убедитесь, что при выполнении условий Д6—Д7 случайные векторы $\boldsymbol{U} = (U_1,\ldots,U_n)$ и $\boldsymbol{R} = (R_1,\ldots,R_n)$, определенные при описании критерия знаковых рангов, независимы.

5* Определим *антиранг* A_i как такой номер, что $|Z_{A_i}| = |Z|_{(i)}$ (т. е. A_i есть индекс того наблюдения, которое соответствует i-му по величине абсолютному значению). Положим $W_i = I_{\{Z_{A_i}>0\}}$. Докажите, что при выполнении гипотезы H_0' и условий Д6—Д7 случайные величины $W_1,\ldots,W_n$ образуют схему Бернулли с $p = 1/2$.

6* Используя представление $T = \sum i\,W_i$, с помощью теорем Линдеберга (П6) и Лебега (П5) установите при справедливости гипотезы H_0' и условий Д6—Д7 сходимость распределения стандартизованной статистики $T^* = (T - \mathsf{M}T)/\sqrt{\mathsf{D}T}$ к закону $\mathcal{N}(0,1)$.

7* Проверьте формулу для $\mu_{m,n}$ из соотношений (11).

Указание. Используйте представление (10).

РЕШЕНИЯ ЗАДАЧ

1. Пусть $Z_{(i_1)} < \ldots < Z_{(i_p)}$ — положительные порядковые статистики. Тогда T является суммой рангов этих статистик относительно их абсолютных значений.

Изобразим круг с центром в начале координат и радиусом $Z_{(i_1)}$ (рис. 4). Тогда ранг $Z_{(i_1)}$ равен числу точек $Z_{(j)}$ в круге, включая $Z_{(i_1)}$, поскольку мы ранжируем расстояния от 0. Во второй сумме из (12) выделим слагаемые с $i = i_1$. Заметим, что полусуммы $(Z_{(i_1)} + Z_{(j)})/2$ $(1 \leqslant j \leqslant i)$ будут положительны только для $Z_{(j)}$, попавших в круг. Перебор всех $i = i_1,\ldots,i_p$ завершает доказательство.

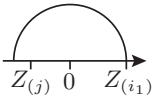


Рис. 4

2. Обозначим главный минор порядка n через D_n. Условием положительной определенности матрицы $\boldsymbol{\Sigma}$ является положительность всех D_i, $i = 1,\ldots,n$ (см. П10). Разложив D_n сначала по первой строке, а затем — по первому столбцу, получим рекуррентное соотношение

$$D_{n+2} = D_{n+1} - \rho^2 D_n, \tag{13}$$

которое представляет собой разностное уравнение 2-го порядка с постоянными коэффициентами (см. [42, с. 672]). Поскольку $D_1 = 1$ и $D_2 = 1 - \rho^2$, выводим из (13), что $D_0 = D_1 = 1$ (*начальные условия*). Так как в уравнении (13) участвует ρ^2, то его решения при ρ и $-\rho$ совпадают. Поэтому можно считать, что $\rho \geqslant 0$.

Если $0 \leqslant \rho < 1/2$, то характеристический многочлен $f(\lambda) = \lambda^2 - \lambda + \rho^2$ уравнения (13) имеет действительные корни $\lambda_{1,2} = \frac{1}{2}\,(1 \pm \sqrt{1 - 4\rho^2})$, причем $0 \leqslant \lambda_2 < \lambda_1 \leqslant 1$. Общее решение уравнения (13) в этом случае представляется в виде $a\lambda_1^n + b\lambda_2^n$, где константы a и b находятся из начальных условий. Ответ таков:

$$D_n = \left[(1 - \lambda_2)\lambda_1^n - (1 - \lambda_1)\lambda_2^n\right] / \,(\lambda_1 - \lambda_2). \tag{14}$$

Неравенство $0 \leqslant \lambda_2 < \lambda_1 \leqslant 1$ влечет неравенства $1 - \lambda_2 > 1 - \lambda_1$ и $\lambda_1^n > \lambda_2^n$. Поэтому правая часть равенства (14) положительна при всех n.

Если $\rho = 1/2$, то $\lambda_{1,2} = 1/2$. Общее решение уравнения (13) ищется в виде $(a+bn)\,\lambda_{1,2}^n$. Начальные условия выполняются для $D_n = (1 + n)\,2^{-n} > 0$.

Отметим, что если $\rho > 1/2$, то при достаточно большом n минор D_n станет *отрицательным*. Действительно, формула (14) и в этом случае дает решение уравнения (13), только теперь λ_1 и λ_2 — комплексные числа. Записав их в тригонометрической форме, используя формулу для синуса суммы, нетрудно вывести, что

$$D_n = \frac{\rho^n}{H(\rho)} \sin[(n+1)\,\arcsin H(\rho)], \quad \text{где } H(\rho) = \frac{\sqrt{4\rho^2 - 1}}{2\rho}\,.$$

Функция $H(\rho)$ монотонно отображает луч $(1/2, +\infty)$ на интервал $(0, 1)$. Введем обозначение $\varphi = \arcsin H(\rho)$. Тогда $\varphi \in (0,\pi/2)$. Очевидно, что при n^*, равном целой части от π/φ, аргумент синуса $(n^* + 1)\varphi$ впервые попадает в интервал $(\pi,2\pi)$, где D_n становится отрицательным.

3. Согласно формуле из П9, искомая вероятность равна

$$\int\limits_0^\infty \int\limits_0^\infty \frac{1}{2\pi\sqrt{1-\rho^2}}\,\exp\left\{-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right\} dxdy.$$

Переходя к полярным координатам $x = r\cos\varphi$, $y = r\sin\varphi$ с якобианом замены r, получим

$$\frac{1}{2\pi\sqrt{1-\rho^2}} \int\limits_0^{\pi/2} \left[\int\limits_0^\infty \exp\left\{-\frac{(1 - \rho\sin 2\varphi)}{2(1-\rho^2)}\,r^2\right\} r\,dr\right] d\varphi.$$

Вычислим внутренний интеграл и введем замену $\psi = 2\varphi$:

$$\frac{\sqrt{1-\rho^2}}{4\pi} \int\limits_0^\pi \frac{1}{1-\rho\sin\psi}\, d\psi.$$

Это табличный интеграл (см. [20, с. 66, 144]): при $|\rho| < 1$

$$\int\limits_0^\pi \frac{1}{1-\rho\sin\psi}\, d\psi = \frac{2}{\sqrt{1-\rho^2}}\ \operatorname{arctg}\ \frac{\operatorname{tg}\frac{\psi}{2}-\rho}{\sqrt{1-\rho^2}}\Bigg|_0^\pi = \frac{\pi + 2\ \arcsin\rho}{\sqrt{1-\rho^2}}\,.$$

4. Для $U_i = I_{\{X_i>0\}}$ по лемме о независимости из § 3 гл. 1 имеем

$$\mathsf{P}(U_i = u_i, |X_i| \leqslant x_i, i = 1,\ldots,n) = \prod_{i=1}^n \mathsf{P}(U_i = u_i, |X_i| \leqslant x_i).$$

Покажем, что перемножаемые вероятности также распадаются в произведения. Пусть $u_i = 1$. Тогда в силу непрерывности и симметрии относительно нуля функция распределения $F(x)$ случайной величины X_i имеем:

$$\mathsf{P}(U_i = 1, |X_i| \leqslant x) = \mathsf{P}(0 \leqslant X_i \leqslant x) = F(x) - F(0) =$$
$$= F(x) - \frac{1}{2} = \frac{1}{2}\,(2F(x) - 1) = \mathsf{P}(U_i = 1)\,\mathsf{P}(|X_i| \leqslant x).$$

Аналогично доказывается, что для любого x верно равенство

$$\mathsf{P}(U_i = 0, |X_i| \leqslant x) = \mathsf{P}(U_i = 0)\,\mathsf{P}(|X_i| \leqslant x).$$

Таким образом, $\boldsymbol{U} = (U_1,\ldots,U_n)$ и $(|X_1|,\ldots,|X_n|)$ независимы. Так как $\boldsymbol{R} = (R_1,\ldots,R_n)$ — вектор-функция от $(|X_1|,\ldots,|X_n|)$, то $\boldsymbol{U}$ и $\boldsymbol{R}$ также независимы.

5. Вектор антирангов $\boldsymbol{A} = (A_1,\ldots,A_n)$ — вектор-функция от $\boldsymbol{R}$. Действительно, образуем $(2 \times n)$-матрицу из столбцов (i,R_i) и переставим столбцы в порядке возрастания R_i. Тогда $\boldsymbol{A}$ — первая строка в полученной матрице:

$$\begin{pmatrix} 1 & \cdots & n \\ R_1 & \cdots & R_n \end{pmatrix} \to \begin{pmatrix} A_1 & \cdots & A_n \\ 1 & \cdots & n \end{pmatrix}.$$

$$\left[\text{Например,}\quad \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 1 & 4 \end{pmatrix} \to \begin{pmatrix} 3 & 1 & 2 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix}.\right]$$

Ввиду задачи 4 случайные векторы $\boldsymbol{U}$ и $\boldsymbol{A}$ независимы. Используя это и формулу полной вероятности (П7), запишем:

$$\mathsf{P}(\boldsymbol{W} = \boldsymbol{w}) = \sum_{\boldsymbol{a}} \mathsf{P}(\boldsymbol{W} = \boldsymbol{w} \mid \boldsymbol{A} = \boldsymbol{a})\,\mathsf{P}(\boldsymbol{A} = \boldsymbol{a}) =$$
$$= \sum_{\boldsymbol{a}} \mathsf{P}(I_{\{Z_{a_i}>0\}} = w_i, i = 1,\ldots,n)\,\mathsf{P}(\boldsymbol{A} = \boldsymbol{a}) =$$
$$= \left(\frac{1}{2}\right)^n \sum_{\boldsymbol{a}} \mathsf{P}(\boldsymbol{A} = \boldsymbol{a}) = 2^{-n},$$

что и требовалось установить.

6. Согласно определениям статистики T и случайных величин W_i имеем:

$$T = \sum_{i=1}^{n} R_i\, U_i = \sum_{i=1}^{n} R_i\, I_{\{Z_i>0\}} = \sum_{i=1}^{n} i\, I_{\{Z_{A_i}>0\}} = \sum_{i=1}^{n} i\, W_i.$$

Другими словами, T — линейная комбинация независимых и одинаково распределенных бернуллиевских случайных величин W_i. Отсюда по свойствам из П2 немедленно получаем, что

$$\mathbf{M}T = \sum_{i=1}^{n} i\, \mathbf{M}W_i = \frac{1}{2}\sum_{i=1}^{n} i = n(n+1)/4,$$
$$\mathbf{D}T = \sum_{i=1}^{n} i^2\, \mathbf{D}W_i = \frac{1}{4}\sum_{i=1}^{n} i^2 = n(n+1)(2n+1)/24.$$

Для установления асимптотической нормальности статистики T центрируем W_i (перейдем к $\xi_i = W_i - \mathbf{M}W_i = W_i - 1/2$) и используем следующее утверждение.

Теорема 1. *Пусть $\xi_1, \xi_2, \ldots$ — независимые и одинаково распределенные случайные величины, причем $\mathbf{M}\xi_1 = 0$, $0 < \sigma^2 = \mathbf{D}\xi_1 < \infty$. Рассмотрим $S_n = \sum\limits_{i=1}^{n} c_i\xi_i$, где $\{c_i\}$ — числовая последовательность. Если*

$$r_n = \frac{\max\{|c_1|, \ldots, |c_n|\}}{\sqrt{c_1^2 + \ldots + c_n^2}} \to 0 \quad \textit{при } n \to \infty,$$

то $S_n \,/\sqrt{\mathbf{D}S_n} \xrightarrow{d} Z \sim \mathcal{N}(0,1)$, где $\mathbf{D}S_n = \sigma^2(c_1^2 + \ldots + c_n^2)$.

Доказательство. Проверим условие Линдеберга (см. П6):

$$\frac{1}{\sigma^2 \sum c_i^2} \sum_{i=1}^{n} \mathbf{M}\left[c_i^2\xi_i^2\, I\left(|\xi_i| > \varepsilon\sigma\, \frac{\sqrt{\sum c_i^2}}{|c_i|}\right)\right] \leqslant$$
$$\leqslant \frac{1}{\sigma^2 \sum c_i^2} \sum_{i=1}^{n} \mathbf{M}\left[c_i^2\xi_1^2\, I\left(|\xi_1| > \varepsilon\sigma/r_n\right)\right] = \frac{1}{\sigma^2}\, \mathbf{M}\left[\xi_1^2\, I\left(|\xi_1| > \varepsilon\sigma/r_n\right)\right].$$

При всех n случайные величины под знаком последнего математического ожидания мажорируются величиной ξ_1^2 с $\mathbf{M}\xi_1^2 = \sigma^2 < \infty$. Так как $\varepsilon\sigma/r_n \to \infty$, то они сходятся к 0 при всех ω. Чтобы завершить доказательство, остается применить теорему Лебега о мажорируемой сходимости (П5). ■

7. Ввиду одинаковой распределенности случайных величин δ_i согласно формуле (10) имеем

$$\mathbf{M}T_n = 1 + (n-1)\, \mathbf{M}(\delta_1 \,|\, S_n = m). \tag{15}$$

Вычислим условное математическое ожидание (см. П7):

$$\mathbf{M}(\delta_1 \,|\, S_n = m) = \mathbf{P}(\delta_1 = 1 \,|\, S_n = m) =$$
$$= \frac{\mathbf{P}(\delta_1 = 1, S_n = m)}{\mathbf{P}(S_n = m)} = \frac{\mathbf{P}(\zeta_1 \neq \zeta_2, \zeta_3 + \ldots + \zeta_n = m-1)}{\mathbf{P}(\zeta_1 + \ldots + \zeta_n = m)} =$$
$$= \frac{2p(1-p)\, C_{n-2}^{m-1} p^{m-1}(1-p)^{l-1}}{C_n^m\, p^m (1-p)^l} = \frac{2\, C_{n-2}^{m-1}}{C_n^m} = \frac{2lm}{n(n-1)}\,,$$

где $l = n - m$. Остается только подставить его в (15).

ОТВЕТЫ НА ВОПРОСЫ

1. Аналогично вычислениям в примере 3 из (10) имеем

$$\mathbf{M}T_n = 1 + (n-1)\mathbf{M}\delta_1 = 1 + (n-1)\mathbf{P}(\zeta_1 \neq \zeta_2) = 1 + 2pq(n-1),$$

где $q = 1 - p$. Согласно свойству 3 дисперсии (П2) запишем

$$\mathbf{D}T_n = (n-1)\,\mathbf{D}\delta_1 + 2(n-2)\,\mathbf{cov}\,(\delta_1, \delta_2)\,. \tag{16}$$

В соответствии с определением случайных величин δ_i

$$\mathbf{D}\delta_1 = \mathbf{M}\delta_1^2 - (\mathbf{M}\delta_1)^2 = \mathbf{M}\delta_1 - 4p^2q^2 = 2pq(1-2pq),$$
$$\mathbf{cov}\,(\delta_1, \delta_2) = \mathbf{P}(\delta_1 = 1, \delta_2 = 1) - (\mathbf{M}\delta_1)^2 =$$
$$= \mathbf{P}(\zeta_1 = 0, \zeta_2 = 1, \zeta_3 = 0) + \mathbf{P}(\zeta_1 = 1, \zeta_2 = 0, \zeta_3 = 1) -$$
$$- 4p^2q^2 = q^2p + p^2q - 4p^2q^2 = pq(1-4pq).$$

Подставляя в формулу (16), получим

$$\mathbf{D}T_n = 2pq[(n-1)(1-2pq) + (n-2)(1-4pq)].$$

2. При $p = 1/2$ по формулам, выведенным в предыдущем ответе, $\mathbf{M}T_n = (n+1)/2$ и $\mathbf{D}T_n = (n-1)/4$. Для $n = 100$ и $T_n = 60$ величина $(T_n - \mathbf{M}T_n)/\sqrt{\mathbf{D}T_n} \approx 1{,}91$. По таблице Т2 искомая вероятность равна 0,028 (сравните с вопросом 1 гл. 12).

3. Условная вероятность любой последовательности из m единиц и $l = n - m$ нулей при условии $\{S_n = m\}$ одинакова и равна $p^m q^l / \mathbf{P}(S_n = m) = 1/C_n^m$. Пусть $k = 2i + 1$. Тогда имеется либо i серий из «0» и $i + 1$ серия из «1», либо $i + 1$ серия из «0» и i серий из «1». В первом случае разбиение l нулей на i (непустых) групп можно осуществить C_{l-1}^{i-1} способами, m единиц на $i + 1$ группу — C_{m-1}^{i} способами (см. вопрос 3 гл. 10).

Глава 16

НЕСКОЛЬКО НЕЗАВИСИМЫХ ВЫБОРОК

В этой главе критерий ранговых сумм Уилкоксона—Манна—Уитни из § 5 гл. 14 обобщается на случай, когда данные состоят из нескольких рядов наблюдений (*обработок*), которые рассматриваются как реализации *независимых между собой* выборок. Исходная гипотеза H_0 говорит об отсутствии различия в обработках, т. е. предполагается, что все наблюдения можно считать одной выборкой из общей совокупности.

Не в совокупности ищи единства, но более — в единообразии разделения.

Козьма Прутков

§ 1. ОДНОФАКТОРНАЯ МОДЕЛЬ

Данные. Данные состоят из $N = \sum_{j=1}^{k} n_j$ наблюдений x_{ij}, по n_j наблюдений в j-й выборке (обработке), $j = 1, \ldots, k$. Будем считать их реализацией случайных величин X_{ij}, где

$$X_{ij} = \mu + \beta_j + \varepsilon_{ij}, \quad i = 1, \ldots, n_j; \;\; j = 1, \ldots, k. \tag{1}$$

Здесь μ — (неизвестное) общее среднее, β_j — (неизвестный) эффект от воздействия фактора для j-й выборки, ε_{ij} — случайная ошибка. Положим $\mu_j = \mu + \beta_j$.

Обработки			
1	2	...	k
x_{11}	x_{12}	...	x_{1k}
x_{21}	x_{22}	...	x_{2k}
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_{n_1 1}$	$x_{n_2 2}$	...	$x_{n_k k}$

Допущения

Д1. *Все ошибки ε_{ij} независимы.*

Д2. *Все ε_{ij} имеют одинаковое непрерывное (неизвестное) распределение.*

Для проверки *гипотезы однородности*

$H_0 \colon \mu_1 = \ldots = \mu_k$

можно использовать критерий Краскела—Уоллиса (§ 2) или критерий Джонкхиера (§ 3).

§ 2. КРИТЕРИЙ КРАСКЕЛА—УОЛЛИСА

Критерий Краскела—Уоллиса (см. [88, с. 131]) применяется для проверки гипотезы H_0 против альтернативы

$H_1 \colon$ не все μ_j равны между собой.

Выполним **следующие шаги**.

1. Ранжируем все N наблюдений вместе от меньшего к большему. Пусть R_{ij} обозначает ранг наблюдения X_{ij} в этой совместной ранжировке.

2. Положим для $j = 1, \ldots, k$

$$S_j = \sum_{i=1}^{n_j} R_{ij},\ R_{\cdot j} = S_j/n_j,\ R_{\cdot\cdot} = \frac{1}{N}\sum R_{ij} = \frac{1}{N}\ \frac{N(N+1)}{2} = \frac{N+1}{2}\,.$$

Таким образом, $R_{\cdot j}$ — это средний ранг наблюдений X_{ij}, относящихся к обработке j, $R_{\cdot\cdot}$ — общий средний ранг.

3. Найдем значение *статистики критерия Краскела—Уоллиса* W, определяемой формулой

$$W = \frac{12}{N(N+1)} \sum_{j=1}^{k} n_j (R_{\cdot j} - R_{\cdot\cdot})^2 = \left[\frac{12}{N(N+1)} \sum_{j=1}^{k} S_j^2/n_j\right] - 3\,(N+1).$$

Если гипотеза H_0 верна, то все k выборок берутся из общей совокупности, поэтому величины $R_{\cdot j}$ не должны сильно отличаться от $R_{\cdot\cdot}$. Это объясняет, почему большие значения статистики W противоречат гипотезе H_0.

Малые выборки. Гипотезу H_0 следует отвергнуть, если наблюдаемое значение w статистики W окажется больше или равно w_α, где критическая граница w_α для заданного уровня значимости α (см. § 1 гл. 12) определяется по таблице $A.7$ книги [88] (к сожалению, только для $k = 3$ и $1 \leqslant n_j \leqslant 5$).

Большие выборки. Если гипотеза H_0 верна, то при

$$\min\{n_1, \ldots, n_k\} \to \infty$$

статистика W имеет в качестве предельного закона распределение хи-квадрат с $k-1$ степенями свободы (см. [86, с. 190]). Приближенный критерий уровня α таков: отклонить гипотезу H_0, если $w \geqslant z_{1-\alpha}$, где $z_{1-\alpha}$ — это $(1-\alpha)$-квантиль χ^2_{k-1}-распределения (см. таблицу Т3); в противном случае — принять гипотезу H_0.

Поправка. Для выборок среднего размера точность приближения может оказаться недостаточной. Например, согласно примечанию переводчика на с. 132 книги [88], при $\alpha = 5\%$ для $k = 2$ и $n_1 = n_2 = 4, 5, 6$ относительная погрешность ошибки I рода превосходит 33%. Следующая поправка статистики W, предложенная Р. Иманом и Дж. Давенпортом (1976 г.), позволяет существенно уменьшить эту погрешность (для указанного случая — в среднем в 5–6 раз). Именно, в качестве статистики для проверки однород-

ности k выборок возьмем

$$\widetilde{W} = \frac{1}{2}\, W\,[(N-k)/(N-1-W)\,+\,1]\,. \tag{2}$$

Приближенное критическое значение уровня α для нее равно

$$\widetilde{w}_\alpha = \frac{1}{2}\,[z_{1-\alpha}\,+\,(k-1)f_{1-\alpha}]\,, \tag{3}$$

где $z_{1-\alpha}$ и $f_{1-\alpha}$ — $(1-\alpha)$-квантили, соответственно, закона χ^2_{k-1} (см. табл. Т3) и распределения Фишера—Снедекора $F_{k-1,N-k}$ (см. табл. Т5).

Совпадения. Если среди x_{ij} встречаются одинаковые значения, то для вычисления W надо брать средние ранги, а затем заменить W на

$$W' = W/\gamma, \quad \text{где } \gamma = 1 - \frac{1}{N(N^2-1)} \sum_{m=1}^{g} l_m(l_m^2-1),$$

g — число групп совпадений, l_m — количество элементов в m-й группе. Не совпадающие с другими x_{ij} считаются группой размера 1.

Сравнение обработок. Для того чтобы узнать, какие из обработок отличаются друг от друга, О. Данн (1964 г.) предложил следующий приближенный критерий уровня α: принять решение $\mu_r \neq \mu_s$, если

$$|R_{\,\cdot r} - R_{\,\cdot s}| \geqslant C_{rs} = x_p \left[\frac{N(N+1)}{12}\right]^{1/2} (1/n_r + 1/n_s)^{1/2}, \tag{4}$$

где $p = 1-\alpha/[k(k-1)]$, x_p — p-квантиль стандартного нормального закона $\mathcal{N}(0,1)$ (см. табл. Т2).

Оценка контраста. Для значимо различающихся обработок с номерами r и s представляет интерес *контраст* $\Delta_{rs} = \mu_r - \mu_s$. В качестве *первичной оценки* для него возьмем оценку параметра сдвига, задаваемую формулой (10) гл. 14:

$$V_{rs} = MED\,\{X_{ir} - X_{ls},\, 1 \leqslant i \leqslant n_r, 1 \leqslant l \leqslant n_s\}, \quad r \neq s; \quad V_{rr} = 0.$$

(Отметим, что достаточно подсчитать лишь $k(k-1)/2$ значений статистик V_{rs} для $r < s$, поскольку $V_{sr} = -V_{rs}$.)

Далее, вычислим *взвешенные суммы*

$$W_r = \sum_{s=1}^{k} n_s V_{rs} \Big/ \sum_{s=1}^{k} n_s = \frac{1}{N} \sum_{s=1}^{k} n_s V_{rs}, \quad r = 1,\ldots,k. \tag{5}$$

Наконец, определим *уточненную оценку контраста* как

$$\widehat{\Delta}_{rs} = W_r - W_s. \tag{6}$$

Пример 1. ***Содержание влаги*** **[83, с. 368].** Были взяты 14 образцов некоторого продукта, которые случайным образом разбили на пять групп заданных размеров. Все группы хранились в разных условиях, а после хранения у всех образцов определили содержание влаги. Данные (в %) приведены в следующей таблице (в скобках указан соответствующий ранг R_{ij}):

Условия хранения продукта				
1	2	3	4	5
7,8 (7) 8,3 (10,5) 7,6 (6) 8,4 (12) 8,3 (10,5)	5,4 (1) 7,4 (5) 7,1 (3,5)	8,1 (9) 6,4 (2)	7,9 (8) 9,5 (13) 10,0 (14)	7,1 (3,5)
$S_1 = 46$	$S_2 = 9{,}5$	$S_3 = 11$	$S_4 = 35$	$S_5 = 3{,}5$
$R_{\cdot 1} = 9{,}2$	$R_{\cdot 2} = 3{,}17$	$R_{\cdot 3} = 5{,}5$	$R_{\cdot 4} = 11{,}67$	$R_{\cdot 5} = 3{,}5$

Статистика Краскела—Уоллиса W для этих данных принимает значение 8,39. Учитывая два совпадения, получаем $W' \approx 8{,}43$. Для закона χ^2_4 по табл. Т3 линейной интерполяцией находим, что приближенный фактический уровень значимости $\alpha_0 \approx 8\%$.

Так как размеры выборок малы, сделаем поправку Имана и Давенпорта. В соответствии с формулой (2) статистика $\widetilde{W} \approx 12{,}51$. Зададим уровень значимости $\alpha = 5\%$. Найдем квантили $z_{1-\alpha}$ и $f_{1-\alpha}$, участвующие в формуле (3). Из табл. Т3 для $k - 1 = 4$ берем $z_{1-\alpha} = 9{,}49$. Для $k_1 = 4$ и $k_2 = 9$ согласно табл. Т5 имеем $f_{1-\alpha} \approx 3{,}63$. Отсюда $\widetilde{w}_\alpha \approx 12{,}01 < 12{,}51$. Следовательно, гипотеза однородности отвергается даже на уровне значимости 5%. (Отметим, что относительная ошибка найденного ранее приближенного фактического уровня значимости $\alpha_0 = 8\%$ составляет $(8 - 5)/5 = 60\%$.)

Для выяснения того, какие же из способов хранения значимо отличаются друг от друга, применим приближенный метод Данна (см. формулу (4)). Поскольку n_j малы, зададим не 5%-ный, а больший уровень значимости.*) Возьмем, скажем, $\alpha = 0{,}15$. Тогда для $p = 0{,}9925$ из [10, с. 116] извлекаем $x_p \approx 2{,}43$. Все значения C_{rs} $(r < s)$ указаны в таблице:

r, s	1, 2	1, 3	1, 4	1, 5	2, 3	2, 4	2, 5	3, 4	3, 5	4, 5
$\lvert R_{\cdot r} - R_{\cdot s}\rvert$	6,03	3,70	2,47	5,70	2,33	8,50	0,33	6,17	2,00	8,17
C_{rs}	7,42	8,51	7,42	11,14	9,28	8,30	11,74	9,28	12,45	11,74

*) Когда данных мало, функция дискретного распределения статистики $|R_{\cdot r} - R_{\cdot s}|$ растет большими «скачка́ми». Поэтому неразумно устанавливать слишком жесткие условия.

Неравенство (4) имеет место только для пары $(r, s) = (2, 4)$. Таким образом, при вероятности ошибочного решения 0,15 способы хранения 2 и 4 различаются значимо.

s	Первичные оценки		n_s
1	$V_{41} = 1{,}2$	$V_{21} = -1{,}2$	5
2	$V_{42} = 2{,}5$	$V_{22} = 0$	3
3	$V_{43} = 1{,}7$	$V_{23} = -0{,}85$	2
4	$V_{44} = 0$	$V_{24} = -2{,}5$	3
5	$V_{45} = 2{,}4$	$V_{25} = 0$	1
	$W_4 = 1{,}38$	$W_2 = -1{,}09$	

Оценим контраст Δ_{42}. Приведем значения необходимых для этого первичных оценок V_{rs} и их взвешенных средних W_r. В данном случае значение уточненной оценки контраста $\widehat{\Delta}_{42} = W_4 - W_2 = 2{,}47$ мало отличается от значения первичной: $V_{42} = 2{,}5$.

Комментарии

1) Распределение статистики W при условии справедливости гипотезы H_0 можно получить из того, что в этом случае все $N!/(n_1! \dots n_k!)$ возможных наборов по n_1 рангов для первой выборки, …, n_k рангов для k-й выборки равновероятны (см. формулу 3 гл. 10). На каждом наборе подсчитывается значение W и заносится в таблицу.

Вопрос 1. Чему равна $\mathbf{P}(W \geqslant 2)$ при $k = n_1 = n_2 = 2$?

2) Надо иметь в виду, что неоднородность некоторой пары выборок может быть замаскирована присутствием других выборок в таблице данных (см. [88, с. 135]). Как заметил К. Габриэль (1969 г.), статистика W имеет тот недостаток, что ее значение w_{sub}, вычисленное для некоторого подмножества выборок, может превзойти значение w_{tot} для всех выборок. Например, пусть выборке с номером 1 соответствуют ранги 8, 9, 10, 11; выборке с номером 2 — 1, 2, 6, 7; выборке с номером 3 — 3, 4, 5, 12. Тогда для выборок с номерами 1 и 2 ($k = 2$) $w_{sub} \approx 5{,}333$, а для всех трех выборок $w_{tot} \approx 4{,}769$. (Отметим, что этот же изъян присущ и статистике Фридмана из следующей главы.)

3) Приближенные критические границы C_{rs} в формуле (4) рассчитаны в предположении, что гипотеза однородности H_0 верна. Когда она не верна, способность метода Данна обнаруживать значимо различные обработки резко уменьшается с ростом k. Это хорошо видно из примера 1, где $k = 5$. Хотя на уровне 0,15 мы и приняли решение $\mu_2 \neq \mu_4$, уже на уровне 0,1 ($x_p = 2{,}58$ и $C_{24} = 8{,}81 > 8{,}50$) мы не смогли бы сделать этого.*) В защиту

*) Критерий ранговых сумм Уилкоксона—Манна—Уитни, примененный к обработкам 2 и 4 для проверки гипотезы H_0 против односторонней альтернативы $\mu_2 < \mu_4$, имеет фактический уровень значимости $\alpha_0 = 0{,}05$.

столь консервативного подхода приведем (см. [88, с. 145]) описание ситуации, где встречается другая крайность — тенденция различать на самом деле неразличимое.

Статистика может доказать что угодно, даже правду.

Ноэл Мойнихан

«Рекламируя по телевидению свое новое лекарство как панацею, каждая фирма непременно заявляет, что при испытаниях оно показало себя эффективнее всех известных препаратов. Целью таких испытаний (если, конечно, они не откровенно подделаны) является вовсе не помощь изготовителю в принятии решения, а лишь предоставление ему возможности оптимистически объявить о проведенном сравнении, дабы произвести впечатление на публику и увеличить продажу. Правильнее было бы потребовать от него использовать для статистического анализа метод множественных сравнений. Чем более одинаково неэффективны были бы хваленые лекарства, тем труднее было бы сделать убедительный вывод и тем большее число испытаний пришлось бы провести, чтобы прийти к желаемому заключению».

4) При оценке контрастов с помощью первичных оценок V_{rs} мы сталкиваемся с тем неприятным обстоятельством, что они не удовлетворяют линейным соотношениям, которые выполняются для самих контрастов. Так, $\Delta_{42} = \Delta_{41} + \Delta_{12}$, но в общем случае $V_{42} \neq V_{41} + V_{12}$ ($2{,}5 \neq 1{,}2 + 1{,}2 = 2{,}4$ в примере 1). На эту несогласованность величин V_{rs} обратил внимание Э. Леман. Э. Спетволль (1968 г.) предложил уточненные оценки $\widehat{\Delta}_{rs}$ (см. формулы (5) и (6)), которые не только согласованы, но и состоятельны, когда n_r и n_s стремятся к бесконечности, а остальные n_j фиксированы. Недостатком оценки $\widehat{\Delta}_{rs}$ является ее зависимость от выборок с номерами, отличными от r и s.

5) В случае, когда функция F_ε распределения ошибок ε_{ij} в модели (1) является нормальной, асимптотическая эффективность критерия Краскела—Уоллиса по отношению к оптимальному F-критерию однофакторного дисперсионного анализа (статистика которого задается равенством (9) в примере 2 ниже) равна величине $E(F_\varepsilon)$ из формулы (11) гл. 14.

Если справедливо допущение, что наблюдения X_{ij} имеют *нормальное распределение* (или очень похожее на него), то можно воспользоваться критериями из следующего примера.

Пример 2. ***Проверка однородности независимых нормальных выборок.*** Пусть все $X_{ij} \sim \mathcal{N}(\mu_j, \sigma_j^2)$ $(i = 1, \ldots, n_j, j = 1, \ldots, k)$ независимы, причем параметры μ_j и σ_j неизвестны. Несмещенными оценками для μ_j и σ_j^2 являются (см. пример 3 гл. 6)

$$X_{\cdot j} = \frac{1}{n_j} \sum_{i=1}^{n_j} X_{ij} \quad \text{и} \quad S_j^2 = \frac{1}{n_j - 1} \sum_{i=1}^{n_j} (X_{ij} - X_{\cdot j})^2. \tag{7}$$

Положим $N = n_1 + \ldots + n_k$. Для проверки гипотезы

H': $\sigma_1 = \ldots = \sigma_k$, $\mu_1, \ldots, \mu_k$ — любые,

обычно используется **критерий Бартлетта**, статистикой которого служит отношение взвешенных среднего арифметического и среднего геометрического величин $S_1^2, \ldots, S_k^2$:

$$B = \left(\frac{1}{N} \sum_{j=1}^{k} n_j S_j^2 \right) \Bigg/ \sqrt[N]{\prod_{j=1}^{k} (S_j^2)^{n_j}}. \tag{8}$$

Если выполняется гипотеза H' и все $n_j > 3$, то статистика

$$B^* = \gamma^{-1} N \ln B, \quad \text{где } \gamma = 1 + \frac{1}{3(k-1)} \left[\left(\sum_{j=1}^{k} \frac{1}{n_j} \right) - \frac{1}{N} \right],$$

приближенно имеет χ^2_{k-1}-распределение (см. [10, с. 47]). Можно показать, что критерий Бартлетта обобщает критерий Фишера из примера 1 гл. 14 (задача 6).

Вопрос 2. Почему большие значения B противоречат H'?

Когда гипотеза H' принимается, для установления однородности выборок остается убедиться, что верна гипотеза

$$H'' \colon \mu_1 = \ldots = \mu_k.$$

Для ее проверки используется ***F*-критерий однофакторного дисперсионного анализа** (см. § 4 гл. 21), основанный на отношении

$$R = \frac{\dfrac{1}{k-1} \displaystyle\sum_{j=1}^{k} n_j (X_{\cdot j} - X_{\cdot\cdot})^2}{\dfrac{1}{N-k} \displaystyle\sum_{j=1}^{k} \sum_{i=1}^{n_j} (X_{ij} - X_{\cdot j})^2}, \quad \text{где } X_{\cdot\cdot} = \frac{1}{N} \sum_{j=1}^{k} \sum_{i=1}^{n_j} X_{ij}, \tag{9}$$

которое при справедливости гипотезы H'' распределено в точности по закону $F_{k-1,N-k}$ (см. пример 1 гл. 14) для любых $n_j > 1$.

Доказательство. По условию $\sigma_1 = \ldots = \sigma_k = \sigma$. Так как статистика R не зависит от σ, то без ограничения общности будем считать, что $\sigma = 1$. Для фиксированного j рассмотрим случайную величину S_j^2 из формулы (7). По теореме 1 гл. 11 имеем

$$(n_j - 1)\, S_j^2 = \sum_{i=1}^{n_j} (X_{ij} - X_{\cdot j})^2 \sim \chi^2_{n_j - 1}. \tag{10}$$

Так как выборки независимы и хи-квадрат является частным случаем гамма-распределения, из соотношения (10) и леммы 1 гл. 4 вытекает, что

$$V_{int} = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (X_{ij} - X_{\cdot j})^2 \sim \chi^2_{N-k}. \tag{11}$$

V_{int}: от *англ.* variability — изменчивость, interior — внутренний.

Здесь V_{int} — *мера общей изменчивости внутри выборок.*

С другой стороны, случайные величины $X_{\cdot j} \sim \mathcal{N}(\mu_j, 1/n_j)$ независимы между собой. Поэтому (см. задачу 4) при справедливости

гипотезы H'' статистика

$$V_{out} = \sum_{j=1}^{k} n_j (X_{\cdot j} - X_{\cdot\cdot})^2 \sim \chi^2_{k-1}. \tag{12}$$

V_{out}: outside (*англ.*) — внешний.

Здесь V_{out} — *мера разброса между выборками.*

Ввиду независимости выборок и теоремы 1 гл. 11 $X_{\cdot 1}, \ldots, X_{\cdot k}$ не зависят от $S_1, \ldots, S_k$. Поскольку $X_{\cdot\cdot} = \dfrac{1}{N} \sum\limits_{j=1}^{k} n_j X_{\cdot j}$, т. е. является линейной комбинацией $X_{\cdot j}$, из леммы о независимости из § 3 гл. 1 вытекает независимость V_{int} и V_{out}. Использование определения закона Фишера—Снедекора завершает доказательство. ■

Замечание 1. Для любых X_{ij} верно тождество (задача 3)

$$V_{tot} = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (X_{ij} - X_{\cdot\cdot})^2 = V_{int} + V_{out}. \tag{13}$$

V_{tot}: total (*англ.*) — общий.

Таким образом, *общая изменчивость (разброс, дисперсия)* V_{tot} распадается на слагаемые, каждое из которых представляет свой источник изменчивости данных. Отсюда и происходит название «*дисперсионный анализ*».

Отметим, что статистика критерия Краскела—Уоллиса W является нормированной величиной V_{out} для рангов R_{ij}, статистика F-критерия R из формулы (9) — это отношение нормированных величин V_{out} и V_{int}.

Вопрос 3. Как связана R при $k = 2$ с T из примера 1 гл. 14?

Замечание 2. Критерий Бартлетта *весьма чувствителен* даже к небольшим отклонениям распределения элементов выборок от нормального. Так, допустим, что все наблюдения распределены по закону Стьюдента t_7 с 7 степенями свободы, которое очень похоже на нормальный закон $\mathcal{N}(0,1)$ (см. рис. 7 гл. 11). В следующей таблице приведены некоторые значения (взятые из [10, с. 114, 174]) функций распределения этих законов и разности $\Delta(x)$ между ними.

x	0,0	0,5	1,0	1,5	2,0	2,5	3,0
t_7	0,5	0,683	0,825	0,911	0,957	0,980	0,990
$\mathcal{N}(0,1)$	0,5	0,691	0,841	0,933	0,977	0,994	0,999
$\Delta(x)$	0	0,008	0,016	0,022	0,020	0,014	0,009

Наибольшее отличие $\Delta_{max} = 0{,}022$ достигается при $x \approx 1{,}5$.

Обозначим через b_α критическое значение уровня α статистики Бартлетта B, т. е. $\mathbf{P}(B \geqslant b_\alpha) = \alpha$ при справедливости гипотезы H'. Г. Бокс (1953 г.) установил, что даже для больших выборок при замене распределения $\mathcal{N}(0,1)$ на закон t_7 эта вероятность *меняется драматически* (на рис. 1 изображен сдвиг плотности B при такой

замене). Например, для уровня значимости $\alpha = 5\%$ получаем следующую картину:

Законы	Количество выборок		
	$k=2$	$k=5$	$k=10$
t_7	17%	32%	49%
$\mathcal{N}(0,1)$	5%	5%	5%

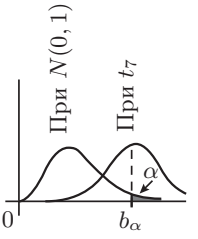


Рис. 1

Статистик, уверенный в том, что использует уровень 5%, в действительности может иметь дело с уровнем 49%! Приведем цитату на затронутую тему из [84, с. 38]:

> «Предполагалось, что отклонения от идеальных моделей можно игнорировать как несущественные; что статистические процедуры, оптимальные в строгой модели, останутся примерно таковыми и в приближенной модели. К сожалению, оказалось, что такие надежды зачастую не имеют под собой никакой почвы; даже безобидные отклонения часто имеют следствием эффекты гораздо более сильные, нежели это предвидело большинство статистиков».

§ 3. КРИТЕРИЙ ДЖОНКХИЕРА

На практике часто встречается ситуация, когда исследователь пытается выявить значимое возрастание (или убывание) уровня интересующего его фактора от выборки к выборке. В этом случае надо применять не критерий Краскела—Уоллиса, а более чувствительный *критерий Джонкхиера* ([88, с. 136]).*)

Он используется для проверки гипотезы однородности H_0 против *альтернативы возрастания влияния фактора*

$$H_2\colon \mu_1 \leqslant \ldots \leqslant \mu_k, \tag{14}$$

где хотя бы одно из неравенств строгое.

Выполняются **следующие шаги**.

1. Вычисляются $k(k-1)/2$ значений статистики Манна—Уитни U_{rs}, $1 \leqslant r < s \leqslant k$ (см. § 5 гл. 14), где

$$U_{rs} = \sum_{i=1}^{n_r} \sum_{l=1}^{n_s} I_{\{X_{ir} < X_{ls}\}}. \tag{15}$$

2. В качестве *статистики критерия Джонкхиера* берется

$$J = \sum_{r<s} U_{rs} = \sum_{r=1}^{k-1} \sum_{s=r+1}^{k} U_{rs}. \tag{16}$$

*) Критерий был предложен в работах Т. Терпстры (1952) и независимо А. Джонкхиера (1954) (ссылки на работы см. в [88, с. 140]).

Малые выборки. Гипотеза H_0 отвергается, если наблюдаемое значение статистики J окажется не меньше критической величины t_α, которую для $k = 3$ и $2 \leqslant n_j \leqslant 8$ можно найти в таблице $A.8$ книги [88].

Большие выборки. Пусть $J^* = (J - \mathbf{M}J)/\sqrt{\mathbf{D}J}$, где

$$\mathbf{M}J = \sum_{r<s} \mathbf{M}U_{rs} = \frac{1}{2}\sum_{r<s} n_r n_s = \frac{1}{4}\left(N^2 - \sum_{j=1}^{k} n_j^2\right),$$

$$\mathbf{D}J = \frac{1}{72}\left[N^2(2N+3) - \sum_{j=1}^{k} n_j^2(2n_j+3)\right], \quad N = \sum_{j=1}^{k} n_j.$$

Если верна гипотеза H_0, то статистика J^* имеет асимптотическое распределение $\mathcal{N}(0,1)$ при $\min\{n_1,\ldots,n_k\} \to \infty$ (см. [86, с. 199]). Обозначим через $x_{1-\alpha}$ квантиль уровня $1-\alpha$ для закона $\mathcal{N}(0,1)$. В случае, когда наблюдаемое значение J^* больше или равно $x_{1-\alpha}$, гипотеза H_0 отвергается, в противном случае — принимается.

Совпадения. Для учета совпадений следует заменить индикаторы в (15) на $I_{\{X_{ir}<X_{ls}\}} + \frac{1}{2}\, I_{\{X_{ir}=X_{ls}\}}$, чтобы в случае равенства значений сумма дополнительно увеличивалась на $\frac{1}{2}$.

Сравнение обработок с контрольной. Такая задача возникает, например, при исследовании эффективности ряда новых методов лечения (лекарств), предназначенных для улучшения принятого ранее стандартного метода. (Разумеется, позже можно сравнить друг с другом те обработки, которые оказались значимо лучше контрольной.)

Пусть роль контроля играет обработка 1. *Приближенный критерий Данна* уровня α выглядит так: следует принять решение $\mu_j > \mu_1$, если

$$|R_{\cdot j} - R_{\cdot 1}| \geqslant D_j = x_p \left[\frac{N(N+1)}{12}\right]^{1/2} (1/n_j + 1/n_1)^{1/2}, \tag{17}$$

где $p = 1 - \alpha/(k-1)$, x_p — p-квантиль закона $\mathcal{N}(0,1)$ (табл. Т2).

Пример 3. ***Роль мотивации*** [88, с. 137]**.** П. Хандел (1969 г.) исследовал влияние чистой мотивации (знания цели работы) на выполнение монотонных производственных операций (вытачивание металлических заготовок определенных форм и размеров). 18 мужчин были случайным образом разделены на 3 группы. Рабочие, попавшие в контрольную группу A, не имели информации о требуемой производительности, в группе B они получили лишь общее представление о том, что должны делать, наконец,

в группе C рабочие имели точную информацию о задании и могли контролировать себя по графику, лежащему перед ними. В таблице приведены числа заготовок, обработанных каждым из рабочих за время эксперимента (в скобках указаны ранги R_{ij}).*)

Группа A	Группа B	Группа C
40 (5,5)	38 (2,5)	48 (18)
35 (1)	40 (5,5)	40 (5,5)
38 (2,5)	47 (17)	45 (15)
43 (10,5)	44 (13)	43 (10,5)
44 (13)	40 (5,5)	46 (16)
41 (8)	42 (9)	44 (13)
$S_1 = 40{,}5$	$S_2 = 52{,}5$	$S_3 = 78$
$R_{\cdot 1} = 6{,}75$	$R_{\cdot 2} = 8{,}75$	$R_{\cdot 3} = 13$

Поскольку мы ожидаем таких отклонений от H_0, при которых производительность растет с осведомленностью, применим критерий Джонкхиера. По формуле (15) с учетом совпадений получаем

$$U_{12} = 22,\ U_{13} = 30{,}5,\ U_{23} = 26{,}5.$$

Согласно (16) имеем $J = 22 + 30{,}5 + 26{,}5 = 79$.

По [88, табл. A.8] находим, что фактический уровень значимости $\alpha_0 = 0{,}023$. Теперь применим приближение для больших выборок и сравним с тем, что дал точный критерий. Значение $J^* \approx 2{,}02$. В табл. Т2 ему соответствует уровень 0,022.

Сравним группы B и C с контрольной. Из приведенной выше таблицы имеем $R_{\cdot 2} - R_{\cdot 1} = 2$, $R_{\cdot 3} - R_{\cdot 1} = 6{,}25$. Для $\alpha = 0{,}05$ квантиль $x_p = 1{,}96$, $D_2 = D_3 \approx 6{,}04$. Поэтому группа C значимо отличается от контрольной, а группа B — нет.

Наконец, оценим контраст Δ_{31}. Из таблицы данных находим, что $V_{12} = -1{,}5$, $V_{13} = -4$ и $V_{23} = -3$. По формулам (5) и (6) вычисляем $W_1 = -11/6$, $W_3 = 7/3$ и $\widehat{\Delta}_{31} = W_3 - W_1 = 25/6 \approx 4{,}17 \neq \neq V_{31} = 4$.

Комментарии

1) Нетрудно видеть, что величину J можно вычислить по совместной ранжировке всех N наблюдений. Таким образом, хотя для подсчета J и не нужна совместная ранжировка, зная ее и не зная самих x_{ij}, можно восстановить значение J. Поэтому распределение случайной величины J при условии справедливости гипотезы H_0 можно найти тем же способом, что и распределение статистики Краскела—Уоллиса W: все $N!/(n_1! \dots n_k!)$ возможных наборов рангов при выполнении гипотезы H_0 равновероятны; для каждого из них вычисляется значение J.

*) Ранжировка нужна для сравнения групп B и C с группой A на основе (17).

2) Для обеспечения состоятельности критерия Джонкхиера против альтернативы H_2 (см. условие (14)) достаточно, чтобы $N \to \infty$ и $n_j/N \to \gamma_j$, где $0 < \gamma_j < 1$, $j = 1, \ldots, k$.

§ 4. БЛУЖДАНИЕ НА ПЛОСКОСТИ И В ПРОСТРАНСТВЕ

Материал этого параграфа продолжает тему поведения траекторий случайных блужданий из § 6 гл. 14, только теперь частица будет перемещаться по точкам с целыми координатами на плоскости или в пространстве (блуждание по k-мерной целочисленной решетке рассматривается в § 4 гл. 17).

Случайным будем называть такое блуждание, при котором частица переходит в одну из $2k$ соседних по осям координат точек с вероятностью $\dfrac{1}{2k}$ независимо от своего положения (на рис. 2 приведена возможная траектория блуждания*) для $k = 2$). Пусть, как и в § 6 гл. 14, u_{2n} — это *вероятность вернуться в начало координат на* $2n$*-м шаге*, f_{2n} — *вероятность того, что первое возвращение в начало координат произошло в момент* $2n$.

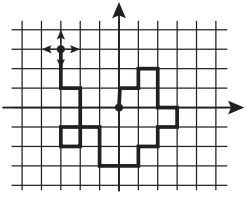


Рис. 2

Лемма. При $n \geqslant 1$ верно равенство $u_{2n} = \sum\limits_{i=1}^{n} f_{2i} u_{2n-2i}$ (здесь $u_0 = 1$).

Доказательство. Общее число путей за время $2n$ очевидно равно $(2k)^{2n}$. Попадание в начало в момент $2n$ может быть либо первым возвращением, либо первое возвращение произошло в некоторый момент $2i < 2n$ и далее через $(2n-2i)$ шагов частица вновь вернется в начало. Вероятность последнего события при фиксированном i равна $f_{2i}u_{2n-2i}$, так как имеется $(2k)^{2i} f_{2i}$ путей длины $2i$, оканчивающихся с первым возвращением, и $(2k)^{2n-2i} u_{2n-2i}$ путей из начала в начало длины $(2n - 2i)$. При разных i эти события несовместны. Складывая их вероятности, выводим доказываемую формулу. ■

Установим ряд интересных результатов с ее помощью.

Теорема 1. Пусть $k = 1$. Обозначим через δ_{2n} время, в течение которого блуждающая по прямой частица, совершая $2n$ шагов, находилась правее нуля. Тогда при $n \geqslant 1$ $\beta_{2i,2n} \equiv \mathbf{P}(\delta_{2n} = 2i) = u_{2i}\, u_{2n-2i}$, $i = 0,1,\ldots,n$ (см. теорему 3 гл. 14).

Доказательство. § 6 гл. 14.) Согласно теореме 1 гл. 14 $\beta_{2n,2n} = u_{2n}$. В силу симметрии имеем также $\beta_{0,2n} = u_{2n}$. Поэтому достаточно доказать теорему для $1 \leqslant i \leqslant n - 1$.

*) В отличие от терминологии § 6 гл. 14, траекторией (путем) теперь станем называть не развертку во времени, а ломаную в $\mathbb{R}^k$, соединяющую последовательные положения блуждающей частицы.

Пусть для такого i в течение ровно $2i$ из $2n$ шагов частица находилась правее нуля. При этом первое возвращение в нуль должно осуществиться в некоторый момент времени $2r < 2n$, и имеются две возможности: либо частица до этого момента частица находилась правее нуля, либо она была левее нуля. В первом случае $1 \leqslant r \leqslant i$, и на участке пути после $2r$ частица находилась правее нуля в течение $(2i - 2r)$ шагов из $(2n - 2r)$. Всего таких путей

$$\frac{1}{2}\, 2^{2r} f_{2r} \cdot 2^{2n-2r} \beta_{2i-2r,2n-2r}.$$

Во втором случае на участке пути после $2r$ частица находилась правее нуля в течение $2i$ шагов из $(2n-2r)$, откуда $r \leqslant n-i$. Таких путей

$$\frac{1}{2}\, 2^{2r} f_{2r} \cdot 2^{2n-2r} \beta_{2i,2n-2r}.$$

Следовательно, при $1 \leqslant i \leqslant n-1$ имеем соотношение

$$\beta_{2i,2n} = \frac{1}{2} \sum_{r=1}^{i} f_{2r} \beta_{2i-2r,2n-2r} + \frac{1}{2} \sum_{r=1}^{n-i} f_{2r} \beta_{2i,2n-2r}. \tag{18}$$

Применим теперь индукцию. Теорема, очевидно, верна при $n = 1$. Предположим, что она справедлива для путей, длина которых меньше $2n$. Тогда формула (18) сведется к равенству

$$\beta_{2i,2n} = \frac{1}{2}\, u_{2n-2i} \sum_{r=1}^{i} f_{2r} u_{2i-2r} + \frac{1}{2}\, u_{2i} \sum_{r=1}^{n-i} f_{2r} u_{2n-2i-2r}.$$

С учетом леммы получаем, что первая сумма равна u_{2i}, тогда как вторая сумма равна u_{2n-2i}, поэтому теорема верна и для путей длины $2n$, что и требовалось доказать. ■

Рассмотрим степенные ряды с коэффициентами u_{2n} и f_{2n}:

$$U(z) = \sum_{n=0}^{\infty} u_{2n} z^{2n}, \quad F(z) = \sum_{n=1}^{\infty} f_{2n} z^{2n}, \quad 0 \leqslant z < 1.$$

Связь между функциями $U(z)$ и $F(z)$ устанавливает

Теорема 2. Имеет место соотношение $F(z) = 1 - 1/U(z)$.

Доказательство. Используем условие $u_0 = 1$ и лемму:

$$\begin{aligned} U(z) - 1 &= \sum_{n=1}^{\infty} u_{2n}\, z^{2n} = \sum_{n=1}^{\infty} z^{2n} \left(\sum_{i=1}^{n} f_{2i}\, u_{2n-2i} \right) = \\ &= \sum_{i=1}^{\infty} f_{2i}\, z^{2i} \sum_{n=i}^{\infty} u_{2n-2i}\, z^{2n-2i} = \\ &= \left(\sum_{i=1}^{\infty} f_{2i}\, z^{2i} \right) \left(\sum_{n=0}^{\infty} u_{2n}\, z^{2n} \right) = F(z)\, U(z). \end{aligned}$$

Изменение порядка суммирования законно в силу неотрицательности членов ряда. ■

Так как $U(z)$ и $F(z)$ — степенные ряды с положительными коэффициентами, то их пределы при $z \to 1$ (конечные или бесконечные) равны, соответственно, $\sum u_{2n}$ и $\sum f_{2n}$ (см. [33, с. 57]). В силу теоремы 2 расходимость ряда $\sum u_{2n}$ равносильна тому, что $\sum f_{2n} = 1$. Последнее равенство означает, что блуждающая частица рано или поздно вернется в начало координат с вероятностью 1. Такое блуждание называют *возвратным*. Докажем замечательный результат, впервые опубликованный Д. Пойа в 1921 г.

Д. Пойа (1887–1985), американский математик

Теорема Пойа. Случайное блуждание на прямой и плоскости возвратно, а в трехмерном пространстве — невозвратно.

Доказательство. Исследуем сходимость ряда $\sum u_{2n}$ в двумерном и трехмерном случаях (возвратность одномерного блуждания была доказана ранее в § 6 гл. 14).

Найдем u_{2n} для блуждания на плоскости. Общее число путей длины $2n$ равно 4^{2n}. Для того, чтобы в момент $2n$ частица снова оказалась в начале координат, число шагов верх должно совпадать с числом шагов вниз, а число шагов вправо — с числом шагов влево. Поэтому, если i — это число шагов вверх, то число шагов вниз равно i, а число шагов вправо так же, как и число шагов влево, равно $n - i$ (всего $2n$ шагов). Представим, что шаг — это шарик, направление — ящик, и мы случайно раскладываем $2n$ шариков по четырем ящикам. В соответствии в формулой (3) гл. 10 число таких размещений шариков равно

$$\frac{(2n)!}{i!\,i!\,(n-i)!\,(n-i)!} = C_{2n}^{n}\left(C_n^i\right)^2. \tag{19}$$

Поскольку i может принимать значения от 0 до n, то

$$u_{2n} = 4^{-2n}C_{2n}^{n}\sum_{i=0}^{n}\left(C_n^i\right)^2 = 4^{-2n}C_{2n}^{n}\sum_{i=0}^{n}C_n^i\,C_n^{n-i} = \left(2^{-2n}C_{2n}^{n}\right)^2. \tag{20}$$

Здесь мы воспользовались тождеством

$$\sum_{i=0}^{n}C_n^i\,C_n^{n-i} = C_{2n}^{n}, \tag{21}$$

вытекающим из сравнения коэффициентов при t^n в обеих частях раскрытого по биному Ньютона равенства $(1+t)^{2n} = (1+t)^n(1+t)^n$. (Другое доказательство формулы (21) будет получено в примере 4 ниже.)

Применение формулы Стирлинга к соотношению (20) (см. формулу (14) гл. 14) дает асимптотику

$$u_{2n} \sim \left(1/\sqrt{\pi n}\right)^2 = \frac{1}{\pi n} \quad \text{при } n \to \infty,$$

откуда следует, что ряд $\sum u_{2n}$ расходится, т. е. блуждание на плоскости возвратно. Аналогично в трехмерном случае имеем

$$u_{2n} = 6^{-2n} \sum_{0 \leqslant i+j \leqslant n} \frac{(2n)!}{i!\, i!\, j!\, j!\, (n-i-j)!\, (n-i-j)!} = \\ = 2^{-2n} C_{2n}^{n} \sum_{0 \leqslant i+j \leqslant n} \left[3^{-n} \frac{n!}{i!\, j!\, (n-i-j)!} \right]^2. \tag{22}$$

Здесь в квадратных скобках стоят вероятности r_{ij} *наблюдать при случайном размещении* n *различимых шариков по трем ящикам в первом ящике* i*, во втором* $-$ j *и в третьем* $-$ $(n-i-j)$ *шариков* (см. § 5 гл. 10). Поэтому

$$\sum_{0 \leqslant i+j \leqslant n} r_{ij}^2 \leqslant \max_{0 \leqslant i+j \leqslant n} r_{ij} \sum_{0 \leqslant i+j \leqslant n} r_{ij} = \max_{0 \leqslant i+j \leqslant n} r_{ij}. \tag{23}$$

Вероятности r_{ij} достигают своего максимального значения при $i_0 = j_0 \sim n/3$ (задача 2). Используя формулу Стирлинга, получаем

$$\max_{0 \leqslant i+j \leqslant n} r_{ij} \sim \frac{c}{n}, \quad \text{где } c = \frac{3\sqrt{3}}{2\pi}. \tag{24}$$

Так как $2^{-2n} C_{2n}^{n} \sim 1/\sqrt{\pi n}$, из соотношений (22)–(24) имеем, что в трехмерном случае u_{2n} по порядку не превосходит $n^{-3/2}$. Следовательно, ряд $\sum u_{2n}$ сходится, и блуждание невозвратно. При этом вероятность *возвращения когда-нибудь* $F(1) = \sum\limits_{n=1}^{\infty} f_{2n}$ приближенно равна 0,35 (см. § 4 гл. 17). ■

Приведем небольшой отрывок из [81, с. 374] о теореме Пойа.

«Прежде всего, почти очевидно, что из этой теоремы вытекает, что в *одномерном и двумерном случаях с вероятностью* 1 *частица бесконечное число раз пройдет через каждое возможное положение*, однако в трехмерном случае это неверно. Таким образом, для двух измерений в известном смысле справедливо утверждение «все дороги ведут в Рим».

С другой стороны, рассмотрим *две* частицы, совершающие независимые случайные блуждания, причем перемещения их происходят одновременно. Встретятся ли они когда-нибудь? Чтобы упростить изложение, мы определим *расстояние* между двумя возможными положениями как наименьшее число шагов, ведущих из одного положения в другое. (Это расстояние равно сумме абсолютных величин разностей координат.) Если две частицы продвигаются на один шаг каждая, то расстояние между ними либо остается тем же, либо изменяется на две единицы, и поэтому расстояние между частицами будет либо всегда четным, либо всегда нечетным. Во втором случае наши две частицы никогда не смогут занять одно и то же положение. В первом случае легко видеть, что вероятность их встречи на n-м шаге равна вероятности того, что первая частица за $2n$ шагов достигнет начального положения второй частицы.

Следовательно, теорема Пойа утверждает, что в двумерном (но не в трехмерном) случае две частицы наверняка бесконечное число раз будут занимать одно и то же положение. Если начальное расстояние между двумя частицами нечетно, то аналогичное рассуждение показывает, что они будут бесконечно много раз занимать соседние положения. Если назвать это встречей, то теорема утверждает, что *в одномерном и двумерном случаях две частицы с достоверностью встретятся бесконечное число раз, однако в трехмерном случае они с положительной вероятностью никогда не встретятся».*

Пример 4. ***Гипергеометрическое распределение*** [81, с. 63]**.** В урне находятся M шаров черного цвета и $(N-M)$ шаров белого цвета. Случайным образом *без возвращения* извлекается группа из n шаров. Тогда вероятность того, что в ней будет *ровно m черных шаров* задается формулой

$$p(m,n,M,N) = C_M^m\, C_{N-M}^{n-m}\,/C_N^n\,, \quad m = 0,1,\ldots,n. \tag{25}$$

Доказательство. Занумеруем черные шары числами от 1 до M, белые — числами от $(M+1)$ до N. Пусть черные шары появились при извлечениях с индексами $1 \leqslant i_1 < i_2 < \ldots < i_m \leqslant n$. Эти индексы можно выбрать C_n^m способами. Для фиксированного набора индексов количество вариантов выбора номеров шаров равно $A_M^m\, A_{N-M}^{n-m}$, где $A_M^m = M(M-1)\ldots(M-m+1) = m!\, C_M^m$ обозначает *число размещений из M по m.*

Так как количество всех элементарных событий равно A_N^n, находим, что

$$p(m,n,M,N) = \frac{C_n^m\, A_M^m\, A_{N-M}^{n-m}}{A_N^n} = \tag{26}$$

$$= \frac{C_n^m\, m!\, C_M^m\, (n-m)!\, C_{N-M}^{n-m}}{n!\, C_N^n} = \frac{C_M^m\, C_{N-M}^{n-m}}{C_N^n}\,. \qquad \blacksquare$$

Поскольку сумма по всем m вероятностей (25) равна 1, взяв $m = i$, $N = 2n$ и $M = n$, получим еще одно доказательство тождества (21).

Название «гипергеометрическое распределение» происходит от *гипергеометрического ряда* (см. [42, с. 279])

$$G(a,b,c,z) = 1 + \frac{ab}{c}\; z + \frac{1}{2!}\,\frac{a(a+1)b(b+1)}{c(c+1)}\; z^2 + \ldots, \tag{27}$$

который при $a = 1$ и $b = c$ сводится к сумме геометрической прогрессии. Нетрудно проверить, что

$$p(m,n,M,N) = \frac{A_{N-M}^n}{A_N^n}\left[\frac{1}{m!}\;\frac{A_M^m\, A_n^m}{A_{N-M-n+m}^m}\right],$$

где величина в квадратных скобках совпадает с коэффициентом при z^m ряда (27), у которого $a = -M$, $b = -n$ и $c = N-M-n+1$.

Любопытно, что многие элементарные функции выражаются через функцию G:

$$(1+z)^n = G(-n,1,1,-z), \qquad \ln(1+z) = z\,G(1,1,2,-z),$$

$$\arcsin z = z\,G\left(1/2,1/2,3/2,z^2\right), \quad \operatorname{arctg} z = z\,G\left(1/2,1,3/2,-z^2\right).$$

Модель случайного выбора без возвращения применяется при *выборочном контроле* качества продукции. В партии из N изделий дефектные изделия играют роль черных шаров. Их число M неизвестно. Пусть в контрольной выборке размера n было обнаружено m дефектных изделий. Формула (25) позволяет сделать выводы относительно истинного значения M.

Еще одним примером использования данной модели может служить *оценка размера популяции* по данным повторного отлова. Из озера вылавливают M рыб, помечают их краской и выпускают обратно. При повторном отлове n рыб m из них оказались помеченными. Как оценить общее число N рыб в озере (задача 1)?

Наконец, отметим, что из формулы (26) вытекает сходимость

$$p(m,n,M,N) \to C_n^m\, p^m (1-p)^{n-m}$$

при $N \to \infty$ и $M/N \to p$, где $0 < p < 1$. Другими словами, для больших N и M практически нет различия между выбором без возвращения и выбором с возвращением.

ЗАДАЧИ

Если вы не добились успеха сразу, попытайтесь еще и еще раз. А потом успокойтесь и живите в свое удовольствие.

Уильям Клод Филдс

1. Пусть единственное наблюдение X_1 представляет собой число помеченных рыб при повторном отлове из примера 4. Оцените общее число рыб в озере методом а) моментов, б) максимального правдоподобия (см. § 2 и § 4 из гл. 9).

2*. Проверьте, что i_0 и j_0, при которых вероятность r_{ij} максимальна (см. пояснение к формуле (22)), принадлежат отрезку $\left[\dfrac{n}{3}-1,\ \dfrac{n}{3}+1\right]$.

Указание. Рассмотрите соседей точки (i_0,j_0) по осям.

3*. Получите тождество (13).

4*. Подправьте доказательство теоремы 1 гл. 11 так, чтобы вывести, что статистика V_{out}, заданная формулой (12), имеет распределение χ^2_{k-1}.

5*. Покажите, что наибольшее значение статистики критерия Краскела—Уоллиса W равно $(N^3 - \sum n_j^3)/[N(N+1)]$.

6*. Установите, что критерий Бартлетта при $k = 2$ равносилен двустороннему критерию Фишера из примера 1 гл. 14.

РЕШЕНИЯ ЗАДАЧ

1. а) Вычислим момент $\mathbf{M}X_1 = \sum\limits_{m=1}^{n} m\, C_M^m\, C_{N-M}^{n-m}\,/C_N^n$:

$$\mathbf{M}X_1 = M \sum_{m=1}^{n} C_{M-1}^{m-1}\, C_{(N-1)-(M-1)}^{(n-1)-(m-1)} \,/\big(C_{N-1}^{n-1}\cdot N/n\big) = nM/N.$$

Отсюда находим оценку метода моментов $\widehat{N} = [nM/X_1]$, где $[\,\cdot\,]$ обозначает *целую часть числа*. Доля m/n помеченных рыб в выборке примерно равна их доле M/N в озере.

б) Рассмотрим отношение соседних вероятностей

$$\frac{p(m,n,M,N)}{p(m,n,M,N-1)} = \frac{(N-M)(N-n)}{(N-M-n+m)N}\,.$$

Простые выкладки показывают, что правая часть больше 1, когда $mN < nM$ и меньше 1, когда $mN > nM$. Поэтому $p(m,n,M,N)$ имеет максимум также при $\widehat{N} = [nM/m]$.

2. Обозначим через i_0 и j_0 те i и j, на которых достигается наибольшее значение функции

$$T(i,j) \equiv T_{ij} = 3^n\, r_{ij} = \frac{n!}{i!\,j!\,(n-i-j)!} \quad \text{при } 0 \leqslant i+j \leqslant n.$$

Сразу можно выписать следующие четыре неравенства:

$$\frac{n!}{(i_0-1)!\,j_0!\,(n-i_0-j_0+1)!} \leqslant T_{i_0j_0}, \quad \frac{n!}{(i_0+1)!\,j_0!\,(n-i_0-j_0-1)!} \leqslant T_{i_0j_0},$$
$$\frac{n!}{i_0!\,(j_0-1)!\,(n-i_0-j_0+1)!} \leqslant T_{i_0j_0}, \quad \frac{n!}{i_0!\,(j_0+1)!\,(n-i_0-j_0-1)!} \leqslant T_{i_0j_0}.$$

Они сводятся к двум таким:

$$n - j_0 - 1 \leqslant 2i_0 \leqslant n - j_0 + 1,$$
$$n - i_0 - 1 \leqslant 2j_0 \leqslant n - i_0 + 1.$$

Подставляя в первое неравенство оценки сверху и снизу для j_0 из второго, получаем для i_0 искомые границы. В силу симметрии они верны и для j_0.

3. Для доказательства тождества (13) потребуется теорема Гюйгенса. Прежде, чем ее сформулировать, дадим несколько простых определений.

Х. Гюйгенс (1629–1695), нидерландский астроном и механик.

Пусть в $\mathbb{R}^k$ заданы точки (векторы) $\boldsymbol{x}_i$ с приписанными им массами m_i. Положим $m = \sum m_i$. *Центром масс* называется точка $\boldsymbol{c} = \dfrac{1}{m}\sum m_i\boldsymbol{x}_i$. Для нее, очевидно, выполняется равенство $\sum m_i(\boldsymbol{x}_i - \boldsymbol{c}) = \boldsymbol{0}$. Величину $I_{\boldsymbol{a}} = \sum m_i|\boldsymbol{x}_i - \boldsymbol{a}|^2$ называют *моментом инерции* относительно точки $\boldsymbol{a}$.

Теорема Гюйгенса. $I_{\boldsymbol{a}} = I_{\boldsymbol{c}} + m|\boldsymbol{c}-\boldsymbol{a}|^2$.

Доказательство. Используем очевидные равенства (см. П10) $|\boldsymbol{x}|^2 = \boldsymbol{x}^T\boldsymbol{x}$ и $|\boldsymbol{x}+\boldsymbol{y}|^2 = |\boldsymbol{x}|^2 + 2\boldsymbol{x}^T\boldsymbol{y} + |\boldsymbol{y}|^2$. Пусть $\boldsymbol{y}_i = \boldsymbol{x}_i - \boldsymbol{c}$. Тогда

$$I_{\boldsymbol{a}} = \sum m_i|\boldsymbol{y}_i + (\boldsymbol{c}-\boldsymbol{a})|^2 = I_{\boldsymbol{c}} + 2\,(\sum m_i\boldsymbol{y}_i)^T(\boldsymbol{c}-\boldsymbol{a}) + m|\boldsymbol{c}-\boldsymbol{a}|^2.$$

Но $\sum m_i\boldsymbol{y}_i = \boldsymbol{0}$, так как $\boldsymbol{c}$ — центр масс. Следовательно, второе слагаемое пропадает. ■

Примером применения этой теоремы может служить вычисление момента инерции тонкого обруча радиуса r и массы m относительно оси, проходящей через точку на ободе и перпендикулярной к плоскости обруча (рис. 3). Согласно теореме он равен $2mr^2$.

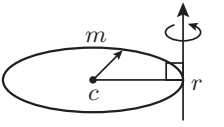


Рис. 3

Докажем теперь тождество (13). Для этого запишем теорему Гюйгенса отдельно для j-й выборки, полагая $m_i = 1$, $m = n_j$, $a = X_{\cdot\cdot}$ и $c = X_{\cdot j}$:

$$\sum_{i=1}^{n_j}(X_{ij} - X_{\cdot\cdot})^2 = \sum_{i=1}^{n_j}(X_{ij} - X_{\cdot j})^2 + n_j(X_{\cdot j} - X_{\cdot\cdot})^2.$$

Остается только просуммировать по j от 1 до k.

4. Возьмем $p_j = n_j/N$ $(j = 1,\ldots,k)$, где $N = \sum n_j$. Тогда, по определению, $X_{\cdot\cdot} = \sum p_j X_{\cdot j}$. Если все $n_j = 1$, то годится доказательство теоремы 1 гл. 11. Обобщим его на случай, когда есть $n_j > 1$. Мы знаем, что $X_{\cdot j} \sim \mathcal{N}(\mu_j, 1/n_j)$ и независимы между собой, кроме того, мы предполагаем, что верна гипотеза $H''\colon \mu_1 = \ldots = \mu_k$. Обозначим через μ это общее среднее.

Станем дополнять до ортогональной матрицы не строку $(1/\sqrt{k},\ldots,1/\sqrt{k})$, а строку $(\sqrt{p_1},\ldots,\sqrt{p_k})$. В качестве $\boldsymbol{Y}$ возьмем вектор с компонентами $Y_j = \sqrt{n_j}\,(X_{\cdot j} - \mu)$. Тогда случайная величина $Y_j \sim \mathcal{N}(0,1)$ и независимы. Рассмотрим $\boldsymbol{Z} = \boldsymbol{C}\boldsymbol{Y}$. При умножении последней строки $\boldsymbol{C}$ на $\boldsymbol{Y}$ получается равенство

$$Z_k = \sqrt{p_1}\,Y_1 + \ldots + \sqrt{p_k}\,Y_k. \tag{28}$$

Из формулы (4) гл. 11 следует, что $\sum\limits_{j=1}^{k} Z_j^2 = \sum\limits_{j=1}^{k} Y_j^2$. В силу теоремы Гюйгенса (при $m_j = n_j$, $m = N$, $a = \mu$, $c = X_{\cdot\cdot}$) имеем

$$\sum_{j=1}^{k} n_j(X_{\cdot j} - X_{\cdot\cdot})^2 = \sum_{j=1}^{k} n_j(X_{\cdot j} - \mu)^2 - N(X_{\cdot\cdot} - \mu)^2. \tag{29}$$

Используя определение Y_j и учитывая равенство (28), запишем правую часть (29) в виде

$$\sum_{j=1}^{k} Y_j^2 - \Bigl(\sum_{j=1}^{k}\sqrt{p_j}\,Y_j\Bigr)^2 = \sum_{j=1}^{k} Z_j^2 - Z_k^2 = \sum_{j=1}^{k-1} Z_j^2.$$

Согласно лемме 1 гл. 11 вектор $\boldsymbol{Z}$ имеет независимые $\mathcal{N}(0,1)$ компоненты. ■

5. Подставим ранги R_{ij} в тождество (13) вместо X_{ij}:

$$\sum_{j=1}^{k} n_j(R_{\cdot j}-R_{\cdot\cdot})^2 = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(R_{ij}-R_{\cdot\cdot})^2 - \sum_{j=1}^{k}\sum_{i=1}^{n_j}(R_{ij}-R_{\cdot j})^2. \quad (30)$$

Поскольку первая двойная сумма не зависит от разбиения рангов $\{1,2,\ldots,N\}$ на группы, минимизируем вторую двойную сумму. Покажем, что ее минимум достигается на любом из разбиений, у которых ранги, расставленные по возрастанию в каждой группе, идут подряд. Для этого нам потребуется

Теорема о межточечных расстояниях (см. [64, с. 28]). Пусть $\boldsymbol{c}$ — центр масс точек $\boldsymbol{x}_i$ в $\mathbb{R}^k$ с массами m_i. Тогда

$$I_{\boldsymbol{c}} = \sum_i m_i|\boldsymbol{x}_i-\boldsymbol{c}|^2 = \frac{1}{m}\sum_{i<j} m_i m_j|\boldsymbol{x}_i-\boldsymbol{x}_j|^2, \quad \text{где } m=\sum_i m_i.$$

Доказательство. Обозначим через $\boldsymbol{y}_i = \boldsymbol{x}_i - \boldsymbol{c}$. При этом

$$|\boldsymbol{x}_i-\boldsymbol{x}_j|^2 = |\boldsymbol{y}_i-\boldsymbol{y}_j|^2 = |\boldsymbol{y}_i|^2+|\boldsymbol{y}_j|^2-2\,\boldsymbol{y}_i^T\boldsymbol{y}_j.$$

Используя определение момента инерции относительно центра масс $\boldsymbol{c}$, запишем

$$\sum_i\sum_j m_i m_j\left(|\boldsymbol{y}_i|^2+|\boldsymbol{y}_j|^2\right) = \sum_i m_i\sum_j\left(m_j|\boldsymbol{y}_i|^2+m_j|\boldsymbol{y}_j|^2\right) =$$
$$= \sum_i m_i\left(m|\boldsymbol{y}_i|^2+I_{\boldsymbol{c}}\right) = 2mI_{\boldsymbol{c}}.$$

С другой стороны,

$$\sum_i\sum_j m_i m_j\boldsymbol{y}_i^T\boldsymbol{y}_j = \sum_i m_i\boldsymbol{y}_i^T\left(\sum_j m_j\boldsymbol{y}_j\right) = 0.$$

Отсюда получаем соотношение

$$2mI_{\boldsymbol{c}} = \sum_i\sum_j m_i m_j|\boldsymbol{x}_i-\boldsymbol{x}_j|^2 = 2\sum_{i<j} m_i m_j|\boldsymbol{x}_i-\boldsymbol{x}_j|^2,$$

которое и требовалось установить. ■

Рассмотрим j-ю группу. Сумма $\sum\limits_{i=1}^{n_j}(R_{ij}-R_{\cdot j})^2$ равна моменту инерции относительно $R_{\cdot j}$ точек R_{ij} с массами 1. Убедимся, что он минимален, когда ранги идут подряд. (Это интуитивно понятно, так как только в таком случае n_j точек с целочисленными координатами образуют наиболее компактную группу.) Действительно, для любого упорядоченного по возрастанию набора из n_j рангов соответствующие межточечные расстояния могут быть только больше, чем у набора, где ранги идут подряд (рис. 4).

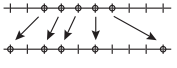


Рис. 4

Вычислим это минимальное значение. Очевидно, что момент инерции не зависит от выбора начала координат, поэтому можно считать, что ранги равны $1,2,\ldots,n_j$. Положим для краткости $n = n_j$. Тогда по теореме Гюйгенса для $m_j = 1$, $m = n$, $a = 0$ и $c = (n+1)/2$

$$I_{R_{\cdot j}} = \sum_{i=1}^{n} \left(i - \frac{n+1}{2}\right)^2 = \sum_{i=1}^{n} i^2 - n\left(\frac{n+1}{2}\right)^2 =$$
$$= n(n+1)(2n+1)/6 - n(n+1)^2/4 =$$
$$= n(n+1)(n-1)/12 = (n^3-n)/12.$$

Остается только подставить полученные результаты в формулу (30) и учесть, что $N = n_1 + \ldots + n_k$.

6. Элементарные выкладки показывают, что статистика $F = S_1^2/S_2^2$ критерия Фишера и статистика B критерия Бартлетта при $k = 2$ связаны соотношением

$$B = [(nF+m)/N]^N F^{-n}, \text{ где } N = n+m.$$

При любом $B_0 > 1$ (см. ответ на вопрос 2) это уравнение имеет два действительных корня $F_1 < 1$ и $F_2 > 1$ (если $B_0 = 1$, то $F_1 = F_2 = 1$). Таким образом, событие $\{B > B_0\}$ эквивалентно объединению несовместных событий $\{F < F_1\}$ и $\{F > F_2\}$ (рис. 5).

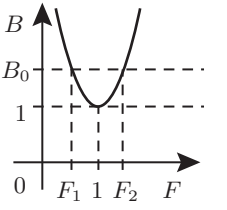


Рис. 5

ОТВЕТЫ НА ВОПРОСЫ

1. С точностью до перестановки столбцов возможны только 3 варианта распределения $N = 4$ рангов:

$$\begin{pmatrix}1&3\\2&4\end{pmatrix}, \quad \begin{pmatrix}1&2\\3&4\end{pmatrix}, \quad \begin{pmatrix}1&2\\4&3\end{pmatrix}.$$

Легко вычислить, что значения статистики W для них равны 2,4, 0,6 и 0 соответственно. Поэтому $\mathbf{P}(W \geqslant 2) = 1/3$.

2. Заметим, что всегда $B \geqslant 1$. Действительно, применим неравенство Иенсена (П4) к выпуклой вниз функции $y = e^x$ и случайной величине ξ, принимающей значения $x_i = \ln s_i^2$ с вероятностями $p_i = n_i/N$:

$$\left[(s_1^2)^{n_1} \cdot \ldots \cdot (s_k^2)^{n_k}\right]^{1/N} = \exp\left\{\frac{1}{N}\left(n_1 \ln s_1^2 + \ldots + n_k \ln s_k^2\right)\right\} =$$
$$= \exp\{\textstyle\sum x_i\, p_i\} = e^{\mathbf{M}\xi} \leqslant \mathbf{M}e^{\xi} = \textstyle\sum e^{x_i} p_i = \dfrac{1}{N}\sum n_i s_i^2.$$

Таким образом, взвешенное среднее арифметическое не меньше, чем среднее геометрическое с теми же весами.

Когда, скажем, s_1^2 увеличивается, а остальные s_j^2 остаются фиксированными, среднее арифметическое растет линейно

по s_1^2, а среднее геометрическое — как степенная функция с показателем $\frac{n_1}{N} < 1$, т. е. *медленнее*. Если гипотеза H' не верна (одна из дисперсий σ_j^2 существенно больше остальных), то соответствующая оценка s_j^2 также будет отличаться от других, что приведет к значению статистики B, значимо превосходящему 1.

3. При $k = 2$ статистика R, задаваемая формулой (9), представляется в следующем виде:

$$R = \frac{\frac{1}{k-1}\, V_{out}}{\frac{1}{N-2}\, V_{int}} = \frac{V_{out}}{\frac{1}{n+m-2}\, V_{int}}\,.$$

В свою очередь, согласно примеру 1 гл. 14 имеем формулу

$$T^2 = \frac{\frac{nm}{n+m}\left(\overline{X} - \overline{Y}\right)^2}{S_{tot}^2}\,, \tag{31}$$

где S_{tot}^2 обозначает несмещенную оценку дисперсии одного наблюдения, построенную на основе объединенной выборки:

$$S_{tot}^2 = \frac{1}{n+m-2}\left[(n-1)\,S_1^2 \,+\, (m-1)\,S_2^2\right] =$$
$$= \frac{1}{n+m-2}\left[\sum (X_i - \overline{X})^2 \,+\, \sum (Y_j - \overline{Y})^2\right] = \frac{1}{n+m-2}\, V_{int}.$$

Следовательно, знаменатели у R и T^2 совпадают. Покажем, что совпадают и числители. Величина V_{out} по определению равна моменту инерции масс n и m, расположенных в точках $\overline{X}$ и $\overline{Y}$, относительно общего центра масс (см. решение задачи 3). В силу теоремы о межточечных расстояниях из решения задачи 5 этот момент равен числителю правой части равенства (31).

Глава 17

МНОГОКРАТНЫЕ НАБЛЮДЕНИЯ

В этой главе мы обобщим схему парных повторных наблюдений из § 1 гл. 15 на случай, когда данные представляют собой k-кратные повторные наблюдения, $k \geqslant 2$.

Пять тысяч раз твердит одно и то же!

Фамусов в «Горе от ума»
А. С. Грибоедова

§ 1. ДВУХФАКТОРНАЯ МОДЕЛЬ

Данные. В каждом из n блоков содержится по одному наблюдению x_{ij} на каждую из k обработок. Будем считать наблюдения реализацией случайных величин X_{ij} в модели

$$X_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \quad i = 1, \ldots, n; \;\; j = 1, \ldots, k. \tag{1}$$

Блоки	Обработки			
	1	2	...	k
1	x_{11}	x_{12}	...	x_{1k}
2	x_{21}	x_{22}	...	x_{2k}
⋮	⋮	⋮	⋮	⋮
n	x_{n1}	x_{n2}	...	x_{nk}

Здесь μ — неизвестное общее среднее, α_i — эффект блока i (неизвестный мешающий параметр), β_j — эффект обработки j (интересующий нас параметр), ε_{ij} — случайная ошибка.*)

Пусть справедливы те же самые, что и в гл. 16,

Допущения

Д1. *Все ошибки ε_{ij} независимы.*

Д2. *Все ε_{ij} имеют одинаковое непрерывное (неизвестное) распределение.*

*) Заметим, что параметризация (1) избыточна и не позволяет однозначно восстановить параметры μ, α_i, β_j. Например, положив $\alpha_i' = \alpha_i + c$, $\beta_j' = \beta_j - c$ при произвольном c, мы имеем тождество $\mu + \alpha_i + \beta_j = \mu + \alpha_i' + \beta_j'$.

§ 2. КРИТЕРИЙ ФРИДМАНА

Для проверки гипотезы

$$H_0\colon \beta_1 = \ldots = \beta_k \tag{2}$$

против альтернативы

$$H_1\colon \text{не все } \beta_j \text{ равны между собой} \tag{3}$$

применяется *критерий Фридмана* (см. [88, с. 155]). Выполним **следующие шаги**.

1. Отдельно для каждого i-го блока (строки таблицы) ранжируем k наблюдений внутри него от меньшего к большему. Обозначим через Q_{ij} ранг X_{ij} в совместной ранжировке $X_{i1},\ldots,X_{ik}$.

2. Для $j = 1,\ldots,k$ положим

$$T_j = \sum_{i=1}^{n} Q_{ij}, \quad Q_{\cdot j} = T_j/n, \quad Q_{\cdot\cdot} = \frac{1}{nk}\cdot n\cdot\frac{k(k+1)}{2} = \frac{k+1}{2}\,. \tag{4}$$

Здесь $Q_{\cdot j}$ — это средний ранг по всем n блокам наблюдений, относящихся к j-й обработке (столбцу таблицы), $Q_{\cdot\cdot}$ — средний ранг по всей таблице.

3. В качестве *статистики критерия Фридмана* возьмем

$$F = \frac{12\,n}{k(k+1)}\sum_{j=1}^{k}(Q_{\cdot j} - Q_{\cdot\cdot})^2 = \left[\frac{12}{nk(k+1)}\sum_{j=1}^{k}T_j^2\right] - 3n(k+1).$$

Если гипотеза H_0 верна, то все $Q_{\cdot j}$ должны быть близки к $Q_{\cdot\cdot} = (k+1)/2$. Если β_j не равны, то $Q_{\cdot j}$ будут различаться сильнее, а поэтому некоторые из $(Q_{\cdot j} - Q_{\cdot\cdot})^2$ окажутся больше других, что приведет к бóльшим значениям F.

Малые выборки. Гипотеза H_0 отвергается, если значение f статистики F превысит критическую границу, определяемую по таблице $A.15$ из [88] ($k = 3$, $n \leqslant 13$; $k = 4$, $n \leqslant 8$; $k = 5$, $n \leqslant 5$).

Большие выборки. Если гипотеза H_0 верна, то статистика F сходится по распределению при $n \to \infty$ к χ^2_{k-1} (см. [86, с. 203]). Приближенный критерий уровня α таков: следует отклонить H_0, если $f \geqslant z_{1-\alpha}$, где $z_{1-\alpha}$ — это $(1-\alpha)$-квантиль закона χ^2_{k-1} (см. таблицу Т3); иначе — принять гипотезу H_0.

Поправка. Для небольших выборок это приближение не является удовлетворительным. Так, при $\alpha = 0{,}01$ для $k = 6$ и $n = 6$ относительная погрешность ошибки I рода составляет 63% (см. [88, с. 132]). Следующая поправка Р. Имана и Дж. Давенпорта (1980 г.) значительно снижает эту погрешность (в данном случае — до 3%).

Положим

$l = 0, m = 2,$ если $7 < k < 20$, $n > 12$;

$l = 1, m = k - 1,$ если $k \leqslant 7$, $n > 7$;

$l = 1, m = (k-1)(n-1),$ если $k > 5$, $2 \leqslant n \leqslant 6$.

Другие возможные случаи см. в [88, с. 12]. Введем статистику

$$\tilde{F} = (lF + mG)/2, \quad \text{где } G = [(n-1)F]/[n(k-1) - F]. \tag{5}$$

Для статистики $\tilde{F}$ приближенное критическое значение уровня α равно $(lz_{1-\alpha} + mf_{1-\alpha})/2$, где $z_{1-\alpha}$ и $f_{1-\alpha}$ — это $(1-\alpha)$-квантили соответственно распределения χ^2_{k-1} (табл. Т3) и закона Фишера—Снедекора $F_{k-1,\,(k-1)(n-1)}$ (табл. Т5).

Совпадения. Если среди $x_{i1}, \ldots, x_{ik}$ есть одинаковые значения, то следует вычислять средние ранги, а затем заменить статистику F на

$$F' = \frac{12 \sum\limits_{j=1}^{k} (T_j - nQ_{\cdot\cdot})^2}{nk(k+1) - \dfrac{1}{k-1} \sum\limits_{i=1}^{n} \left\{ \left(\sum\limits_{m=1}^{g_i} l_{im}^3 \right) - k \right\}}, \tag{6}$$

где g_i — это число групп совпадений в блоке i, l_{im} — размер m-й группы совпадений в блоке i, при этом x_{ij}, отличающиеся от других внутри блока, считаются группой размера 1.

Сравнение обработок. Для определения, какие из обработок отличаются друг от друга, применяется следующий приближенный критерий уровня α: принять решение $\beta_u \neq \beta_v$, если

$$|T_u - T_v| \geqslant C_{k,\alpha} \sqrt{nk(k+1)/12}, \tag{7}$$

где T_j определены в формуле (4), а $C_{k,\alpha}$ представляет собой $(1-\alpha)$-квантиль распределения *размаха* $\xi_{(k)} - \xi_{(1)}$ выборки $(\xi_1, \ldots, \xi_k)$ из $\mathcal{N}(0,1)$. Вот некоторые значения $C_{k,\alpha}$:

k	2	3	4	5	6	7	8
$\alpha = 0{,}1$	2,33	2,90	3,24	3,48	3,66	3,81	3,93
$\alpha = 0{,}05$	2,77	3,31	3,63	3,86	4,03	4,17	4,29
$\alpha = 0{,}01$	3,64	4,12	4,40	4,60	4,76	4,88	4,99

Более подробные таблицы см. в [10, с. 226] или [88, с. 340].

Оценки контрастов. Обозначим через $\Delta_{uv} = \beta_u - \beta_v$ *контраст эффектов обработок* u *и* v. В качестве *первичной оценки контраста* возьмем

$$Z_{uv} = MED\{X_{iu} - X_{iv},\ 1 \leqslant i \leqslant n\}. \tag{8}$$

Поскольку $Z_{uv} = -Z_{vu}$, достаточно найти лишь $k(k-1)/2$ значений Z_{uv} для $u < v$.

Определим *уточненную оценку контраста* $\tilde{\Delta}_{uv}$ так:

$$\tilde{\Delta}_{uv} = Z_{u\cdot} - Z_{v\cdot}, \quad \text{где } Z_{j\cdot} = \frac{1}{k}\sum_{l=1}^{k} Z_{jl}, \quad Z_{jj} = 0. \tag{9}$$

Уточненные оценки не менее эффективны, чем первичные, при этом они еще и согласованы (см. комментарий 4 в § 2 гл. 16). Их недостаток — зависимость $\tilde{\Delta}_{uv}$ от наблюдений, относящимся к остальным $(k-2)$ выборкам.

Пример применения критерия Фридмана к экспериментальным данным содержится в задаче 1.

Комментарии

1) Распределение статистики F при условии справедливости гипотезы H_0 можно получить, опираясь на равновозможность всех $(k!)^n$ ранговых наборов.

Вопрос 1. Чему равна $\mathbf{P}(F \geqslant 4)$ при $k=3$ и $n=2$?

2) П. Элтерен и Г. Нетер (1959 г.) установили, что асимптотическая эффективность (для альтернатив сдвига) критерия Фридмана по отношению к F-критерию двухфакторного дисперсионного анализа (см. пример 1 ниже) равна

$$E^*(F_\varepsilon) = \frac{k}{k+1}\, E(F_\varepsilon),$$

где F_ε обозначает функцию распределения ошибок ε_{ij} из (1), $E(F_\varepsilon)$ задается формулой (11) гл. 14. Некоторые значения $E^*(F_\varepsilon)$ для закона $\mathcal{N}(0,1)$, равномерного распределения и закона Лапласа, приведены в следующей таблице из [88, с. 197]:

Законы (F_ε) \ k	2	3	4	5	10	∞
Нормальный	0,637	0,716	0,764	0,796	0,868	0,995
Равномерный	0,667	0,750	0,800	0,833	0,909	1,000
Лапласа	1,000	1,125	1,200	1,250	1,364	1,500

Пример 1. ***Двухфакторная модель для нормальных наблюдений.*** Допустим, что ошибки ε_{ij} из (1) распределены по закону $\mathcal{N}(0,\sigma^2)$. В этом случае оптимальным критерием для проверки гипотезы H_0 (2) является *F-критерий двухфакторного дисперсионного анализа*, основанный на статистике (см. [8, с. 160])

$$\left[\frac{n}{k-1}\sum_{j=1}^{k}(X_{\cdot j} - X_{\cdot\cdot})^2\right] \Big/ \left[\frac{1}{(n-1)(k-1)}\sum_{i=1}^{n}\sum_{j=1}^{k}(X_{ij} - X_{i\cdot} - X_{\cdot j} + X_{\cdot\cdot})^2\right],$$

где

$$X_{i\cdot} = \frac{1}{k}\sum_{j=1}^{k} X_{ij}, \quad X_{\cdot j} = \frac{1}{n}\sum_{i=1}^{n} X_{ij}, \quad X_{\cdot\cdot} = \frac{1}{nk}\sum_{i=1}^{n}\sum_{j=1}^{k} X_{ij},$$

имеющей при условии справедливости гипотезы H_0 распределение Фишера—Снедекора $F_{k-1,(n-1)(k-1)}$.

В задаче 2 предлагается доказать следующее тождество:

$$\sum_{i=1}^{n}\sum_{j=1}^{k}(X_{ij}-X_{..})^2 = k\sum_{i=1}^{n}(X_{i\cdot}-X_{..})^2 + n\sum_{j=1}^{k}(X_{\cdot j}-X_{..})^2 + \\ + \sum_{i=1}^{n}\sum_{j=1}^{k}(X_{ij}-X_{i\cdot}-X_{\cdot j}+X_{..})^2. \tag{10}$$

Оно показывает, что *общая изменчивость* распадается на части, обусловленные влиянием *эффектов блоков и обработок*, и часть, связанную с изменчивостью *самих данных* (см. замечание 1 гл. 16).

§ 3. КРИТЕРИЙ ПЕЙДЖА

Нередко условия эксперимента таковы, что обработки упорядочены естественным образом, например, по интенсивности стимулов, сложности заданий и т. п. Приводимый ниже критерий Пейджа (см. [88, с. 163]) учитывает информацию, содержащуюся в предполагаемой *упорядоченности* (в отличие от критерия Фридмана, статистика которого F принимает одно и то же значение для всех $k!$ перенумераций обработок).

Для проверки гипотезы H_0 (2) против *альтернативы возрастания эффектов обработок*

$$H_2\colon \beta_1 \leqslant \ldots \leqslant \beta_k, \tag{11}$$

где хотя бы одно из неравенств строгое, вычисляется *статистика критерия Пейджа*

$$L = \sum_{j=1}^{k} jT_j = T_1 + 2T_2 + \ldots + kT_k, \tag{12}$$

где T_j задаются формулой (4).

Малые выборки. Для $k \leqslant 8$ приближенные критические значения l_α статистики L при справедливости H_0 (2) даны в таблице $A.16$ из [88] ($k = 3$, $n \leqslant 20$; $k = 4-8$, $n \leqslant 12$).

Большие выборки. Пусть $L^* = (L - \mathbf{M}L)/\sqrt{\mathbf{D}L}$, где

$$\mathbf{M}L = nk(k+1)^2/4, \quad \mathbf{D}L = n(k-1)k^2(k+1)^2/144. \tag{13}$$

Если верна гипотеза H_0, то распределение статистики L^* сходится при $n \to \infty$ к $\mathcal{N}(0,1)$ (см. [86, с. 207]). Обозначим через $x_{1-\alpha}$ квантиль уровня $1-\alpha$ закона $\mathcal{N}(0,1)$ (см. таблицу Т2). Когда наблюдаемое значение L^* больше или равно $x_{1-\alpha}$, гипотеза H_0 отвергается, в противном случае — принимается.

Совпадения. Используйте средние ранги.

Сравнение обработок с контрольной. Допустим, что роль контроля играет первая обработка. Приближенный критерий уровня α устроен так: принять решение $\beta_j > \beta_1$, если

$$T_j - T_1 \geqslant D_{k,\alpha}\sqrt{nk(k+1)/6}, \tag{14}$$

где величина T_j задана формулой (4), $D_{k,\alpha}$ — $(1-\alpha)$-квантиль распределения максимума из компонент нормального вектора $(\xi_1,\ldots,\xi_{k-1})$ (см. П9), у которого $\xi_j \sim \mathcal{N}(0,1)$ и $\mathbf{M}\xi_i\xi_j = \frac{1}{2}$ при $i \neq j$ (корректность этого определения проверяется в задаче 3). Вот некоторые значения $D_{k,\alpha}$, вычисленные линейной интерполяцией таблицы $A.13$ из [88]:

k	2	3	4	5	6	7	8
$\alpha = 0{,}1$	1,28	1,58	1,74	1,84	1,92	1,98	2,03
$\alpha = 0{,}05$	1,65	1,92	2,06	2,16	2,24	2,29	2,34
$\alpha = 0{,}01$	2,33	2,56	2,69	2,77	2,84	2,89	2,94

Пример 2. ***Прочность волокон хлопка*** (см. [88, с. 165])**.** В опыте, описанном в книге [93, с. 108], изучалось влияние количества калийного удобрения, вносимого в почву (в расчете на K_2O), на разрывную прочность волокон хлопка. При $n = 3$ блоках использовалось $k = 5$ уровней удобрений. С каждой делянки отбирался один образец хлопка, на котором производилось 4 измерения показателя прочности по Прессли. В таблице приведены средние по этим четырем замерам, а в круглых скобках — ранги Q_{ij} внутриблочного ранжирования.

Блоки	Калийное удобрение (кг/га)				
	163	122	82	61	41
1	7,46 (2)	7,17 (1)	7,76 (4)	8,14 (5)	7,62 (3)
2	7,68 (2)	7,57 (1)	7,73 (3)	8,15 (5)	8,00 (4)
3	7,21 (1)	7,80 (3)	7,74 (2)	7,87 (4)	7,93 (5)
	$T_1 = 5$	$T_2 = 5$	$T_3 = 9$	$T_4 = 14$	$T_5 = 12$

Проверим с помощью критерия Пейджа гипотезу об отсутствии влияния количества удобрения на прочность нити против альтернативы убывания прочности с ростом количества удобрения.

$$L = T_1 + 2T_2 + \ldots + 5T_5 = 5 + 2\cdot 5 + 3\cdot 9 + 4\cdot 14 + 5\cdot 12 = 158.$$

Для $\alpha = 0{,}01$ из [88, таблица $A.16$] при $n = 3$ и $k = 5$ находим, что критическое значение $l_\alpha = 155$. Поскольку $158 > 155$, гипотеза (для рассматриваемых уровней количества удобрения) отвергается.

Хотя число блоков n мало, посмотрим, что дает нормальное приближение. По формулам (13) получаем $\mathbf{M}L = 135$ и $\mathbf{D}L = 75$.

Отсюда $L^* \approx 2{,}66$. В соответствии с таблицей Т2 эта величина значимо велика при $\alpha = 0{,}004$, т. е. гипотеза должна быть отвергнута на еще меньшем уровне значимости.

Сравним обработку 4 с обработкой 1. При $\alpha = 0{,}01$ согласно (14) вычисляем границу $D_{k,\alpha}\sqrt{nk(k+1)/6} = 2{,}77\sqrt{15} \approx 10{,}7$. Так как $T_4 - T_1 = 14 - 5 = 9 < 10{,}7$, то на уровне 1% обработки не различаются. Однако, 9 больше, чем граница $2{,}16\sqrt{15} \approx 8{,}37$, соответствующая уровню значимости $\alpha = 5\%$. (Вообще, для небольшого количества данных при сравнении обработок не следует задавать слишком малые уровни, чтобы не пропустить различие.)

Первичные	оценки
$Z_{41} = 0{,}66$	$Z_{11} = 0$
$Z_{42} = 0{,}58$	$Z_{12} = 0{,}11$
$Z_{43} = 0{,}38$	$Z_{13} = -0{,}30$
$Z_{44} = 0$	$Z_{14} = -0{,}66$
$Z_{45} = 0{,}15$	$Z_{15} = -0{,}32$
$Z_{4\cdot} = 0{,}354$	$Z_{1\cdot} = -0{,}234$

Вычислим по формуле (9) уточненную оценку контраста $\tilde{\Delta}_{41}$. Значения необходимых для этого первичных оценок Z_{uv} (см. (8)), а также их средних $Z_{4\cdot}$ и $Z_{1\cdot}$, приведены в таблице. Таким образом, $\widetilde{\Delta}_{41} = Z_{4\cdot} - Z_{1\cdot} = 0{,}588 \neq Z_{41} = 0{,}66$.

Комментарии

1. Распределение статистики L при справедливости гипотезы H_0 можно получить из равновероятности всех $(k!)^n$ наборов рангов.

2. Критерий Пейджа состоятелен против альтернативы возрастания эффектов обработок H_2 (см. (11)).

В заключение, заметим, что однородность нескольких выборок, данные которых *сгруппированы*, можно проверить с помощью критерия хи-квадрат, рассматриваемого в гл. 18.

§ 4. СЧАСТЛИВЫЙ БИЛЕТИК И ВОЗВРАЩЕНИЕ БЛУЖДАНИЯ

Он улетел, но обещал вернуться.

Из мультфильма «Карлсон вернулся»

Покажем на трех примерах, как аппарат характеристических и производящих функций помогает вычислять вероятности довольно сложных событий.

Математика — дело настолько серьезное, что ей всегда нужно немножко занимательности.

Б. Паскаль

Напомним, что *характеристической функцией* случайной величины ξ называется $\psi_\xi(t) = \mathbf{M}e^{it\xi}$. Основные свойства характеристических функций приведены в П9. В отличие от характеристических, *производящие функции* определяются только для случайных величин, принимающих *целые неотрицательные* значения:

$$\varphi_\eta(z) = \mathbf{M}z^\eta = \sum_{n=0}^{\infty} z^n\,\mathbf{P}(\eta = n), \quad \text{где } 0 \leqslant z \leqslant 1. \tag{15}$$

Отметим, что $F(z)$ из теоремы 2 гл. 16 является производящей функцией *времени возвращения блуждающей частицы в начало координат* при $k \leqslant 2$, но не является *вероятностной* производящей функцией при $k = 3$, поскольку в этом случае $F(1) < 1$ из-за того, что время возвращения с положительной вероятностью бесконечно (см. теорему Пойа в § 4 гл. 16).

Назовем шестизначный номер билетика (или талончика) *счастливым*, если сумма трех первых цифр равна сумме трех последних. Пусть η_j обозначает j-ю цифру номера случайно выбранного билетика ($j = 1,\ldots,6$). При этом случайные величины $\eta_1,\ldots,\eta_6$ независимы и одинаково распределены: $\mathbf{P}(\eta_j = m) = 1/10$, $m = 0,1,\ldots,9$. Производящая функция

$$\varphi_{\eta_1}(z) = \frac{1}{10}\,(1 + z + \ldots + z^9) = \frac{1}{10}\,(1 - z^{10})/(1 - z). \tag{16}$$

Нас интересует $\mathbf{P}(\eta_1 + \eta_2 + \eta_3 = \eta_4 + \eta_5 + \eta_6)$. Покажем, что число таких номеров билетиков совпадает с числом номеров, у которых сумма всех шести цифр равна 27. Для этого воспользуемся «принципом отражения» (см. § 6 гл. 14): номеру $(\eta_1,\ldots,\eta_6)$ сопоставим номер $(\eta_1,\eta_2,\eta_3,9-\eta_4,9-\eta_5,9-\eta_6)$. Очевидно, что это взаимно однозначное отображение первой группы номеров на вторую.

Основным свойством производящих (а также характеристических) функций является то, что они перемножаются при сложении независимых случайных величин ξ и η. Действительно, в силу леммы о независимости из § 3 гл. 1 и свойства 5 математического ожидания (П2) имеем:

$$\varphi_{\xi+\eta}(z) = \mathbf{M}z^{\xi+\eta} = \mathbf{M}z^{\xi}z^{\eta} = \mathbf{M}z^{\xi}\cdot\mathbf{M}z^{\eta} = \varphi_{\xi}(z)\,\varphi_{\eta}(z). \tag{17}$$

Положим $S_6 = \eta_1 + \ldots + \eta_6$. Из формул (16) и (17) выводим, что

$$\varphi_{S_6}(z) = 10^{-6}\,(1 - z^{10})^6\,(1 - z)^{-6}. \tag{18}$$

В соответствии с принципом отражения вероятность того, что билет окажется счастливым, равна $\mathbf{P}(S_6 = 27)$, а это есть коэффициент при z^{27} в разложении функции $\varphi_{S_6}(z)$ по степеням z. Для его вычисления используем разложение бинома в ряд Тейлора

$$(1 + z)^x = 1 + C_x^1\,z + C_x^2\,z^2 + \ldots, \quad |z| < 1,\ x \in \mathbb{R}, \tag{19}$$

где $C_x^k = \dfrac{x(x-1)\ldots(x-k+1)}{k!}$ — *обобщенный биномиальный коэффициент*. Тогда правая часть равенства (18) записывается в виде

$$10^{-6}\,(1 - C_6^1\,z^{10} + C_6^2\,z^{20} - \ldots)\,(1 - C_{-6}^1\,z + C_{-6}^2\,z^2 - \ldots).$$

Как счастье своенравно!
Софья в «Горе от ума»
А. С. Грибоедова

Перемножая почленно ряды, видим, что коэффициент при z^{27} равен $10^{-6}\,(-1\cdot C_{-6}^{27} + C_6^1\cdot C_{-6}^{17} - C_6^2\cdot C_{-6}^7)$. Нетрудно подсчитать, что искомая вероятность есть $55252\cdot 10^{-6} \approx 1/18$.

В качестве второго примера найдем вероятность возвращения в 0 одномерного *несимметричного* случайного блуждания $S_n = \xi_1 + \ldots + \xi_n$, у которого «шаги» ξ_k независимы и одинаково распределены: $\mathbf{P}(\xi_k = 1) = p$ и $\mathbf{P}(\xi_k = -1) = 1-p \equiv q$, где $0 < p < 1$. Эта модель ранее появлялась в § 4 гл. 13 при рассмотрении задачи о разорении игрока.

При $p \neq \dfrac{1}{2}$ в силу закона больших чисел (П6) случайное блуждание имеет «снос» (рис. 1), вследствие чего интуитивно понятно, что вероятность возвращения меньше 1. Вычислим, как она зависит от p.

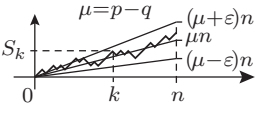


Рис. 1

Как и в гл. 16, будем использовать обозначение u_{2n} для $\mathbf{P}(S_{2n} = 0)$ и f_{2n} для *вероятности вернуться в* 0 *впервые в момент* $2n$. Очевидно, что $u_{2n} = C_{2n}^n p^n q^n$. Пусть $U(z) = \sum\limits_{n=0}^{\infty} u_{2n} z^{2n}$, $F(z) = \sum\limits_{n=1}^{\infty} f_{2n} z^{2n}$, $0 \leqslant z \leqslant 1$. Нетрудно убедиться в том, что формула, связывающая u_{2n} и f_{2n}, приведенная в утверждении леммы из § 4 гл. 16, остается верной и для несимметричного блуждания. Поэтому справедливо также и получаемое с ее помощью равенство $F(z) = 1 - 1/U(z)$ (теорема 2 гл. 16). Вычислим $U(z)$. Для этого понадобится тождество $C_{2n}^n = (-1)^n C_{-1/2}^n 4^n$ (коэффициент $C_{-1/2}^n$ определен выше).

Вопрос 2.
Почему оно выполняется?

Применяя разложение (19) с $x = -1/2$, находим

$$U(z) = \sum_{n=0}^{\infty} (-1)^n C_{-1/2}^n \, 4^n p^n q^n z^{2n} = \left(1 - 4pqz^2\right)^{-1/2}.$$

Следовательно, *вероятность возвращения в* 0 *когда-нибудь* равна

$$F(1) = 1 - 1/U(1) = 1 - \sqrt{1 - 4pq} = 1 - |1 - 2p| = 1 - |p - q|.$$

В частности, при $p = 0{,}9$ имеем $F(1) = 0{,}2$. Заметим, что $F(1) = f_2 + f_4 + \ldots$, но $f_2 = 2pq = 0{,}18$ (рис. 2). Таким образом, если блуждающая частица не вернулась в 0 сразу, то шансы на возвращение в дальнейшем составляют всего лишь 0,02.

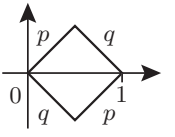


Рис. 2

Наконец, вычислим вероятность возвращения в начало координат для *симметричного* случайного блуждания по k-мерной целочисленной решетке (продолжение материала § 4 гл. 16). Здесь потребуются не только степенные ряды $U(z)$ и $F(z)$, но и характеристические функции.

Вопрос 3.
Как *вероятность невозвращения* $1-F(1)=|p-q|$ связана с вероятностью никогда не проиграть бесконечно богатому противнику (см. § 4 гл. 13)?

Обозначим через $\boldsymbol{S}_n$ положение блуждающей частицы после n шагов. Тогда $\boldsymbol{S}_n = \boldsymbol{\xi}_1 + \ldots + \boldsymbol{\xi}_n$, где k-мерные векторы шагов $\boldsymbol{\xi}_i$ независимы и имеют вид $(0, \ldots, 0, \pm 1, 0, \ldots, 0)$, причем каждое из $2k$ направлений равновероятно.

Прежде всего рассмотрим одномерный случай. Вычислим характеристическую функцию (см. П9) шага ξ_1:

$$\psi_{\xi_1}(t) = \mathbf{M}e^{it\xi_1} = \frac{1}{2}\, e^{-it} + \frac{1}{2}\, e^{it} = \cos t.$$

С учетом свойства 2 из П9 $\psi_{S_n}(t) = \sum e^{itl}\, \mathbf{P}(S_n = l) = \cos^n t$. В силу периодичности при $l \neq m$ функции $e^{it(l-m)}$ на $[-\pi, \pi]$

$$\mathbf{P}(S_n = m) = \frac{1}{2\pi} \int\limits_{-\pi}^{\pi} \psi_{S_n}(t)\, e^{-itm} dt = \frac{1}{2\pi} \int\limits_{-\pi}^{\pi} \cos^n t\; e^{-itm}\, dt.$$

Аналогично в k-мерном случае для векторов $\boldsymbol{t} = (t_1, \ldots, t_k)$ и $\boldsymbol{m} = (m_1, \ldots, m_k)$ справедливы формулы

$$\psi_{\boldsymbol{\xi}_1}(\boldsymbol{t}) = \mathbf{M}e^{i\boldsymbol{t}^T\boldsymbol{\xi}_1} = \frac{1}{2k} \sum_{j=1}^{k} \left(e^{-it_j} + e^{it_j}\right) = \frac{1}{k} \sum_{j=1}^{k} \cos t_j,$$

$$\mathbf{P}(\boldsymbol{S}_n = \boldsymbol{m}) = (2\pi)^{-k} \int\limits_{-\pi}^{\pi} \ldots \int\limits_{-\pi}^{\pi} \left[\frac{1}{k} \sum_{j=1}^{k} \cos t_j\right]^n e^{-\boldsymbol{t}^T\boldsymbol{m}}\, d\boldsymbol{t}.$$

Положим в последнем равенстве $\boldsymbol{m} = \mathbf{0}$. Тогда

$$u_n = \mathbf{P}(\boldsymbol{S}_n = \mathbf{0}) = (2\pi)^{-k} \int\limits_{-\pi}^{\pi} \ldots \int\limits_{-\pi}^{\pi} \left[\frac{1}{k} \sum_{j=1}^{k} \cos t_j\right]^n d\boldsymbol{t}.$$

Суммируя далее под знаком интеграла геометрическую прогрессию, получим следующее представление:

$$U(z) = \sum_{n=0}^{\infty} u_n z^n = (2\pi)^{-k} \int\limits_{-\pi}^{\pi} \ldots \int\limits_{-\pi}^{\pi} \left[1 - \frac{z}{k} \sum_{j=1}^{k} \cos t_j\right]^{-1} d\boldsymbol{t}.$$

Согласно теореме 2 гл. 16 возвратность блуждания равносильна расходимости интеграла

$$U(1) = (2\pi)^{-k} \int\limits_{-\pi}^{\pi} \ldots \int\limits_{-\pi}^{\pi} \left[1 - \frac{1}{k} \sum_{j=1}^{k} \cos t_j\right]^{-1} d\boldsymbol{t}. \tag{20}$$

Подынтегральная функция непрерывна всюду внутри k-мерного куба $[-\pi, \pi] \times \ldots \times [-\pi, \pi]$ кроме начала координат. Поэтому расходимость интеграла (20) эквивалентна расходимости интеграла по k-мерному шару сколь угодно малого радиуса ε с центром в $\mathbf{0}$. Перейдем к полярным координатам $(r, \varphi_1, \ldots, \varphi_{k-1})$ по формулам (8) гл. 3 с якобианом замены $J = r^{k-1} \sin^{k-2} \varphi_1 \sin^{k-3} \varphi_2 \ldots \sin \varphi_{k-2}$. Разложение в ряд Тейлора функции $\cos t = 1 - \frac{1}{2}\, t^2 + o(t^2)$ при $t \to 0$ позволяет заменить подинтегральную функцию вблизи $\mathbf{0}$ на

$\left[\frac{1}{2}\,(t_1^2+\ldots+t_k^2)\right]^{-1} = \left[\frac{1}{2}\,r^2\right]^{-1}$. Таким образом, интеграл по шару эквивалентен произведению одномерных интегралов

$$\int\limits_0^{\varepsilon} \frac{r^{k-1}}{\frac{1}{2}\,r^2}\,dr \left[\int\limits_0^{\pi} |\sin^{k-2}\varphi_1|\,d\varphi_1 \ldots \int\limits_0^{\pi} |\sin\varphi_{k-2}|\,d\varphi_{k-2} \int\limits_0^{2\pi} d\varphi_{k-1}\right],$$

где выражение в квадратных скобках равно некоторой положительной константе. Следовательно, сходимость или расходимость кратного интеграла определяется поведением одномерного интеграла $\int\limits_0^{\varepsilon} r^{k-3}\,dr$, который, очевидно, расходится при $k \leqslant 2$ и сходится при бóльших k. ■

Вероятность возвращения когда-нибудь $F(1) = 1 - 1/U(1)$ можно найти, численно подсчитав интеграл в правой части формулы (20). Она приближенно равна 0,35 при $k = 3$ и 0,2 при $k = 4$.

ЗАДАЧИ

1. Д. Хэбб и К. Уильямс (см. [88, с. 171]) разработали тест эстакадного лабиринта для сравнительной оценки «сообразительности» животных. Он состоит из 12 заданий. В приведенной ниже таблице даны средние числа ошибок при выполнении этих заданий крысами, кроликами и кошками (в скобках указаны ранги Q_{ij} внутри каждой строки).

Самое прекрасное, что мы можем испытать, это ощущение тайны. Она есть источник всякого подлинного искусства и всякой науки.

А. Эйнштейн

Номер задания	Животные		
	Крысы	Кролики	Кошки
1	1,5 (2)	1,7 (3)	0,3 (1)
2	1,1 (2)	1,5 (3)	1,0 (1)
3	1,8 (1)	8,1 (3)	3,6 (2)
4	1,9 (3)	1,3 (2)	0,0 (1)
5	4,3 (3)	4,0 (2)	0,6 (1)
6	2,0 (1)	4,6 (2)	5,5 (3)
7	8,4 (3)	4,0 (2)	1,0 (1)
8	6,6 (3)	5,1 (2)	3,1 (1)
9	2,4 (2)	2,5 (3)	0,1 (1)
10	6,5 (2)	6,9 (3)	1,6 (1)
11	2,6 (2)	2,5 (1)	4,3 (3)
12	6,5 (2)	6,8 (3)	1,0 (1)

Есть ли животные, которые значимо различаются?

2.* Докажите тождество (10).

3* Установите, при каких значениях коэффициента $\rho \in [-1,1]$ матрица $\boldsymbol{\Sigma} = \|\sigma_{ij}\|_{n\times n}$, где

$$\sigma_{ij} = \begin{cases} 1, & \text{если } i = j, \\ \rho, & \text{если } i \neq j, \end{cases}$$

неотрицательно определена (см. П10).
Указание. Используйте представление $\boldsymbol{\Sigma} = (1-\rho)\boldsymbol{E} + \rho\boldsymbol{I}$, где $\boldsymbol{E}$ — единичная матрица, а $\boldsymbol{I}$ — матрица, все элементы которой равны 1, и примените преобразование, приводящее $\boldsymbol{I}$ к главным осям.

РЕШЕНИЯ ЗАДАЧ

Вредно даже читать о предмете прежде, чем сам не поразмышляешь о нем. Ибо вместе с новым материалом в голову прокрадывается чужая точка зрения на него и чужое отношение к нему, и это тем вероятнее, что леность и апатия внушают избавиться от усилий мышления и принимать готовые мысли и давать им ход. Эта привычка затем вкореняется, и тогда мысли все уж идут обычной дорожкой, подобно ручейкам, отведенным в канавы: найти собственную, новую мысль тогда уже вдвойне трудно. Это в значительной мере обусловливает недостаток оригинальности у ученых.

А. Шопенгауэр

1. Применим критерий Фридмана. Суммы рангов по столбцам T_j равны 26, 29 и 17 соответственно. Поэтому статистика критерия $F = 6{,}5$. Согласно табл. Т3 критическим значением $z_{1-\alpha}$ на уровне $\alpha = 0{,}05$ распределения хи-квадрат с $k-1 = 2$ степенями свободы является 5,99. Так как $6{,}5 > 5{,}99$, то гипотеза H_0 об одинаковой сообразительности отвергается на данном уровне значимости.

Поскольку число блоков $n = 12$ невелико, вычислим также поправку Имана и Давенпорта. Новая статистика $G = 4{,}086$. В данном случае $k \leqslant 7$ и $n > 7$, поэтому $l = 1$ и $m = 2$. Отсюда $\tilde{F} = \dfrac{1}{2}(lF + mG) = 7{,}336$. Критическое значение $f_{1-\alpha}$ закона Фишера—Снедекора $F_{2,\,22}$ уровня $\alpha = 0{,}05$, вычисленное линейной интерполяцией по аргументу $1/k_2$ табл. Т5, равно 3,44. Критическая граница $\dfrac{1}{2}(lz_{1-\alpha} + mf_{1-\alpha}) = 6{,}435$. Поскольку $\tilde{F} = 7{,}336 > 6{,}435$, гипотеза H_0 отклоняется.

Проведем сравнение видов животных (обработок) на уровне $\alpha = 5\%$. Тогда $C_{k,\alpha}\sqrt{nk(k+1)/12} = 3{,}31\sqrt{12} \approx 11{,}5$. Так как $|T_2 - T_3| = |29 - 17| = 12 > 11{,}5$, то кошки сообразительнее кроликов. Однако, $|T_1 - T_3| = |26 - 17| = 9 < 11{,}5$, поэтому нельзя утверждать то же самое относительно крыс.

2. Представим левую часть соотношения (10) в виде

$$\sum_{i=1}^{n}\sum_{j=1}^{k}(X_{ij} - X_{\cdot\cdot})^2 = \sum_{i=1}^{n}\sum_{j=1}^{k}(a_i + b_j + c_{ij})^2,$$

где $a_i = X_{i\cdot} - X_{\cdot\cdot}$, $b_j = X_{\cdot j} - X_{\cdot\cdot}$ и $c_{ij} = X_{ij} - X_{i\cdot} - X_{\cdot j} + X_{\cdot\cdot}$. Для доказательства тождества (10) достаточно вывести, что

$$\sum_{i,j}(a_i + b_j + c_{ij})^2 = \sum_{i,j}a_i^2 + \sum_{i,j}b_i^2 + \sum_{i,j}c_{ij}^2 = k\sum_{i}a_i^2 + n\sum_{j}b_i^2 + \sum_{i,j}c_{ij}^2.$$

Проверим, что $\sum\limits_{i,j} a_i b_j = \sum\limits_{i,j} a_i c_{ij} = \sum\limits_{i,j} b_j c_{ij} = 0$. В самом деле,

$$\sum_{i,j} a_i b_j = \sum_{i=1}^{n} a_i \sum_{j=1}^{k} (X_{\cdot j} - X_{\cdot\cdot}) = \sum_{i=1}^{n} a_i \left[\sum_{j=1}^{k} X_{\cdot j} - kX_{\cdot\cdot} \right] = 0,$$

$$\sum_{i,j} a_i c_{ij} = \sum_{i=1}^{n} a_i \sum_{j=1}^{k} (X_{ij} - X_{\cdot j} - X_{i\cdot} + X_{\cdot\cdot}) =$$

$$= \sum_{i=1}^{n} a_i \left[\sum_{j=1}^{k} X_{ij} - \sum_{j=1}^{k} X_{\cdot j} - k(X_{i\cdot} - X_{\cdot\cdot}) \right] =$$

$$= \sum_{i=1}^{n} a_i \left[kX_{i\cdot} - kX_{\cdot\cdot} - k(X_{i\cdot} - X_{\cdot\cdot}) \right] = 0.$$

Аналогично проверяется равенство $\sum\limits_{i,j} b_j c_{ij} = 0$.

3. Собственному значению $\lambda_1 = n$ матрицы $\boldsymbol{I}$ соответствует собственный вектор $(1,\ldots,1)$. Так как ранг матрицы $\boldsymbol{I}$, очевидно, равен 1, то все остальные собственные значения $\lambda_2 = \ldots = \lambda_n = 0$. Пусть $\boldsymbol{C}$ — ортогональная матрица преобразования, приводящего $\boldsymbol{I}$ к главным осям: $\boldsymbol{C}^T \boldsymbol{I} \boldsymbol{C} = \boldsymbol{\Lambda}$. Заметим, что матрица

$$\boldsymbol{B} = \boldsymbol{C}^T \boldsymbol{\Sigma} \boldsymbol{C} = (1-\rho)\, \boldsymbol{C}^T \boldsymbol{E} \boldsymbol{C} + \rho\, \boldsymbol{C}^T \boldsymbol{I} \boldsymbol{C} = (1-\rho)\boldsymbol{E} + \rho\, \boldsymbol{\Lambda}$$

подобна $\boldsymbol{\Sigma}$ (см. П10). Поэтому согласно следствию 2 из П10 $D_n = \det \boldsymbol{\Sigma} = \det \boldsymbol{B}$. Но $\boldsymbol{B}$ — диагональная матрица, у которой $b_{11} = 1 - \rho + \rho n$, $b_{22} = \ldots = b_{nn} = 1 - \rho$. Следовательно, $D_n = (1 - \rho + \rho n)(1 - \rho)^{n-1}$. Условие неотрицательности всех миноров D_i равносильно неравенству $-\dfrac{1}{n-1} \leqslant \rho \leqslant 1$. Для неотрицательной определенности матрицы $\boldsymbol{\Sigma}$ при любых n необходимо и достаточно, чтобы $0 \leqslant \rho \leqslant 1$.

ОТВЕТЫ НА ВОПРОСЫ

1. Возможно $(3!)^2 = 36$ ранговых наборов. Поскольку статистика F не изменяется при перенумерации блоков и обработок, достаточно рассмотреть только 6 вариантов:

Чтобы правильно задать вопрос, нужно знать бóльшую часть ответа.
Р. Шекли

$$\begin{pmatrix}1&2&3\\1&2&3\end{pmatrix}, \quad \begin{pmatrix}1&2&3\\1&3&2\end{pmatrix}, \quad \begin{pmatrix}1&2&3\\2&1&3\end{pmatrix},$$
$$\begin{pmatrix}1&2&3\\2&3&1\end{pmatrix}, \quad \begin{pmatrix}1&2&3\\3&1&2\end{pmatrix}, \quad \begin{pmatrix}1&2&3\\3&2&1\end{pmatrix}.$$

Легко вычислить, что значения статистики F для них равны 4, 3, 3, 1, 1, и 0 соответственно. Поэтому $\mathsf{P}(F \geqslant 4) = 1/6$.

2) Согласно определению обобщенного коэффициента

$$C_{-1/2}^{n}=\frac{-\frac{1}{2}\left(-\frac{1}{2}-1\right)\dots\left(-\frac{1}{2}-n+1\right)}{n!}=\frac{(-1)^n\cdot 1\cdot 3\cdot\ldots\cdot(2n-1)}{2^n n!}=$$

$$=\frac{(-1)^n(2n)!}{2^n n!\cdot 2\cdot 4\cdot\ldots\cdot 2n}=\frac{(-1)^n(2n)!}{4^n(n!)^2}=(-1)^n\,C_{2n}^{n}\,4^{-n}.$$

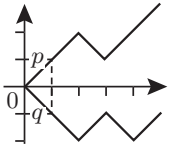


Рис. 3

3. Совершим один шаг блуждания. Тогда частица окажется либо в точке 1 с вероятностью p, либо в точке -1 с вероятностью q (рис. 3). Пусть для определенности $p > 1/2$. Тогда в первом случае невозвращение равносильно неразорению игрока с начальным капиталом 1 при игре против «бесконечно богатого» противника. Его вероятность (см. § 4 гл. 13) равна $1-\lambda = 1-q/p$. Во втором случае возвращение происходит с вероятностью 1. Окончательно имеем, что вероятность невозвращения есть $p(1-q/p)+q\cdot 0 = p-q$.

Глава 18

СГРУППИРОВАННЫЕ ДАННЫЕ

Нередко данные, находящиеся в распоряжении исследователя, представляют собой таблицу количеств попаданий наблюдений в некоторые множества. В этой главе будут рассмотрены методы, позволяющие анализировать такие данные. Все они имеют в качестве предельного закона для статистики критерия *распределение хи-квадрат*, определенное в примере 3 гл. 11.

Эти методы весьма универсальны, но одновременно довольно грубы из-за потери информации при группировке. Их можно рекомендовать для применения на предварительной стадии статистического анализа.

Ба! Знакомые все лица!
Фамусов в «Горе от ума»
А. С. Грибоедова

§ 1. ПРОСТАЯ ГИПОТЕЗА

Пусть $\xi_1, \ldots, \xi_n$ — выборка (см. § 1 гл. 4) из закона с функцией распределения $F(x)$. Разобьем множество значений ξ_1 на N промежутков (возможно, бесконечных) $\Delta_j = (a_j, b_j]$, $j = 1, \ldots, N$ (рис. 1).*) Положим $p_j = \mathbf{P}(\xi_1 \in \Delta_j)$, а случайные величины ν_j — равными количеству элементов выборки в Δ_j ($\nu_1 + \ldots + \nu_N = n$).

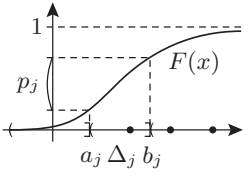


Рис. 1

Функция F неизвестна. Проверяется гипотеза

$$H_0 \colon F(x) = F_0(x),$$

где F_0 — заданная функция распределения. Если гипотеза верна, то согласно закону больших чисел (П6) *частоты попадания в промежутки* $\widehat{p}_j = \nu_j / n$ при достаточно больших n должны быть близки к соответствующим вероятностям $p_j^0 = F_0(b_j) - F_0(a_j)$.

В качестве *меры отклонения* от гипотезы H_0 Карл Пирсон в 1900 г. предложил статистику

$$X_n^2 = n \sum_{j=1}^{N} \frac{1}{p_j^0} \left(\widehat{p}_j - p_j^0 \right)^2 = \sum_{j=1}^{N} \frac{\left(\nu_j - n p_j^0 \right)^2}{n p_j^0} \,. \tag{1}$$

Замечание 1. Первое представление в формуле (1) показывает, что X_n^2 есть взвешенная сумма квадратов отклонений частот от

*) Если множество значений ξ_1 является интервалом, то $a_j = b_{j-1}$.

гипотетических вероятностей. Для фиксированного промежутка в силу центральной предельной теоремы (П6) каждое отклонение асимптотически нормально (см. § 4 гл. 7) и имеет порядок малости $1/\sqrt{n}$. Множитель n перед суммой необходим для того, чтобы предельное распределение статистики не вырождалось в 0. Поскольку складываются квадраты отклонений с весами, обратно пропорциональными гипотетическим вероятностям (чтобы «уравнять» слагаемые между собой), представляется правдоподобным, что предельным законом будет распределение хи-квадрат — сумма квадратов независимых и одинаково распределенных по закону $\mathcal{N}(0,1)$ случайных величин.

Теорема 1. Если $0 < p_j^0 < 1$, $j = 1,\ldots,N$, то при $n \to \infty$

$$X_n^2 \xrightarrow{d} \zeta \sim \chi^2_{N-1}.$$

Вопрос 1. Почему число степеней свободы предельного закона не совпадает с числом слагаемых в суммах из (1)?

Доказательство. Раскладывая независимые «шарики» ξ_i $(i = 1,\ldots,n)$ по «ящикам» Δ_j $(j = 1,\ldots,N)$ с вероятностями p_j^0 попадания в j-й «ящик» (см. § 5 гл. 10), получим

$$\mathbf{P}(\nu_1 = l_1,\ldots,\nu_N = l_N) = \frac{n!}{l_1!\ldots l_N!}\ (p_1^0)^{l_1}\ldots(p_N^0)^{l_N},$$

если все $l_j \geqslant 0$ и $l_1 + \ldots + l_N = n$, иначе вероятность равна 0.

Используя известную формулу возведения суммы в n-ю степень

$$(a_1 + \ldots + a_N)^n = \sum_{\substack{l_1\geqslant 0,\ldots,l_N\geqslant 0,\\ l_1+\ldots+l_N=n}} \frac{n!}{l_1!\ldots l_N!}\ a_1^{l_1}\ldots a_N^{l_N},$$

находим, что характеристическая функция (см. П9) случайного вектора $\boldsymbol{\nu} = (\nu_1,\ldots,\nu_N)$ имеет вид

$$\psi_{\boldsymbol{\nu}}(\boldsymbol{t}) = \mathbf{M}e^{i\boldsymbol{t}^T\boldsymbol{\nu}} = \left(p_1^0 e^{it_1} + \ldots + p_N^0 e^{it_N}\right)^n, \quad \boldsymbol{t} = (t_1,\ldots,t_N). \tag{2}$$

Нетрудно убедиться, что для преобразованного случайного вектора $\boldsymbol{\nu}^* = (\nu_1^*,\ldots,\nu_N^*)$ с компонентами $\nu_j^* = (\nu_j - np_j^0)/\sqrt{n}$ характеристическая функция выглядит так:

$$\psi_{\boldsymbol{\nu}^*}(\boldsymbol{t}) = e^{-i\sqrt{n}\,\boldsymbol{t}^T\boldsymbol{p}^{\mathbf{0}}}\left[1 + \sum_{j=1}^N p_j^0\left(e^{it_j/\sqrt{n}} - 1\right)\right]^n, \quad \boldsymbol{p}^{\mathbf{0}} = \left(p_1^0,\ldots,p_N^0\right).$$

Логарифмируя и раскладывая при $\varepsilon \to 0$ в ряды Тейлора функции $\ln(1+\varepsilon) = \varepsilon - \varepsilon^2/2 + O(\varepsilon^3)$ и $e^{i\varepsilon} = 1 + i\varepsilon - \varepsilon^2/2 + O(\varepsilon^3)$ (см. [82, с. 573]), получаем:

$$\begin{aligned}
\ln\psi_{\boldsymbol{\nu}^*}(\boldsymbol{t}) &= -i\sqrt{n}\,\boldsymbol{t}^T\boldsymbol{p}^{\mathbf{0}} + n\sum_{j=1}^N p_j^0\left(e^{it_j/\sqrt{n}} - 1\right) - \\
&- \frac{n}{2}\left[\sum_{j=1}^N p_j^0\left(e^{it_j/\sqrt{n}} - 1\right)\right]^2 + O\left(1/\sqrt{n}\right) = \\
&= -\frac{1}{2}\sum_{j=1}^N p_j^0 t_j^2 + \frac{1}{2}\left(\sum_{j=1}^N p_j^0 t_j\right)^2 + O\left(1/\sqrt{n}\right) = -\frac{1}{2}\,\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t} + O\left(1/\sqrt{n}\right),
\end{aligned}$$

где (см. П10)

$$\boldsymbol{\Sigma} = \|\sigma_{jk}\|_{N\times N},\ \sigma_{jk} = \begin{cases} p_j^0\,(1-p_j^0) & \text{при } k=j, \\ -p_j^0\,p_k^0 & \text{при } k\neq j. \end{cases} \tag{3}$$

Отсюда следует, что предел $\psi_{\boldsymbol{\nu}^*}(\boldsymbol{t})$ при $n\to\infty$ есть характеристическая функция $\exp\left\{-\frac{1}{2}\,\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t}\right\}$ многомерного нормального закона $\mathcal{N}(\mathbf{0},\boldsymbol{\Sigma})$ (см. П9). (Неотрицательная определенность матрицы $\boldsymbol{\Sigma}$ устанавливается в задаче 5.) По теореме непрерывности из П9 распределение случайной величины $\boldsymbol{\nu}^*$ сходится к указанному закону.

Заметим, что ковариационная матрица $\boldsymbol{\Sigma}$ *вырождена* (П10). Причиной этого является линейная зависимость компонент вектора $\boldsymbol{\nu}^*$:

$$\nu_1^* + \ldots + \nu_N^* = 0. \tag{4}$$

Однако, ее подматрица $\boldsymbol{A}$ размера $(N-1)\times(N-1)$ уже не вырождена. Действительно, нетрудно убедиться, что обратной к ней служит матрица

$$\boldsymbol{A}^{-1} \equiv \boldsymbol{B} = \|b_{jk}\|_{(N-1)\times(N-1)},\ b_{jk} = \begin{cases} 1/p_j^0 + 1/p_N^0 & \text{при } k=j, \\ 1/p_N^0 & \text{при } k\neq j. \end{cases}$$

Таким образом, для подвектора $\boldsymbol{c} = \left(\nu_1^*,\ldots,\nu_{N-1}^*\right)$ предельным будет невырожденный нормальный закон $\mathcal{N}(\mathbf{0},\boldsymbol{A})$. Согласно последнему утверждению из П9 и свойству 3 сходимости из П5

$$\boldsymbol{c}\boldsymbol{B}\boldsymbol{c}^T \xrightarrow{d} \zeta \sim \chi^2_{N-1} \quad \text{при } n\to\infty. \tag{5}$$

С другой стороны, из формул (1) и (4) имеем

$$X_n^2 = \sum_{j=1}^{N} \frac{1}{p_j^0}\left(\nu_j^*\right)^2 = \sum_{j=1}^{N-1} \frac{1}{p_j^0}\left(\nu_j^*\right)^2 + \frac{1}{p_N^0}\left(\nu_1^* + \ldots + \nu_{N-1}^*\right)^2.$$

Но правая часть совпадает с $\boldsymbol{c}\boldsymbol{B}\boldsymbol{c}^T$, что с учетом сходимости (5) завершает доказательство теоремы 1. ■

Как отмечено в [32, с. 111], приближение распределения статистики X_n^2 с помощью закона χ^2_{N-1} является достаточно точным при $n \geqslant 50$ и $np_j^0 \geqslant 5$ для всех $j = 1,\ldots,N$.

Замечание 2. Последнее условие предназначено для того, чтобы обеспечивать возможность попадания хотя бы нескольких наблюдений ξ_i в каждый из промежутков Δ_j. Это необходимо для пригодности лежащего в основе теоремы 1 нормального приближения для распределения величин $\sqrt{n}\,(\widehat{p}_j - p_j^0)$: чем больше для заданного j *ожидаемое количество попаданий* np_j^0, тем приближение точнее. Поэтому число промежутков N не должно быть слишком большим. Однако, его не следует брать и очень малым, так как в этом случае

набор вероятностей $p_1^0,\ldots,p_N^0$ недостаточно хорошо представляет гипотетическую функцию распределения $F_0(x)$. Обычно на практике берут $N \approx \log_2 n$.

Когда N выбрано, возникает вопрос, каким образом задавать промежутки $\Delta_j = (a_j, b_j]$. Если областью возможных значений случайной величины ξ_1 служит ограниченный интервал, то можно разбить его на *равные по длине части*. Альтернативным выбором (годящимся для неограниченных областей значений ξ_1) является разбиение действительной прямой на *равновероятные промежутки*, у которых $a_j = b_{j-1}$, а правые границы b_j находятся из уравнений $F_0(b_j) = j/N$, $j = 1,\ldots,N$.

Иногда N и p_j^0 не выбираются исследователем, а определяются самой изучаемой проблемой.

Г. И. Мендель (1822–1884), австрийский естествоиспытатель.

Пример 1. ***Генетические законы Менделя*** (см. [35, с. 563])**.** В экспериментах с селекцией гороха (1856–1863) Мендель наблюдал частоты различных видов семян, получаемых при скрещивании растений с круглыми желтыми семенами и растений с морщинистыми зелеными семенами. Эти данные и значения теоретических вероятностей, определяемые в соответствии с законом Менделя независимого расщепления признаков, приведены в следующей таблице:

Тип семян	Частота $\widehat{p}_j$	Вероятность p_j^0
Круглые и желтые	315/556	9/16
Морщинистые и желтые	101/556	3/16
Круглые и зеленые	108/556	3/16
Морщинистые и зеленые	32/556	1/16

Проверим гипотезу H_0 о согласованности частот с теоретическими вероятностями при помощи критерия хи-квадрат. Статистика критерия (см. формулу (1)) $X_n^2 \approx 0{,}47$. Из табл. Т3 получаем, что это значение находится между квантилями уровня 0,05 и 0,1 закона χ_3^2. Таким образом, согласие наблюдений с гипотезой H_0 очень хорошее.

Вопрос 2. Чем подозрителен датчик псевдослучайных чисел, у которого в промежутки $\left(0,\frac{1}{4}\right]$, $\left(\frac{1}{4},\frac{1}{2}\right]$, $\left(\frac{1}{2},\frac{3}{4}\right]$ и $\left(\frac{3}{4},1\right]$ попали соответственно 504, 505, 492 и 499 точек?

§ 2. СЛОЖНАЯ ГИПОТЕЗА

Метод группировки наблюдений с последующим применением критерия хи-квадрат применим и для проверки сложной гипотезы H_0' о принадлежности неизвестной функции распределения элементов выборки некоторому заданному классу функций распределения $\mathcal{F} = \{F(x,\boldsymbol{\theta}),\ \boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^k\}$.

В этом случае общая (при всевозможных $\boldsymbol{\theta} \in \Theta$) область значений ξ_1 также разбивается на N промежутков $\Delta_j = (a_j, b_j]$,

$j = 1, \ldots, N$. Как и ранее ν_j обозначает число элементов выборки в Δ_j. Однако теперь вероятности $\mathbf{P}(\xi_1 \in \Delta_j)$ при H_0' уже не будут заданы однозначно, а представляют собой функции от $\boldsymbol{\theta}$: $p_j(\boldsymbol{\theta}) = F(b_j, \boldsymbol{\theta}) - F(a_j, \boldsymbol{\theta})$ (рис. 2). Из-за этой зависимости от неизвестного параметра нельзя просто подставить $p_j(\boldsymbol{\theta})$ вместо p_j^0 в (1). Р. Фишер (1924 г.) доказал, что если подставить $p_j(\tilde{\boldsymbol{\theta}})$, где $\tilde{\boldsymbol{\theta}}$ — *оценка максимального правдоподобия, основанная на частотах* (определяемая ниже), то при некоторых условиях на класс $\mathcal{F}$ функций распределения (см. [32, с. 115]) статистика

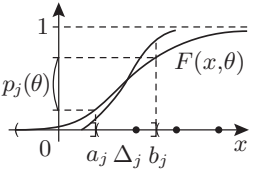


Рис. 2

$$\tilde{X}_n^2 = \sum_{j=1}^{N} \bigl(\nu_j - np_j(\tilde{\boldsymbol{\theta}})\bigr)^2 \Big/ \bigl[np_j(\tilde{\boldsymbol{\theta}})\bigr] \tag{6}$$

будет иметь в качестве предельного закона снова распределение хи-квадрат, только уже с $(N - 1 - k)$ степенями свободы, где k — размерность вектора $\boldsymbol{\theta}$.

Доказательство этой теоремы можно найти в [44, с. 462–470].

Определение. *Значением оценки максимального правдоподобия, основанной на частотах* $\tilde{\boldsymbol{\theta}}$, *служит вектор* $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$, *на котором достигается максимум вероятности*

$$\mathbf{P}(\nu_1 = l_1, \ldots, \nu_N = l_N) = \frac{n!}{l_1! \ldots l_N!}\, [p_1(\boldsymbol{\theta})]^{l_1} \ldots [p_N(\boldsymbol{\theta})]^{l_N}.$$

Это равносильно максимизации по $\boldsymbol{\theta}$ *функции*

$$\sum_{j=1}^{N} l_j \ln p_j(\boldsymbol{\theta}) \tag{7}$$

или (для гладких моделей) решению системы, вообще говоря, нелинейных уравнений

$$\sum_{j=1}^{N} l_j \frac{\partial \ln p_j(\boldsymbol{\theta})}{\partial \theta_m} = 0, \quad m = 1, \ldots, k. \tag{8}$$

Пример 2. ***Критерий*** χ^2 ***для пуассоновской модели*** (см. [32, с. 116])**.** Положим $\pi_m(\theta) = e^{-\theta}\theta^m/m!$, $m \geqslant 0$. Возьмем промежутки $\Delta_j = [j-1, j)$, $j = 1, \ldots, N-1$; $\Delta_N = [N-1, \infty)$. Тогда вероятности $p_j(\theta) = \pi_{m-1}(\theta)$, $j = 1, \ldots, N-1$; $p_N(\theta) = \sum\limits_{m=N-1}^{\infty} \pi_m(\theta)$. Так как θ — скалярный параметр, причем $(d/d\theta) \ln \pi_m(\theta) = m/\theta - 1$, то система (8) сводится к одному уравнению:

$$\sum_{j=0}^{N-2} l_{j+1}\,(j/\theta - 1) + l_N \sum_{m=N-1}^{\infty} (m/\theta - 1)\,\pi_m(\theta) \Big/ \sum_{m=N-1}^{\infty} \pi_m(\theta) = 0.$$

Поскольку $l_1 + \ldots + l_N = n$, отсюда получаем соотношение

$$\theta = \frac{1}{n}\left[\sum_{j=0}^{N-2} j\, l_{j+1} + l_N \sum_{m=N-1}^{\infty} m\,\pi_m(\theta) \Big/ \sum_{m=N-1}^{\infty} \pi_m(\theta)\right]. \tag{9}$$

Первый член в скобках равен сумме всех значений ξ_i, меньших или равных $N-2$. Второй член представляет собой $l_N\,\mathbf{M}(\xi_1 \,|\, \xi_1 \geqslant N-1)$ (см. П7). Он *приближенно* равен сумме всех значений ξ_i, которые больше или равны $N-1$. Поэтому решение $\tilde{\theta}$ уравнения (9) близко к среднему арифметическому $\overline{\xi}$ — оценке максимального правдоподобия параметра θ, построенной по всей выборке.

Вопрос 3. Почему $\overline{\xi}$ — ОМП для пуассоновской модели?

Применим критерий хи-квадрат к данным о падениях самолетов-снарядов в южной части Лондона во время второй мировой войны (см. [81, с. 177]). Опасность попадания в жилые дома вместо военных объектов велика при низкой точности стрельбы (при так называемой *стрельбе по площадной цели*).

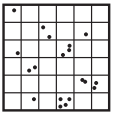


Рис. 3

Карта южной части Лондона была разбита на $n = 24 \times 24 = 576$ небольших участков, каждый площадью 1/4 кв. км. На карте были отмечены места падения самолетов-снарядов (подобно рис. 3). В таблице ниже приведены количества участков l_{j+1} ровно с j падениями, $j = 0{,}1, \ldots, 7$. Так как участков много, а вероятность попадания самолета-снаряда на отдельный участок мала, то при справедливости гипотезы о низкой точности стрельбы можно воспользоваться законом редких событий (см. § 1 гл. 5), согласно которому число попаданий на любой из участков есть (приближенно) пуассоновская случайная величина с некоторым общим для всех участков параметром θ. Мы также предположим, что попадания на разные участки независимы.

Общее число падений $M = \sum j\, l_{j+1} = 537$. Возьмем в качестве начальной оценки неизвестного параметра закона Пуассона *среднее число падений на один участок* $\widehat{\theta} = M/n \approx 0{,}932$. Тогда ожидаемые количества участков ровно с j падениями примерно равны $n\pi_j(\widehat{\theta})$.

j	0	1	2	3	4	5	6	7
l_{j+1}	229	211	93	35	7	0	0	1
$n\pi_j(\widehat{\theta})$	226,7	211,4	98,5	30,6	7,14	1,33	0,21	0,03
$n\pi_j(\tilde{\theta})$	228,6	211,3	97,6	30,1	8,46			

Прежде чем вычислять статистику критерия хи-квадрат, надо объединить последние 4 столбца таблицы для того, чтобы ожидаемое количество оказалось не меньше 5: $l_4 + \ldots + l_7 = 8$ и $n(\pi_4(\widehat{\theta}) + \ldots + \pi_7(\widehat{\theta})) = 8{,}71$.

Теперь заменим начальную оценку $\widehat{\theta}$ на $\tilde{\theta}$, максимизируя по θ функцию (7) на компьютере (удобно вычислять $p_5(\theta)$ по формуле $p_5(\theta) = 1 - p_1(\theta) - \ldots - p_4(\theta)$). Вероятно, проще всего уменьшать θ с шагом $h = 0{,}001$ до тех пор, пока функция возрастает. Ответ таков: $\tilde{\theta} = 0{,}924$ (отличие от $\widehat{\theta}$ составляет всего-навсего 0,008). Соответствующие ожидаемые количества приведены в третьей строке таблицы.

Значение статистики $\tilde{X}_n^2$ (см. формулу (6)) для таких данных равно 1,05. Поскольку $N = 5$ и $k = 1$, предельный закон должен иметь $N - k - 1 = 3$ степени свободы. Из табл. Т3 находим, что значение статистики попадает в интервал (0,58; 2,37), обра-

зованный 10% и 50% квантилями χ^2_3 (с помощью таблицы из [10, с. 140] уточняем, что фактический уровень значимости равен 0,79). Поэтому гипотеза о низкой точности стрельбы принимается. В [81, с. 177] отмечено:

> «Большинство населения верило в тенденцию точек падения скапливаться в нескольких местах. Если бы это было верно, то следовало бы ожидать бóльшую долю участков без попаданий либо с большим числом попаданий и меньшую долю участков промежуточного класса. Приведенная таблица показывает, что точки падения были совершенно случайными, все участки — равноправными; здесь мы имеем поучительную иллюстрацию того установленного факта, что неискушенному человеку случайность представляется регулярностью или стремлением к скоплению.»

Обратим внимание на *необходимость объединения маловероятных промежутков*: если оставить $N = 8$, то $\tilde\theta \approx \widehat\theta = 0{,}932$ и $\tilde X^2_n = 32{,}6$. Это значимо велико для χ^2_6 даже на уровне 10^{-5} (см. [10, с. 144]). Причиной резкого роста значения статистики является малая величина $np_8(\tilde\theta) \approx 0{,}03$, придающая слишком большой вес квадрату отклонения наблюдаемого количества $l_8 = 1$ от ожидаемого количества $np_8(\tilde\theta)$.

Если данные предварительно группируются, то оценить $\boldsymbol\theta$ можно и до группировки наблюдений, например, методом максимального правдоподобия (см. § 4 гл. 9). Однако, как показывает следующий пример, в этом случае статистика $\tilde X^2_n$ будет сходиться, вообще говоря, к *другому предельному закону*.

Пример 3. ***Проверка нормальности по сгруппированным данным.*** Пусть $\xi_i \sim \mathcal{N}(\mu,\sigma^2)$, $i = 1,\ldots,n$, причем оба параметра μ и σ неизвестны. Для разбиения прямой на промежутки $\Delta_j = (a_j, b_j]$, $j = 1,\ldots,N$, оценим неизвестную функцию распределения $\Phi((x-\mu)/\sigma)$ при помощи $\Phi((x-\overline\xi)/S)$, где $\overline\xi = \frac1n\sum\xi_i$ и $S^2 = \frac1n\sum(\xi_i-\overline\xi)^2$. Чтобы вероятности попадания ξ_i в промежутки Δ_j были примерно одинаковы, возьмем в качестве b_j решения уравнений

$$\Phi((x-\overline\xi)/S) = j/N, \quad j = 1,\ldots,N-1,$$

(см. табл. Т2 или приближение Хамакера для Φ^{-1} из § 5 гл. 4).

Далее подсчитаем ν_j — количества попаданий в построенные промежутки. Затем вычислим основанную на частотах оценку максимального правдоподобия $\tilde{\boldsymbol\theta} = (\tilde\theta_1,\tilde\theta_2)$ при помощи численного поиска точки максимума функции (7), исходя из точки с координатами $(\overline\xi,S)$. При этом для нахождения $p_j(\boldsymbol\theta)$ понадобится запрограммировать приближенное вычисление $y = \Phi(x)$, например,

с помощью **алгоритма Морана** (см. [58, с. 282]):

$s = 0$
$t = x * \mathrm{Sqr}(2) / 3$
For $i = 0$ To 12
    $z = i + 0.5$
    $s = s + \mathrm{Sin}(z * t) * \mathrm{Exp}(-z * z / 9) / z$
Next i
$y = 0.5 + s / 3.1415926536$

(Он обеспечивает 9 точных десятичных цифр у $\Phi(x)$ при $|x| \leqslant 7$.)

Важно отметить, что сами оценки $\overline{\xi}$ и S использовать в формуле (6) *нельзя*. В [80, с. 322] указано, что в случае нарушения этого запрета статистика $\tilde{X}_n^2$ не будет (асимптотически) следовать распределению хи-квадрат с $N-3$ степенями свободы: график ее функции распределения пройдет несколько ниже графика функции распределения закона χ^2_{N-3}. Не будет она следовать и распределению хи-квадрат с $N-1$ степенями свободы (как было бы при точно известных параметрах). График ее функции распределения пройдет несколько выше.*)

В качестве иллюстрации на рис. 4 приведены графики функций F_7 и F_9 распределения законов χ^2_7 и χ^2_9 соответственно. Они ограничивают полосу, в которой будет проходить график функции распределения предельного закона для $\tilde{X}_n^2$ при $N = 10$, если для вычисления $p_j(\boldsymbol{\theta})$ использовать оценки $\overline{\xi}$ и S. Согласно табл. Т3 на уровне 0,95 ширина полосы равна $16{,}9 - 14{,}1 = 2{,}8$.

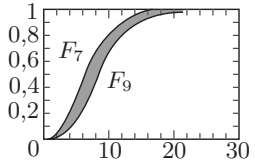


Рис. 4

§ 3. ПРОВЕРКА ОДНОРОДНОСТИ

Допустим, что имеется k *независимых* между собой выборок размеров n_i из распределений F_i, $i = 1, \ldots, k$. Общее число наблюдений $n = n_1 + \ldots + n_k$. Проверим *гипотезу однородности*

$H_0''\colon F_1 = \ldots = F_k$

с помощью критерия хи-квадрат. Для этого сгруппируем данные: разобьем общую для всех выборок область значений наблюдений на промежутки Δ_j, $j = 1, \ldots, N$, и для каждой пары индексов (i, j) подсчитаем величину ν_{ij} — *количество попаданий элементов i-й выборки в j-й промежуток* (рис. 5). В результате получим $k \times N$ таблицу (рис. 6), которую и будем анализировать в дальнейшем.

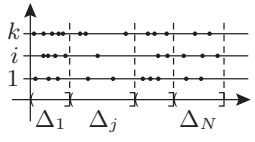


Рис. 5

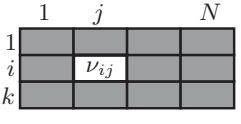


Рис. 6

Иногда данные с самого начала имеют дискретную структуру: в опытах наблюдается некоторый переменный признак, принимающий конечное число N значений (см. пример 4 ниже).

*) Как показали Чернов и Леман в 1954 г. (см. [13, с. 284]), статистика $\tilde{X}_n^2$ асимптотически распределена как сумма $\xi_1^2 + \ldots + \xi_{N-3}^2 + \gamma_1 \xi_{N-2}^2 + \gamma_2 \xi_{N-1}^2$, где ξ_i — независимые $\mathcal{N}(0,1)$-случайные величины; числа γ_1 и γ_2 лежат между 0 и 1 и зависят от проверяемого закона и способа разбиения на промежутки области возможных значений наблюдений.

Если гипотеза H_0'' верна, то ожидаемое количество наблюдений в ячейке с индексами i и j равно $n_i p_j$, где $\boldsymbol{p} = (p_1, \ldots, p_N)$ обозначает (неизвестный) вектор вероятностей попадания в промежутки Δ_j при справедливости гипотезы H_0''. Естественной оценкой для p_j служит $\widehat{p}_j = (\nu_{1j} + \ldots + \nu_{kj})/n$ — общая по всем выборкам частота попаданий в Δ_j (см. задачу 6). Тогда статистика

$$\widehat{X}_n^2 = \sum_{i=1}^{k} \sum_{j=1}^{N} \left(\nu_{ij} - n_i \widehat{p}_j\right)^2 / (n_i \widehat{p}_j) \tag{10}$$

измеряет отклонение наблюдаемых количеств от ожидаемых. Если справедлива гипотеза H_0'', то, как доказано в [44, с. 483], статистика $\widehat{X}_n^2$ сходится по распределению к хи-квадрат случайной величине с $(k-1)(N-1)$ степенями свободы при $\min\{n_1, \ldots, n_k\} \to \infty$.

Следующий любопытный пример из [72, с. 132] показывает, что к выводам, основанным на применении этого предельного результата, следует относиться с известной осторожностью.

Факты — упрямая вещь, но статистика гораздо сговорчивее.

Лоренс Питер

Пример 4. ***Парадокс критерия хи-квадрат*** **[72, с. 132].** Ниже приведены три таблицы, в которых отражено действие некоторого лекарства (способа лечения) только на мужчин, только на женщин и, наконец, на больных обоего пола (объединенные результаты).

Мужчины	B	$\overline{B}$
A	700	800
$\overline{A}$	80	130

Женщины	B	$\overline{B}$
A	150	70
$\overline{A}$	400	280

Вместе	B	$\overline{B}$
A	850	870
$\overline{A}$	480	410

Здесь A — принимавшие лекарство, $\overline{A}$ — не принимавшие лекарство, B — выздоровевшие, $\overline{B}$ — не выздоровевшие.

Заметим, что среди принимавших лекарство мужчин доля выздоровевших $700/(700+800) \approx 0{,}467$ больше, чем $80/(80+130) \approx \approx 0{,}381$ — доля выздоровевших среди мужчин, не принимавших лекарство. Такая же картина и у женщин: $150/220 \approx 0{,}682 > > 400/680 \approx 0{,}588$.

Рассчитано, что петербуржец, проживающий на солнцепеке, выигрывает двадцать процентов здоровья.

Козьма Прутков

Статистики $\widehat{X}_n^2$ (см. формулу (10)) для таблиц данных мужчин и женщин принимает значения 5,456 и 6,125. Из [10, с. 141] (см. также табл. Т3) для закона хи-квадрат с 1 степенью свободы находим, что фактические уровни значимости равны соответственно 0,020 и 0,013. Это говорит о существенности различия вероятностей выздоровления между теми, кто принимал лекарство и теми, кто его не принимал.

С другой стороны, как это ни странно, из таблицы с объединенными результатами следует, что доля выздоровевших больше среди тех людей, которые лекарство *не принимали* (!): $480/870 \approx 0{,}539 > 850/1720 \approx 0{,}494$, причем статистика $\widehat{X}_n^2$ для третьей таблицы равна 4,782, что значимо велико на уровне 0,029.

В [72, с. 133] Г. Секей пишет:

«Аналогично, новое лекарство может оказаться эффективным в каждом из десяти различных госпиталей, но объединение результатов укажет на то, что это лекарство либо бесполезно, либо вредно».

Статистика — самая точная из всех лженаук.

Джин Ко

Причина парадокса заключается в непропорциональном представительстве в разных категориях: мужчины выздоравливают хуже, но лекарство испытывалось в основном на них.

Кроме того, число мужчин (210), не принимавших лекарство, недостаточно велико: согласно таблице, приведенной в книге Дж. Флейс «*Статистические методы для изучения таблиц долей и пропорций*», вероятность β ошибки II рода, для таких данных равна 50%. Чтобы обеспечить $\beta = 10\%$, необходимо иметь не менее 475 пациентов в этой категории.

ЗАДАЧИ

Опыт — лучший учитель.

1. Проверьте первый столбец табл. Т1 на равномерность с помощью критерия хи-квадрат.

2. В [10, с. 21] проводится анализ 2000 четырехзначных псевдослучайных чисел из книги М. Кадырова «Таблицы случайных чисел» (Ташкент, 1936). Первая цифра оказалась нулем у 160, тройкой — у 247, шестеркой — у 191, девяткой — у 185 чисел (остальные 1217 чисел начинались с других цифр). Стоит ли пользоваться такой таблицей?

3. Ниже приведены данные о количестве студентов двух групп, решивших в течение месяца занятий 0, 1–7, 8–15 и более 15 задач. Проверьте гипотезу о том, что студенты обеих групп одинаково активно решают задачи.

Число задач	0	1–7	8–15	> 15
Группа 1	9	8	5	4
Группа 2	3	5	9	11

4. Выведите теорему 1 при $N = 2$ непосредственно из центральной предельной теоремы (П6).

5*. Докажите неотрицательную определенность матрицы $\boldsymbol{\Sigma}$, задаваемой формулой (3), а) вычислив главные миноры (см. П10), б) установив, что она является ковариационной матрицей случайного вектора $\boldsymbol{\nu}^*$ из доказательства теоремы 1.

6*. Покажите при помощи метода неопределенных множителей Лагранжа (см. [46, с. 271]), что оценка $\widehat{p}_j$ из § 3 максимизирует функцию правдоподобия сгруппированной выборки при условии $p_1 + \ldots + p_N = 1$.

РЕШЕНИЯ ЗАДАЧ

Мало хотеть — надо уметь.

1. Поскольку длина столбца $n = 20$ возьмем $N = 4 \approx \log_2 n$ промежутков. При справедливости гипотезы равномерности равные

по вероятностям попадания промежутки имеют одинаковую длину $1/N = 1/4$. Ниже приведены подсчитанные по таблице Т1 значения ν_j, $j = 1,\ldots,N$. Все ожидаемые количества $np_j = 5$. Отсюда $X_n^2 = (0^2{+}3^2{+}1^2{+}2^2)/5 = 2{,}8$. Согласно таблице Т3 критическим значением закона χ_3^2 на уровне $\alpha = 5\%$ является 7,82. Так как $2{,}8 < 7{,}82$, то гипотеза равномерности принимается.

	0–24	25–49	50–74	75–99
ν_j	5	2	6	7

2. $X_n^2 = (40^2{+}47^2{+}9^2{+}15^2)/200{+}17^2/1200 = 20{,}816$. В соответствии с табл. Т3 эта величина значимо велика для χ_4^2 даже на уровне 0,001. Авторы [10] (с. 21) заключают: «Рекомендацию случайных чисел М. Кадырова для статистических расчетов едва ли можно признать оправданной».

3. Нетрудно по формуле (10) подсчитать, что $\widehat{X}_n^2 \approx 8{,}039$. Согласно табл. Т3 5%-ное критическое значение χ_3^2 равно 7,82. Поскольку $8{,}039 > 7{,}82$, проверяемая гипотеза отвергается.

Обратим внимание на то, что фактический уровень значимости (см. § 1 гл. 12) в данном случае равен 4,5% ([10, с. 141]), т. е. на уровне, скажем, 4% гипотезу пришлось бы принимать. При внимательном просмотре данных, замечаем, что количества ν_{ij} в группе 1 убывают, а в группе 2 возрастают. Без всякой науки различие групп представляется очевидным, в то время как критерий хи-квадрат едва смог его уловить. Эта задача демонстрирует недостаточную чувствительность критерия (особенно для небольших выборок).

4. Положим $q = 1 - p$ и представим статистику X_n^2 в виде

$$\frac{(\nu_1 - np)^2}{np} + \frac{(\nu_2 - nq)^2}{nq} = \frac{(\nu_1 - np)^2}{np} + \frac{[(n - \nu_1) - n(1-p)]^2}{nq} =$$
$$= \frac{(\nu_1 - np)^2}{np} + \frac{(np - \nu_1)^2}{nq} = \frac{(\nu_1 - np)^2}{npq} = \left[\frac{\nu_1 - np}{\sqrt{npq}}\right]^2.$$

В силу центральной предельной теоремы и свойства 3 сходимости из П5 правая часть сходится по распределению к χ_1^2.

5. **а)** Вычислим $\det \boldsymbol{\Sigma}$ для произвольных действительных $p_1^0,\ldots,p_N^0$. Для этого заметим, что $\boldsymbol{\Sigma} = \boldsymbol{D}(\boldsymbol{E} - \boldsymbol{P})$, где $\boldsymbol{D}$ — диагональная матрица с элементами $p_1^0,\ldots,p_N^0$ на главной диагонали, $\boldsymbol{E}$ — единичная матрица, $\boldsymbol{P}$ — матрица, все строки которой равны $(p_1^0,\ldots,p_N^0)^T$. Согласно свойствам определителей $\det \boldsymbol{\Sigma} = \det \boldsymbol{D} \cdot \det(\boldsymbol{E} - \boldsymbol{P}) = p_1^0 \ldots p_N^0 \det(\boldsymbol{E} - \boldsymbol{P})$. Чтобы вычислить $\det(\boldsymbol{E} - \boldsymbol{P})$, применим к $\boldsymbol{P}$ преобразование подобия с *ортогональной* матрицей $\boldsymbol{C}$, у которой первым столбцом

является вектор $(1/\sqrt{N},\ldots,1/\sqrt{N})$ (см. теорему 1 гл. 11):

$$\boldsymbol{C}^T\boldsymbol{P}\boldsymbol{C} = \begin{pmatrix} \frac{1}{\sqrt{N}} & \ldots & \frac{1}{\sqrt{N}} \\ * & * & * \\ * & * & * \end{pmatrix} \begin{pmatrix} p_1^0 & \ldots & p_N^0 \\ \ldots & \ldots & \ldots \\ p_1^0 & \ldots & p_N^0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{N}} & * & * \\ \ldots & * & * \\ \frac{1}{\sqrt{N}} & * & * \end{pmatrix} =$$

$$= \begin{pmatrix} p_1^0\sqrt{N} & \ldots & p_N^0\sqrt{N} \\ 0 & \ldots & 0 \\ 0 & \ldots & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{N}} & * & * \\ \ldots & * & * \\ \frac{1}{\sqrt{N}} & * & * \end{pmatrix} = \begin{pmatrix} \sum p_j^0 & * & * \\ 0 & \ldots & 0 \\ 0 & \ldots & 0 \end{pmatrix}$$

(перемножая $\boldsymbol{C}^T$ и $\boldsymbol{P}$, мы использовали ортогональность всех строк матрицы $\boldsymbol{C}^T$, кроме первой, вектору $(1,\ldots,1)$). Отсюда

$$\boldsymbol{C}^T(\boldsymbol{E}-\boldsymbol{P})\boldsymbol{C} = \boldsymbol{E} - \boldsymbol{C}^T\boldsymbol{P}\boldsymbol{C} = \begin{pmatrix} 1-\sum p_j^0 & * & * & * \\ 0 & 1 & 0 & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ 0 & \ldots & 0 & 1 \end{pmatrix}.$$

Поскольку матрица в правой части является верхнетреугольной, то ее определитель равен $1-\sum p_j^0$. Следовательно, $\det\boldsymbol{\Sigma} = (1-\sum p_j^0)\prod p_j^0$ для любых действительных p_j^0.

Если p_j^0 — это вероятности попадания в промежутки Δ_j ($p_j^0 > 0$, $\sum p_j^0 = 1$), то все главные миноры $\boldsymbol{\Sigma}$ положительны, за исключением последнего, который равен 0.

Так как матрица $\boldsymbol{P}$ не симметрична, то нельзя сослаться на теорему о приведении к главным осям (см. решение задачи 3 гл. 17). Однако ее можно с помощью *невырожденного* преобразования $\boldsymbol{T}$ привести к *жордановой нормальной форме*: $\boldsymbol{J} = \boldsymbol{T}^{-1}\boldsymbol{P}\boldsymbol{T}$, у которой на главной диагонали стоят собственные значения матрицы $\boldsymbol{P}$, над главной диагональю — 0 или 1, а на остальных местах — только нули (см. [49, с. 142]). Легко видеть, что $\sum p_j^0$ — собственное значение матрицы $\boldsymbol{P}$, соответствующее собственному вектору $(1,\ldots,1)$. Поскольку ранг матрицы $\boldsymbol{P}$ равен 1, то все другие собственные значения этой матрицы нулевые. Далее рассуждаем аналогично.

К. Жордан (1838–1922), французский математик.

б) Вычислим моменты первого и второго порядка компонент вектора $\boldsymbol{\nu} = (\nu_1,\ldots,\nu_N)$, где $\nu_j = \sum\limits_{l=1}^{n} I_{\{\xi_l\in\Delta_j\}}$, используя то, что элементы выборки ξ_l $(l=1,\ldots,n)$ независимы и одинаково распределены, а промежутки Δ_j не пересекаются:

$$\mathsf{M}\nu_j = \mathsf{M}\sum_l I_{\{\xi_l\in\Delta_j\}} = \sum_l \mathsf{P}(\xi_l\in\Delta_j) = n\,\mathsf{P}(\xi_1\in\Delta_j) = np_j^0,$$

$$\mathsf{M}\nu_j^2 = \mathsf{M}\sum_{l,\,m} I_{\{\xi_l\in\Delta_j\}}I_{\{\xi_m\in\Delta_j\}} = \sum_{l,\,m}\mathsf{M}I_{\{\xi_l\in\Delta_j,\,\xi_m\in\Delta_j\}} =$$

$$= \left(\sum_{m=l} + \sum_{m\ne l}\right)\mathsf{P}(\xi_l\in\Delta_j,\xi_m\in\Delta_j) = np_j^0 + n(n-1)(p_j^0)^2,$$

$$\mathsf{M}\nu_j\nu_k = \left(\sum_{m=l} + \sum_{m\ne l}\right)\mathsf{P}(\xi_l\in\Delta_j,\xi_m\in\Delta_k) = 0 + n(n-1)p_j^0p_k^0.$$

Согласно определениям дисперсии и ковариации, находим

$$\mathbf{D}\nu_j = \mathbf{M}\nu_j^2 - (\mathbf{M}\nu_j)^2 = np_j^0(1-p_j^0),$$
$$\mathbf{cov}\,(\nu_j,\nu_k) = \mathbf{M}\nu_j\nu_k - \mathbf{M}\nu_j\cdot\mathbf{M}\nu_k = -np_j^0p_k^0.$$

В силу свойств дисперсии и ковариации 1 и 4 из П2 матрица $\boldsymbol{\Sigma}$ служит ковариационной матрицей случайного вектора $\boldsymbol{\nu}^*$.

Моменты компонент случайного вектора $\boldsymbol{\nu}$ также можно получить дифференцированием характеристической функции $\psi_{\boldsymbol{\nu}}(\boldsymbol{t})$, $\boldsymbol{t}=(t_1,\ldots,t_N)$, задаваемой формулой (2), на основании следующего утверждения (см. [65, с. 165]).

Утверждение. Если у случайного вектора $\boldsymbol{\xi}$ конечны моменты $\mathbf{M}|\xi_j|^m$ при всех $j=1,\ldots,N$, то существуют *смешанные моменты* $\alpha_{l_1,\ldots,l_N}=\mathbf{M}\,\xi_1^{l_1}\ldots\xi_N^{l_N}$ для всех $l_j\geqslant 0$, $l_1+\ldots+l_N\leqslant m$. В этом случае характеристическая функция $\psi_{\boldsymbol{\xi}}(\boldsymbol{t})$ имеет непрерывные частные производные до порядка m включительно, причем

$$\alpha_{l_1,\ldots,l_N} = (-i)^{l_1+\ldots+l_N}\left.\frac{\partial^{l_1+\ldots+l_N}\psi_{\boldsymbol{\xi}}(\boldsymbol{t})}{\partial t_1^{l_1}\ldots\partial t_N^{l_N}}\right|_{t_1=\ldots=t_N=0}.$$

Положим $f(\boldsymbol{t}) = p_1^0e^{it_1}+\ldots+p_N^0e^{it_N}$. Согласно (2) $\psi_{\boldsymbol{\nu}}(\boldsymbol{t})=f^n(\boldsymbol{t})$. Вычислим частные производные характеристической функции в $\mathbf{0}$:

$$\frac{\partial\psi_{\boldsymbol{\nu}}(\mathbf{0})}{\partial t_j} = nf^{n-1}\,ip_j^0e^{it_j}\big|_{\boldsymbol{t}=\mathbf{0}} = inp_j^0,$$
$$\frac{\partial^2\psi_{\boldsymbol{\nu}}(\mathbf{0})}{\partial t_j^2} = inp_j^0\left[(n-1)f^{n-2}ip_j^0e^{it_j}+f^{n-1}ie^{it_j}\right]\big|_{\boldsymbol{t}=\mathbf{0}} =$$
$$= -np_j^0[(n-1)p_j^0+1] = -np_j^0-n(n-1)(p_j^0)^2,$$
$$\frac{\partial^2\psi_{\boldsymbol{\nu}}(\mathbf{0})}{\partial t_j\partial t_k} = n(n-1)f^{n-2}ip_k^0e^{it_k}\,ip_j^0e^{it_j}\big|_{\boldsymbol{t}=\mathbf{0}} = -n(n-1)p_j^0p_k^0.$$

Поскольку $|\nu_j|\leqslant n$, то $\mathbf{M}\,|\nu_j|^m<\infty$ для любого m. Применяя утверждение, находим интересующие нас моменты первого и второго порядка.

6. Раскладывая $n=n_1+\ldots+n_k$ наблюдений по $k\times N$ ($i=1,\ldots,k$; $j=1,\ldots,N$) ячейкам таблицы (см. рис. 6), видим, что *функцией правдоподобия сгруппированной выборки* служит

$$L(\boldsymbol{p}) = c\prod_{i,\,j}p_j^{\nu_{ij}} = c\prod_j p_j^{l_j},\quad \text{где } c=n!\Big/\prod_{i,j}\nu_{ij}!\ \text{ и } l_j=\sum_i\nu_{ij}.$$

Поэтому задача равносильна максимизации по переменным $p_1,\ldots,p_N$ функции

$$f(\boldsymbol{p})=\sum_j l_j\ln p_j \quad \text{при условии} \quad g(\boldsymbol{p})=1-\sum_j p_j=0.$$

Составим *функцию Лагранжа* $F(\boldsymbol{p},\lambda) = f(\boldsymbol{p}) + \lambda\, g(\boldsymbol{p})$ и запишем систему уравнений для поиска экстремальных точек:

$$\frac{\partial F(\boldsymbol{p},\lambda)}{\partial p_j} = \frac{l_j}{p_j} - \lambda = 0, \quad j = 1,\ldots,N; \quad \frac{\partial F(\boldsymbol{p},\lambda)}{\partial \lambda} = g(\boldsymbol{p}) = 0.$$

Из N первых уравнений находим, что $p_j = l_j/\lambda$. Подставляя их в последнее уравнение, получим $\lambda = l_1 + \ldots + l_N = n$, откуда $\widehat{p}_j = l_j/n$, что и требовалось установить.

ОТВЕТЫ НА ВОПРОСЫ

И сам я догадаюсь.
Чацкий в «Горе от ума»
А. С. Грибоедова

1. Поскольку $\nu_1 + \ldots + \nu_N = n$, то слагаемые в суммах из формулы (1) зависимы между собой.

2. Во-первых, для качественной проверки равномерности 2000 чисел разбиение отрезка $[0,1]$ всего на 4 промежутка явно недостаточно: $\log_2 2000 \approx 11$.

Во-вторых, статистика X_n^2, определяемая формулой (1), принимает значение $(4^2 + 5^2 + 8^2 + 1^2)/500 = 0{,}212$. Это значение *подозрительно мало*: вероятность того, что X_n^2 окажется таким или еще меньше, согласно табл. Т3 для χ_3^2 примерно равна 0,025.

Понятно, что при несовпадении истинных вероятностей p_j попадания в промежутки с гипотетическими p_j^0 статистика X_n^2 имеет тенденцию к росту (в [32, с. 113] доказано, что критерий хи-квадрат будет состоятельным (см. § 1 гл. 13) против таких альтернатив). Поэтому обычно обращают внимание только на большие значения статистики.

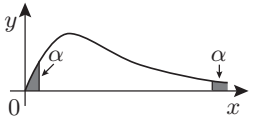


Рис. 7

Однако, иногда важно, чтобы датчик генерировал псевдослучайные числа, которые больше похожи на реализацию выбираемых наудачу из отрезка $[0,1]$ точек, чем последовательности, *равномерные по Вейлю* (см. § 5 гл. 3). У выбираемых наудачу точек есть определенный случайный разброс. Нетрудно предложить способы более регулярного заполнения отрезка $[0,1]$ (зависимыми) случайными точками, у которых частоты попадания в промежутки будут меньше отличаться от соответствующих вероятностей, чем при выборе наудачу. (Например, такой: первую точку берем наудачу, вторую выбираем наудачу из бóльшего по длине из двух отрезков, на которые разбила $[0,1]$ первая точка, и т. д.) Для таких датчиков значения статистики X_n^2 будут попадать в область «левого хвоста» закона χ^2 (рис. 7).

3. Функцией правдоподобия пуассоновской выборки служит $L(\lambda) = \lambda^{\sum x_i} e^{-\lambda n} / \prod x_i!$ (см. § 1 гл. 5). Уравнение правдоподобия $\dfrac{d}{d\lambda} \ln L(\lambda) = \dfrac{1}{\lambda} \sum x_i - n = 0$ имеет корень $\tilde{\lambda} = \overline{x}$.

Часть V
АНАЛИЗ МНОГОМЕРНЫХ ДАННЫХ

Эта часть книги содержит базовые сведения по статистическому анализу таблиц вида «объекты—признаки». В ней рассматриваются методы классификации многомерных данных (гл. 19), корреляционный анализ признаков (гл. 20) и модель линейной регрессии (гл. 21).

Глава 19

КЛАССИФИКАЦИЯ

Навостри ум свой математикой, если не найдешь для этого никакого иного средства, остерегайся только классификации букашек, поверхностное знание которой совершенно бесполезно, а точное уводит в бесконечность. И не забывай, что число фибр твоего мозга, их складок и извилин конечно. Там, где сидит какая-нибудь история бабочки, нашлось бы, может быть, место для биографий Плутарха, которые могли бы вдохновить тебя.

Г. К. Лихтенберг

При статистическом анализе таблицы данных, состоящей из нескольких столбцов (признаков), необходимо иметь в виду *эффект существенной многомерности*, из-за которого к верным выводам можно прийти лишь при одновременном учете всей совокупности взаимосвязанных признаков. Так, точки и кружки на рис. 1 почти не отличаются друг от друга по каждой из координат в отдельности, но очевидным образом разделяются по новому признаку — сумме координат.

Похожий случай приводится в [1, с. 15]: попытка различить два типа потребительского поведения семей сначала по одному признаку (расходы на питание), потом по другому (расходы на промышленные товары и услуги) не дала результата, в то время как одновременный учет обоих признаков позволил обнаружить значимое различие между анализируемыми совокупностями семей. (См. также пример 1 в гл. 23.)

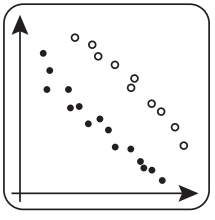


Рис. 1

Рассмотрим еще один пример, показывающий, что удачная классификация может даже привести к появлению нового направления исследований (см. [4, с. 35]).

«С давних пор астрономы знали о различной светимости звезд, т. е. о различной их «истинной яркости». В конце XIX в. были открыты также различные спектральные классы звезд, попросту говоря — различный цвет их излучения (от красного до голубого). До 1913 г. эти характеристики существовали в представлении ученых раздельно, но вот (независимо друг от друга) датский астроном Герцшпрунг и американец Ресселл сопоставили их между собой и построили двумерную проекцию объектов-звезд на плоскость признаков спектр — светимость. Результаты оказались неожиданными (рис. 2).

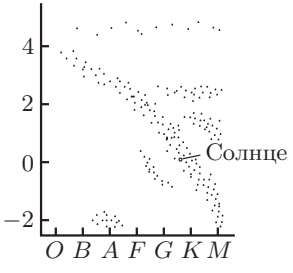


Рис. 2

Астрономы увидели, что звезды не распределены в пространстве этих признаков равномерно, а образуют несколько ярко выраженных кластеров, причем стало возможным предсказать эволюцию звезд по значениям их основных характеристик. С тех пор диаграмма Герцшпрунга—Ресселла стала одним из важных инструментов в работе современных астрономов.»

Если число признаков $m > 3$, то разбиение множества объектов на компактные группы (так называемые *кластеры*)*) может оказаться непростой задачей. Данная глава посвящена знакомству с некоторыми подходами к ее решению.

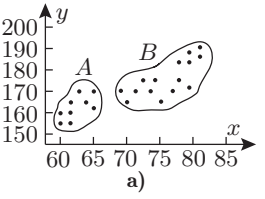
§ 1. НОРМИРОВКА, РАССТОЯНИЯ И КЛАССЫ

Статистическая однородность — понятие, базисное для статистики; общепринято, что какую-либо обработку статистических данных (усреднение, установление связей и т. д.) надо производить только в однородных группах наблюдений.

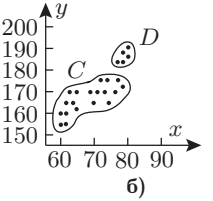
И. Д. Мандель, «Кластерный анализ»

Разбиение объектов на классы может в значительной степени *зависеть от выбора единиц измерения* (масштабов шкал) признаков: килограммы или фунты, сантиметры или дюймы.

Пример 1 ([52, с. 26]). Студенты группы записывают свои вес (x) и рост (y). По этим данным на плоскости строится *диаграмма рассеяния* («облако» точек с координатами (x_i,y_i), $i = 1,\ldots,n$, где n — число студентов в группе). Масштабы по осям задаются произвольно. На рис. 3, *а* девушки (A) довольно четко отделяются от юношей (B). На рис. 3, *б* шкала на оси веса сжата вдвое. При этом более естественным представляется уже деление на высоких юношей (D) и всех остальных студентов (C).



а)



б)

Рис. 3

В приведенном примере при классификации не следует считать расстоянием между объектами евклидово расстояние между соответствующими точками (x_i, y_i) на плоскости, так как признаки имеют разные единицы измерения. Требуется предварительная нормировка показателей, переводящая их в безразмерные величины. Перечислим наиболее распространенные типы нормировки одномерных наблюдений $Z_1,\ldots,Z_n$.

Типы нормировки

N1) $Z'_i = (Z_i - Z_{min})/(Z_{max} - Z_{min})$.

N2) $Z'_i = (Z_i - \overline{Z})/S$, где $\overline{Z} = \frac{1}{n}\sum Z_i$ — среднее арифметическое, $S^2 = \frac{1}{n}\sum(Z_i - \overline{Z})^2$ — выборочная дисперсия.

N3) $Z'_i = (Z_i - MED)/MAD$, где MED — выборочная медиана (см. § 2 гл. 7), MAD**) — *(нормированная) медиана абсолютных отклонений* от MED:

$$MAD = \frac{1}{\Phi^{-1}(3/4)}\ MED\{|Z_i - MED|,\ i = 1,\ldots,n\},$$

где $\Phi^{-1}(x)$ — функция, обратная к функции распределения закона $\mathcal{N}(0,1)$.***) Такое преобразование менее подвержено влиянию выделяющихся значений Z_i.

*) Разные варианты уточнения этого понятия приведены ниже.

**) Сокращение MAD происходит от английского наименования *Median of Absolute Deviations*.

***) Множитель $1/\Phi^{-1}(3/4) \approx 1{,}483$ обеспечивает для выборки из закона $\mathcal{N}(\mu,\sigma^2)$ сходимость $MAD \xrightarrow{d} \sigma$ при $n \to \infty$ (см. П5).

Помимо типа нормировки решающее влияние на результат классификации оказывает *выбор меры близости* между m-мерными точками. Приведем основные способы задания расстояния d_{ij} от точки $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \ldots, x_{im})$ до точки $\boldsymbol{x}_j = (x_{j1}, x_{j2}, \ldots, x_{jm})$.

Расстояния между объектами

D1) *Метрика города* (рис. 4)*): $d_{ij} = \sum\limits_{l=1}^{m} |x_{il} - x_{jl}|$. При использовании метрики города хорошо выделяются классы, имеющие вид «облака», вытянутого вдоль оси некоторого признака.

В случае, когда координаты объектов принимают только значения 0 и 1, это расстояние равно количеству несовпадающих координат, т. е. длине пути по ребрам единичного m-мерного куба из одной вершины в другую (*метрика Хемминга*).

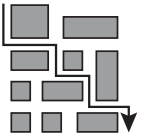


Рис. 4

D2) *Евклидова метрика*: $d_{ij} = \left(\sum\limits_{l=1}^{m} (x_{il} - x_{jl})^2 \right)^{1/2}$.

D3) *Метрика Чебышёва*: $d_{ij} = \max\limits_{1 \leqslant l \leqslant m} |x_{il} - x_{jl}|$.

Все три расстояния являются частными случаями (соответственно при $p = 1$, 2 и ∞) так называемого *расстояния Минковского*

$$d_{ij} = \left(\sum_{l=1}^{m} |x_{il} - x_{jl}|^p \right)^{1/p}.$$

Известно (см. [90, с. 208]), что при любом $p \geqslant 1$ для расстояния Минковского выполняется *неравенство треугольника*: $d_{ij} \leqslant d_{ik} + d_{kj}$. На рис. 5 изображены единичные «шары» B_p^m для $p = 1$, 2, ∞ при $m = 2$. Отношение объемов $\rho_m = V(B_1^m)/V(B_\infty^m) = 1/m!$ при увеличении размерности быстро уменьшается: $\rho_2 = 1/2$, $\rho_5 = 1/120$, $\rho_{10} = 1/3628800 \approx 3 \cdot 10^{-7}$.

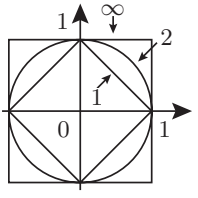


Рис. 5

Вопрос 1. Почему метрика Чебышёва отвечает значению $p = \infty$?

Иногда матрица расстояний (мер близости) d_{ij} между объектами задается непосредственно: например, как таблица экспертных оценок сходства объектов или как матрица прямых измерений близости (скажем, размеров межотраслевых поставок). В этом случае снимается проблема выбора типа нормировки и расстояния. (Однако, заметим, что для некоторых из рассматриваемых ниже методов классификации требуются сами координаты объектов, а не только расстояния d_{ij} между ними.)

На основе заданного расстояния между объектами можно уточнить, какие множества называются *группами однородных объектов* или *классами*. Выделим некоторые

Типы классов

C1) Класс типа ядра [60] (в [56, с. 235] такой класс называется *сгущением*). Все расстояния между объектами внутри

*) Ее также называют *метрикой city-block* или *манхеттенской*.

класса меньше любого из расстояний между объектами класса и остальной частью множества объектов. На рис. 6 сгущениями являются A и B. Остальные пары множеств не разделяются с помощью этого определения.

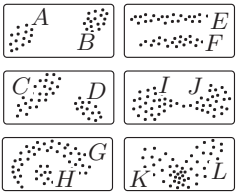


Рис. 6

C2) Кластер (*сгущение в среднем* [56]). Среднее расстояние внутри класса меньше среднего расстояния объектов класса до всех остальных. Множества C и D теперь разделяются, но у E (G) среднее внутреннее расстояние больше, чем среднее расстояние между E и F (G и H).

C3) Класс типа ленты [60] (*слабое сгущение* [56]). Существует $\tau > 0$ такое, что для любого $\boldsymbol{x}_i$ из класса S найдется такой объект $\boldsymbol{x}_j \in S$, что $d_{ij} \leqslant \tau$, а для всех $\boldsymbol{x}_k \notin S$ справедливо неравенство $d_{ik} > \tau$. В смысле этого определения на рис. 6 разделяются все пары множеств кроме I и J, K и L.

C4) Класс с центром. Существует порог $R > 0$ и некоторая точка $\boldsymbol{x}^*$ в пространстве, занимаемом объектами класса S (в частности, элемент этого множества) такие, что все объекты из S и только они содержатся в шаре радиуса R с центром в $\boldsymbol{x}^*$. Часто в качестве $\boldsymbol{x}^*$ выступает центр масс класса S, т. е. координаты центра определяются как средние значения признаков у объектов класса. Множества I и J являются классами с центром, а E, F и G — нет.

Обратим внимание на то, что накладывающиеся множества K и L не разделяются при помощи перечисленных определений классов. Тем не менее, в примере 4 из § 5 предлагается способ проведения разделяющей границы между подобными множествами на основе статистической модели случайного выбора из одной из k многомерных нормальных совокупностей.

Вообще, классификация (кластер-анализ) отличается от других разделов статистики большой *зависимостью* результатов расчетов от содержательных установок исследователя.

Научное исследование — это искусство, а правила в искусстве, если они слишком жестки, приносят больше вреда, чем пользы.
Дж. Томсон, «Дух науки»

§ 2. ЭВРИСТИЧЕСКИЕ МЕТОДЫ

Подавляющая часть классификаций на практике проводится именно эвристическими методами. Это объясняется относительной простотой и содержательной ясностью таких алгоритмов, возможностью вмешательства в их работу путем изменения одного или нескольких параметров, смысл которых обычно понятен, и невысокой трудоемкостью алгоритмов.

A1) Связные компоненты. Все объекты разбиваются на классы *типа ленты, или слабого сгущения* (тип C3 в § 1), где задаваемый параметр $\tau \in (\min d_{ij}, \max d_{ij})$. В этой постановке задача классификации эквивалентна нахождению связных компонент

графа (вершины графа i и j соединены ребром, если $d_{ij} \leqslant \tau$). (Алгоритм выделения связных компонент методом *поиска в глубину* излагается в § 9.) Для выбора величины τ полезно построить *гистограмму межобъектных расстояний* (высота прямоугольника над промежутком Δ_l на рис. 7 пропорциональна количеству d_{ij} в Δ_l). При хорошей структурированности данных гистограмма, как правило, имеет два выделяющихся максимума: при $d_{ij} \approx d_{int}$ (*типичное внутриклассовое расстояние*) и при $d_{ij} \approx d_{out}$ (*типичное межклассовое расстояние*). Часто удачным выбором τ оказывается значение $(d_{int} + d_{out})/2$.

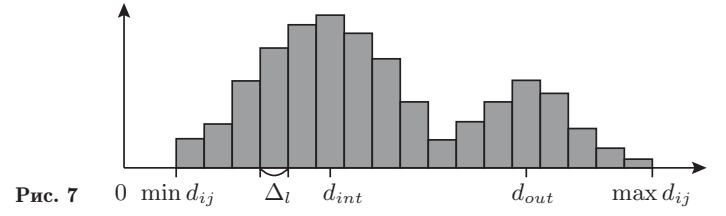


Рис. 7

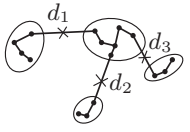


Рис. 8

А2) Кратчайший незамкнутый путь (КНП). Его также называют *минимальным покрывающим деревом* или *каркасом*. Соединяются ребром две ближайшие точки, затем среди оставшихся отыскивается точка, ближайшая к любой из уже соединенных точек, и присоединяется к ним и т. д. до исчерпания всех точек. Р. Прим в 1957 г. доказал, что построенный таким способом граф имеет минимальную общую длину ребер среди всевозможных соединений, связывающих все вершины (см. [28, с. 60]).

В найденном КНП затем отбрасывают $k-1$ самых длинных дуг и получают k классов (рис. 8).*) Метод позволяет выделять классы произвольной формы.

Заметим, что если разрешается добавлять в граф новые вершины (точки Штейнера), то можно добиться меньшей, чем у КНП, длины пути. Уже для трех точек A, B и C на плоскости таких, что наибольший из углов $\triangle ABC$ меньше 120°, существует точка O внутри треугольника, для которой минимальна сумма $|OA| + |OB| + |OC|$.

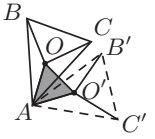


Рис. 9

Доказательство [64, с. 77]. Пусть O — некоторая точка внутри $\triangle ABC$. При повороте на 60° вокруг вершины A точки O, B и C перейдут в точки O', B' и C', соответственно (рис. 9). Так как $\triangle AOO'$ равносторонний, то $|AO| = |OO'|$. Отрезок OC переходит в отрезок $O'C'$. Поэтому $|OC| = |O'C'|$. Таким образом, сумма $|OA| + |OB| + |OC| = |OO'| + |OB| + |O'C'|$, т. е. равна длине ломаной $BOO'C'$. Она минимальна, когда ломаная является отрезком. Поскольку $\angle AOO' = 60°$, для этого необходимо, чтобы $\angle BOA = 120°$. Другими словами, сторона AB должна быть видна из оптимальной

*) Число k задает исследователь. На практике обычно $2 \leqslant k \leqslant 5$.

точки под углом 120°. Из симметрии то же самое должно быть верно и для двух других сторон $\triangle ABC$. ■

Вопрос 2. Как построить оптимальную точку O при помощи циркуля и линейки?

(Постройте на стороне BC, как на основании, равносторонний треугольник и опишите вокруг него окружность.)

Для произвольного расположения вершин графа на плоскости и любого числа точек Штейнера эффективные алгоритмы построения кратчайшего соединения неизвестны (согласно [28, с. 63]).

А3) Метод k-средних*) предназначен для выделения классов типа С4 («класс с центром»). Приведем **два варианта**:

(a) *Алгоритм Г. Болла и Д. Холла* (1965 г., см. [52, с. 110]). Случайно выбираются k объектов (*эталонов*); каждый объект присоединяется к ближайшему эталону (тем самым образуются k классов); в качестве новых эталонов принимаются центры масс классов.**) После пересчета объекты снова распределяются по ближайшим эталонам и т. д. Критерием окончания алгоритма служит стабилизация центров масс всех классов.

Вопрос 3. Где на плоскости находится точка, сумма расстояний от которой до вершин выпуклого четырехугольника минимальна?

Вместо случайно выбираемых эталонов лучше использовать k наиболее удаленных объектов: сначала отыскиваются два самых удаленных друг от друга объекта, затем l-й эталон ($l = 3, \ldots, k$) определяется как наиболее удаленный в среднем от уже имеющихся.

(b) *Алгоритм Дж. Мак-Кина* (1967 г., см. [52, с. 98]). Он отличается от метода Болла и Холла тем, что при просмотре списка объектов пересчет центра масс класса происходит после присоединения к нему каждого очередного объекта.

Отметим, что алгоритм Мак-Кина связан с функционалом качества разбиения $F1$ из § 5.

А4) Алгоритм «Форель». Случайный объект объявляется центром класса; все объекты, находящиеся от него на расстоянии не большем R, входят в первый класс. В нем определяется центр масс, который объявляется новым центром класса и т. д. до стабилизации центра. Затем все объекты, попавшие в первый класс изымаются, и процедура повторяется с новым случайным центром.

«*Форель*»: Первоначальное название ФОРЭЛ — ФОРмальный ЭЛемент. Предложен В. Н. Елкиной и Н. Г. Загоруйко в 1966 г., см. [1, с. 222].

Можно скомбинировать алгоритмы А4 и А2 ([52, с. 67)]: при небольшом R по алгоритму А4 находят $k' > k$ классов; их центры соединяют КНП, из которого удаляют $k - 1$ самых длинных ребер и получают k классов. При этом образуются классы более сложной формы, чем m-мерные шары (рис. 10). Здесь важна идея двухэтапности классификации: сначала выделить заведомо компактные маленькие группы, затем произвести их объединение. Так можно успешно классифицировать довольно большие массивы информации (сотни объектов).

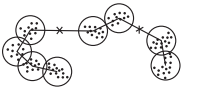


Рис. 10

А5) Метод потенциальных ям. Предположим, что каждый объект $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{im})$ создает вокруг себя поле притяжения

*) Это название, ставшее популярным, введено Дж. Мак-Кином.

**) Считается, что каждому объекту приписана масса 1.

с некоторой весовой функцией, например, гладким *квартическим ядром*

$$W_i(\boldsymbol{x}) = \bigl[1 - (r_i/R)^2\bigr]^2 I_{\{r_i \leqslant R\}},$$

где $r_i = |\boldsymbol{x} - \boldsymbol{x}_i|$, а параметр $R > 0$ задает эффективный размер области притяжения. Все вместе объекты создают потенциальное поле $U(\boldsymbol{x}) = -\sum W_i(\boldsymbol{x})$. Классам соответствуют потенциальные ямы: объект $\boldsymbol{x}_i$ относится к яме, в которую он «скатывается» при свободном движении. Практически приходится, стартовав с $\boldsymbol{x}_i$, запускать некоторый алгоритм (локальной) минимизации.

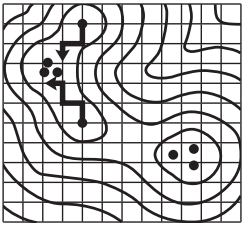


Рис. 11

Одним из простейших методов минимизации функции m переменных $U(\boldsymbol{x})$ является циклический перебор $2m$ точек, соседних с текущей по осям с шагом $\pm h$ (h — точность поиска точки минимума). Если значение $U(\boldsymbol{x})$ в какой-то из них меньше, чем в текущей, происходит перемещение в нее (противоположную по этой оси точку можно не проверять) (рис. 11). Для ускорения полезно запоминать для каждой оси знак успешного перемещения по ней во время предыдущего цикла и сначала пробовать сдвигаться в том же направлении. Критерием окончания служит отсутствие перемещений за время цикла.

Обозначим через $\boldsymbol{y}_i$ конечную точку пути из $\boldsymbol{x}_i$. Объекты $\boldsymbol{x}_i$ и $\boldsymbol{x}_j$ отнесем к одному классу, если $\sum\limits_{l=1}^{m} |y_{il} - y_{jl}| \leqslant 2mh$ (т. е. $\boldsymbol{y}_i$ и $\boldsymbol{y}_j$ близки в *метрике города* D1 из § 1). При этом число перемещений при поиске локального минимума (длина пути) характеризует удаленность $\boldsymbol{x}_i$ от «центра» класса (дна ямы).

Из-за необходимости проведения численной минимизации для каждого $\boldsymbol{x}_i$ метод рекомендуется применять для классификации небольшого числа (нескольких десятков) объектов.

§ 3. ИЕРАРХИЧЕСКИЕ ПРОЦЕДУРЫ

Мне завещал отец:
Во-первых, угождать всем людям без изъятья;
Хозяину, где доведется жить,
Начальнику, с кем буду я служить,
Слуге его, который чистит платья,
Швейцару, дворнику, для избежанья зла,
Собаке дворника, чтоб ласкова была.
Молчалин в «Горе от ума» А. С. Грибоедова

Общая схема этих процедур такова: сначала каждый объект считается отдельным классом; на первом шаге объединяются два ближайших объекта, которые образуют новый класс (если сразу несколько объектов (классов) одинаково близки, то выбирается одна случайная пара); вычисляются *меры отдаленности* ρ (см. ниже)*) от этого класса до всех остальных классов, и размерность матрицы межклассовых мер отдаленности сокращается на единицу; шаги процедуры повторяются до тех пор, пока все объекты не объединятся в один класс.

*) Мы не называем их расстояниями из-за того, что не для всех мер отдаленности выполняется неравенство треугольника.

Наиболее известны следующие две процедуры (рис. 12):

Р1) Метод «ближнего соседа»: $\rho_{min} = \min\limits_{\boldsymbol{x}_i \in S_k,\, \boldsymbol{x}_j \in S_l} d_{ij}$,

Р2) Метод «дальнего соседа»: $\rho_{max} = \max\limits_{\boldsymbol{x}_i \in S_k,\, \boldsymbol{x}_j \in S_l} d_{ij}$.

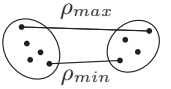
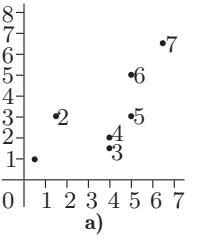


Рис. 12

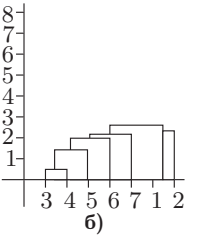
Решение о том, какое из разбиений на классы, получаемых при проведении иерархической процедуры, наиболее содержательно, принимается на основе анализа так называемой *дендрограммы*: по горизонтали откладываются номера объектов, а по вертикали — значения мер отдаленности $\rho(S_k,S_l)$, при которых происходили объединения классов S_k и S_l.

Dendron (*греч.*) — дерево.

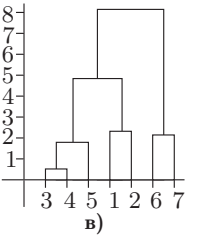
На основе данных, представленных на рис. 13, *а*, построены дендрограммы для процедур Р1 (рис. 13, *б*) и Р2 (рис. 13, *в*). Хорошо видна следующая особенность метода «ближнего соседа» — *цепочечный эффект*: независимо от формы кластера к нему присоединяются ближайшие к границе объекты. Метод «дальнего соседа» не приводит к подобному эффекту. Отметим, что разбиение на классы на основе Р1 эквивалентно разбиению, получаемому методами А1 и А2. Подробнее о сравнении разных методов классификации см. в § 7.



а)



б)



в)

Рис. 13

Рассмотрим некоторые другие иерархической процедуры.

Р3) Метод средней связи: $\rho_{ave} = \dfrac{1}{n_k n_l} \sum\limits_{\boldsymbol{x}_i \in S_k} \sum\limits_{\boldsymbol{x}_j \in S_l} d_{ij}$ (здесь n_k и n_l — количества объектов в классах S_k и S_l).

А. Н. Колмогоровым было предложено изящное обобщение метода Р3, основанное на понятии *степенного среднего* чисел $c_1 > 0, \ldots, c_n > 0$

$$\overline{c}_\tau = \left(\frac{1}{n} \sum_{i=1}^{n} c_i^\tau \right)^{1/\tau}. \tag{1}$$

Очевидно, $\overline{c}_1 = \dfrac{1}{n} \sum c_i$ — среднее арифметическое. Устремляя τ к $+\infty$, $-\infty$ и 0, получаем, соответственно, $\overline{c}_{+\infty} = \max c_i$, $\overline{c}_{-\infty} = \min c_i$, $\overline{c}_0 = (\prod c_i)^{1/n}$ — среднее геометрическое (проверьте!). Таким образом, *мера отдаленности Колмогорова*

$$\rho_K(\tau) = \left[\frac{1}{n_k n_l} \sum_{\boldsymbol{x}_i \in S_k} \sum_{\boldsymbol{x}_j \in S_l} d_{ij}^\tau \right]^{1/\tau} \tag{2}$$

включает в себя в качестве частных случаев ρ_{min}, ρ_{max} и ρ_{ave}.

Average (*англ.*) — средний.

Р4) Метод центров масс: $\rho_{center} = |\overline{\boldsymbol{x}}_k - \overline{\boldsymbol{x}}_l|^2$, где $\overline{\boldsymbol{x}}_k$ и $\overline{\boldsymbol{x}}_l$ обозначают центры масс k-го и l-го классов.

Недостатком этого метода является возможность появления *инверсий* — нарушений монотонности увеличения уровня при построении дендрограммы (т. е. объединение классов на некотором шаге процедуры осуществляется при более низком значении меры отдаленности, чем на более раннем шаге).

Вопрос 4.
Может ли возникнуть инверсия при классификации методом Р4 трех объектов?

P5) Метод Уорда*): $\rho_W = \dfrac{n_k n_l}{n_k + n_l} |\overline{\boldsymbol{x}}_k - \overline{\boldsymbol{x}}_l|^2$. Как вытекает из приведенной ниже теоремы 1, множитель $\dfrac{n_k n_l}{n_k + n_l}$ позволяет предотвратить появление инверсий.

Замечание 1. Покажем, что верно равенство

$$\rho_W = \sum_{\boldsymbol{x}_i \in S_k \cup S_l} |\boldsymbol{x}_i - \overline{\boldsymbol{x}}_{k,l}|^2 - \sum_{\boldsymbol{x}_i \in S_k} |\boldsymbol{x}_i - \overline{\boldsymbol{x}}_k|^2 - \sum_{\boldsymbol{x}_i \in S_l} |\boldsymbol{x}_i - \overline{\boldsymbol{x}}_l|^2, \qquad (3)$$

где $\overline{\boldsymbol{x}}_{k,l}$ — центр масс $S_k \cup S_l$. Таким образом, ρ_W представляет собой *прирост общей внутриклассовой инерции* V_{int} (см. формулу (6) в § 5) *при замене классов* S_k *и* S_l *на их объединение.*

Доказательство. В силу теоремы Гюйгенса (см. решение задачи 3 гл. 16) правая часть (3) равна $n_k|\overline{\boldsymbol{x}}_k - \overline{\boldsymbol{x}}_{k,l}|^2 + n_l|\overline{\boldsymbol{x}}_l - \overline{\boldsymbol{x}}_{k,l}|^2$. Сгруппируем массы S_k и S_l, поместив их в точки $\overline{\boldsymbol{x}}_k$ и $\overline{\boldsymbol{x}}_l$ соответственно. Общий центр масс $\overline{\boldsymbol{x}}_{k,l}$ при группировке масс не меняется (убедитесь!). Для завершения доказательства остается воспользоваться теоремой о межточечных расстояниях из решения задачи 5 гл. 16. ■

Для вычисления (на очередном шаге иерархической процедуры) мер отдаленности между вновь образованным классом и всеми другими оставшимися классами удобно воспользоваться общей для методов P1—P5 **формулой Г. Ланса и У. Уильямса**:

$$\rho(S_0, S_1 \cup S_2) = C_1\rho_{01} + C_2\rho_{02} + C_3\rho_{12} + C_4|\rho_{01} - \rho_{02}|, \qquad (4)$$

где $\rho_{01} = \rho(S_0, S_1)$ и т. п. Коэффициенты C_1—C_4 для процедур P1—P5 приведены в следующей таблице:

Номер	Название	C_1	C_2	C_3	C_4
P1	Ближний сосед	$1/2$	$1/2$	0	$-1/2$
P2	Дальний сосед	$1/2$	$1/2$	0	$1/2$
P3	Средняя связь	$\dfrac{n_1}{n_1+n_2}$	$\dfrac{n_2}{n_1+n_2}$	0	0
P4	Центры масс	$\dfrac{n_1}{n_1+n_2}$	$\dfrac{n_2}{n_1+n_2}$	$-\dfrac{n_1 n_2}{(n_1+n_2)^2}$	0
P5	Метод Уорда	$\dfrac{n_0+n_1}{n_0+n_1+n_2}$	$\dfrac{n_0+n_2}{n_0+n_1+n_2}$	$-\dfrac{n_0}{n_0+n_1+n_2}$	0

Для методов P1 и P2 формула (4) вытекает из тождеств

$$\min\{\alpha,\beta\} = \frac{1}{2}(\alpha+\beta) - \frac{1}{2}|\alpha-\beta|, \quad \max\{\alpha,\beta\} = \frac{1}{2}(\alpha+\beta) + \frac{1}{2}|\alpha-\beta|.$$

Для процедур P3—P5 формула (4) выводится в задаче 1.

*) Предложен Дж. Уордом (J. Ward) в 1963 г.

Замечание 2. При использовании меры отдаленности Колмогорова, определяемой равенством (2), величина $\rho_K^\tau(\tau)$, очевидно, удовлетворяет условию (4) с теми же коэффициентами, что и $\rho_{ave} = \rho_K(1)$.

Замечание 3. Формула Ланса—Уильямса (4) была обобщена на более широкий класс иерархических процедур М. Жамбю:

$$\rho(S_0, S_1 \cup S_2) = C_1\rho_{01} + C_2\rho_{02} + C_3\rho_{12} + C_4|\rho_{01} - \rho_{02}| + \\ + C_5 h(S_0) + C_6 h(S_1) + C_7 h(S_2), \tag{5}$$

где $h(S_k)$ — значение меры отдаленности, при котором произошло образование класса S_k (т. е. его уровень на дендрограмме). В частности, при

$$C_1 = (m_0 + m_1)/M, \quad C_2 = (m_0 + m_2)/M, \quad C_3 = (m_1 + m_2)/M,$$
$$C_4 = 0, \quad C_5 = -m_0/M, \quad C_6 = -m_1/M, \quad C_7 = -m_2/M,$$

где $M = m_0 + m_1 + m_2$, получим меру отдаленности $\rho_J(S_k,S_l)$, равную *моменту инерции объединения* $S_k \cup S_l$ *относительно его центра масс* (см. [30, с. 117]).

Следующие условия (Г. Миллиган, 1979 г.) обеспечивают отсутствие инверсий у процедур, удовлетворяющих формуле (5).

Теорема 1. Пусть

а) коэффициенты C_1, C_2, C_5, C_6, C_7 неотрицательны,
б) $C_4 \geqslant -\min\{C_1,C_2\}$,
в) $C_1 + C_2 + C_3 \geqslant 1$.

Тогда иерархическая процедура, задаваемая формулой (5), не имеет инверсий.

Контрпример. Для *метода центров масс* Р4 нарушается условие в):

$$C_1 + C_2 + C_3 = 1 - n_1n_2/(n_1 + n_2)^2 < 1.$$

§ 4. БЫСТРЫЕ АЛГОРИТМЫ

В 1978 г. М. Брюиношем было введено важное понятие *редуктивности (сводимости)* мер отдаленности. Оно заключается в том, что для произвольного $\delta > 0$ условия $\rho(S_1, S_2) \leqslant \delta$ и $\rho(S_0, S_1 \cup S_2) < \delta$ влекут выполнение хотя бы одного из неравенств $\rho(S_0, S_1) < \delta$ или $\rho(S_0, S_2) < \delta$ (см. [52, с. 55]). Иными словами, δ-окрестность множества, полученного объединением двух δ-близких классов (пунктир на рис. 14), включена в объединение δ-окрестностей этих двух классов (сплошная линия на рис. 14). Можно дать другую формулировку свойства редуктивности: если $\rho(S_1, S_2) \leqslant \delta$, а $\rho(S_0, S_1) \geqslant \delta$, $\rho(S_0, S_2) \geqslant \delta$, то $\rho(S_0, S_1 \cup S_2) \geqslant \delta$.

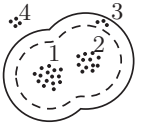
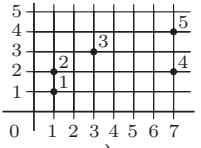


Рис. 14

Вопрос 5. Обладают ли свойством редуктивности а) ρ_{min}, б) ρ_W?

Для редуктивных мер отдаленности можно построить *быстрые алгоритмы* (см. [1, с. 263]), позволяющие классифицировать тысячи объектов. Разберем их **принципиальную схему**.

При k-м шаге иерархической процедуры приходится искать минимальный элемент в матрице мер отдаленностей классов, т. е. минимум из $N_k = (n-k+1)(n-k)/2$ чисел. Ясно, что минимальный элемент находится среди элементов, не превосходящих некоторый *порог* δ. При соответствующем выборе δ массив таких элементов имеет размер намного меньший, чем N_k. В этом и состоит основная причина большой скорости работы алгоритмов и экономии компьютерной памяти.



а)

	1	2	3	4	5
1	0	①	4	7	9
2		0	③	6	8
3			0	5	5
4				0	②
5					0

б)

	3	4	5	6
3	0	5	5	③
4		0	②	6
5			0	8
6				0

в)

	3	6	7
3	0	③	5
6		0	6
7			0

г)

	7	8
7	0	5
8		0

д)

Рис. 15

Обозначим через U_δ множество пар (i,j) номеров классов таких, что $\rho_{ij} \leqslant \delta$, а через V_δ — массив из номеров классов, присутствующих в U_δ.

На первом этапе попарно объединяются классы из V_δ до тех пор, пока текущее U_δ не станет пустым. Затем берется новое значение порога $\delta' > \delta$ и формируются множества $U_{\delta'}$ и $V_{\delta'}$ и т. д. до полного слияния всех объектов в один класс.

Рассмотрим работу алгоритма на примере данных рис. 15, *а*. В качестве расстояния между объектами используем метрику города D1 из § 1, а мерой отдаленности классов будем считать ρ_{min}. Зададим $\delta = 3$. Тогда $U_\delta = \{(1,2),(2,3),(4,5)\}$ и $V_\delta = \{1,2,3,4,5\}$ (рис. 15, *б*). После объединения наиболее близких объектов с номерами 1 и 2 в новый класс под номером 6 матрица межклассовых мер отдаленности преобразуется к виду, приведенному на рис. 15, *в*. При этом изменятся U_δ и V_δ: $U_\delta = \{(3,6),(4,5)\}$, $V_\delta = \{3,4,5,6\}$ и т. д. На шаге, соответствующем рис. 15, *д*, уже нет пар классов, для которых $\rho_{min} \leqslant 3$. Поэтому приходится увеличить порог до $\delta' = 5$, что и позволяет завершить процедуру.

При выполнении свойства редуктивности рассмотренный алгоритм строит точно такую же дендрограмму, что и обычная иерархическая процедура (для присоединения класса 3 к классам 1,2 на рис. 14 достаточно просмотреть окрестности 1 и 2, а не проверять связь 1,2–4).

Эффективность алгоритма существенно зависит от выбора последовательности порогов $\delta < \delta' < \delta'' < \ldots$. Если δ слишком велико, то U_δ включает почти все пары классов, и мы практически возвращаемся к традиционной иерархической процедуре. Если же величина порога слишком мала, то придется часто формировать множества U_δ и V_δ. Обычно используют следующие **два способа выбора порогов**.

Первый способ *(на основе гистограммы ρ_{ij})*. Порог δ находится из условия достаточности выделенной компьютерной памяти для хранения всех $\rho_{ij} \leqslant \delta$.

Пусть, например, число объектов равно 100. Тогда на первом шаге имеется $N_1 = 100 \cdot 99/2 = 4950$ пар классов. При этом задано, что их количество в памяти не должно превышать 200. Предположим, что первые четыре интервала гистограммы содержат 180 пар, а пятый —

еще 40 пар. В таком случае можно поместить в память только 180 пар, причем порог δ будет равен наибольшему значению ρ_{ij} этих пар.

Второй способ *(изменяющееся среднее)*. Вычисляется среднее арифметическое всех ρ_{ij}. Пусть это будет δ_1. Если возможно размещение в памяти всех $\rho_{ij} \leqslant \delta_1$, то полагается $\delta = \delta_1$. Иначе удаляются пары, у которых $\rho_{ij} > \delta_1$, а для оставшихся снова подсчитывается среднее арифметическое δ_2 и т. д. до тех пор, пока все $\rho_{ij} \leqslant \delta_k$ не уместятся в заданной компьютерной памяти.

Замечание 4. Е. Диде и В. Моро (1984 г.) показали, что если слегка ужесточить условия теоремы 1, заменив неравенство в) на неравенство в′) $C_1 + C_2 + \min\{0,C_3\} \geqslant 1$, то меры отдаленности, удовлетворяющие (5), будут редуктивными.

§ 5. ФУНКЦИОНАЛЫ КАЧЕСТВА РАЗБИЕНИЯ

Материал этого параграфа довольно сложен.

Будем считать, что *число классов k задано* (случай, когда оно неизвестно, обсуждается в § 6). Пусть n_l обозначает количество объектов (m-мерных точек) $\boldsymbol{x}_i$ ($i = 1,\ldots,n$) в классе S_l ($l = 1,\ldots,k$).

При разбиении на классы типа «класс с центром» (С4 в § 1) в основе многих характеристик качества лежит формула

Формула (6) немедленно вытекает из ее одномерного аналога (13) в гл. 1б.

$$\sum_{i=1}^{n} |\boldsymbol{x}_i - \overline{\boldsymbol{x}}|^2 = \sum_{l=1}^{k} \sum_{\boldsymbol{x}_i \in S_l} |\boldsymbol{x}_i - \overline{\boldsymbol{x}}_l|^2 + \sum_{l=1}^{k} n_l |\overline{\boldsymbol{x}}_l - \overline{\boldsymbol{x}}|^2. \qquad (6)$$

Здесь слева от равенства стоит V_{tot} — *момент инерции относительно общего центра масс $\overline{\boldsymbol{x}}$ всех объектов*; первое слагаемое справа, обозначаемое через V_{int}, представляет собой *сумму моментов инерции I_l классов S_l относительно их центров масс $\overline{\boldsymbol{x}}_l$*; второе слагаемое V_{out} характеризует *межклассовый разброс* (это момент инерции системы, у которой точки каждого класса стянуты в его центр масс).

Перечислим некоторые **функционалы качества** F, определенные на множестве всевозможных разбиений S объектов на k классов (чем $F(S)$ меньше, тем разбиение S лучше).

$$F1 = \sum_{l=1}^{k} I_l \equiv V_{int} \qquad (\textit{общая внутриклассовая инерция}).$$

Основной результат Дж. Мак-Кина (1967 г.) состоит в доказательстве того, что его вариант алгоритма k-средних (А3 в § 2) приводит к локальному минимуму функционала $F1$. Точную минимизацию $F1$ с использованием динамического программирования осуществил Р. Дженсен в 1969 г. (см. [26]).

Замечание 5. Как следует из замечания 1, метод Уорда (Р5 из § 3) обеспечивает наименьшее увеличение функционала $F1$ при каждом

шаге процедуры. Однако, эта пошаговая оптимальность алгоритма, вообще говоря, не влечет его оптимальности в том же смысле для любого наперед заданного числа классов.

$$F2 = \sum_{l=1}^{k} \frac{1}{n_l}\, I_l \qquad (\textit{сумма внутриклассовых дисперсий}).$$

$$F3 = \sum_{l=1}^{k} \sum_{\boldsymbol{x}_i, \boldsymbol{x}_j \in S_l} |\boldsymbol{x}_i - \boldsymbol{x}_j|^2 = 2 \sum_{l=1}^{k} n_l I_l.$$

Последнее равенство здесь справедливо в силу теоремы о межточечных расстояниях (см. решение задачи 5 гл. 16).

$$F4 = \left[\frac{1}{n-k}\, V_{int}\right] \Big/ \left[\frac{1}{k-1}\, V_{out}\right]$$

(сравните с формулой (9) гл. 16).

Ряд функционалов качества разбиения связан с обобщающим формулу (6) матричным тождеством (задача 2)

$$\boldsymbol{W}_{tot} = \boldsymbol{W}_{int} + \boldsymbol{W}_{out}. \tag{7}$$

Здесь $\boldsymbol{W}_{tot} = \sum\limits_{i=1}^{n} (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^T$ — *общая матрица рассеяния*; $\boldsymbol{W}_{int}$ — *матрица внутриклассового разброса*: $\boldsymbol{W}_{int} = \sum\limits_{l=1}^{k} \boldsymbol{W}_l$, где $\boldsymbol{W}_l = \sum\limits_{\boldsymbol{x}_i \in S_l} (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_l)(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_l)^T$ — *матрица рассеяния l-го класса*; $\boldsymbol{W}_{out} = \sum\limits_{l=1}^{k} n_l (\overline{\boldsymbol{x}}_l - \overline{\boldsymbol{x}})(\overline{\boldsymbol{x}}_l - \overline{\boldsymbol{x}})^T$ — *матрица разброса между классами*. Формула (6) получается из тождества (7) применением к участвующим в ней матрицам операции взятия следа (см. П10). В частности,

$$F1 = V_{int} = \operatorname{tr} \boldsymbol{W}_{int}.$$
$$F5 = \det \boldsymbol{W}_{int}.$$
$$F6 = \operatorname{tr}\left(\boldsymbol{W}_{int} \boldsymbol{W}_{tot}^{-1}\right).$$
$$F7 = \det\left(\boldsymbol{W}_{int} \boldsymbol{W}_{tot}^{-1}\right) = \det \boldsymbol{W}_{int} / \det \boldsymbol{W}_{tot}.$$

Функционалы $F6$—$F7$ изучались Х. Фридманом и Дж. Рубином в 1967 г. Они *инвариантны* по отношению к невырожденным линейным преобразованиям координат объектов. Действительно, если $\widetilde{\boldsymbol{x}}_i = \boldsymbol{B}\boldsymbol{x}_i$, где $\boldsymbol{B}$ — невырожденная $(m \times m)$-матрица, то $\widetilde{\boldsymbol{W}}_{int} = \boldsymbol{B}\boldsymbol{W}_{int}\boldsymbol{B}^T$ и $\widetilde{\boldsymbol{W}}_{tot} = \boldsymbol{B}\boldsymbol{W}_{tot}\boldsymbol{B}^T$ (см. свойство 5 дисперсии и ковариации из П2). При этом матрица

$$\widetilde{\boldsymbol{W}}_{int} \widetilde{\boldsymbol{W}}_{tot}^{-1} = \boldsymbol{B}\boldsymbol{W}_{int}\boldsymbol{B}^T \left(\boldsymbol{B}^T\right)^{-1} \boldsymbol{W}_{tot}^{-1} \boldsymbol{B}^{-1} = \boldsymbol{B}\boldsymbol{W}_{int}\boldsymbol{W}_{tot}^{-1}\boldsymbol{B}^{-1}$$

подобна матрице $\boldsymbol{W}_{int}\boldsymbol{W}_{tot}^{-1}$, поэтому согласно следствию 2 из П10 у них одинаковые следы и определители.

В целях последующего применения к задаче классификации нормальных совокупностей (пример 3) рассмотрим

Пример 2. ***Оценки максимального правдоподобия для параметров многомерной нормальной модели*** **[32, с. 66].** Предположим, что случайные m-мерные векторы $\boldsymbol{\xi}_i \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ (см. П9) и независимы ($i = 1, \ldots, n$). Пусть вектор средних значений $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_m)$ и *невырожденная* ковариационная $(m \times m)$-матрица $\boldsymbol{\Sigma}$ неизвестны, $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Общее количество неизвестных параметров с учетом симметричности $\boldsymbol{\Sigma}$ равно $m + m(m+1)/2$. Согласно П9 плотность случайного вектора $\boldsymbol{\xi}_i$ задается формулой

$$p(\boldsymbol{x}, \boldsymbol{\theta}) = (2\pi)^{-m/2} (\det \boldsymbol{\Sigma})^{-1/2} \exp\left\{ -\frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \right\}. \quad (8)$$

В силу независимости случайных векторов $\boldsymbol{\xi}_i$ логарифм функции правдоподобия (§ 4 гл. 9) имеет вид

$$\ln L(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ln p(\boldsymbol{x}_i, \boldsymbol{\theta}) =$$
$$= -\frac{1}{2} \left[n \ln \det \boldsymbol{\Sigma} + \sum_{i=1}^{n} (\boldsymbol{x}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}) \right] + \text{const}. \quad (9)$$

Максимизация $L(\boldsymbol{\theta})$ эквивалентна минимизации выражения в квадратных скобках в формуле (9). Положим $\overline{\boldsymbol{x}} = (\boldsymbol{x}_1 + \ldots + \boldsymbol{x}_n)/n$. Разобьем сумму в равенстве (9) на два слагаемых на основе тождества, обобщающего теорему Гюйгенса (см. решение задачи 3 гл. 16), которое предлагается проверить в задаче 3:

$$\sum_{i=1}^{n} (\boldsymbol{x}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}) =$$
$$= n (\overline{\boldsymbol{x}} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\overline{\boldsymbol{x}} - \boldsymbol{\mu}) + \sum_{i=1}^{n} (\boldsymbol{x}_i - \overline{\boldsymbol{x}})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_i - \overline{\boldsymbol{x}}). \quad (10)$$

Первое слагаемое неотрицательно для любой $\boldsymbol{\Sigma}$ в силу положительной определенности матрицы $\boldsymbol{\Sigma}^{-1}$ и равно 0 только при $\boldsymbol{\mu} = \overline{\boldsymbol{x}}$.

Вопрос 6. Как доказать, что $\boldsymbol{\Sigma}^{-1}$ положительно определена?

Второе слагаемое от $\boldsymbol{\mu}$ не зависит. Следовательно, достаточно минимизировать по $\boldsymbol{\Sigma}$ выражение

$$n \ln \det \boldsymbol{\Sigma} + \sum_{i=1}^{n} (\boldsymbol{x}_i - \overline{\boldsymbol{x}})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_i - \overline{\boldsymbol{x}}). \quad (11)$$

Введем *выборочную ковариационную матрицу*

$$\widehat{\boldsymbol{\Sigma}} = \widehat{\boldsymbol{\Sigma}}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \frac{1}{n} \sum_{i=1}^{n} (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^T. \quad (12)$$

На основе легко проверяемого тождества $\boldsymbol{y}^T \boldsymbol{B} \boldsymbol{y} = \operatorname{tr}(\boldsymbol{y}\boldsymbol{y}^T \boldsymbol{B})$ и линейности оператора tr, получаем

$$\sum_{i=1}^{n} (\boldsymbol{x}_i - \overline{\boldsymbol{x}})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_i - \overline{\boldsymbol{x}}) = n \operatorname{tr}(\widehat{\boldsymbol{\Sigma}} \boldsymbol{\Sigma}^{-1}). \quad (13)$$

Учитывая формулу (13), выражение (11) можно переписать в виде

$$n\left[\ln\det\boldsymbol{\Sigma}+\operatorname{tr}\left(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}\right)\right]. \tag{14}$$

Минимизация суммы (14) по $\boldsymbol{\Sigma}$ эквивалентна минимизации функции

$$\psi(\boldsymbol{\Sigma})=\operatorname{tr}\left(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}\right)-m-\ln\,\det\left(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}\right)$$

(постоянные m и $\ln\det\widehat{\boldsymbol{\Sigma}}$ добавлены для дальнейших преобразований).

Так как $\widehat{\boldsymbol{\Sigma}}$ и $\boldsymbol{\Sigma}$ — ковариационные матрицы, то (согласно замечанию из П10) они неотрицательно определены, а $\boldsymbol{\Sigma}$ даже положительно определена в силу невырожденности. Обозначим через $\lambda_1\geqslant\lambda_2\geqslant\ldots\geqslant\lambda_m\geqslant 0$ собственные значения пары матриц $(\widehat{\boldsymbol{\Sigma}},\boldsymbol{\Sigma})$, т. е. решения характеристического уравнения $\det\left(\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1}-\lambda\boldsymbol{E}\right)=0$ (см. П10). Тогда

$$\psi(\boldsymbol{\Sigma})=\sum_{l=1}^{m}(\lambda_l-1-\ln\lambda_l).$$

Поскольку $\lambda-1-\ln\lambda\geqslant 0$ при $\lambda\geqslant 0$, из последнего соотношения получаем, что $\psi(\boldsymbol{\Sigma})\geqslant 0$, причем равенство имеет место только при $\lambda_1=\ldots=\lambda_m=1$, т. е. для $\boldsymbol{\Sigma}=\widehat{\boldsymbol{\Sigma}}$.

Вопрос 7. Почему условие $\lambda_1=\ldots=\lambda_m=1$ влечет равенство $\boldsymbol{\Sigma}=\widehat{\boldsymbol{\Sigma}}$?

Таким образом, оценками максимального правдоподобия параметров $\boldsymbol{\mu}$ и $\boldsymbol{\Sigma}$ являются, соответственно, статистики $\overline{\boldsymbol{x}}$ и $\widehat{\boldsymbol{\Sigma}}$.

Пример 3. ***Многомерные нормальные наблюдения и функционалы качества разбиения на классы*** **[1, с. 169].** Предположим, что каждое наблюдение $\boldsymbol{\xi}_i$ $(i=1,\ldots,n)$ извлечено из одного из k нормальных законов $\mathcal{N}(\boldsymbol{\mu}_l,\boldsymbol{\Sigma}_l)$ $(l=1,\ldots,k)$. При этом $\boldsymbol{\mu}_l$ и $\boldsymbol{\Sigma}_l$, вообще говоря, неизвестны. Задача исследователя — определить, какие n_1 из n наблюдений извлечены из класса $\mathcal{N}(\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_1)$, какие n_2 из n наблюдений извлечены из класса $\mathcal{N}(\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_2)$ и т. д. Числа n_l в общем случае также неизвестны.

Введем векторный *параметр разбиения* $\boldsymbol{\gamma}=(\gamma_1,\ldots,\gamma_n)$, где γ_i — номер класса (от 1 до k), к которому относится наблюдение $\boldsymbol{\xi}_i$. Тогда задачу разбиения на классы можно сформулировать как задачу оценивания неизвестных параметров $\gamma_1,\ldots,\gamma_n$ при «мешающих» неизвестных параметрах $\boldsymbol{\mu}_l$ и $\boldsymbol{\Sigma}_l$ $(l=1,\ldots,k)$. Обозначим весь набор неизвестных параметров через $\boldsymbol{\theta}$, т. е. $\boldsymbol{\theta}=(\boldsymbol{\gamma},\boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_1,\ldots,\boldsymbol{\Sigma}_k)$. Тогда логарифм функции правдоподобия (см. формулу (9)) будет равен (с точностью до сдвига на константу) величине

$$-\frac{1}{2}\sum_{l=1}^{k}\left[n_l(\boldsymbol{\gamma})\ln\det\boldsymbol{\Sigma}_l+\sum_{\boldsymbol{x}_i\in S_l(\boldsymbol{\gamma})}(\boldsymbol{x}_i-\boldsymbol{\mu}_l)^T\boldsymbol{\Sigma}_l^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}_l)\right]. \tag{15}$$

Максимизируем выражение (15) по $\boldsymbol{\theta}$. Пусть $\widehat{\boldsymbol{\theta}}$ обозначает точку максимума.*)

Рассуждения, аналогичные проведенным в примере 2, показывают, что оценками максимального правдоподобия для параметров $\boldsymbol{\mu}_l$ при фиксированных $\boldsymbol{\gamma}, \boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_k$ являются *центры масс классов*

$$\overline{\boldsymbol{x}}_l = \overline{\boldsymbol{x}}_l(\boldsymbol{\gamma}) = \frac{1}{n_l} \sum_{\boldsymbol{x}_i \in S_l(\boldsymbol{\gamma})} \boldsymbol{x}_i, \quad (l = 1, \ldots, k).$$

Подставляя их в формулу (15) и учитывая тождество (10), получаем, что достаточно минимизировать по $\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_k$ и $\boldsymbol{\gamma}$ функцию

$$\sum_{l=1}^{k} \left[n_l \ln \det \boldsymbol{\Sigma}_l + \sum_{\boldsymbol{x}_i \in S_l} (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_l)^T \boldsymbol{\Sigma}_l^{-1} (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_l) \right], \tag{16}$$

которая, принимая во внимание равенства (12) и (13), совпадает с

$$\sum_{l=1}^{k} \left[n_l \ln \det \boldsymbol{\Sigma}_l + n_l \operatorname{tr} (\widehat{\boldsymbol{\Sigma}}_l \boldsymbol{\Sigma}_l^{-1}) \right], \tag{17}$$

где $\widehat{\boldsymbol{\Sigma}}_l = \dfrac{1}{n_l} \sum\limits_{\boldsymbol{x}_i \in S_l} (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_l)(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_l)^T$ — *выборочная ковариационная матрица класса* S_l. Отметим, что $\widehat{\boldsymbol{\Sigma}}_l = \dfrac{1}{n_l} \boldsymbol{W}_l$, где $\boldsymbol{W}_l$ — это матрица рассеяния класса S_l (см. пояснение к формуле (7)).

Анализ выражений (16) и (17) в некоторых частных случаях приводит к **ряду интересных выводов**.

Случай 1. $\boldsymbol{\Sigma}_l = \sigma^2 \boldsymbol{E}$ $(l = 1, \ldots, k)$, причем параметр σ известен. Тогда минимизация (16) по $\boldsymbol{\gamma}$ равносильна минимизации функционала качества разбиения на классы $F1 = V_{int}$.

Случай 2. $\boldsymbol{\Sigma}_l = \boldsymbol{\Sigma}$ $(l = 1, \ldots, k)$, где матрица $\boldsymbol{\Sigma}$ известна. В этом случае минимизируемым функционалом является

$$F8 = \sum_{l=1}^{k} \sum_{\boldsymbol{x}_i \in S_l} d(\boldsymbol{x}_i, \overline{\boldsymbol{x}}_l),$$

где под расстоянием между точками $\boldsymbol{x}$ и $\boldsymbol{y}$ подразумевается

$$d(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x} - \boldsymbol{y})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{y}) \quad (\textit{расстояние Махаланобиса}). \tag{18}$$

Случай 3. $\boldsymbol{\Sigma}_l = \boldsymbol{\Sigma}$ $(l = 1, \ldots, k)$, где матрица $\boldsymbol{\Sigma}$ неизвестна. Тогда выражение (17) можно представить в виде

$$n \left[\ln \det \boldsymbol{\Sigma} + \operatorname{tr} (\widehat{\boldsymbol{\Sigma}}_{int} \boldsymbol{\Sigma}^{-1}) \right], \tag{19}$$

где $\widehat{\boldsymbol{\Sigma}}_{int} = \dfrac{1}{n} \sum\limits_{l=1}^{k} n_l \widehat{\boldsymbol{\Sigma}}_l = \dfrac{1}{n} \sum\limits_{l=1}^{k} \boldsymbol{W}_l = \dfrac{1}{n} \boldsymbol{W}_{int}$. (20)

*) Оговоримся сразу, что состоятельность оценки $\widehat{\boldsymbol{\theta}}$ остается под сомнением, поскольку размерность вектора $\boldsymbol{\theta}$ превосходит общее число наблюдений n.

Точно так же, как и в примере 2, находим, что матрица $\widehat{\boldsymbol{\Sigma}}_{int}$ минимизирует сумму (19). Подставив ее в формулу (19), замечаем, что второе слагаемое превращается в константу, а минимизация первого по $\boldsymbol{\gamma}$ равносильна минимизации функционала $F5 = \det \boldsymbol{W}_{int}$.

Замечание 6. Если матрицы ковариаций $\boldsymbol{\Sigma}_l$ $(l = 1, \ldots, k)$ совпадают, то можно обобщить одномерный F-критерий однофакторного дисперсионного анализа (см. формулу (9) гл. 16) на многомерную нормальную модель, например, так: для проверки *гипотезы о равенстве средних* $H_0\colon \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \ldots = \boldsymbol{\mu}_k$ в качестве статистик критерия будем использовать монотонные функции от $F7 = \det \boldsymbol{W}_{int} / \det \boldsymbol{W}_{tot}$. Положим для краткости $L \equiv F7$. Известно (см. [8, с. 49]), что при справедливости гипотезы H_0

а) $\dfrac{n-k-1}{k-1}\,\dfrac{1-\sqrt{L}}{\sqrt{L}} \sim F_{2k-2,2(n-k-1)}$, если $m = 2$ и $k \geqslant 2$;

б) $\dfrac{n-k-m+1}{m}\,\dfrac{1-L}{L} \sim F_{m,n-k-m+1}$, если $m \geqslant 1$ и $k = 2$;

в) $\dfrac{n-k-m+1}{m}\,\dfrac{1-\sqrt{L}}{\sqrt{L}} \sim F_{2m,2(n-k-m+1)}$, если $m \geqslant 1$ и $k = 3$.

Для других значений m и k и больших n можно указать хорошую аппроксимацию, предложенную М. Бартлеттом (1947 г.): статистика

$$-\left(n - 1 - \frac{m+k}{2}\right) \ln L$$

имеет приближенно χ^2-распределение с $m(k-1)$ степенями свободы.

Обобщение на многомерный случай F-критерия двухфакторного дисперсионного анализа (см. пример 1 гл. 17) приведено в [8, с. 160].

Случай 4. Матрицы ковариаций $\boldsymbol{\Sigma}_l$ не равны между собой и неизвестны. Минимизируя отдельно каждое слагаемое в сумме (17), находим, что оценками максимального правдоподобия для $\boldsymbol{\Sigma}_l$ служат выборочные ковариационные матрицы $\widehat{\boldsymbol{\Sigma}}_l$. Подставляя их в формулу (17), превращаем вторые слагаемые в квадратных скобках в константы. При этом минимизируемый по $\boldsymbol{\gamma}$ функционал приобретает вид

$$F9 = \sum_{l=1}^{k} n_l \ln \det \boldsymbol{\Sigma}_l.$$

Заметим, что функционал $F9$ равносилен (взвешенному) среднему геометрическому определителей матриц $\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_k$, в то время как функционал $F5$ эквивалентен определителю матрицы, равной (взвешенному) среднему арифметическому этих же матриц (см. формулу (20)).

	1	2	3	4	5	6
1	129	64	95	17,5	11,2	13,8
2	154	74	76	20,0	14,2	16,5
3	170	87	71	17,9	12,3	15,9
4	188	94	73	19,5	13,3	14,8
5	161	81	55	17,1	12,1	13,0
6	164	90	58	17,5	12,7	14,7
7	203	109	65	20,7	14,0	16,8
8	178	97	57	17,3	12,8	14,3
9	212	114	65	20,5	14,3	15,5
10	221	123	62	21,2	15,2	17,0
11	183	97	52	19,3	12,9	13,5
12	212	112	65	19,7	14,2	16,0
13	220	117	70	19,8	14,3	15,6
14	216	113	72	20,5	14,4	17,7
15	216	112	75	19,6	14,0	16,4

	1	2	3	4	5	6
16	205	110	68	20,8	14,1	16,4
17	228	122	78	22,5	14,2	17,8
18	218	112	65	20,3	13,9	17,0
19	190	93	78	19,7	13,2	14,0
20	212	111	73	20,5	13,7	16,6
21	201	105	70	19,8	14,3	15,9
22	196	106	67	18,5	12,6	14,2
23	158	71	71	16,7	12,5	13,3
24	255	126	86	21,4	15,0	18,0
25	234	113	83	21,3	14,8	17,0
26	205	105	70	19,0	12,4	14,9
27	186	97	62	19,0	13,2	14,2
28	241	119	87	21,0	14,7	18,3
29	220	111	88	22,5	15,4	18,0
30	242	120	85	19,9	15,3	17,6

	1	2	3	4	5	6
31	199	105	73	23,4	15,0	19,1
32	227	117	77	25,0	15,3	18,6
33	228	122	82	24,7	15,0	18,5
34	232	123	83	25,3	16,8	15,5
35	231	121	78	23,5	16,5	19,6
36	215	118	74	25,7	15,7	19,0
37	184	100	69	23,3	15,8	19,7
38	175	94	73	22,2	14,8	17,0
39	239	124	77	25,0	16,8	27,0
40	203	109	70	23,3	15,0	18,7
41	226	118	72	26,0	16,0	19,4
42	226	119	77	26,5	16,8	19,3
43	210	103	72	20,5	14,0	16,7

Рис. 16

Пример 4. ***Разделение многомерных нормальных законов (дискриминантный анализ).*** В таблице на рис. 16 указаны размеры челюстей и зубов тридцати собак (номера 1–30), двенадцати волков (номера 31–42) и ископаемого черепа неизвестного животного (номер 43), найденного в четвертичном слое (по данным Де Бониса, приведенным в [30, с. 25]). Рис. 17 демонстрирует измеряемые характеристики: 1 — длина черепа, 2 — длина верхней челюсти, 3 — ширина верхней челюсти; следующие измерения относятся к зубам: 4 — длина верхнего карнивора, 5 — длина первого верхнего моляра, 6 — ширина первого верхнего моляра. Требуется решить, к какому из классов (собак или волков) следует отнести неизвестное животное.

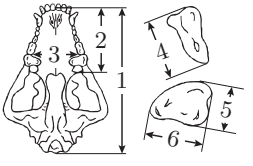


Рис. 17

Для того, чтобы это сделать, воспользуемся статистической моделью случайного выбора единственного наблюдения из некоторого многомерного нормального закона, причем заранее известно, что этот закон является одним из k заданных законов $\mathcal{N}(\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)$, где $l = 1, \ldots, k$.

В соответствии с **принципом максимального правдоподобия** будем считать *областью притяжения* закона $\mathcal{N}(\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)$ $(l = 1, \ldots, k)$ множество таких точек $\boldsymbol{x} \in \mathbb{R}^m$, где плотность распределения $\mathcal{N}(\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)$ больше других. Это равносильно тому, что величина

$$h_l(\boldsymbol{x}) = \ln \det \boldsymbol{\Sigma}_l + (\boldsymbol{x} - \boldsymbol{\mu}_l)^T \boldsymbol{\Sigma}_l^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_l)$$

имеет *наименьшее значение* среди $h_1, \ldots, h_k$ (см. формулу (8)). Рис. 18 иллюстрирует получаемое при $m = 1$ и $k = 2$ разбиение $\mathbb{R}$ на две области притяжения.

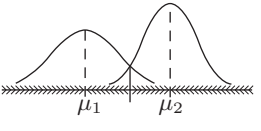


Рис. 18

Замечание 7. В случае $k = 2$ данный способ выбора между законами с произвольными (известными) плотностями p_1 и p_2 минимизирует сумму $\alpha + \beta$ вероятностей ошибок I и II рода при проверке гипотезы $H_1\colon p(\boldsymbol{x}) = p_1(\boldsymbol{x})$ (см. § 1 гл. 13) против альтернативы $H_2\colon p(\boldsymbol{x}) = p_2(\boldsymbol{x})$ по единственному наблюдению $\boldsymbol{x}_0 \in \mathbb{R}^m$.

Доказательство. Обозначим через G критическое множество критерия для проверки гипотезы H_1 против альтернативы H_2. Тогда

$$\alpha + \beta = \int\limits_G p_1(\boldsymbol{x})\,d\boldsymbol{x} + \int\limits_{\mathbb{R}^m \setminus G} p_2(\boldsymbol{x})\,d\boldsymbol{x} = 1 + \int\limits_G (p_1(\boldsymbol{x}) - p_2(\boldsymbol{x}))\,d\boldsymbol{x}.$$

Интеграл в правой части принимает наименьшее значение в случае, когда множество G состоит из всех точек, где подынтегральная функция неположительна, т. е. $p_2(\boldsymbol{x}) \geqslant p_1(\boldsymbol{x})$. ■

Отметим, что оптимальное множество G принадлежит (при $c = 1$) семейству критических множеств из критерия Неймана—Пирсона (см. § 2 гл. 13)

$$G_c = \left\{ \boldsymbol{x} \in \mathbb{R}^m\colon \frac{p_2(\boldsymbol{x})}{p_1(\boldsymbol{x})} \geqslant c \right\}.$$

Для $\alpha + \beta$ можно получить оценку сверху (см. [90, с. 392]):

$$\alpha + \beta \leqslant e^{-B}, \quad \text{где } B = -\ln \int \sqrt{p_1(\boldsymbol{x})\,p_2(\boldsymbol{x})}\,d\boldsymbol{x}.$$

Здесь $B = B(p_1,p_2)$ — *расстояние Бхаттачария* (см. [1, с. 67]). Оно инвариантно относительно взаимно однозначных преобразований координат и обращается в нуль при $p_1 = p_2$. Известно, что для двух нормальных законов

$$B = \frac{1}{2}\,\ln \frac{\det\left(\dfrac{\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2}{2}\right)}{\sqrt{(\det \boldsymbol{\Sigma}_1)(\det \boldsymbol{\Sigma}_2)}} + \frac{1}{8}\,(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^T \left(\frac{\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2}{2}\right)^{-1} (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1).$$

В частности, при $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ видим, что $B = d(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)/8$, где $d(\boldsymbol{x},\boldsymbol{y})$ обозначает расстояние Махаланобиса (см. формулу (18)).

Применим эти результаты к данным о собаках ($l = 1$) и волках ($l = 2$). Чтобы уравнять влияние размеров черепа и зубов (первые на порядок больше последних), нормируем столбцы таблицы, приведенной на рис. 16, с помощью устойчивого к выделяющимся наблюдениям преобразования

$Z' = (Z - MED)/MAD$ (нормировка N3 из § 1).

Так как истинные $\boldsymbol{\mu}_l$ и $\boldsymbol{\Sigma}_l$ неизвестны, используем вместо них оценки максимального правдоподобия $\overline{\boldsymbol{x}}_l$ и $\widehat{\boldsymbol{\Sigma}}_l$. [Все промежуточные вычисления (нормировку, подсчет средних и ковариаций,

умножение матриц, нахождение обратных матриц и определителей) удобно проводить с использованием встроенных функций программы Excel.]

В результате подсчетов получим значения $h_1 = -2{,}94$ и $h_2 = 39{,}7$. Поскольку $h_1 < h_2$, в соответствии с принципом максимального правдоподобия неизвестное ископаемое животное, скорее всего, было собакой. При этом значение $B = 4{,}08$, что приводит к оценке суммы вероятностей ошибок I и II рода величиной $e^{-B} \approx 0{,}017$.

Этот вывод подтверждает и *диаграмма рассеяния* (рис. 19) нормированных признаков 1 (длина черепа) и 4 (длина верхнего карнивора). Первый откладывался по оси X, второй — по оси Y. Волки обозначены треугольниками, собаки — точками, неизвестное животное — крестиком. На основе диаграммы можно заключить, что главным признаком, с помощью которого различаются собаки и волки, является *размер зубов*.

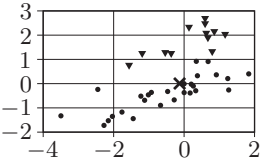


Рис. 19

Бабушка, а почему у тебя такие большие зубы?

Красная Шапочка

§ 6. НЕИЗВЕСТНОЕ ЧИСЛО КЛАССОВ

В случае, когда число классов k заранее не задано, функционал качества разбиения выбирают, чаще всего, в виде некоторой алгебраической комбинации (взвешенной суммы, произведения) двух функционалов $F(S)$ и $G(S)$. Первый, как правило, характеризует внутриклассовый разброс и является убывающей функцией от k. Второй сдерживает тенденцию к излишней детализации и возрастает при увеличении k.

Основываясь на общей концепции степенных средних, разработанной А. Н. Колмогоровым (см. [52, с. 90]), можно взять $F = \overline{y}_\tau$ и $G = 1/\overline{z}_\tau$ (см. формулу (1)), где

$$y_i = \left[\frac{1}{\nu(\boldsymbol{x}_i)} \sum_{\boldsymbol{x}_j \in S(\boldsymbol{x}_i)} d_{ij}^{\tau} \right]^{1/\tau} \quad \text{и} \quad z_i = \frac{\nu(\boldsymbol{x}_i)}{n} \qquad (i = 1, \ldots, n).$$

Здесь $S(\boldsymbol{x}_i)$ обозначает класс, включающий объект $\boldsymbol{x}_i$, $\nu(\boldsymbol{x}_i)$ — число объектов в $S(\boldsymbol{x}_i)$, n — общее количество объектов. Выбор параметра τ находится в распоряжении исследователя.

Величина $\overline{z}_\tau$ называется *мерой концентрации* точек, соответствующей разбиению S. Пусть n_l — это число объектов в классе S_l, $p_l = \dfrac{n_l}{n}$ $(l = 1, \ldots, k)$. Тогда нетрудно проверить, что $\overline{z}_{-1} = 1/k$, $\overline{z}_{+\infty} = \max p_l$, $\overline{z}_{-\infty} = \min p_l$, $-\log_2 \overline{z}_0 = -\sum p_l \log_2 p_l$ — энтропия разбиения (см. § 5 гл. 12),

$$\overline{z}_1 = \frac{1}{n} \sum_{i=1}^{n} z_i = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\nu(\boldsymbol{x}_i)}{n} \right) = \frac{1}{n^2} \sum_{l=1}^{k} n_l^2 = \sum_{l=1}^{k} p_l^2. \qquad (21)$$

Заметим, что при любом τ мера концентрации имеет минимум, равный $1/n$, при разбиении на n одноточечных классов, и максимум, равный 1, при объединении всех объектов в один общий класс.

Пример 5. ***Задача аппроксимации матрицы связей*** **[1, с. 173], [52, с. 103], [56].** Предположим, что данные представляют собой *матрицу связей* $\boldsymbol{A} = \|a_{ij}\|_{n\times n}$ между объектами (чем больше величина a_{ij}, тем связь сильнее). В качестве a_{ij} могут выступать, скажем, объемы межотраслевых поставок или коэффициенты корреляции признаков.*)

Элементами *булевых матриц* служат нули и единицы. Они названы так в честь Дж. Буля (1815–1864) — английского логика.

Требуется найти такое разбиение S с булевой матрицей $\boldsymbol{B} = \|b_{ij}\|_{n\times n}$, которое бы в наибольшей степени соответствовало матрице $\boldsymbol{A}$. Элементы b_{ij} определяются так: $b_{ij} = 1$, если объекты $\boldsymbol{x}_i$ и $\boldsymbol{x}_j$ принадлежат одному классу, $b_{ij} = 0$ — если разным.

Как сопоставить матрицы $\boldsymbol{A}$ и $\boldsymbol{B}$ друг с другом? Ведь a_{ij} — действительные числа, а b_{ij} — булевы переменные. Б. Г. Миркин (см. [56]) предложил взвешивать b_{ij}, вводя некоторые коэффициенты сдвига μ и масштаба $\sigma > 0$, и использовать в качестве меры точности аппроксимации величину

$$\Delta(S,\mu,\sigma) = \sum_{i,j=1}^{n} (a_{ij} - \mu - \sigma b_{ij})^2. \tag{22}$$

Минимизацию $\Delta(S,\mu,\sigma)$ надо производить по S, μ и σ. Для упрощения задачи зафиксируем μ и σ. Тогда поиск минимума функции (22) по S равносилен минимизации функционала

$$H(S) = -\sum_{i,j=1}^{n} (a_{ij} - \rho) b_{ij}, \quad \text{где } \rho = \mu + \sigma/2. \tag{23}$$

Действительно, поскольку $b_{ij}^2 = b_{ij}$, можно записать отдельное слагаемое из суммы в формуле (22) в виде

$$(a_{ij} - \mu)^2 - 2\sigma(a_{ij} - \mu)b_{ij} + \sigma^2 b_{ij}^2 = \text{const} - 2\sigma(a_{ij} - \rho)b_{ij}, \tag{24}$$

что и доказывает равносильность.

Величина ρ играет роль *порога существенности* для величины связей. В самом деле, чтобы минимизировать правую часть равенства (24) по b_{ij} надо, очевидно, положить $b_{ij} = 1$, если $a_{ij} > \rho$, и $b_{ij} = 0$, если $a_{ij} \leqslant \rho$. Однако, сами b_{ij} порождаются разбиением S, поэтому мы не имеем права назначать им произвольные булевы значения. Тем не менее, для уменьшения $H(S)$ при $a_{ij} > \rho$ выгодно поместить $\boldsymbol{x}_i$ и $\boldsymbol{x}_j$ в один класс.

Переходя к внутриклассовым связям (для которых $b_{ij} = 1$), имеем

$$H(S) = -\sum_{l=1}^{k} \sum_{\boldsymbol{x}_i,\boldsymbol{x}_j\in S_l} (a_{ij} - \rho) = -\sum_{l=1}^{k} \sum_{\boldsymbol{x}_i,\boldsymbol{x}_j\in S_l} a_{ij} + \rho \sum_{l=1}^{k} n_l^2,$$

где n_l — число объектов в классе S_l. Это представление показывает, что $H(S)$ обеспечивает компромисс между общей суммой

*) Если заданы расстояния d_{ij} между объектами, то можно формально преобразовать их в связи, например, по формуле $a_{ij} = 1/(1 + d_{ij})$.

внутриклассовых связей и (взвешенной) мерой концентрации $\overline{z}_1$ (см. формулу (21)).

Разбиение, минимизирующее $H(S)$, обладает тем свойством, что получаемые классы S_l оказываются *кластерами* (сгущениями в среднем) в смысле определения С2 из § 1: средняя связь $a(S_l) = \sum\limits_{\boldsymbol{x}_i,\boldsymbol{x}_j\in S_l} a_{ij}/n_l^2$ не меньше, чем средняя связь вовне $a(S_l,\overline{S}_l)$ и средняя связь между любыми классами $a(S_l,S_m)$. Точнее, для любых $l\neq m$ имеем $a(S_l,\overline{S}_l)$, $a(S_l,S_m)\leqslant\rho\leqslant a(S_l)$ (см. [1, с. 174]).

Это оптимальное разбиение может быть найдено с помощью иерархической процедуры, на очередном шаге которой объединяются те классы S_l и S_m, для которых сумма связей $\sum\limits_{\boldsymbol{x}_i\in S_l,\boldsymbol{x}_j\in S_m}(a_{ij}-\rho)$ максимальна и положительна. Процесс объединения заканчивается, когда такие суммы для всех $l\neq m$ становятся неположительными.

Смысл параметров μ и σ проясняет минимизация по ним $\Delta(S,\mu,\sigma)$ при фиксированном разбиении S. Вычислив частные производные по μ и σ функции

$$f(\mu,\sigma)=\sum_{(i,j):\,b_{ij}=1}(a_{ij}-\mu-\sigma)^2+\sum_{(i,j):\,b_{ij}=0}(a_{ij}-\mu)^2,$$

получим систему уравнений относительно оптимальных $\tilde{\mu}$ и $\tilde{\sigma}$:

$$\sum_{(i,j):\,b_{ij}=1}(a_{ij}-\tilde{\mu}-\tilde{\sigma})+\sum_{(i,j):\,b_{ij}=0}(a_{ij}-\tilde{\mu})=0,\qquad\sum_{(i,j):\,b_{ij}=1}(a_{ij}-\tilde{\mu}-\tilde{\sigma})=0.$$

Подставляя второе уравнение в первое, находим, что $\tilde{\mu}$ равно *средней межкластерной связи*. Из второго уравнения вытекает, что $\tilde{\mu}+\tilde{\sigma}$ — это *средняя внутрикластерная связь*. При этом порог $\tilde{\rho}$ — их полусумма, $\tilde{\sigma}$ — характеристика контрастности связей (ср. с рис. 7).

Недостатком рассмотренного метода является наличие единого порога ρ для всех классов, что не соответствует ситуации, когда присутствуют классы существенно различных размеров (о путях его преодоления рассказывается в [1, с. 175]).

§ 7. СРАВНЕНИЕ МЕТОДОВ

Определяющее значение при выборе метода классификации имеет общее число объектов n. Если оно велико (сотни или тысячи), то необходимо применять эвристический алгоритм А4 («Форель»), скомбинированный с А2 (кратчайший незамкнутый путь) (§ 2), или быстрые иерархические процедуры для редуктивных мер отдаленности (§ 4).

Для $n\leqslant 200$ в [52, с. 108–117] приводятся результаты экспериментального сравнения нескольких методов классификации

(в частности, А3 из § 2 и Р1—Р5 из § 3) путем их применения к моделированным совокупностям объектов.

Объекты генерировались как точки в m-мерном единичном гиперкубе ($3 \leqslant m \leqslant 7$). Задавалось число классов k ($3 \leqslant k \leqslant 7$) и наудачу в гиперкубе выбирались центры классов $\boldsymbol{y}_l$ ($l = 1, \ldots, k$). Для каждого класса с помощью датчика случайных чисел моделировались m длин сторон гиперпараллелепипеда с центром в $\boldsymbol{y}_l$, в который затем случайно бросались n_l точек ($30 \leqslant n_l \leqslant 50$). При фиксированных m и k каждое разбиение генерировалось 50 раз и показатели усреднялись (всего было обработано около 2000 выборок).

Кроме того, генерировались «шумящие» объекты, равномерно распределенные в гиперкубе. Их количество составляло заданный процент p от n ($10\% \leqslant p \leqslant 30\%$). Эти объекты предназначались только для того, чтобы «сбивать с толку» алгоритмы классификации, но, естественно, не участвовали в расчете показателей качества классификации (одним из таких показателей было расстояние Хемминга (см. метрику D1 из § 1) между моделированным разбиением и разбиением, которое построил алгоритм).

Изложим кратко **основные выводы**. Наилучшей (почти идеальной) по восстанавливаемости разбиения проявила себя иерархическая процедура Р5 («метод Уорда»). Следом за ней идут процедура Р2 («дальнего соседа») и алгоритм А3 (метод k-средних Болла и Д. Холла) (случайный выбор эталонов в алгоритме А3 показал себя как крайне неудачный). Самой плохой оказалась процедура Р1 («ближнего соседа»).

По уровню устойчивости к шуму лидерами стали алгоритмы А3 и Р5. Замыкает список снова Р1.

Не следует воспринимать эти результаты как приговор методу «ближнего соседа» Р1, имеющему цепочечный эффект. Речь может идти лишь о том, что на первичном этапе классификации, не обладая информацией о структуре классов, лучше применять менее чувствительные методы.

В случае, когда классы имеют сложную форму, скажем, относятся к типу С3 из § 1 («класс типа ленты или слабое сгущение»), именно алгоритм Р1 позволит правильно произвести разбиение.

Всякий необходимо причиняет пользу, употребленный на своем месте. Напротив того: упражнения лучшего танцмейстера в химии неуместны; советы опытного астронома в танцах глупы.

Козьма Прутков

Г. Миллиган и М. Купер в 1985 г. опубликовали результаты экспериментального сравнения в похожих условиях тридцати методов классификации (см. [52, с. 157]), в число которых вошли алгоритмы минимизации функционалов (см. § 5) $F4 = \left[\frac{1}{n-k}\ \mathrm{tr}\,\boldsymbol{W}_{int}\right] \Big/ \left[\frac{1}{k-1}\ \mathrm{tr}\,\boldsymbol{W}_{out}\right]$ (Калинский и Харабаш, 1974) и инвариантного функционала $F7 = \det \boldsymbol{W}_{int} / \det \boldsymbol{W}_{tot}$

(Фридман и Рубин, 1967). Первый оказался самым лучшим, а второй — в группе самых плохих. Дело, скорее всего, в удачной нормировке матриц рассеяния у $F4$ (сравните с формулой (9) гл. 16).

Вред или польза действия обусловливаются совокупностью обстоятельств.
Козьма Прутков

§ 8. ПРЕДСТАВЛЕНИЕ РЕЗУЛЬТАТОВ

После проведения классификации важно в удобной форме представить ее результаты. Приведем список важнейших (согласно [52, с. 159]) **характеристик классификации**.

1. Распределение номеров объектов по номерам классов.
2. Гистограмма межобъектных расстояний (подобная изображенной на рис. 7).
3. Средние внутриклассовые расстояния.
4. Матрица средних межклассовых расстояний.
5. Визуальное представление данных на плоскости двух (в пространстве трех) «наиболее информативных» признаков.
6. Дендрограмма для иерархических процедур.
7. Средние значения и размахи во всех классах для каждого признака.

Последний пункт наиболее принципиален, так как в подавляющем большинстве случаев интерпретация классов происходит по средним значениям признаков в них. Сопоставление же средних значений для заданного признака наиболее просто осуществляется, если классы не имеют наложения проекций. *Степень разделенности* классов по каждой оси можно охарактеризовать с помощью коэффициента

$$\gamma = 1 - \sum L_j / \sum R_l,$$

где R_l — размах по заданному признаку l-го класса, а $L_1, L_2, \ldots$ — длины наложений проекций классов на ось признака (рис. 20). Если $\gamma = 1$, то классы полностью разделимы. Чем ближе γ к 0, тем больше наложение проекций классов друг на друга.

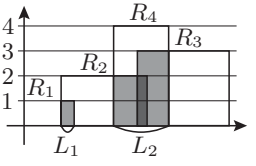


Рис. 20

§ 9. ПОИСК В ГЛУБИНУ

Основой для выделения связных компонент графа в методе A1 из § 2 служит алгоритм обхода всех вершин неориентированного связного графа G, именуемый *поиском в глубину* (сокращенно ПГ). Изложим его, следуя [28, с. 323].*)

В процессе поиска в глубину вершинам графа G присваиваются номера (*ПГ-номера*), а ребра помечаются. В начале ребра не помечены, вершины не имеют ПГ-номеров. Начинаем с произвольной

*) В [28] приведены также подробные описания (с примерами) многих других полезных алгоритмов поиска на графах, скажем, *поиска в ширину* (с. 37) или *метода Дейкстры* нахождения кратчайшего пути между двумя заданными вершинами графа (с. 342).

вершины v. Присваиваем ей ПГ-номер $\text{ПГ}(v) = 1$ и выбираем произвольное ребро vw. Ребро vw помечается, как «прямое», а вершина w получает (из v) ПГ-номер $\text{ПГ}(w) = 2$. После этого переходим в вершину w. Пусть в результате выполнения нескольких шагов этого процесса пришли в вершину x, при этом k — последний присвоенный ПГ-номер. Далее действуем в зависимости от ситуации следующим образом.

1. Имеется непомеченное ребро xy. Если у вершины y уже есть ПГ-номер, то ребро xy помечаем как «обратное» и продолжаем поиск непомеченного ребра, выходящего из вершины x. Если же вершина y не имеет ПГ-номера, то полагаем $\text{ПГ}(y) = k + 1$, ребро xy помечаем как «прямое» и переходим в вершину y. Вершина y считается получивший свой ПГ-номер из вершины x.

2. Все ребра, выходящие из x, помечены. Тогда возвращаемся в вершину, из которой x получила свой ПГ-номер.

Процесс закончится, когда все ребра будут помечены и произойдет возвращение в вершину v.

$N_1 = (3, 2, 5)$
$N_2 = (4, 1, 5)$
$N_3 = (1, 5)$
$N_4 = (2, 5)$
$N_5 = (2, 1, 3, 4, 6)$
$N_6 = (5)$

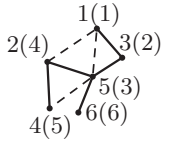


Рис. 21

Для того, чтобы в пункте 2 узнать, из какой вершины получила свой ПГ-номер вершина x, надо запоминать список Q, в который каждая вершина включается в момент получения ПГ-номера и исключается, как только произойдет возвращение из этой вершины. Включение и исключение вершин производятся всегда с конца списка.*)

Пусть граф G задан *списками смежности*, т. е. для каждой вершины x приведен N_x — список всех вершин, соединенных с ней ребрами.

На рис. 21 изображен граф и списки смежности, которыми он задан, а также представлены результаты применения к этому графу поиска в глубину из вершины $v = 1$. Каждой вершине приписан ее номер и ПГ-номер (в скобках). «Прямые» ребра проведены сплошными линиями, а «обратные» — пунктирными.

Отметим, что по построению «прямые» ребра образуют незамкнутый путь (дерево), соединяющий все вершины (связного) графа G. Однако, этот путь, вообще говоря, не является кратчайшим (см. А2 из § 2): на рис. 21 выгоднее соединить вершины 4 и 6, а не 5 и 6.

Можно показать, что трудоемкость и объем памяти для поиска в глубину составляют $O(n + m)$, где n — число вершин графа, m — число ребер (см. [28, с. 325]).

Поиск в глубину используется в качестве составной части во многих алгоритмах. В частности, с его помощью почти без дополнительных вычислительных затрат решается задача *выделения связных компонент графа*. Для этого достаточно один раз просмотреть список вершин графа, выполняя следующие действия. Если просматриваемая вершина i $(i = 1, \ldots, n)$ имеет ПГ-номер, то

*) Программисты называют такой список *стеком*.

перейти к следующей. Иначе положить $v = i$, $\text{ПГ}(v) = k+1$, где k — последний присвоенный ПГ-номер, и применить поиск в глубину. После его окончания (т. е. выделения компоненты, содержащей i) продолжить просмотр списка вершин с $i + 1$. Различать вершины разных компонент можно по их ПГ-номерам, если для каждой компоненты запомнить последний ПГ-номер.

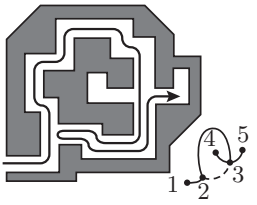


Рис. 22

Замечание 8. После удаления $k - 1$ самых длинных ребер КНП (алгоритм А2 из § 2) автоматическое разбиение на классы можно произвести путем выделения связных компонент оставшегося графа.

Нить, которую дала Ариадна Тезею для блуждания по лабиринту острова Крит, предназначалась не только для того, чтобы найти обратную дорогу, но и позволяла избежать «петель» — бесконечного хождения по замкнутой траектории. Для этого нужно просто поворачивать назад, если нить пересекает дорогу (рис. 22). Петля становится тупиком. На языке теории графов происходит возвращение по «обратным» ребрам, которое разрывает циклы и превращает граф лабиринта в дерево.

ЗАДАЧИ

1.* Выведите формулу (4) Ланса и Уильямса для мер отдаленности иерархических процедур Р3—Р5 из § 3.

2.* Проверьте, что $\boldsymbol{W}_{tot} = \boldsymbol{W}_{int} + \boldsymbol{W}_{out}$ (см. формулу (7)).

3. Получите тождество (10).

Указание. Введите вектор $\boldsymbol{y} = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{x} - \boldsymbol{\mu})$.

4.* Пусть $p_1(\boldsymbol{x})$ и $p_2(\boldsymbol{x})$ — плотности законов $\mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ и $\mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ (см. замечание 7), причем $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$ (*модель Фишера*).

а) Запишите явно уравнение *разделительной гиперплоскости* $h_1 = h_2$.

б) Найдите вероятности ошибок I и II рода α и β при классификации на ее основе.

Ну, между ими я, конечно, зауряд, немножко поотстал, ленив, подумать ужас!

Репетилов в «Горе от ума» А. С. Грибоедова

РЕШЕНИЯ ЗАДАЧ

1. Для процедуры средней связи Р3 с учетом $S_1 \cap S_2 = \varnothing$ имеем

$$[n_0(n_1+n_2)]\,\rho_{ave}(S_0, S_1 \cup S_2) = \sum_{\boldsymbol{x}_i \in S_0} \sum_{\boldsymbol{x}_j \in S_1 \cup S_2} d_{ij} =$$

$$= \sum_{\boldsymbol{x}_i \in S_0} \sum_{\boldsymbol{x}_j \in S_1} d_{ij} + \sum_{\boldsymbol{x}_i \in S_0} \sum_{\boldsymbol{x}_j \in S_2} d_{ij} =$$

$$= n_0 n_1\, \rho_{ave}(S_0, S_1) + n_0 n_2\, \rho_{ave}(S_0, S_2).$$

Разделив обе части на $n_0(n_1 + n_2)$, получим требуемое.

В случае метода центров масс Р4 доказательство немного сложнее. Рассмотрим систему из двух масс n_1 и n_2, которые

находятся в центрах масс $\overline{\boldsymbol{x}}_1$ и $\overline{\boldsymbol{x}}_2$ классов S_1 и S_2 соответственно. По теореме о межточечных расстояниях (см. решение задачи 5 гл. 16) момент инерции этой системы относительно общего центра масс $\overline{\boldsymbol{x}}_{1,2}$ равен

$$I_{\overline{\boldsymbol{x}}_{1,2}} = n_1|\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_{1,2}|^2 + n_2|\overline{\boldsymbol{x}}_2 - \overline{\boldsymbol{x}}_{1,2}|^2 = \frac{n_1 n_2}{n_1+n_2}\,|\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2|^2. \quad (25)$$

В силу теоремы Гюйгенса (см. решение задачи 3 гл. 16) момент инерции относительно центра масс $\overline{\boldsymbol{x}}_0$ класса S_0 имеет вид

$$I_{\overline{\boldsymbol{x}}_0} = n_1|\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_0|^2 + n_2|\overline{\boldsymbol{x}}_2 - \overline{\boldsymbol{x}}_0|^2 = I_{\overline{\boldsymbol{x}}_{1,2}} + (n_1+n_2)|\overline{\boldsymbol{x}}_{1,2} - \overline{\boldsymbol{x}}_0|^2.$$

Отсюда и из формулы (25) следует, что

$$n_1|\overline{\boldsymbol{x}}_1-\overline{\boldsymbol{x}}_0|^2+n_2|\overline{\boldsymbol{x}}_2-\overline{\boldsymbol{x}}_0|^2-\frac{n_1n_2}{n_1+n_2}\,|\overline{\boldsymbol{x}}_1-\overline{\boldsymbol{x}}_2|^2 = (n_1+n_2)|\overline{\boldsymbol{x}}_{1,2}-\overline{\boldsymbol{x}}_0|^2.$$

Остается только разделить обе части на (n_1+n_2).

Поскольку мера отдаленности метода Уорда Р5 задается равенством $\rho_W = \dfrac{n_1n_2}{n_1+n_2}\,|\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2|^2$, формула Ланса—Уильямса для нее легко выводится из формулы для ρ_{center}.

2. Достаточно доказать, что для фиксированного класса S_l выполняется матричное тождество

$$\sum_{\boldsymbol{x}_i\in S_l}(\boldsymbol{x}_i-\overline{\boldsymbol{x}})(\boldsymbol{x}_i-\overline{\boldsymbol{x}})^T = \sum_{\boldsymbol{x}_i\in S_l}(\boldsymbol{x}_i-\overline{\boldsymbol{x}}_l)(\boldsymbol{x}_i-\overline{\boldsymbol{x}}_l)^T + \\ + n_l(\overline{\boldsymbol{x}}_l-\overline{\boldsymbol{x}})(\overline{\boldsymbol{x}}_l-\overline{\boldsymbol{x}})^T.$$

Для элементов на главной диагонали оно вытекает из теоремы Гюйгенса. В противном случае оно следует из тождества $ab = [(a+b)^2-(a-b)^2]/4$ и линейности центра масс.

3. Положим $\boldsymbol{y} = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{x}-\boldsymbol{\mu})$ (см. П10). Ввиду дистрибутивности матричного умножения $\overline{\boldsymbol{y}} = \boldsymbol{\Sigma}^{-1/2}(\overline{\boldsymbol{x}}-\boldsymbol{\mu})$. В новых координатах доказываемое равенство (10) имеет вид

$$\sum_{i=1}^{n}|\boldsymbol{y}_i|^2 = n|\overline{\boldsymbol{y}}|^2 + \sum_{i=1}^{n}|\boldsymbol{y}_i-\overline{\boldsymbol{y}}|^2.$$

Оно справедливо в силу все той же теоремы Гюйгенса.

4. **а)** Уравнение разделяющей классы поверхности $h_1 = h_2$ в данном случае эквивалентно условию

$$(\boldsymbol{x}-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1) = (\boldsymbol{x}-\boldsymbol{\mu}_2)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_2). \quad (26)$$

Сделаем замену $\boldsymbol{y} = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{x}-\boldsymbol{\mu}_1)$. Тогда $\boldsymbol{x} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{y}+\boldsymbol{\mu}_1$ и $\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{x}-\boldsymbol{\mu}_2) = \boldsymbol{y}+\boldsymbol{b}$, где $\boldsymbol{b} = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)$. При этом условие (26) превращается в уравнение

$$|\boldsymbol{y}+\boldsymbol{b}|^2 - |\boldsymbol{y}|^2 = (\boldsymbol{y}+\boldsymbol{b})^T(\boldsymbol{y}+\boldsymbol{b}) - \boldsymbol{y}^T\boldsymbol{y} = 2\boldsymbol{b}^T\boldsymbol{y} + |\boldsymbol{b}|^2 = 0.$$

Это уравнение гиперплоскости в координатах y. Подставив в него выражения для $\boldsymbol{y}$ и $\boldsymbol{b}$, запишем его в координатах x:

$$\boldsymbol{b}^T(2\boldsymbol{y}+\boldsymbol{b}) = (\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T\boldsymbol{\Sigma}^{-1}(2\boldsymbol{x}-\boldsymbol{\mu}_1-\boldsymbol{\mu}_2) = 0. \qquad (27)$$

Положив $\boldsymbol{a} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)$, перепишем (27) в форме

$$\boldsymbol{a}^T(2\boldsymbol{x}-\boldsymbol{\mu}_1-\boldsymbol{\mu}_2) = 0.$$

б) Пусть верна гипотеза $H_1\colon \boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$. Введем случайную величину $Y = \boldsymbol{a}^T(2\boldsymbol{X}-\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)$. Как линейная функция от нормального вектора $\boldsymbol{X}$, случайная величина Y имеет нормальное распределение (П9). Найдем $\mathsf{M}Y$ и $\mathsf{D}Y$ при условии справедливости гипотезы H_1:

$$\mathsf{M}_1 Y = \boldsymbol{a}^T(2\boldsymbol{\mu}_1-\boldsymbol{\mu}_1-\boldsymbol{\mu}_2) = \boldsymbol{a}^T(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2) = d,$$

где $d = d(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2) = (\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)$ — расстояние Махаланобиса между $\boldsymbol{\mu}_1$ и $\boldsymbol{\mu}_2$ (см. формулу (18)); по свойствам дисперсии из П2

$$\mathsf{D}_1 Y = 4\mathsf{D}_1(\boldsymbol{a}^T\boldsymbol{X}) = 4\boldsymbol{a}^T\boldsymbol{\Sigma}\boldsymbol{a} = 4\boldsymbol{a}^T\boldsymbol{\Sigma}[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)] = 4d.$$

Таким образом, $Y \sim \mathcal{N}(d, 4d)$ при выполнении гипотезы H_1. Отсюда

$$\alpha = \mathsf{P}_1(Y \leqslant 0) = \mathsf{P}_1\left(\frac{Y-d}{2\sqrt{d}} \leqslant \frac{0-d}{2\sqrt{d}}\right) = \Phi\left(-\frac{\sqrt{d}}{2}\right),$$

где $\Phi(x)$ обозначает функцию распределения закона $\mathcal{N}(0,1)$.

Аналогично, при гипотезе $H_2\colon \boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$ находим, что случайная величина $Y \sim \mathcal{N}(-d, 4d)$ и

$$\beta = \mathsf{P}_2(Y > 0) = \mathsf{P}_2\left(\frac{Y+d}{2\sqrt{d}} > \frac{0+d}{2\sqrt{d}}\right) = 1-\Phi\left(\frac{\sqrt{d}}{2}\right) = \alpha.$$

Чем более далекими в метрике Махаланобиса являются $\boldsymbol{\mu}_1$ и $\boldsymbol{\mu}_2$, тем меньше вероятность ошибочной классификации.

ОТВЕТЫ НА ВОПРОСЫ

1. Пусть вектор $\boldsymbol{x} = (x_1,\ldots,x_m)$ имеет ровно k компонент величины $M = \max\{|x_1|,\ldots,|x_m|\}$. Тогда

$$(|x_1|^p + \ldots + |x_m|^p)^{1/p} = M\left[k + \varepsilon_1^p + \ldots + \varepsilon_{m-k}^p\right]^{1/p},$$

где $\varepsilon_i < 1$. При $p \to \infty$ выражение в квадратных скобках справа стремится к k, а вся правая часть — к M.

2. Обозначим через D вершину равностороннего треугольника BCD (рис. 23). Искомой точкой служит точка пересечения окружности с отрезком AD внутри $\triangle ABC$. Действительно, $\angle BOC = 120°$, поскольку опирается на дугу BC величины 240°. С другой стороны, $\angle BOD = \angle BCD = 60°$, так как

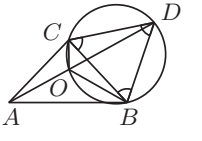


Рис. 23

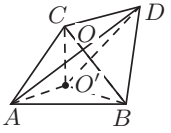


Рис. 24

эти углы опираются на общую дугу BD. Отсюда $\angle AOB = 180^\circ - \angle BOD = 120^\circ$, $\angle AOC = 360^\circ - \angle BOC - \angle AOB = 120^\circ$.

3. Точка пересечения диагоналей O — искомая. Для любой другой точки O' отрезки диагоналей заменяются на ломаные (рис. 24).

4. Может. Для вершин равностороннего треугольника со сторонами длины 2 на втором шаге происходит объединение на уровне $\sqrt{3} < 2$ (рис. 25).

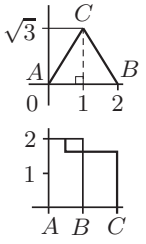


Рис. 25

5. а) Если $\rho_{min}(S_0, S_1) \geqslant \delta$ и $\rho_{min}(S_0, S_2) \geqslant \delta$, то

$$\rho_{min}(S_0, S_1 \cup S_2) = \min\{\rho_{min}(S_0, S_1), \rho_{min}(S_0, S_2)\} \geqslant \delta.$$

б) Пусть $\rho_W(S_1, S_2) \leqslant \delta$, а $\rho_W(S_0, S_1) \geqslant \delta$, $\rho_W(S_0, S_2) \geqslant \delta$. Воспользуемся формулой Ланса и Уильямса:

$$\rho_W(S_0, S_1 \cup S_2) \geqslant [(n_0+n_1)\delta + (n_0+n_2)\delta - n_0\delta]/(n_0+n_1+n_2) = \delta.$$

6. Так как (см. П10) $(\boldsymbol{\Sigma}^{-1})^T = (\boldsymbol{\Sigma}^T)^{-1} = \boldsymbol{\Sigma}^{-1}$, то обратная ковариационная матрица $\boldsymbol{\Sigma}^{-1}$ является симметричной. Пусть $\boldsymbol{C}$ — ортогональная матрица, преобразующая $\boldsymbol{\Sigma}$ к главным осям: $\boldsymbol{C}^T\boldsymbol{\Sigma}\boldsymbol{C} = \boldsymbol{\Lambda}$. Тогда $\boldsymbol{C}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{C} = \boldsymbol{\Lambda}^{-1}$, где $\boldsymbol{\Lambda}^{-1}$ — диагональная матрица с положительными элементами на главной диагонали.

7. Согласно обобщенной проблеме собственных значений (см. П10) для пары $\widehat{\boldsymbol{\Sigma}}$ и $\boldsymbol{\Sigma}$ найдется такая невырожденная матрица $\boldsymbol{D}$, что $\boldsymbol{D}^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{D} = \boldsymbol{\Lambda}$ и $\boldsymbol{D}^T\boldsymbol{\Sigma}\boldsymbol{D} = \boldsymbol{E}$. Если $\boldsymbol{\Lambda} = \boldsymbol{E}$, то $\widehat{\boldsymbol{\Sigma}} = (\boldsymbol{D}^T)^{-1}(\boldsymbol{D})^{-1} = (\boldsymbol{D}\boldsymbol{D}^T)^{-1} = \boldsymbol{\Sigma}$.

Глава 20

КОРРЕЛЯЦИЯ

После того, как (с помощью методов из гл. 19) объекты разбиты на однородные группы (классы), возникает задача изучения *взаимосвязей признаков* внутри отдельного класса. На практике чаще всего встречаются следующие два вида зависимостей: а) объекты образуют «облако» эллиптического типа (рис. 1, *а*), б) объекты располагаются в окрестности некоторой кривой (поверхности) (рис. 1, *б*). В случае а) оба признака являются «полноценными» случайными величинами, и изучению подлежит уровень зависимости (корреляции) между ними. Случай б) соответствует «функциональной» зависимости между признаками, испорченной шумом. Зависимости первого вида изучаются в этой главе методами *корреляционного анализа*. Методы, позволяющие во втором случае построить интересующую исследователя кривую (поверхность), относятся к так называемому *регрессионному анализу*. Они обсуждаются в гл. 21 и § 2 гл. 22.

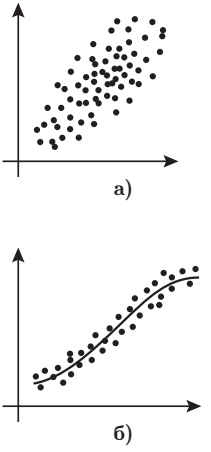


Рис. 1

§ 1. ГЕОМЕТРИЯ ГЛАВНЫХ КОМПОНЕНТ

Допустим, что каждый из n объектов описывается m признаками (координатами), и представим данные (для отдельного класса объектов) в форме таблицы $\boldsymbol{X} = \|x_{il}\|_{n\times m}$. Вычислим для каждого признака (столбца матрицы $\boldsymbol{X}$) *среднее значение* $\overline{x}_l = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_{il}$ и центрируем данные: $x'_{il} = x_{il} - \overline{x}_l$.[*)] Далее в этой главе будем считать x_{il} уже *центрированными*:

Д1. $\overline{x}_l = 0$ для $l = 1,\ldots,m$.

Обозначим через $\widehat{\boldsymbol{\Sigma}} = \|\widehat{\sigma}_{kl}\|_{m\times m}$ *выборочную ковариационную матрицу* (центрированных) признаков: $\widehat{\boldsymbol{\Sigma}} = \dfrac{1}{n}\,\boldsymbol{X}^T\boldsymbol{X}$ (т. е. $\widehat{\sigma}_{kl}$ — выборочная ковариация k-го и l-го столбцов матрицы $\boldsymbol{X}$).

Сократ. А смог бы ты, не глядя на скалу, а рассматривая ее отражение в воде, сказать, как можно было бы взобраться на самую вершину?

А. Реньи. Диалоги

Поскольку $\widehat{\boldsymbol{\Sigma}}$ — матрица ковариаций, она неотрицательно определена (см. П10). Следовательно, существует ортогональная матрица $\boldsymbol{C}$, приводящая $\widehat{\boldsymbol{\Sigma}}$ к главным осям: $\boldsymbol{C}^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{C} = \boldsymbol{\Lambda}$. Здесь $\boldsymbol{\Lambda}$ — диагональная матрица с неотрицательными элементами $\lambda_1 \geqslant \ldots \geqslant \lambda_m$ на главной диагонали, которые являются корнями уравнения $\det(\widehat{\boldsymbol{\Sigma}} - \lambda\boldsymbol{E}) = 0$. Они называются *собственными значениями* матрицы $\widehat{\boldsymbol{\Sigma}}$. Предположим, что все λ_l *положительны*

[*)] Это преобразование не искажает интересующую нас внутреннюю структуру класса, характер взаимосвязей признаков.

и различны (для экспериментальных данных x_{il} это условие выполняется практически всегда). При этом столбцы $\boldsymbol{c}_1,\ldots,\boldsymbol{c}_m$ матрицы $\boldsymbol{C}$ (*главные оси или компоненты*) определяются однозначно с точностью до выбора направления оси (одновременного изменения знака всех координат вектора $\boldsymbol{c}_l$). Они образуют новый ортонормированный базис в $\mathbb{R}^m$ (рис. 2), обладающий **рядом важных свойств**.

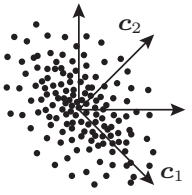


Рис. 2

1) Проекции объектов на первую главную компоненту $\boldsymbol{c}_1$ имеют *наибольшую* выборочную дисперсию среди проекций на всевозможные направления $\boldsymbol{d}$ в пространстве $\mathbb{R}^m$.

   Доказательство. Вектор проекций $\boldsymbol{y}$ на направление $\boldsymbol{d}$ ($\boldsymbol{d}^T\boldsymbol{d}=1$) задается равенством $\boldsymbol{y}=\boldsymbol{X}\boldsymbol{d}$. Ввиду допущения Д1

$$\overline{\boldsymbol{y}}=\frac{1}{n}\sum_{i=1}^{n}y_i=\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{m}x_{il}d_l=\sum_{l=1}^{m}\overline{x}_l d_l=0.$$

   Тогда *выборочная дисперсия проекций на направление* $\boldsymbol{d}$ равна

$$S^2(\boldsymbol{d})=\frac{1}{n}\sum_{i=1}^{n}y_i^2=\frac{1}{n}\,\boldsymbol{y}^T\boldsymbol{y}=\frac{1}{n}\,(\boldsymbol{X}\boldsymbol{d})^T\boldsymbol{X}\boldsymbol{d}=\boldsymbol{d}^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{d}. \tag{1}$$

   Тем самым задача сводится к максимизации по $\boldsymbol{d}$ квадратичной формы $\boldsymbol{d}^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{d}$ при условии $\boldsymbol{d}^T\boldsymbol{d}=1$. Для ее решения применим метод неопределенных множителей Лагранжа (см. [46, с. 271]). Приравнивая нулю частные производные по переменным d_l *функции Лагранжа*

$$F(\boldsymbol{d},\lambda)=\boldsymbol{d}^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{d}-\lambda(\boldsymbol{d}^T\boldsymbol{d}-1)=\sum_{k=1}^{m}\sum_{l=1}^{m}\widehat{\sigma}_{kl}d_kd_l-\lambda\sum_{l=1}^{m}d_l^2+\lambda,$$

   приходим к системе (линейных относительно $d_1,\ldots,d_m$) уравнений

$$\frac{\partial}{\partial d_l}F(\boldsymbol{d},\lambda)=2\sum_{k=1}^{m}\widehat{\sigma}_{kl}d_k-2\lambda d_l=0,\quad l=1,\ldots,m.$$

   Ее можно записать в матричной форме:

$$\widehat{\boldsymbol{\Sigma}}\boldsymbol{d}-\lambda\boldsymbol{d}=\boldsymbol{0}\iff(\widehat{\boldsymbol{\Sigma}}-\lambda\boldsymbol{E})\boldsymbol{d}=\boldsymbol{0}. \tag{2}$$

   Поскольку $\boldsymbol{d}^T\boldsymbol{d}=1$, нас интересуют только ненулевые решения. Для них должно выполняться условие $\det(\widehat{\boldsymbol{\Sigma}}-\lambda\boldsymbol{E})=0$, т. е. искомое направление $\boldsymbol{d}$ обязано быть собственным вектором $\widehat{\boldsymbol{\Sigma}}$, отвечающим собственному значению λ. Умножая соотношение (2) на $\boldsymbol{d}^T$ слева, получим

$$\boldsymbol{d}^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{d}=\lambda\,\boldsymbol{d}^T\boldsymbol{d}=\lambda.$$

   Левая часть есть $S^2(\boldsymbol{d})$ (см. формулу (1)). Следовательно, λ должно равняться наибольшему собственному значению λ_1 матрицы $\widehat{\boldsymbol{\Sigma}}$, а $\boldsymbol{d}=\boldsymbol{c}_1$. ■

2) Второй собственный вектор $\boldsymbol{c}_2$ характеризуется тем, что выборочная дисперсия проекций объектов на ось $\boldsymbol{c}_2$ максимальна среди всех направлений $\boldsymbol{d}$, *ортогональных вектору* $\boldsymbol{c}_1$, т. е. таких, что $\boldsymbol{d}^T\boldsymbol{c}_1 = 0$.

Доказательство. Функция Лагранжа в этом случае выглядит так:

$$F(\boldsymbol{d},\lambda,\mu) = \boldsymbol{d}^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{d} - \lambda(\boldsymbol{d}^T\boldsymbol{d} - 1) - \mu(\boldsymbol{d}^T\boldsymbol{c}_1 - 0).$$

Вычисляя ее частные производные по d_l, приходим к системе

$$\widehat{\boldsymbol{\Sigma}}\boldsymbol{d} - \lambda\boldsymbol{d} - \mu\boldsymbol{c}_1 = \boldsymbol{0}. \tag{3}$$

Транспонируем равенство (3), умножим на $\boldsymbol{c}_1$ справа и воспользуемся тем, что $\boldsymbol{c}_1$ — собственный вектор с собственным значением λ_1 (т. е. $\widehat{\boldsymbol{\Sigma}}\boldsymbol{c}_1 = \lambda_1\boldsymbol{c}_1$), а также условием $\boldsymbol{d}^T\boldsymbol{c}_1 = 0$:

$$\boldsymbol{d}^T\widehat{\boldsymbol{\Sigma}}\boldsymbol{c}_1 - \lambda\,\boldsymbol{d}^T\boldsymbol{c}_1 - \mu\,\boldsymbol{c}_1^T\boldsymbol{c}_1 = 0 \Longleftrightarrow \lambda_1\cdot 0 - \lambda\cdot 0 - \mu\cdot 1 = 0.$$

Отсюда $\mu = 0$. С учетом этого из (3) вытекает, что $\boldsymbol{d}$ — собственный вектор матрицы $\widehat{\boldsymbol{\Sigma}}$ с собственным значением λ. Умножая равенство (3) слева на $\boldsymbol{d}^T$, выводим, что $\lambda = \lambda_2$ и $\boldsymbol{d} = \boldsymbol{c}_2$. ■

3) При $l \geqslant 3$ аналогично устанавливается, что $\boldsymbol{c}_l$ — направление с наибольшей выборочной дисперсией проекций объектов среди направлений, ортогональных векторам $\boldsymbol{c}_1,\ldots,\boldsymbol{c}_{l-1}$.

4) Сумма выборочных дисперсий исходных признаков (столбцов матрицы $\boldsymbol{X}$) $\widehat{\sigma}_{11} + \ldots + \widehat{\sigma}_{mm} = \operatorname{tr}\widehat{\boldsymbol{\Sigma}}$ в силу подобия матриц $\widehat{\boldsymbol{\Sigma}}$ и $\boldsymbol{\Lambda}$ (см. П10) равна $\operatorname{tr}\boldsymbol{\Lambda} = \lambda_1 + \ldots + \lambda_m = S^2(\boldsymbol{c}_1) + \ldots + S^2(\boldsymbol{c}_m)$, т. е. сумме выборочных дисперсий проекций объектов на (новые) главные оси. Эта величина может рассматриваться как *мера общего разброса* объектов относительно их центра масс. Представляет интерес *относительная доля разброса, приходящаяся на k первых главных осей*,

$$\gamma_k = (\lambda_1 + \ldots + \lambda_k)/(\lambda_1 + \ldots + \lambda_m), \quad k \leqslant m.$$

Если эта величина при некотором k достаточно близка к 1, то возможно *уменьшение размерности* пространства признаков за счет перехода от m исходных признаков к k новым признакам — первым главным компонентам. На практике нередко удается ограничиться двумя или тремя компонентами без существенной потери информации. Объекты описываются координатами в новых осях, которым специалисты-прикладники, как правило, могут придать содержательную интерпретацию.

Математика приводит нас к дверям истины, но самих дверей не открывает.

В. Ф. Одоевский

Пример 1 ([30, с. 206])**.** Найдем и интерпретируем главные компоненты для данных примера 4 гл. 19. Напомним, что исходными признаками там были следующие: 1 — длина черепа, 2 — длина верхней челюсти, 3 — ширина верхней челюсти, 4 —

	1	2	3	4	5	6
1	1	0,96	0,35	0,61	0,72	0,59
2		1	0,20	0,66	0,74	0,59
3			1	0,37	0,35	0,35
4				1	0,89	0,76
5					1	0,79
6						1

Рис. 3

$\boldsymbol{c}_1$	$\boldsymbol{c}_2$	$\boldsymbol{c}_3$	$\boldsymbol{c}_4$	$\boldsymbol{c}_5$	$\boldsymbol{c}_6$
0,43	0,23	0,53	0,11	0,05	0,68
0,43	0,38	0,39	0,01	−0,20	−0,69
0,23	−0,89	0,38	−0,02	0,00	−0,13
0,44	−0,07	−0,40	−0,52	−0,58	0,18
0,46	0,02	−0,27	−0,31	0,78	−0,09
0,42	−0,10	−0,44	0,79	−0,09	−0,01

Рис. 4

длина верхнего карнивора, 5 — длина первого верхнего моляра, 6 — ширина первого верхнего моляра.

Очевидно, что при нормировке данных с помощью средних арифметических и стандартных отклонений признаков (N2 из § 1 гл. 19) выборочная ковариационная матрица $\widehat{\boldsymbol{\Sigma}}$ совпадает с выборочной корреляционной матрицей исходных признаков. [Ее нетрудно подсчитать с помощью программы Excel. Результаты вычислений представлены таблицей на рис. 3.]

Обратим внимание, что длина черепа (признак 1) и длина верхней челюсти (признак 2) сильно коррелируют ($\widehat{\rho}_{12} = 0{,}96$), поэтому целесообразно оставить в модели только один из них. Признаки 4, 5 и 6, относящиеся к зубам, также очень тесно связаны между собой, поскольку $\widehat{\rho}_{45} = 0{,}89$, $\widehat{\rho}_{46} = 0{,}76$ и $\widehat{\rho}_{56} = 0{,}79$.

На рис. 4 приведены собственные векторы $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_6$, вычисленные *степенным методом* (см. § 3). Собственные значения λ_k, соответствующие им доли следа $\lambda_k / \sum \lambda_l$ (в %) и накопленные доли γ_k (в %) указаны в следующей таблице:

Номера компонент	1	2	3	4	5	6
Собственные значения	4,100	0,883	0,639	0,259	0,097	0,022
Проценты от следа	68,3	14,7	10,7	4,3	1,6	0,4
Накопленные проценты	68,3	83,0	93,7	98,0	99,6	100,0

На первые 3 компоненты приходится 93,7% полной дисперсии «облака». При этом первая компонента имеет смысл *общего размера*. Это следует из того, что все координаты у $\boldsymbol{c}_1$ одного знака и примерно одинаковы по величине, т. е. при проецировании на эту ось координаты нормированных признаков просто складываются.

Вторая компонента, по существу, отвечает за *ширину верхней челюсти* (признак 3), поскольку третья координата у $\boldsymbol{c}_2$ по абсолютной величине равна $0{,}89 \approx 1$. Эта ось отражает различие в пропорциях челюстей и отличает удлиненные формы от укороченных (гончих и колли от бульдогов и боксеров). На

вторую ось волки проецируются в основном рядом с немецкими овчарками.

Третья ось противопоставляет размеры челюстей размерам зубов: первые три координаты у $\boldsymbol{c}_3$ примерно равны по абсолютной величине последним трем координатам, но противоположны по знаку. Другими словами, третья ось отражает *относительную* (по сравнению с размерами черепа) *величину зубов*. Она позволяет отличить животных с развитыми зубами (волки, немецкие овчарки, доберманы) от собак других пород (сенбернары, мастифы, сеттеры).

5) Пусть M_k — это подпространство, натянутое на главные оси $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_k$. Оказывается, при проецировании объектов на произвольное подпространство L_k размерности k в $\mathbb{R}^m$ геометрическая структура *искажается в наименьшей степени*, если этим подпространством является M_k (см. [1, с. 350]):
а) сумма квадратов расстояний от объектов до их проекций на L_k минимальна, когда $L_k = M_k$ (в этом случае она равна $n(\lambda_{k+1} + \ldots + \lambda_m)$ (доказательство см. в [76, с. 243]));
б) при проецировании на M_k наименее искажается сумма квадратов расстояний между всевозможными парами объектов (для M_k ее изменение составляет $n^2(\lambda_{k+1} + \ldots + \lambda_m)$);
в) когда $L_k = M_k$, в наименьшей степени искажаются расстояния от объектов до их центра масс (совпадающего с началом координат $\mathbf{0}$ ввиду допущения Д1), а также углы между всевозможными парами прямых, соединяющих объекты с $\mathbf{0}$.

Поясним последнее свойство. Рассмотрим *матрицу скалярных произведений* $\boldsymbol{G} = \|g_{ij}\|_{n\times n} = \boldsymbol{X}\boldsymbol{X}^T$. Нетрудно понять геометрический смысл элементов этой матрицы: $g_{ii} = \sum\limits_{l=1}^{m} x_{il}^2$ представляет собой квадрат расстояния от i-го объекта до $\mathbf{0}$, а при $i \neq j$ величина $g_{ij} = \sum\limits_{l=1}^{m} x_{il}x_{jl}$ пропорциональна косинусу угла между прямыми, соединяющими i-й и j-й объекты с началом координат.

Обозначим через $\boldsymbol{Y} = \|y_{il}\|_{n\times m}$ матрицу координат проекций объектов на подпространство L_k. Ей соответствует $\boldsymbol{H} = \|h_{ij}\|_{n\times n} = \boldsymbol{Y}\boldsymbol{Y}^T$. Тогда при $L_k = M_k$ достигается

$$\min_{L_k} |\boldsymbol{G} - \boldsymbol{H}|^2 = n^2(\lambda_{k+1}^2 + \ldots + \lambda_m^2), \tag{4}$$

где $|A|^2 = \sum a_{ij}^2$ — квадрат *евклидовой нормы матрицы* $\boldsymbol{A}$.

Замечание 1. Проецирование на плоскость двух первых компонент часто применяется еще на этапе классификации (выделении однородных групп). Авторы [4, с. 103] считают: «Гипотеза состоит в том, что наибольший разброс данные будут иметь в направлениях, «соединяющих» центры групп, а значит, проекции на старшие главные компоненты обеспечат наилучшую «точку зрения» на данные,

когда группы видны на наибольших расстояниях и не закрывают одна другую».

Замечание 2. Следует иметь в виду, что главные компоненты вычисляются по выборочной ковариационной матрице $\widehat{\boldsymbol{\Sigma}}$ и поэтому *зависят от масштаба признаков*. Скажем, если один из признаков принимает значения от 0 до 100, а другие — от 0 до 10, то независимо от структуры данных первый признак будет отождествляться с первой главной компонентой. Чтобы избежать этого, обычно перед вычислением главных компонент данные нормируют (см. § 1 гл. 19).

§ 2. ЭЛЛИПСОИД РАССЕЯНИЯ

Рассмотрим m-мерный случайный вектор $\boldsymbol{\xi}$ с математическим ожиданием $\mathsf{M}\boldsymbol{\xi} = \mathbf{0}$ и ковариационной матрицей $\boldsymbol{\Sigma}_{\boldsymbol{\xi}}$. (В частности, годится *эмпирическое распределение*[*)] с нулевыми средними арифметическими значений координат признаков (допущение Д1 из § 1) и выборочной ковариационной матрицей $\widehat{\boldsymbol{\Sigma}}$.)

В П10 доказано, что любая ковариационная матрица неотрицательно определена. Потребуем дополнительно, чтобы матрица $\boldsymbol{\Sigma}_{\boldsymbol{\xi}}$ была *невырожденной*. Это равносильно ее положительной определенности, а также положительности всех ее собственных значений $\lambda_1 \geqslant \ldots \geqslant \lambda_m$. Обозначим соответствующие им собственные векторы (направления главных осей) через $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_m$.

Для произвольного направления $\boldsymbol{d}$ в $\mathbb{R}^m$ ($|\boldsymbol{d}| = 1$) случайная величина $\zeta = \boldsymbol{d}^T\boldsymbol{\xi}$ представляет собой *координату проекции* вектора $\boldsymbol{\xi}$ на направление $\boldsymbol{d}$. Найдем такое направление, для которого дисперсия $\mathsf{D}\zeta$ имеет *наибольшее значение*. Запишем:

$$\mathsf{D}\zeta = \mathsf{M}(\boldsymbol{d}^T\boldsymbol{\xi})^2 = \mathsf{M}(\boldsymbol{d}^T\boldsymbol{\xi})(\boldsymbol{d}^T\boldsymbol{\xi})^T = \mathsf{M}\,\boldsymbol{d}^T\boldsymbol{\xi}\boldsymbol{\xi}^T\boldsymbol{d} = \boldsymbol{d}^T\boldsymbol{\Sigma}_{\boldsymbol{\xi}}\boldsymbol{d}.$$

В точности так же, как в § 1 для случая выборочной ковариационной матрицы $\widehat{\boldsymbol{\Sigma}}$, доказывается, что максимум дисперсии $\mathsf{D}\zeta$ достигается на направлении $\boldsymbol{c}_1$.

Аналогично устанавливается, что при $l > 1$ собственный вектор $\boldsymbol{c}_l$ является направлением с наибольшей дисперсией $\mathsf{D}\zeta$ среди направлений, ортогональных векторам $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_{l-1}$.

В силу того, что $\boldsymbol{\Sigma}_{\boldsymbol{\xi}}$ невырождена, существует обратная к ней матрица $\boldsymbol{\Sigma}_{\boldsymbol{\xi}}^{-1}$, которая также положительно определена (см. вопрос 6 гл. 19).

Определение. *Эллипсоидом рассеяния* распределения вектора $\boldsymbol{\xi}$ называется m-мерный эллипсоид

$$\boldsymbol{x}^T\boldsymbol{\Sigma}_{\boldsymbol{\xi}}^{-1}\boldsymbol{x} \leqslant m + 2.$$

[*)] У которого каждому набору $(x_{i1}, \ldots, x_{im})$ координат i-го объекта (т. е. строке таблицы данных $\boldsymbol{X}$) приписана вероятность $1/n$.

Он однозначно выделяется среди всех других эллипсоидов следующим своим свойством (см. [44, с. 333]): если рассмотреть *равномерно распределенный* на нем вектор $\boldsymbol{U}$ (имеющий постоянную плотность внутри эллипсоида и равную 0 вне его), то первые (равные 0) и вторые моменты (т. е. ковариации компонент) вектора $\boldsymbol{U}$ совпадут с моментами вектора $\boldsymbol{\xi}$.

Осями эллипсоида рассеяния служат главные оси матрицы $\boldsymbol{\Sigma_\xi}$ (рис. 5 для $\widehat{\boldsymbol{\Sigma}}$). Длины его полуосей пропорциональны $\sqrt{\lambda_l}$, где λ_l — собственные значения матрицы $\boldsymbol{\Sigma_\xi}$, а m-мерный объем равен константе, умноженной на $(\det \boldsymbol{\Sigma_\xi})^{1/2} = (\lambda_1 \dots \lambda_m)^{1/2}$.

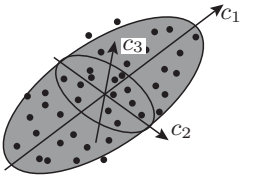


Рис. 5

В заключение параграфа познакомимся с одним из способов сравнения «степеней рассеяния» многомерных распределений, который связан с их эллипсоидами рассеяния.

Пусть m-мерные случайные векторы $\boldsymbol{\xi}$ и $\boldsymbol{\eta}$ имеют распределения с $\mathsf{M}\boldsymbol{\xi} = \mathsf{M}\boldsymbol{\eta} = \mathbf{0}$ и матрицами ковариаций $\boldsymbol{\Sigma_\xi}$, $\boldsymbol{\Sigma_\eta}$.

Определение. Будем говорить, что *среднеквадратическое рассеяние* случайного вектора $\boldsymbol{\xi}$ вокруг начала координат $\mathbf{0}$ не больше, чем рассеяние вектора $\boldsymbol{\eta}$, если для любого $\boldsymbol{d} \in \mathbb{R}^m$ верно неравенство

$$\mathsf{M}(\boldsymbol{d}^T\boldsymbol{\xi})^2 \leqslant \mathsf{M}(\boldsymbol{d}^T\boldsymbol{\eta})^2. \tag{5}$$

Неравенство (5) означает, что дисперсия случайной величины $\boldsymbol{d}^T\boldsymbol{\xi}$ для любого направления $\boldsymbol{d}$ не превосходит дисперсии случайной величины $\boldsymbol{d}^T\boldsymbol{\eta}$. Раскрывая скобки, очевидно, получаем, что неравенство (5) равносильно неотрицательной определенности матрицы $\boldsymbol{\Sigma_\eta} - \boldsymbol{\Sigma_\xi}$.

В [11, с. 102] доказано, что при условии невырожденности матриц $\boldsymbol{\Sigma_\xi}$ и $\boldsymbol{\Sigma_\eta}$ среднеквадратическое рассеяние вектора $\boldsymbol{\xi}$ вокруг начала координат не больше рассеяния вектора $\boldsymbol{\eta}$ тогда и только тогда, когда эллипсоид рассеяния вектора $\boldsymbol{\xi}$ целиком *лежит внутри* эллипсоида рассеяния вектора $\boldsymbol{\eta}$.

Замечание 3. При $m \geqslant 2$ неравенство (5) устанавливает лишь частичный порядок на множестве ковариационных матриц. Например, матрицы $\begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$ и $\begin{pmatrix} 5 & 0 \\ 0 & 1 \end{pmatrix}$ не лучше и не хуже одна другой, поскольку в направлении $\boldsymbol{d} = (1{,}0)$ меньше дисперсия для первой матрицы, а в направлении $\boldsymbol{d} = (0{,}1)$ — для второй (рис. 6).

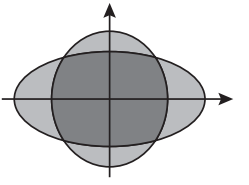


Рис. 6

В подобной ситуации для сравнения «степеней рассеяния» многомерных законов обычно используют такие скалярные характеристики ковариационных матриц, как след $\lambda_1 + \dots + \lambda_m$ или определитель $\lambda_1 \dots \lambda_m$. При этом надо иметь в виду, что выводы для разных характеристик (как и для выше приведенных матриц) могут оказаться *прямо противоположными* (эту тему продолжает задача 1).

§ 3. ВЫЧИСЛЕНИЕ ГЛАВНЫХ КОМПОНЕНТ

Следующий простой алгоритм (*степенной метод*) позволяет вычислить для симметричной (не обязательно неотрицательно определенной) матрицы $\boldsymbol{A}$ несколько *наибольших по модулю* собственных значений λ_l и соответствующих собственных векторов $\boldsymbol{c}_l$ (см. [6, с. 221], [16, с. 166]). Сначала разберем частный случай этого алгоритма — вычисление λ_1 и $\boldsymbol{c}_1$.

Возьмем произвольный ненулевой вектор $\boldsymbol{x}_0$ из $\mathbb{R}^m$. Пусть выполнены два условия: $|\lambda_1| > |\lambda_2| \geqslant \ldots \geqslant |\lambda_m|$ и $\boldsymbol{c}_1^T\boldsymbol{x}_0 \neq 0$. Положим $\boldsymbol{x}_{i+1} = \boldsymbol{A}\boldsymbol{x}_i$ (т. е. $\boldsymbol{x}_i = \boldsymbol{A}^i\boldsymbol{x}_0$, что объясняет название метода). Тогда последовательность значений $t_{i+1} = \boldsymbol{x}_i^T\boldsymbol{x}_{i+1}/|\boldsymbol{x}_i|^2$ сходится к λ_1 со скоростью геометрической прогрессии со знаменателем $\gamma = |\lambda_2/\lambda_1| < 1$.

Доказательство. Обозначим через $b_1,\ldots,b_m$ координаты вектора $\boldsymbol{x}_0$ в базисе $\boldsymbol{c}_1,\ldots,\boldsymbol{c}_m$, т. е. $\boldsymbol{x}_0 = b_1\boldsymbol{c}_1 + \ldots + b_m\boldsymbol{c}_m$. Так как $\boldsymbol{A}\boldsymbol{c}_l = \lambda_l\boldsymbol{c}_l$, то $\boldsymbol{A}^i\boldsymbol{c}_l = \lambda_l^i\boldsymbol{c}_l$. Отсюда $\boldsymbol{x}_i = \lambda_1^i b_1\boldsymbol{c}_1 + \ldots + \lambda_m^i b_m\boldsymbol{c}_m = \lambda_1^i(b_1\boldsymbol{c}_1 + \boldsymbol{\delta}_i)$, где $\boldsymbol{\delta}_i = b_2(\lambda_2/\lambda_1)^i\boldsymbol{c}_2 + \ldots + b_m(\lambda_m/\lambda_1)^i\boldsymbol{c}_m$, причем в силу первого условия $|\boldsymbol{\delta}_i| = O(\gamma^i) \to 0$ при $i \to \infty$.

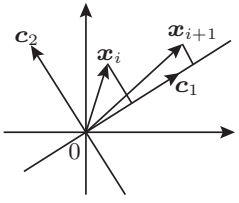


Рис. 7

В частности, когда все $\lambda_l > 1$, это означает, что при каждой итерации длина вектора $\boldsymbol{x}_i$ увеличивается, но быстрее остальных растет его координата по оси $\boldsymbol{c}_1$. При этом направление $\boldsymbol{x}_i$ приближается к направлению первой главной оси (рис. 7).

Формально это можно доказать с помощью известного из линейной алгебры *неравенства Коши—Буняковского—Шварца*: $|\boldsymbol{x}^T\boldsymbol{y}| \leqslant |\boldsymbol{x}||\boldsymbol{y}|$. Согласно второму условию алгоритма $b_1 = \boldsymbol{c}_1^T\boldsymbol{x}_0 \neq 0$. Отсюда $|\boldsymbol{x}_i| = (\boldsymbol{x}_i^T\boldsymbol{x}_i)^{1/2} = |\lambda_1|^i[|b_1| + O(\gamma^i)]$; $\boldsymbol{e}_i = \boldsymbol{x}_i/|\boldsymbol{x}_i| = (\operatorname{sign}\lambda_1)^i(\operatorname{sign} b_1)\boldsymbol{c}_1 + \boldsymbol{r}_i$, где $|\boldsymbol{r}_i| = O(\gamma^i)$; $t_{i+1} = \lambda_1 + O(\gamma^i)$. ■

Замечание 4. Если $|\lambda_1| > 1$, то $|\boldsymbol{x}_i| \to \infty$, а если $|\lambda_1| < 1$, то $|\boldsymbol{x}_i| \to 0$ при $i \to \infty$. При вычислении на компьютере в первом случае возможно переполнение, во втором — исчезновение порядка. Предотвратить эти ситуации позволяет нормировка векторов $\boldsymbol{x}_i$ на каждом шаге.

Учитывая замечание 4, приходим к следующему **алгоритму для вычисления λ_1 и $\boldsymbol{c}_1$.**

1. Положим $i = 0$, $t_0 = 0$ и зададим произвольный ненулевой вектор $\boldsymbol{x}_0$ из $\mathbb{R}^m$. Нормируем $\boldsymbol{x}_0$: пусть $\boldsymbol{e}_0 = \boldsymbol{x}_0/|\boldsymbol{x}_0|$ (обычно сразу берут $\boldsymbol{e}_0 = (1,0,\ldots,0)$).
2. Вычислим $\boldsymbol{x}_{i+1} = \boldsymbol{A}\boldsymbol{e}_i$ и $t_{i+1} = \boldsymbol{e}_i^T\boldsymbol{x}_{i+1}$.
3. Нормируем $\boldsymbol{x}_{i+1}$: $\boldsymbol{e}_{i+1} = \boldsymbol{x}_{i+1}/|\boldsymbol{x}_{i+1}|$.
4. Если $|t_{i+1} - t_i| \leqslant \varepsilon$, где ε — достаточно малое число, то положим $\lambda_1 = t_{i+1}$ и $\boldsymbol{c}_1 = \boldsymbol{e}_{i+1}$ и закончим итерационный процесс. Иначе увеличим i на 1 и перейдем к шагу 2.

Замечание 5. Если при выборе $\boldsymbol{x}_0$ случится, что $\boldsymbol{c}_1^T\boldsymbol{x}_0 = 0$, то через несколько итераций за счет погрешностей округления, как правило, появится ненулевая координата у $\boldsymbol{x}_i$ в направлении $\boldsymbol{c}_1$, и процесс, хотя и с некоторым запаздыванием, выйдет на первое собственное значение. Можно также стартовать с разных $\boldsymbol{x}_0$ и сравнивать результаты.

Пусть $|\lambda_1| > |\lambda_2| > \ldots > |\lambda_k| > |\lambda_{k+1}| \geqslant \ldots \geqslant |\lambda_m|$. Для вычисления $k > 1$ первых главных компонент кроме умножения на $\boldsymbol{A}$ и нормировки понадобится еще ортогонализация. Именно, предположим, что λ_l и $\boldsymbol{c}_l$ уже известны при всех $l \leqslant k-1$. Чтобы найти λ_k и $\boldsymbol{c}_k$, выполним **следующие шаги**.

1. Положим $i = 0$, $t_0 = 0$ и зададим произвольный вектор $\boldsymbol{x}_0$ из $\mathbb{R}^m$ (надо только, чтобы $\boldsymbol{c}_k^T\boldsymbol{x}_0 \neq 0$).
2. Проведем ортогонализацию $\boldsymbol{x}_i$ к векторам $\boldsymbol{c}_1,\ldots,\boldsymbol{c}_{k-1}$: $\boldsymbol{y}_i = \boldsymbol{x}_i - (\boldsymbol{c}_1^T\boldsymbol{x}_i)\,\boldsymbol{c}_1 - \ldots - (\boldsymbol{c}_{k-1}^T\boldsymbol{x}_i)\,\boldsymbol{c}_{k-1}$.
3. Нормируем $\boldsymbol{y}_i$: $\boldsymbol{e}_i = \boldsymbol{y}_i/|\boldsymbol{y}_i|$.
4. Вычислим $\boldsymbol{x}_{i+1} = \boldsymbol{A}\boldsymbol{e}_i$ и $t_{i+1} = \boldsymbol{e}_i^T\boldsymbol{x}_{i+1}$.
5. Нормируем $\boldsymbol{x}_{i+1}$: $\boldsymbol{z}_{i+1} = \boldsymbol{x}_{i+1}/|\boldsymbol{x}_{i+1}|$.
6. Если $|t_{i+1} - t_i| \leqslant \varepsilon$, то положим $\lambda_k = t_{i+1}$, $\boldsymbol{c}_k = \boldsymbol{z}_{i+1}$ и закончим процесс. В противном случае приравняем $\boldsymbol{x}_{i+1} = \boldsymbol{z}_{i+1}$, увеличим i на 1 и перейдем к шагу 2.

В заключение обсудим, насколько следует доверять вычисленным значениям λ_l и $\boldsymbol{c}_l$. Следующий пример показывает, что для числа объектов n порядка нескольких десятков разброс этих значений может оказаться *достаточно большим*.

Пример 2. ***Статистика главных компонент для нормальной модели.*** Предположим, что объекты представляют собой выборку размера n из закона $\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$ с невырожденной матрицей $\boldsymbol{\Sigma}$, у которой все собственные значения λ_l ($l = 1,\ldots,m$) различны. Тогда:

1) собственные значения $\widehat{\lambda}_1,\ldots,\widehat{\lambda}_m$ и отвечающие им собственные векторы $\widehat{\boldsymbol{c}}_1,\ldots,\widehat{\boldsymbol{c}}_m$ выборочной ковариационной матрицы $\widehat{\boldsymbol{\Sigma}}$ являются оценками максимального правдоподобия для соответствующих $\lambda_1,\ldots,\lambda_m$, $\boldsymbol{c}_1,\ldots,\boldsymbol{c}_m$ (см. пример 2 гл. 19), и, следовательно, обладают асимптотической эффективностью (см. § 4 гл. 9);
2) $\sqrt{n}\,(\widehat{\lambda}_l - \lambda_l) \xrightarrow{d} \xi \sim \mathcal{N}(0, 2\lambda_l^2)$ при $n \to \infty$ (см. [1, с. 354]);
3) $\sqrt{n}\,(\widehat{\boldsymbol{c}}_l - \boldsymbol{c}_l) \xrightarrow{d} \boldsymbol{\zeta} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{V})$ при $n \to \infty$, где элементами матрицы $\boldsymbol{V}$ служат $v_{kl} = \lambda_l \sum\limits_{k \neq l} |\boldsymbol{c}_l|^2 \lambda_k/(\lambda_k - \lambda_l)^2$.

Построим на основе свойства 2 асимптотический доверительный интервал (см. § 1 гл. 11) для λ_1 в примере 1. Так как свойство 2 равносильно сходимости $\sqrt{n/2}\,(\widehat{\lambda}_l/\lambda_l - 1) \xrightarrow{d} \xi/(\lambda_l\sqrt{2}) \sim \mathcal{N}(0,1)$, то при достаточно больших n приближенно с вероятностью $1-\alpha$

выполняется неравенство

$$\widehat{\lambda}_l \Big/ \Big(1 + x_{1-\alpha/2}\sqrt{2/n}\Big) < \lambda_l < \widehat{\lambda}_l \Big/ \Big(1 - x_{1-\alpha/2}\sqrt{2/n}\Big),$$

где $x_{1-\alpha/2}$ — квантиль уровня $(1-\alpha/2)$ закона $\mathcal{N}(0,1)$ (см. § 3 гл. 7). Для $\alpha = 5\%$ по таблице Т2 находим, что $x_{1-\alpha/2} = 1{,}96$. При $n = 43$ получаем 95%-ный доверительный интервал для λ_1 от 2,88 до 7,1.

§ 4. ЛИНЕЙНОЕ ШКАЛИРОВАНИЕ

Предположим, что исходной информацией об n объектах служат экспертные данные о *различиях* между ними, выраженных числами d_{ij} (или о *сходствах* a_{ij}).*)

Хотелось бы представить объекты в виде точек некоторого координатного пространства $\mathbb{R}^m$ невысокой размерности (обычно $m = 2$ или 3). При этом *различия* d_{ij} между объектами должны быть *как можно точнее* переданы *евклидовыми расстояниями* δ_{ij} между точками в $\mathbb{R}^m$.

Каждый объект характеризуется координатами соответствующей точки по осям (*шкалам*) построенного пространства. Сами шкалы интерпретируются как новые признаки — те латентные или скрытые факторы, которыми неосознанно руководствуются эксперты, оценивая степень различия объектов. Процесс представления объектов наборами координат в некоторых осях называют *шкалированием*.

Линейное шкалирование Торгерсона.** Метод полезно связать с выделением главных компонент, для чего удобно начать с нетипичного для приложений случая, когда объекты уже представлены точками $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{im})$ в $\mathbb{R}^m$, $i = 1, \ldots, n$, а различия d_{ij} являются евклидовыми расстояниями между ними:

$$d_{ij}^2 = \delta_{ij}^2 = \sum_{l=1}^{m} (x_{il} - x_{jl})^2. \tag{6}$$

Введем матрицу $\boldsymbol{X} = \|x_{il}\|_{n\times m}$. Допустим, что ее элементы x_{il} *центрированы по столбцам* (см. Д1 в § 1). Зададим величины g_{ij}, $i, j = 1, \ldots, n$, формулами

$$g_{ij} = -\frac{1}{2}\left(\delta_{ij}^2 - \delta_{\cdot j}^2 - \delta_{i\cdot}^2 + \delta_{\cdot\cdot}^2\right), \tag{7}$$

где $\delta_{\cdot j}^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n}\delta_{ij}^2$, $\quad \delta_{i\cdot}^2 = \dfrac{1}{n}\sum\limits_{j=1}^{n}\delta_{ij}^2$, $\quad \delta_{\cdot\cdot}^2 = \dfrac{1}{n^2}\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n}\delta_{ij}^2$.

Процедура (7) перехода от матрицы $\boldsymbol{\Delta} = \|\delta_{ij}\|_{n\times n}$ к матрице $\boldsymbol{G} = \|g_{ij}\|_{n\times n}$ называется *двойным центрированием*: нетрудно

*) Сходства можно преобразовать в различия, например, по формулам $d_{ij} = 1/(1 + a_{ij})$ или $d_{ij} = 1 - a_{ij}$ при $0 \leqslant a_{ij} \leqslant 1$.

**) У. Торгерсон предложил этот метод в 1952 г., опираясь на работы Т. Юнга и А. Хаусхолдера (1938, 1941). Его также называют *методом главных проекций*.

убедиться, что средние значения каждой строки и каждого столбца матрицы $\boldsymbol{G}$ равны 0.

Метод Торгерсона опирается на следующие две теоремы.

Теорема 1. Имеет место равенство $\boldsymbol{G} = \boldsymbol{X}\boldsymbol{X}^T$ (т. е. $\boldsymbol{G}$ совпадает с матрицей скалярных произведений, встречавшейся ранее в § 1).

Доказательство (см. также [25, с. 79]). Надо проверить, что

$$g_{ij} = \sum_{l=1}^{m} x_{il}x_{jl}, \quad i,j = 1,\ldots,n. \tag{8}$$

Раскрывая скобки в правой части равенства (6), получим формулу

$$\delta_{ij} = \sum_{l=1}^{m} x_{il}^2 + \sum_{l=1}^{m} x_{jl}^2 - 2\sum_{l=1}^{m} x_{il}x_{jl}. \tag{9}$$

Далее, для любого фиксированного j в силу теоремы Гюйгенса (см. решение задачи 3 гл. 16) имеем представление

$$\frac{1}{n}\sum_{i=1}^{n}(x_{il} - x_{jl})^2 = \frac{1}{n}\sum_{i=1}^{n}(x_{il} - 0)^2 + (x_{jl} - 0)^2.$$

(Здесь $m_i = 1/n$, $m = 1$, $a = x_{jl}$ и $c = \overline{x}_l = 0$ с учетом допущения Д1 из § 1.) Суммируя по $l = 1,\ldots,m$, приходим к равенству

$$\delta_{\cdot j}^2 = \sum_{l=1}^{m} x_{\cdot l}^2 + \sum_{l=1}^{m} x_{jl}^2, \quad \text{где } x_{\cdot l}^2 = \frac{1}{n}\sum_{i=1}^{n} x_{il}^2. \tag{10}$$

Точно так же, как формула (10), выводится соотношение

$$\delta_{i\cdot}^2 = \sum_{l=1}^{m} x_{\cdot l}^2 + \sum_{l=1}^{m} x_{il}^2. \tag{11}$$

Припишем каждой точке (строке) $\boldsymbol{x}_i = (x_{i1},\ldots,x_{im})$, $i = 1,\ldots,n$, массу 1. По допущению Д1 их центром масс будет начало координат $\mathbf{0}$. Согласно теореме о межточечных расстояниях (см. решение задачи 5 гл. 16) запишем:

$$I_{\mathbf{0}} = \sum_{i=1}^{n} |\boldsymbol{x}_i|^2 = \frac{1}{n}\sum_{i<j} |\boldsymbol{x}_i - \boldsymbol{x}_j|^2 = \frac{1}{2n}\sum_{i=1}^{n}\sum_{j=1}^{n} |\boldsymbol{x}_i - \boldsymbol{x}_j|^2,$$

где $I_{\mathbf{0}}$ — момент инерции точек относительно $\mathbf{0}$. Последнее соотношение эквивалентно равенству

$$\delta_{\cdot\cdot}^2 = \frac{2}{n}\, I_{\mathbf{0}} = \frac{2}{n}\sum_{i=1}^{n}\sum_{l=1}^{m} x_{il}^2 = 2\sum_{l=1}^{m} x_{\cdot l}^2. \tag{12}$$

Подставляя соотношения (9)–(12) в формулу (7), получаем равенство (8). ■

Теорема 2 (см. [8, с. 189])**.** *Отличные от нуля* собственные значения μ_l у матрицы скалярных произведений $\boldsymbol{G} = \boldsymbol{X}\boldsymbol{X}^T = \|g_{ij}\|_{n\times n}$ и матрицы рассеяния $\boldsymbol{W} = \boldsymbol{X}^T\boldsymbol{X} = \|w_{ij}\|_{m\times m}$ одинаковы. При

этом соответствующие собственные векторы $\boldsymbol{y}_l$ и $\boldsymbol{z}_l$ ($\boldsymbol{G}\boldsymbol{y}_l = \mu_l \boldsymbol{y}_l$, $\boldsymbol{W}\boldsymbol{z}_l = \mu_l \boldsymbol{z}_l$) связаны простым соотношением

$$\sqrt{\mu_l}\,\boldsymbol{y}_l = \boldsymbol{X}\boldsymbol{z}_l. \tag{13}$$

Поскольку выборочная ковариационная матрица $\widehat{\boldsymbol{\Sigma}}$ и матрица рассеяния $\boldsymbol{W}$ отличаются лишь постоянным множителем: $\widehat{\boldsymbol{\Sigma}} = \dfrac{1}{n}\boldsymbol{W}$, $\boldsymbol{z}_l$ совпадают с собственными векторами $\boldsymbol{c}_l$ матрицы $\widehat{\boldsymbol{\Sigma}}$, а $\mu_l = n\lambda_l$.

Заметим, что справа в равенстве (13) стоит вектор координат проекций точек на l-ю главную компоненту $\boldsymbol{z}_l$. Таким образом, *не зная самих координат* x_{il}, можно на основе различий $d_{ij} = \delta_{ij}$ восстановить конфигурацию точек в пространстве $\mathbb{R}^m$ с помощью следующей **процедуры шкалирования**:

1. Двойным центрированием матрицы различий $\boldsymbol{\Delta}$ вычислить матрицу $\boldsymbol{G}$.
2. Степенным методом найти для матрицы $\boldsymbol{G}$ первые m отличных от нуля (положительных) собственных значений μ_l и соответствующих им собственных векторов $\boldsymbol{y}_l$.
3. Из (13) определить координаты точек в новом базисе пространства $\mathbb{R}^m$, состоящем из главных осей $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_m$.

Замечание 6. Сама матрица $\boldsymbol{X}$ *не восстанавливается* по матрице $\boldsymbol{G}$ однозначно. Действительно, пусть координаты точек $\boldsymbol{x}_i$ преобразуются в новые координаты по формуле $\tilde{\boldsymbol{x}}_i = \boldsymbol{C}\boldsymbol{x}_i$, где $\boldsymbol{C}$ — любая *ортогональная* матрица (см. П10). Так как векторы $\boldsymbol{x}_i$ — это столбцы матрицы $\boldsymbol{X}^T$, то $\tilde{\boldsymbol{X}}^T = \boldsymbol{C}\boldsymbol{X}^T$ или $\tilde{\boldsymbol{X}} = \boldsymbol{X}\boldsymbol{C}^T$. Тогда

$$\tilde{\boldsymbol{G}} = \tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^T = (\boldsymbol{X}\boldsymbol{C}^T)(\boldsymbol{C}\boldsymbol{X}^T) = \boldsymbol{X}(\boldsymbol{C}^T\boldsymbol{C})\boldsymbol{X}^T = \boldsymbol{X}\boldsymbol{X}^T = \boldsymbol{G},$$

т. е. матрица $\boldsymbol{G}$ сохраняется при ортогональном преобразовании координат.[*)]

С другой стороны, матрица рассеяния $\boldsymbol{W} = \boldsymbol{X}^T\boldsymbol{X}$ меняется:

$$\tilde{\boldsymbol{W}} = \tilde{\boldsymbol{X}}^T\tilde{\boldsymbol{X}} = (\boldsymbol{C}\boldsymbol{X}^T)(\boldsymbol{X}\boldsymbol{C}^T) = \boldsymbol{C}(\boldsymbol{X}^T\boldsymbol{X})\boldsymbol{C}^T = \boldsymbol{C}\boldsymbol{W}\boldsymbol{C}^T.$$

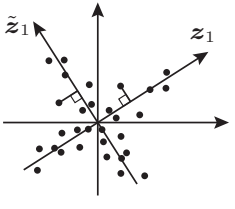


Рис. 8

Новая матрица $\tilde{\boldsymbol{W}}$ подобна $\boldsymbol{W}$ (см. П10). Следовательно, их собственные значения μ_l совпадают. Новые собственные векторы $\tilde{\boldsymbol{z}}_l$ связаны с $\boldsymbol{z}_l$ тем же преобразованием, что и координаты точек: $\tilde{\boldsymbol{z}}_l = \boldsymbol{C}\boldsymbol{z}_l$ (рис. 8). В самом деле,

$$\begin{aligned}\tilde{\boldsymbol{W}}\tilde{\boldsymbol{z}}_l &= (\boldsymbol{C}\boldsymbol{W}\boldsymbol{C}^T)(\boldsymbol{C}\boldsymbol{z}_l) = \boldsymbol{C}(\boldsymbol{W}\boldsymbol{z}_l) = \\ &= \boldsymbol{C}(\mu_l\boldsymbol{z}_l) = \mu_l(\boldsymbol{C}\boldsymbol{z}_l) = \mu_l\tilde{\boldsymbol{z}}_l.\end{aligned}$$

При этом правая часть соотношения (13) сохраняется:

$$\tilde{\boldsymbol{X}}\tilde{\boldsymbol{z}}_l = (\boldsymbol{X}\boldsymbol{C}^T)(\boldsymbol{C}\boldsymbol{z}_l) = \boldsymbol{X}\boldsymbol{z}_l.$$

Подход Торгерсона состоит в применении описанной выше процедуры шкалирования к *произвольной симметричной* матрице

[*)] По этой причине исследователи иногда производят дополнительный (субъективный) поворот системы координат, чтобы содержательно интерпретировать признаки, соответствующие повернутым осям.

различий $\boldsymbol{D} = \|d_{ij}\|_{n\times n}$. Размерность m координатного пространства задается исследователем. Обозначим через $\boldsymbol{G}^*$ результат двойного центрирования матрицы $\boldsymbol{D}$.*)

Замечание 7. В случае, когда различия d_{ij} не являются евклидовыми расстояниями, матрица $\boldsymbol{G}^*$ может не быть неотрицательно определенной. Прием, который обычно используется для преодоления этого, заключается в переходе к модели с так называемой *аддитивной константой*: $d'_{ij} = d_{ij} + a$, где $a > 0$. Наименьшее значение a, обеспечивающее выполнение *неравенства треугольника* для всевозможных троек d'_{ij}, очевидно, определяется по формуле

$$\widehat{a} = \max_{1\leqslant i,j,k\leqslant n} \max\{d_{ik} - d_{ij} - d_{jk}, 0\}. \tag{14}$$

Но из выполнения неравенства треугольника еще не следует неотрицательная определенность матрицы $\boldsymbol{G}^*$, вычисленной на основе d'_{ij} (см. пример 3 ниже). Покажем, что, увеличивая a, можно добиться того, чтобы d'_{ij} оказались евклидовыми расстояниями для некоторой конфигурации точек в $\mathbb{R}^m$, если размерность m также достаточно велика. Действительно, для любой размерности k можно указать $k+1$ точку в $\mathbb{R}^k$, которые находятся друг от друга на одном и том же расстоянии h (они называются *вершинами правильного симплекса*). Таковыми будут, скажем, $\boldsymbol{u}_0 = (0,\ldots,0)$ и $\boldsymbol{u}_i = (u_{i1},\ldots,u_{ik})$ $(i = 1,\ldots,k)$, где

$$u_{ij} = \begin{cases} h(\sqrt{k+1} + k - 1)/(k\sqrt{2}), & \text{если } i = j, \\ h(\sqrt{k+1} - 1)/(k\sqrt{2}), & \text{если } i \neq j, \end{cases} \tag{15}$$

(рис. 9 для $k = 2$). При очень большом значении a объекты будут располагаться вблизи вершин правильного симплекса с ребрами длины a в пространстве $\mathbb{R}^{n-1}$, где n — число объектов.

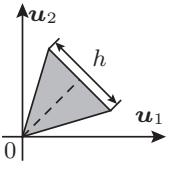


Рис. 9

С другой стороны, при увеличении a все более нивелируются *относительные* отличия между d'_{ij} по сравнению с отличиями между исходными d_{ij}. Кроме того, увеличивается число значимых по величине положительных μ_l, т. е. растет размерность m пространства, в котором располагается конфигурация точек, достаточно точно воспроизводящая d'_{ij}. Поэтому разумнее не добиваться неотрицательной определенности матрицы $\boldsymbol{G}^*$, а ограничиться тем, чтобы несколько ее наибольших собственных значений μ_l были положительными, а все остальные собственные значения — малыми по модулю.

Пример 3. Проиллюстрируем вышесказанное на примере анализа корреляционной матрицы, изображенной на рис. 3.

*) При умножении всех d_{ij} на $c > 0$ элементы и собственные значения $\boldsymbol{G}^*$ умножаются на c^2, собственные векторы остаются прежними, а задаваемые формулой (13) координаты объектов в построенных шкалах умножается на c. Из-за этого шкалирование Торгерсона называют *линейным*.

Положим $d_{ij} = 100 \times (1 - \widehat{\rho}_{ij})$. Вычисленные степенным методом из § 3 собственные значения матрицы $\boldsymbol{G}^*$ и приходящиеся на них проценты от полной дисперсии (следа матрицы) приведены в следующей таблице:

Номера компонент	1	2	3	4	5	6
Собственные значения	3820	1321	317	−187	12,5	0,0
Проценты от следа	72,3	25,0	6,0	−3,5	0,2	0,0
Накопленные проценты	72,3	97,3	103,3	99,8	100,0	100,0

Поскольку $\mu_4 = -187$, матрица $\boldsymbol{G}^*$ не является неотрицательно определенной. На рис. 10 показана конфигурация точек на плоскости, возникающая при выборе $m = 2$. Таблица на рис. 11 показывает, насколько точно евклидовы расстояния между точками плоскости (они указаны в скобках) соответствуют d_{ij}.

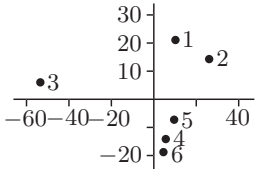


Рис. 10

	1	2	3	4	5	6
1	0	4 (18)	65 (66)	39 (36)	28 (29)	41 (41)
2		0	80 (80)	34 (35)	26 (27)	41 (39)
3			0	63 (62)	65 (65)	65 (64)
4				0	11 (8)	24 (5)
5					0	21 (13)
6						0

Рис. 11

В целом точность хорошая за исключением расстояния 1–2 и расстояний между точкой 6 и точками 4, 5 (первое слишком большое, вторые — слишком малые). Для объяснения этого факта заметим, что для $d_{23} = 80$, $d_{21} = d_{12} = 4$ и $d_{13} = 65$ не выполняется неравенство треугольника: $80 > 4 + 65$. Чтобы его обеспечить, пришлось увеличить расстояние между точками 1 и 2 до 18.

В свою очередь, точки 4, 5 и 6 должны находиться на расстоянии примерно 65 от точки 3, при этом каждая из них должна быть (почти) равноудалена от точек 1 и 2. Это означает, что точки 4, 5 и 6 обязаны располагаться рядом, а именно — вблизи точки пересечения окружности с центром в 3 и радиусом 65 и серединного перпендикуляра к отрезку с концами в точках 1 и 2. Получаем противоречие с тем, что $d_{46} = 24$ и $d_{56} = 21$.

По формуле (14) на компьютере была подсчитана наименьшая аддитивная константа $\widehat{a} = 11$, обеспечивающая выполнение неравенства треугольника для всех троек $d'_{ij} = d_{ij} + \widehat{a}$. Таким d'_{ij} соответствует матрица $\boldsymbol{G}^*$ со следующим спектром:

Номера компонент	1	2	3	4	5	6
Собственные значения	4958	2094	663	190	−93,3	0,0
Проценты от следа	63,5	26,8	8,5	2,4	−1,2	0,0
Накопленные проценты	63,5	90,3	98,8	101,2	100,0	100,0

Здесь уже положительны μ_1—μ_4, а доля следа, приходящаяся на отрицательное собственное значение, уменьшилась с 3,5% до 1,2%.

Как показывает рис. 12, точки только слегка отдалились от начала координат, конфигурация качественно не изменилась. Однако на первые две оси теперь приходится всего 90,3%, а не 97,3% следа матрицы $\boldsymbol{G}^*$, как раньше (дисперсия «размазывается» по спектру). Это проявляется в плохом воспроизведении d'_{ij} для признаков 4, 5 и 6 (см. таблицу на рис. 13).

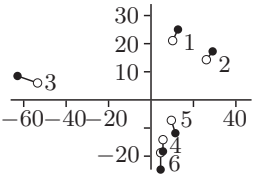
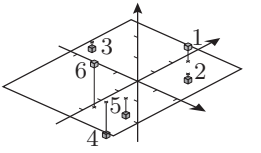


Рис. 12

	1	2	3	4	5	6
1	0	15 (18)	76 (76)	50 (45)	39 (37)	52 (50)
2		0	91 (91)	45 (44)	37 (34)	52 (48)
3			0	74 (72)	76 (75)	76 (73)
4				0	22 (10)	35 (5)
5					0	32 (14)
6						0

Рис. 13

Преодолеть этот недостаток можно за счет увеличения размерности m до 3. Как демонстрирует рис. 14, третья координата как раз и отвечает за разделение признака 6 и признаков 4 и 5.



1) (13,26,5)
2) (29,18,−1)
3) (−62,9,−1)
4) (5,−19,−15)
5) (11,−11,−6)
6) (4,−24,19)

Рис. 14

Соответствие пространственных расстояний и d'_{ij} (за исключением пары 4–5) довольно хорошее (рис. 15).

	1	2	3	4	5	6
1	0	15 (19)	76 (76)	50 (50)	39 (38)	52 (52)
2		0	91 (91)	45 (46)	37 (34)	52 (52)
3			0	74 (74)	76 (75)	76 (76)
4				0	22 (13)	35 (35)
5					0	32 (29)
6						0

Рис. 15

Чтобы добиться неотрицательной определенности матрицы $\boldsymbol{G}^*$, понадобится увеличить аддитивную константу a до 17. Спектр выглядит так:

Номера компонент	1	2	3	4	5	6
Собственные значения	5631	2568	904	335	7,3	0,0
Проценты от следа	59,6	27,2	9,6	3,6	0,1	0,0
Накопленные проценты	59,6	86,8	96,4	99,9	100,0	100,0

При $m = 4$ все расстояния между точками (при округлении до ближайшего целого числа) совпадают с $d'_{ij} = d_{ij} + 17$.

Пример 4. ***Анализ межотраслевых связей*** [78, с. 47]**.** Данные были взяты из книги [34] известного американского экономиста А. Картер. Они представляют собой подсчитанные на основе межотраслевого баланса американской экономики *коэффициенты прямых затрат* p_{ij}, равные отношению затрат i-й отрасли на закупку продукции у j-й отрасли к общей величине затрат

i-й отрасли. Симметризация r_{ij} дает в качестве мер близости $a_{ij} = r_{ij} + r_{ji}$. Переход от a_{ij} к различиям d_{ij} осуществлялся по формуле $d_{ij} = (\max a_{ij} - a_{ij})/(\max a_{ij} - \min a_{ij})$.

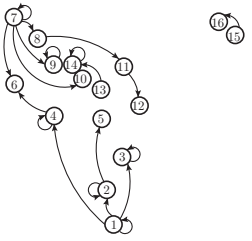


Рис. 16

На рис. 16 представлена структура межотраслевых связей, полученная методом многомерного шкалирования (чтобы не загромождать рисунок, отмечены только отрасли, которые производят значительные закупки или поставки). Сильные связи обозначены стрелками.

Основу структуры связей составляет обеспечение первичными ресурсами ресурсоемких отраслей. В первую очередь это поставки аграрного, лесного и рыбного сектора (1) в пищевую (2), табачную (3) и текстильную (4) отрасли. Пищевая отрасль (2) снабжает обувное производство (5). Продукция текстильной промышленности (4) нужна при переработке металлолома (6).

Вторую основную ветвь составляют поставки чугуна и стали (7) в группу металлопотребляющих отраслей: металлообрабатывающую промышленность (8), автомобильную промышленность (9), железнодорожный транспорт и суда (10) и переработку металлолома (6). Металлообрабатывающую промышленность (8) отдает значительную часть своей продукции в строительство (11), которое в свою очередь делает вклад в сектор недвижимости и помещений (12). Еще две очевидные ветви — это связь добычи руд цветных металлов (13) с обработкой цветных металлов (14) и нефтедобывающей промышленности (15) с нефтеперерабатывающей (16).

§ 5. ШКАЛИРОВАНИЕ ИНДИВИДУАЛЬНЫХ РАЗЛИЧИЙ

Во многих исследованиях, особенно в социальных и поведенческих науках, данные получают от нескольких источников информации, например, от нескольких субъектов или при разных экспериментальных условиях. Недостатком метода Торгерсона является то, что входной информацией для него служит единственная матрица различий $\boldsymbol{D} = \|d_{ij}\|_{n\times n}$ (n — число объектов), полученная *осреднением* данных всех субъектов. При осреднении теряется информация, отражающая различие во мнениях субъектов, которая может оказаться весьма ценной для понимания сути изучаемого явления. Подход, учитывающий эти различия, называется *анализом точек зрения* или *шкалированием индивидуальных различий* ([78, с. 85]).

Рассмотрим часто применяемую при таком подходе

Модель взвешенных факторов

Обозначим через $\boldsymbol{X} = \|x_{il}\|_{n\times m}$ матрицу координат объектов (стимулов), где x_{il} — неизвестные параметры модели. Будем называть ее *групповой*. Предполагается, что у каждого субъекта s ($s = 1, \ldots, k$) существует своя, индивидуальная матрица координат

объектов $\boldsymbol{X}_s = \|x_{ils}\|_{n\times m}$, элементы которой связаны с элементами групповой матрицы $\boldsymbol{X}$ следующим образом:

$$x_{ils} = \sqrt{w_{ls}}\, x_{il}, \tag{16}$$

где w_{ls} — (неизвестный) *вес* координаты l для субъекта s. Величина $\sqrt{w_{ls}}$ представляет собой коэффициент растяжения по l-й оси координат объектов в пространстве субъекта s по отношению к групповым координатам x_{il} (см. пример 5 ниже). Таким образом, всего в модели имеется $(nm + mk)$ неизвестных параметров: nm групповых координат x_{il} и mk весов w_{ls}.

Иначе можно трактовать модель взвешенных факторов как определение субъектами расстояний между объектами не на основе обычной, а на основе *взвешенной евклидовой метрики*:

$$\delta_{ijs} = \left[\sum_{l=1}^{m} (x_{ils} - x_{jls})^2\right]^{1/2} = \left[\sum_{l=1}^{m} w_{ls}\,(x_{il} - x_{jl})^2\right]^{1/2}.$$

Пример 5. ***Зрительное восприятие букв*** **[78, с. 91].** Испытуемые оценивали степень сходства между 19 печатными буквами русского языка. На рис. 17, *а* изображена полученная конфигурация в групповом пространстве, а рис. 17, *б* дает представление о расположении экспертов (субъектов) на плоскости весовых коэффициентов.

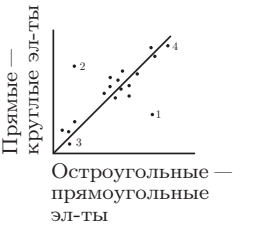
Эксперты различаются по тому, какой вес они придают каждой из осей. Одна часть экспертов (они находятся ниже диагонали на рис. 17, *б*) придает больший вес первой оси, противопоставляющей буквы с остроугольными элементами буквам с прямоугольными элементами. Другая — второй оси, отвечающей за преобладание круглых элементов букв над прямоугольными. На рис. 17, *б* такие эксперты располагаются выше диагонали.

При этом для разных экспертов степень различия весов неодинакова. Большинство экспертов группируется вблизи диагонали. Для них различие между весами осей несущественно.

С другой стороны, для эксперта с номером 1 веса сильно отличаются: первая ось для него намного важнее второй. Для эксперта с номером 2 картина прямо противоположная. На рис. 18 представлены субъективные пространства восприятия для экспертов 1 (рис. 18, *а*) и 2 (рис. 18, *б*), получаемые растяжением или сжатием групповой конфигурации с коэффициентами, равными квадратным корням из соответствующих весов.

Кроме различия по отношению двух весов, эксперты отличаются между собой по сумме обоих весов. Непосредственный анализ исходных данных показывает, что экспертам, которые при сравнении букв обоим факторам приписывают малые веса, соответствует

а)



б)

Рис. 17

а)

б)

Рис. 18

большое количество ответов «похожи», а экспертам, которые приписывают обоим факторам большие веса, соответствует наименьшее число ответов «похожи». Например, эксперт с номером 3 при сравнении 153 пар букв ответил «похожи» 117 раз, а эксперт 4 — только 3 раза.

Для каждого субъекта s введем *диагональную* $(m \times m)$-матрицу $\boldsymbol{W}_s$ с элементами $\sqrt{w_{ls}}$ ($l = 1,\ldots,m$, s фиксировано) и запишем соотношения (16) в матричной форме: $\boldsymbol{X}_s = \boldsymbol{X}\boldsymbol{W}_s$. Тогда матрица скалярных произведений субъекта s представляется в виде

$$\boldsymbol{G}_s = \|g_{ijs}\|_{n\times n} = \boldsymbol{X}_s\boldsymbol{X}_s^T = \boldsymbol{X}\boldsymbol{W}_s^2\boldsymbol{X}^T \ \text{(т. е. } g_{ijs} = \sum_{l=1}^{m} x_{il}x_{jl}w_{ls}).$$

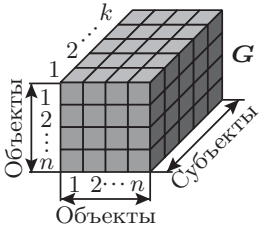


Рис. 19

Все вместе матрицы $\boldsymbol{G}_s$, $s = 1,\ldots,k$, образуют трехмерный массив $\boldsymbol{G}$ (рис. 19).

С другой стороны, для каждого субъекта s можно двойным центрированием (7) его матрицы различий $\boldsymbol{D}_s = \|d_{ijs}\|_{n\times n}$ вычислить $\boldsymbol{G}_s^* = \|g_{ijs}^*\|_{n\times n}$. Объединяя все матрицы $\boldsymbol{G}_s^*$, получим массив $\boldsymbol{G}^*$.

Напомним (см. формулу (4) в § 1), что одним из свойств оптимальности главных компонент является то, что матрица скалярных произведений координат объектов *искажается в наименьшей степени* при проецировании на подпространство, порожденное несколькими первыми компонентами. Поэтому в качестве *меры соответствия* параметров модели x_{il} и w_{ls} данным $\boldsymbol{D}_s$ $(s = 1,\ldots,k)$ возьмем

$$F = |\boldsymbol{G}^* - \boldsymbol{G}|^2 = \sum_s |\boldsymbol{G}_s^* - \boldsymbol{G}_s|^2 = \sum_{i,j,s} (g_{ijs}^* - \sum_l x_{il}x_{jl}w_{ls})^2.$$

Опишем алгоритм численной минимизации F, опирающийся на разработанный в 1970 г. Дж. Кэрролом и Дж. Чанг

Метод канонической декомпозиции[*)] (см. [78, с. 96]).

Наряду с групповой матрицей $\boldsymbol{X}$ введем матрицы $\boldsymbol{Y} = \|y_{jl}\|_{n\times m}$ и $\boldsymbol{Z} = \|z_{ls}\|_{m\times k}$, где $y_{jl} = x_{jl}$ и $z_{ls} = w_{ls}$. Игнорируя совпадение y_{jl} с x_{jl}, будем считать $\boldsymbol{X}$, $\boldsymbol{Y}$ и $\boldsymbol{Z}$ наборами независимых переменных. В новых обозначениях минимизируемая функция F запишется так:

$$F = F(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z}) = \sum_{i,j,s} (g_{ijs}^* - \sum_l x_{il}y_{jl}z_{ls})^2. \qquad (17)$$

Процедура минимизации включает **следующие пункты**.

1. В качестве начальных значений для $\boldsymbol{X}$ и $\boldsymbol{Y}$ можно использовать координаты $\widehat{\boldsymbol{X}}_{med}$, получаемые применением метода Торгерсона к матрице

   $$\widehat{\boldsymbol{D}}_{med} = MED\{\boldsymbol{D}_1,\ldots,\boldsymbol{D}_k\}.$$

[*)] Минимизация нелинейной функции F представляется как последовательное решение ряда линейных систем.

(Медиана предпочтительнее среднего арифметического: она уменьшает влияние на начальную оценку групповых координат тех субъектов, чьи d_{ijs} сильно отличаются от других.)

2. Фиксируем $\boldsymbol{X}$, $\boldsymbol{Y}$ и минимизируем F сначала по $\boldsymbol{Z}$. Для этого положим $u = n(i-1)+j$ (так что $u = 1,\ldots,n^2$). Определяя $a_{ul} = x_{il}y_{jl}$ и $b_{us} = g^*_{ijs}$, перепишем функцию (17) в виде

$$F = \sum_{u,s}(b_{us} - \sum_{l} a_{ul}\, z_{ls})^2 = |\boldsymbol{B} - \boldsymbol{A}\boldsymbol{Z}|^2,$$

где $\boldsymbol{A} = \|a_{ul}\|_{n^2\times m}$, $\boldsymbol{B} = \|b_{us}\|_{n^2\times k}$. Согласно *методу наименьших квадратов* минимизирующие правую часть веса $\widehat{\boldsymbol{Z}}$ могут быть найдены по формуле $\widehat{\boldsymbol{Z}} = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{B}$.*)

Доказательство. Пусть $\boldsymbol{b}_s$ и $\boldsymbol{z}_s$ $(s = 1,\ldots,k)$ обозначают столбцы матриц $\boldsymbol{B}$ и $\boldsymbol{Z}$ соответственно. Тогда

$$|\boldsymbol{B} - \boldsymbol{A}\boldsymbol{Z}|^2 = \sum_{s}|\boldsymbol{b}_s - \boldsymbol{A}\boldsymbol{z}_s|^2.$$

Ввиду формулы (7) гл. 21, s-е слагаемое достигает минимума при $\widehat{\boldsymbol{z}}_s = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{b}_s$. Остается объединить все векторы $\boldsymbol{b}_s$ в матрицу $\boldsymbol{B}$. ■

3. Попытаемся улучшить оценку для $\boldsymbol{Y}$ при фиксированных $\boldsymbol{X}$ и $\widehat{\boldsymbol{Z}}$. Положим $v = n(i-1)+s$ (так что $v = 1,\ldots,nk$). Определяя $p_{lv} = x_{il}\widehat{z}_{ls}$ и $q_{jv} = g^*_{ijs}$, минимизируем

$$F = \sum_{j,v}(q_{jv} - \sum_{l} y_{jl}\, p_{lv})^2 = |\boldsymbol{Q} - \boldsymbol{Y}\boldsymbol{P}|^2,$$

где $\boldsymbol{P} = \|p_{lv}\|_{m\times nk}$, $\boldsymbol{Q} = \|q_{jv}\|_{n\times nk}$. Оценка наименьших квадратов для $\boldsymbol{Y}$ имеет вид $\widehat{\boldsymbol{Y}} = \boldsymbol{Q}\boldsymbol{P}^T(\boldsymbol{P}\boldsymbol{P}^T)^{-1}$.

Вопрос 1. Почему верна эта формула?

4. Аналогично, фиксируя $\widehat{\boldsymbol{Y}}$ и $\widehat{\boldsymbol{Z}}$, попробуем улучшить оценку для $\boldsymbol{X}$. Вводя обозначения $t = n(j-1)+s$ (так что $t = 1,\ldots,nk$), $\boldsymbol{R} = \|r_{lt}\|_{m\times nk}$, $\boldsymbol{H} = \|h_{it}\|_{n\times nk}$, где $r_{lt} = \widehat{y}_{jl}\widehat{z}_{ls}$ и $h_{it} = g^*_{ijs}$, получим

$$F = \sum_{i,t}(h_{it} - \sum_{l} x_{il}\, r_{lt})^2 = |\boldsymbol{H} - \boldsymbol{X}\boldsymbol{R}|^2,$$

что дает оценку $\widehat{\boldsymbol{X}} = \boldsymbol{H}\boldsymbol{R}^T(\boldsymbol{R}\boldsymbol{R}^T)^{-1}$.

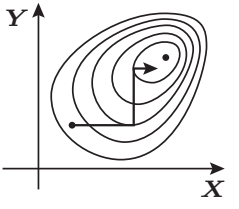


Рис. 20

Пункты 2–4 составляют *один шаг* итерационной процедуры. С каждым шагом функция F уменьшается. При этом происходит точная минимизация по $\boldsymbol{X}$ ($\boldsymbol{Y}$ или $\boldsymbol{Z}$) при фиксированных значениях остальных наборов переменных.**) Процесс минимизации завершается, когда уменьшение F станет несущественным.

*) Вместо вычисления $(\boldsymbol{A}^T\boldsymbol{A})^{-1}$ быстрее решить *методом Холецкого* (см. П10) систему линейных уравнений $(\boldsymbol{A}^T\boldsymbol{A})\widehat{\boldsymbol{Z}} = \boldsymbol{A}^T\boldsymbol{B}$.

**) По существу, осуществляется *покоординатный спуск*, только в качестве координат выступают не числа, а матрицы (рис. 20).

Так как переменные x_{il} и x_{jl} были разделены, итоговые матрицы $\boldsymbol{X}$ и $\boldsymbol{Y}$ могут отличаться одна от другой. Для преодоления этого несоответствия с содержательным смыслом x_{il} после завершения поиска минимума полагают $\boldsymbol{Y} = \boldsymbol{X}$ и еще раз пересчитывают оценку матрицы весов $\widehat{\boldsymbol{Z}}$.

Можно было бы приравнивать матрицы на каждой итерации. В этом случае пришлось бы пересчитывать только две матрицы вместо трех. Однако свойства такого алгоритма не изучены, и нельзя утверждать, что он сходится хотя бы к локальному минимуму.

Исследования рассмотренного алгоритма канонической декомпозиции с помощью моделирования по методу Монте-Карло показали его высокую работоспособность ([78, с. 97]).

Пример 6. ***Восприятие на слух болгарских согласных*** [78, с. 118]**.** В болгарском Институте обучения иностранных студентов болгарскому языку Гергановым Е. Н., Терехиной А. Ю. и Фрумкиной Р. М. проводились исследования различий восприятия согласных звуков между болгарами и иностранными студентами. Были выбраны 4 группы испытуемых по 50 человек: болгар, испанцев, вьетнамцев и арабов. Пары из 21 согласных звуков в сочетании со звуком Ъ (БЪ—ВЪ) были записаны на магнитную пленку — всего 210 пар. Испытуемые должны были внимательно слушать и отмечать, похожи ли два звука в паре. В результате для каждого испытуемого была получена матрица сходств. Совместный анализ данных всех испытуемых позволил построить *шестимерное* стимульное пространство, представленное на рис. 21.

Рис. 21

Первая ось (рис. 21, *а*) делит все согласные на сибилянты (Ч, Ш, Ж, ДЖ, Ц, С, ДЗ, З) и остальные. Вторая ось разделяет сибилянты по частоте: низкочастотные — шипящие (Ч, Ш, Ж, ДЖ) и высокочастотные — свистящие (З, С, Ц, ДЗ). Третья ось (рис. 21, *б*) выделяет среди всех согласных группу сонорных (Н, М, Р, Л). Четвертая ось отвечает за место образования звуков: от губных (Ф, В, П, Б, М) к зубным (С, З, Ц, ДЗ), далее — к переднеязычным (Т, Д) и, наконец, — к заднеязычным (К, Г). Пятая ось (рис. 21, *в*) выделяет дрожащий звук Р. Интересной является шестая ось, она соответствует признаку «глухость—звонкость», но «количество звонкости» определяется как бы в относительной степени. Из рис. 21, *в* хорошо видно, что противопоставление глухости—звонкости реализуется *попарно*: З располагается выше, чем С; Ж — выше, чем Ш и т. д.

На рис. 22 представлены профили весов, с которыми учитывали эти 6 признаков 4 группы испытуемых — носителей разных языков. Все испытуемые, особенно группа испанцев, при сравнении звуков наибольшее значение придавали признаку «сибилянтность». Для испанцев характерен также большой вес для признака «сонорность». Наиболее низкие веса у арабов, они в наименьшей степени учитывали признаки «шипящие—свистящие», «сонорность»

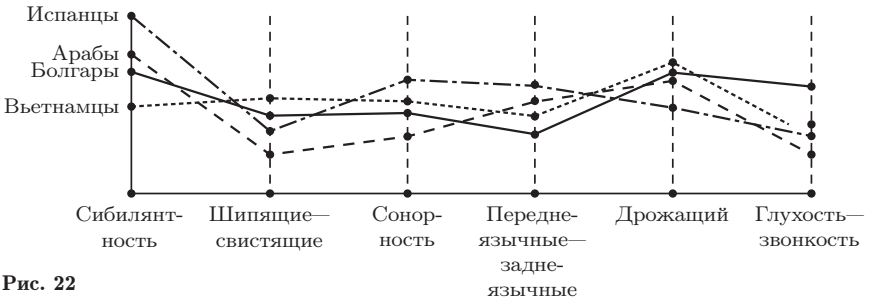


Рис. 22

и «глухость—звонкость». Болгары, в меньшей степени, чем все остальные, учитывали признак «переднеязычные—заднеязычные», а ориентировались при сравнении в большой степени на признак «глухость—звонкость».

По-видимому, различие в восприятии звуков носителями разных языков объясняется различием фонетических систем, на которые опираются их родные языки.

§ 6. НЕЛИНЕЙНЫЕ МЕТОДЫ ПОНИЖЕНИЯ РАЗМЕРНОСТИ

Круговое движение первее прямолинейного: оно проще и более совершенно.

Аристотель

Ортогональное проецирование на подпространство первых главных компонент, осуществляемое при линейном шкалировании Торгерсона, хорошо передает структуру данных, когда множество точек имеет большой разброс по одним направлениям и совсем небольшой — по другим. Если же «облако» точек не похоже на многомерный эллипсоид (его конфигурация существенно нелинейна), то точки, находящиеся на разных концах конфигурации, могут при ортогональном проецировании накладываться друг на друга.

Пример 7. ***Проецирование вершин многомерного симплекса*** **[78, с. 35].** На рис. 23, *а* изображены проекции на плоскость двух первых главных компонент 16 групп точек по 4 точки в группе. Каждая группа генерировалась с помощью датчика случайных чисел вблизи одной из 16 вершин правильного симплекса в $\mathbb{R}^{15}$ (см. формулу (15)). Точки из окрестности одной вершины обозначены одинаковыми буквами.

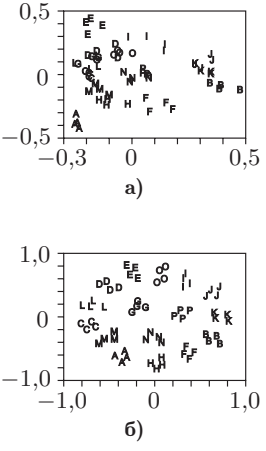


Рис. 23

На первую (из пятнадцати) компоненту приходится 8,3% следа ковариационной матрицы, на вторую — 7,6%, на обе — лишь 15,9%. Видим, что проецирование на плоскость двух первых главных компонент не позволяет сохранить разделение на группы.

Для анализа подобных сложных конфигураций применяются *нелинейные методы понижения размерности*. Они заключаются в минимизации некоторой функции F, выражающей *суммарное*

расхождение между заданными различиями объектов d_{ij} и расстояниями δ_{ij} между образами объектов в подпространстве небольшой размерности.

В мире не происходит ничего, в чем бы не был виден смысл какого-нибудь максимума или минимума.

Л. Эйлер

Наиболее простой из таких функций является

$$F_0 = \sum_{i<j} (\delta_{ij} - d_{ij})^2.$$

Широкое распространение получил **метод Сэммона**[*]:

$$F_1 = \frac{1}{C_1} \sum_{i<j} (\delta_{ij} - d_{ij})^2 / d_{ij}, \quad \text{где } C_1 = \sum_{i<j} d_{ij}, \tag{18}$$

который обладает свойством более точно передавать небольшие различия и менее точно — большие, так как при отображении больших расстояний допустимы бóльшие ошибки. (Деление на константу C_1 нужно для инвариантности F_1 к изменению масштаба величин d_{ij}.)

На рис. 23, *б* изображен результат применения метода Сэммона к данным примера 7. Значение F_1 удалось уменьшить с 0,44 до 0,13. Видим, что разделение точек на группы сохраняется на плоскости.

Для поиска минимума функции F прежде всего строится начальная конфигурация точек в пространстве небольшой размерности m. Это могут быть просто проекции в подпространство, образованное какими-либо из m координатных осей, или в подпространство m первых главных компонент, или, наконец, случайная конфигурация, в которой nm координат получены с помощью датчика случайных чисел.

Исходя из начальной конфигурации, отыскивается конфигурация точек, на которой достигается (локальный) минимум функции F. Для этого обычно применяется

Метод сопряженных градиентов (см. [6, с. 284]).

1. Определяется направление $\boldsymbol{p}_t$ в пространстве $\mathbb{R}^{nm}$ по формулам

   $$\boldsymbol{p}_1 = -\boldsymbol{g}_1, \quad \boldsymbol{p}_t = -\boldsymbol{g}_t + \beta_t \boldsymbol{p}_{t-1} \text{ при } t = 2,3,\ldots,$$

   где $\boldsymbol{g}_t = \left(\dfrac{\partial F}{\partial x_{11}}, \ldots, \dfrac{\partial F}{\partial x_{nm}} \right)$ обозначает градиент минимизируемой функции F, $\boldsymbol{p}_{t-1}$ — направление на предыдущем шаге, коэффициент $\beta_t = |\boldsymbol{g}_t|^2 / |\boldsymbol{g}_{t-1}|^2$.

   [В задаче 2 предлагается проверить, что координаты градиента функции F_1, заданной формулой (18), удовлетворяют соотношениям

   $$\frac{\partial F}{\partial x_{il}} = \frac{2}{C_1} \sum_{\substack{j=1 \\ j \neq i}}^{n} \left(\frac{1}{d_{ij}} - \frac{1}{\delta_{ij}} \right) (x_{il} - x_{jl}), \tag{19}$$

[*] Предложен Дж. Сэммоном в 1969 г., см. [78, с. 35].

где $i = 1, \ldots, n$; $l = 1, \ldots, m$. Считается, что все объекты различны ($d_{ij} \neq 0$), а в случае совпадения образов объектов в $\mathbb{R}^m$ ($\delta_{ij} = 0$) соответствующее слагаемое полагается равным некоторой константе $c \neq 0$, чтобы образы могли «разойтись» в процессе минимизации.]

2. Производится перемещение до точки минимума по выбранному направлению (например, с помощью рассматриваемого ниже алгоритма *золотого сечения*)

$$\boldsymbol{x}_{t+1} = \boldsymbol{x}_t + \alpha_t \boldsymbol{p}_t, \text{ где } F(\boldsymbol{x}_t + \alpha_t \boldsymbol{p}_t) = \min_{\alpha \geqslant 0} F(\boldsymbol{x}_t + \alpha \boldsymbol{p}_t).$$

3. Если $|\boldsymbol{g}_{t+1}| \leqslant \varepsilon$, где ε — достаточно малое число (скажем, $\varepsilon = 0{,}001$), то поиск заканчивается, иначе t увеличивается на 1, и происходит возвращение к пункту 1.

Впервые метод сопряженных градиентов был опубликован в качестве численного алгоритма для решения системы линейных уравнений $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ с положительно определенной матрицей $\boldsymbol{A}$. Эта задача равносильна минимизации по $\boldsymbol{x}$ квадратичной функции $F(\boldsymbol{x}) = \frac{1}{2}\,\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{x}$. Оказывается, что (при отсутствии погрешностей округления) метод сопряженных градиентов позволяет найти точку минимума $F(\boldsymbol{x})$ не более, чем за k итераций, где k — размерность $\boldsymbol{x}$ (см. [18, с. 200]). При этом направления одномерной минимизации $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_k$ являются *взаимно сопряженными* относительно матрицы $\boldsymbol{A}$: $\boldsymbol{p}_i^T\boldsymbol{A}\boldsymbol{p}_j = 0$ при всех $i \neq j$, что объясняет название метода.

При решении задач понижения размерности *метод сопряженных градиентов* обладает по сравнению с другими методами численной минимизации **рядом преимуществ**:

1) на каждом шаге осуществляется минимизация по выбранному направлению, что ценно ввиду большого числа nm оптимизируемых переменных;
2) в отличие от *метода Ньютона* (см. [18, с. 141]) не требуется память для хранения и не расходуется время на вычисление матрицы вторых производных размера $nm \times nm$;
3) в отличие от *наискорейшего спуска* (см. пример 8) метод сопряженных градиентов обычно удовлетворительно работает при минимизации «овражных» функций.

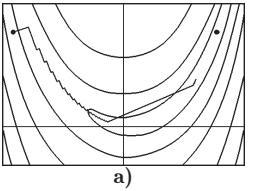
Пример 8. *Критика метода наискорейшего спуска*. Иногда в качестве направления минимизации на очередном шаге рекомендуют выбирать *антиградиент* $-\boldsymbol{g}_t$, поскольку в этом направлении функция убывает быстрее всего (см. [45, с. 478]).*) Однако, эта

*) В частности, для поверхности («горы»), задаваемой гладкой функцией двух переменных, антиградиент указывает направление (локально) наискорейшего спуска.

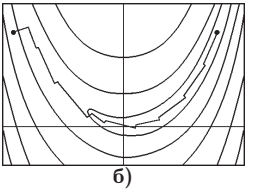
«жадность» метода приводит к огромному числу смен направления при движении вдоль «дна оврага». Наглядный пример (см. [18, с. 140]) дает поиск минимума *функции Розенброка*

$$F(x,y) = 100(y - x^2)^2 + (1 - x)^2,$$

которая определяет в трехмерном пространстве поверхность типа серповидного ущелья с точкой минимума (1,1). На рис. 24, *а* показан процесс минимизации $F(x,y)$ методом наискорейшего спуска из точки $(-6/5,1)$ (гладкие кривые — это линии уровня функции, отрезки ломаной соответствуют шагам спуска). Алгоритм мог бы надолго «застрять» вблизи начала координат, если бы не помогла счастливая случайность — на одной из итераций процедура одномерного поиска неожиданно нашла второй минимум по направлению. Затем было выполнено еще несколько сотен шагов, но ощутимого изменения целевой функции это не дало. Счет был прерван после 1000 итераций вдали от искомого решения.



а)



б)

Рис. 24

В свою очередь, метод сопряженных градиентов позволил дойти до точки минимума за два десятка шагов (рис. 24, *б*). Хорошо виден его циклический характер в случае двух переменных: движение вдоль дна ущелья подправляется перемещением по антиградиенту.

Для полноты изложения приведем один из алгоритмов минимизации по направлению, в основе которого лежит

Метод золотого сечения (см. [18, с. 120]).

Алгоритм включает **следующие пункты**.

1. Необходимо проверить, что $\boldsymbol{p}_t$ — это направление убывания F (определение $\boldsymbol{p}_t$ см. выше). Для этого зададим малый шаг $\delta > 0$ (скажем, $\delta = 0{,}01$) и вычислим при $\alpha = \delta$ значение

   $$\varphi(\alpha) = F(\boldsymbol{x}_t + \alpha \boldsymbol{p}_t).$$

   Если $\varphi(\delta) > \varphi(0)$, то заменим направление $\boldsymbol{p}_t$ на антиградиент $-\boldsymbol{g}_t$ и будем уменьшать величину шага делением пополам до тех пор, пока не добьемся убывания функции.
2. Определим отрезок $[a, b]$, содержащий точку минимума функции $\varphi(\alpha)$. Положим $a = 0$, а для нахождения b будем увеличивать шаг вдвое до тех пор, пока функция не прекратит убывать.

Забросить ключ
от дома и уйти,
Не как изгнанник, что
бредет без смысла,
Но выбирая свой
маршрут разумно,
Меняя скорость, если
склон сменился.
У. Х. Оден, «The Journey» (1928)

Для заданных выше F_0 или F_1 это обязательно произойдет, поскольку они стремятся к $+\infty$ при $|\boldsymbol{x}| \to \infty$.

При поиске *точки минимума* α^* (непрерывной) функции $\varphi(\alpha)$ методом золотого сечения предполагается, что φ *унимодальна*, т. е. строго убывает слева от α^* и строго возрастает справа от α^*. Для такой функции положение точки минимума можно уточнить, вычислив φ в *двух* внутренних точках отрезка $[a, b]$. Например, так как на рис. 25 $\varphi(r) < \varphi(s)$, то α^* обязана находиться на отрезке $[a, s]$. (Если нет унимодальности, то в результате применения метода золотого сечения будет найден один из локальных минимумов.)

3. На отрезке $[a, b]$ возьмем две пробные точки $r = a + (1-\varkappa)(b-a)$ и $s = a + \varkappa(b-a)$, где $\varkappa = (\sqrt{5}-1)/2 \approx 0{,}618$ — *золотое сечение* (см. решение задачи 5 гл. 1),[*)] и вычислим значения $\varphi(r)$ и $\varphi(s)$.
4. Если $\varphi(r) < \varphi(s)$, то полагаем $\alpha_t = r$, $a' = a$, $r' = a + (1-\varkappa)(s-a)$, $s' = r$, $b' = s$ и вычисляем $\varphi(r')$. В противном случае полагаем $\alpha_t = s$, $a' = r$, $r' = s$, $s' = a + \varkappa(b-r)$, $b' = b$ и вычисляем $\varphi(s')$.
5. Если $b' - a' \leqslant \varepsilon$ (скажем, $\varepsilon = 0{,}01$), то заканчиваем поиск, иначе убираем у всех переменных штрихи и возвращаемся к пункту 4.[**)]

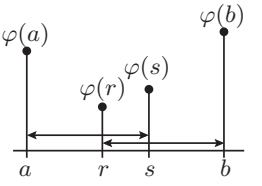


Рис. 25

Замечание 8. В отличие от линейного шкалирования Торгерсона (§ 4), метод Сэммона позволяет обрабатывать матрицы различий *с пропусками*. Для этого суммирование в формуле (18) достаточно проводить только для тех пар объектов, у которых различия измерены. Экспериментально установлено, что качество восстановления конфигурации будет почти таким же, как для полной матрицы, даже если доля пропусков составляет около 30%.

Замечание 9. *О заполнении пропусков.*[***)] В таблицах реальных данных, как правило, встречаются пропуски. Если обрабатывать только часть таблицы, состоящую из строк без пропусков, то может пропасть много важной информации.

С другой стороны, многим методам заполнения пропусков (средневыборочными значениями, по регрессии (см. гл. 21) и т. п.) присущи следующие **два принципиальных недостатка**.

1. Параметры алгоритмов заполнения обычно вычисляются по присутствующим данным, что вносит зависимость между наблюдениями.
2. Распределение данных после заполнения отличается от истинного (возникает смесь истинного и вырожденных распределений с вырождением на гиперплоскостях, на которых располагаются предсказываемые значения). В частности, заполнение средними приводит к искусственному увеличению доли объектов со значениями признаков в центре совокупности.
   Эти недостатки влекут смещенность и несостоятельность оценок параметров моделей при обработке заполненной таблицы данных.

[*)] Замечательно, что точки r и s осуществляют золотые сечения не только отрезка $[a, b]$, но и отрезков $[a, s]$ и $[r, b]$ соответственно. Это позволяет вычислять всего одно значение φ при следующей итерации.

[**)] При каждом выполнении пункта 4 длина отрезка локализации точки минимума уменьшается в $1 + \varkappa \approx 1{,}618$ раз. Всего за 10 повторов первоначальный отрезок сожмется в $(1 + \varkappa)^{10} \approx 123$ раза.

[***)] Материал замечания основан на дополнении Никифорова А. М. к переводу книги Литтл Р. Дж. А., Рубин Д. Б. «Статистический анализ данных с пропусками», — М.: Финансы и статистика, 1990.

В качестве альтернативы рассмотрим

Метод локального заполнения

Пусть у i-й строки таблицы $\boldsymbol{X} = \|x_{il}\|_{n\times m}$ есть данные в столбцах с номерами $l_1,\ldots,l_k$, $k < m$. Обозначим через S_i множество строк $\boldsymbol{X}$, не имеющих пропусков в столбцах $l_1,\ldots,l_k$. Для заполнения пропуска в j-й ячейке i-й строки выделим в S_i подмножество R_{ij} тех строк, у которых нет пропуска в j-й ячейке. Если R_{ij} пусто, то заполнение невозможно (i-ю строку придется исключить из обработки). В противном случае найдем в R_{ij} строку, ближайшую к i-й в смысле, скажем, евклидова расстояния в k-мерном подпространстве, порожденном столбцами с номерами $l_1,\ldots,l_k$ (если таких строк окажется несколько, то выберем одну из них случайно). Заполним пропуск значением из j-й ячейки найденной строки.

Какими **статистическими свойствами** обладает этот метод? Будем рассматривать строки таблицы как реализации случайных векторов $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n$. Пусть случайная величина M_{il} — *индикатор наличия пропуска* в l-м столбце i-й строки. Положим $\boldsymbol{M}_i = (M_{i1},\ldots,M_{im})$, $i = 1,\ldots,n$. Допустим, что

1. случайные векторы $(\boldsymbol{X}_1,\boldsymbol{M}_1),\ldots,(\boldsymbol{X}_n,\boldsymbol{M}_n)$ независимы и одинаково распределены;
2. $\boldsymbol{M}_1$ не зависит от $\boldsymbol{X}_1$ (т. е. пропуски не зависят от значений);
3. распределение $\boldsymbol{X}_1$ имеет m-мерную плотность (см. П8);
4. $\mathsf{P}(M_{1l} = 1) = p_l < 1$ для всех $l = 1,\ldots,m$.

Обозначим через $F(\boldsymbol{x}) = F(x_1,\ldots,x_m)$ *функцию распределения* вектора $\boldsymbol{X}_1 = (X_{11},\ldots,X_{1m})$: $F(\boldsymbol{x}) = \mathsf{P}(X_{1l} \leqslant x_l,\ l = 1,\ldots,m)$, а через $\tilde{F}_n(\boldsymbol{x})$ — *эмпирическую функцию распределения*, построенную по полученной методом локального заполнения таблице $\tilde{\boldsymbol{X}}$:

$$\tilde{F}_n(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n} I_{\{\tilde{X}_{il}\leqslant x_l,\ l=1,\ldots,m\}}.$$

Тогда при выполнении указанных выше условий 1–4

$$\sup_{\boldsymbol{x}}\left|\tilde{F}_n(\boldsymbol{x}) - F(\boldsymbol{x})\right| \xrightarrow{\text{п. н.}} 0 \quad \text{при } n\to\infty.$$

Приведенное утверждение означает, что при неограниченном увеличении размера выборки локальное заполнение обеспечивает совпадение распределения заполненной выборки с истинным. Из него вытекает, что оценки, *непрерывные в равномерной метрике* (см. [11, с. 26]), состоятельные для полных данных, будут состоятельны и для данных с пропусками после локального заполнения.

§ 7. РАНГОВАЯ КОРРЕЛЯЦИЯ

Нередко на практике представляет интерес *гипотеза независимости признаков* ξ и η:

$H_0\colon F_{\xi,\eta}(x,y) = F_\xi(x)F_\eta(y)$ при всех x,y,

где $F_{\xi,\eta}(x,y) = \mathbf{P}(\xi \leqslant x, \eta \leqslant y)$ — это функция распределения случайного вектора (ξ,η) (см. П8), а $F_\xi(x)$ и $F_\eta(y)$ — функции распределения его компонент.

Для ее проверки применяют ранговые критерии. Они не зависят от конкретного вида функций F_ξ и F_η при условии их непрерывности. Кроме того, эти критерии робастны (устойчивы) (см. § 4 гл. 8) к выделяющимся наблюдениям («выбросам»), которые обычно присутствуют в крупных массивах реальных данных. Рассмотрим сначала наиболее часто используемый

Критерий Спирмена

Обозначим через R_i *ранг* (т. е. номер в порядке возрастания) наблюдения ξ_i среди $\xi_1,\ldots,\xi_n$, а через S_i ранг η_i среди $\eta_1,\ldots,\eta_n$. Таким образом, наблюдения порождают n пар рангов $(R_1,S_1),\ldots,(R_n,S_n)$. Статистикой критерия Спирмена служит *выборочный коэффициент корреляции* ρ_S ранговых наборов $(R_1,\ldots,R_n)$ и $(S_1,\ldots,S_n)$, определяемый формулой

$$\rho_S = \sum_{i=1}^n (R_i - \overline{R})(S_i - \overline{S}) \Big/ \left[\sum_{i=1}^n (R_i - \overline{R})^2 \sum_{i=1}^n (S_i - \overline{S})^2\right]^{1/2}. \qquad (20)$$

В этой формуле $\overline{R} = \overline{S} = \dfrac{1}{n}\sum\limits_{i=1}^n i = \dfrac{n+1}{2}$. С учетом легко доказываемого по индукции равенства $\sum\limits_{i=1}^n i^2 = \dfrac{n(n+1)(2n+1)}{6}$ имеем

$$\sum_{i=1}^n (R_i - \overline{R})^2 = \sum_{i=1}^n (S_i - \overline{S})^2 = \sum_{i=1}^n \left(i - \frac{n+1}{2}\right)^2 = \frac{n^3-n}{12}.$$

Переставив пары (R_i,S_i) в порядке возрастания первой компоненты, получим набор $(1,T_1),\ldots,(n,T_n)$. Тогда статистика (20) запишется в виде

$$\rho_S = \frac{12}{n^3-n}\sum_{i=1}^n \left(i - \frac{n+1}{2}\right)\left(T_i - \frac{n+1}{2}\right). \qquad (21)$$

Таким образом, ρ_S — линейная функция от рангов T_i. Правую часть равенства (21) можно также представить (задача 3) в виде

$$\rho_S = 1 - \frac{6}{n^3-n}\sum_{i=1}^n (i-T_i)^2 = 1 - \frac{6}{n(n-1)(n+1)}\sum_{i=1}^n (R_i - S_i)^2, \quad (22)$$

который наиболее удобен для вычислений.

Совпадения. Пример 9 показывает, что формула (22) пригодна для подсчета ρ_S *только в случае отсутствия совпадений* (т. е. в случае, когда среди значений наблюдений $\xi_1,\ldots,\xi_n$ $(\eta_1,\ldots,\eta_n)$ нет одинаковых). Если совпадения есть, то при ранжировании им следует присваивать *средние ранги*[*)] и затем вычислять ρ_S на основе формулы (20).[**)]

Пример 9 ([2, с. 108])**.** Десять однородных предприятий были проранжированы вначале по *степени прогрессивности их оргструктур* (признак ξ), затем по *эффективности их функционирования в отчетном году* (признак η). В результате были получены следующие две ранжировки: 1; 2,5; 2,5; 4,5; 4,5; 6,5; 6,5; 8; 9,5; 9,5 и 1; 2; 4,5; 4,5; 4,5; 4,5; 8; 8; 8; 10. Для этих данных правая часть равенства (22) равна 0,921, а коэффициент ρ_S, вычисленный по формуле (20), имеет значение 0,917. С помощью табл. Т6 устанавливаем, что при $n = 10$ критической границей для ρ_S на уровне значимости 0,001 служит величина 0,879. Поскольку $0{,}917 > 0{,}879$, корреляционную связь между признаками ξ и η следует признать значимой.

Исследуем некоторые свойства статистики ρ_S при справедливости гипотезы H_0. Множество рангов $(T_1,\ldots,T_n)$ — это некоторая перестановка множества $(1,\ldots,n)$. При выполнении гипотезы H_0 все $n!$ таких перестановок равновероятны. Поэтому для любого $1 \leqslant i \leqslant n$

$$\mathbf{M}T_i = \sum_{k=1}^{n} k\,\mathbf{P}(T_i = k) = \sum_{k=1}^{n} k\,\frac{(n-1)!}{n!} = \frac{n+1}{2}\,.$$

Из формулы (21) немедленно получаем, что $\mathbf{M}\rho_S = 0$ при выполнении гипотезы H_0. Нетрудно установить, что $\mathbf{D}\rho_S = 1/(n-1)$ (задача 4). Так как ρ_S — коэффициент корреляции, то согласно следствию из неравенства Коши—Буняковского (П4) всегда $-1 \leqslant \rho_S \leqslant 1$. Крайние значения достигаются: при полном соответствии наборов рангов ($R_i = S_i$, $i = 1,\ldots,n$) имеем $\rho_S = 1$, а при противоположных рангах ($T_i = n - i + 1$) получаем $\rho_S = -1$.

Вопрос 2. Почему верно последнее утверждение?

Достаточно близкие к 1 (или -1) значения ρ_S противоречат гипотезе H_0. Критические границы при односторонней альтернативе $H_1\colon \boldsymbol{\rho}(\xi,\eta) > 0$ на нескольких уровнях значимости для $n \leqslant 50$ можно найти в табл. Т6.

Для $n > 50$ годится нормальное приближение, основанное на сходимости

$$\rho_S \Big/ \sqrt{\mathbf{D}\rho_S} \xrightarrow{d} Z \sim \mathcal{N}(0,1) \quad \text{при } n \to \infty$$

в случае справедливости гипотезы H_0 (доказательство см. в [86, с. 227]).

*) Так, для выборки 2, 5, 5, 7 получаем ранжировку 1; 2,5; 2,5; 4.

**) Для подсчета выборочного коэффициента корреляции ранжировок (20) можно воспользоваться функцией «Коррел» из Excel.

Поправка. Для небольших выборок это приближение не является удовлетворительным. Р. Иман и У. Коновер в 1978 г. предложили следующую поправку, значительно повышающую точность аппроксимации (см. [88, с. 10]). Положим

$$\tilde{\rho}_S = \frac{1}{2}\,\rho_S\left(\sqrt{n-1} + \sqrt{(n-2)/\,(1-\rho_S^2)}\right).$$

С помощью табл. Т2 и Т4 вычислим $z_\alpha = (x_{1-\alpha} + y_{1-\alpha})/2$, где $x_{1-\alpha}$ и $y_{1-\alpha}$ обозначают, соответственно, квантили уровня $(1-\alpha)$ закона $\mathcal{N}(0,1)$ и распределения Стьюдента с $(n-2)$ степенями свободы (см. § 2 гл. 11). Если $\tilde{\rho}_S \geqslant z_\alpha$, то гипотеза H_0 отвергается в пользу альтернативы $H_1\colon \boldsymbol{\rho}(\xi,\eta) > 0$, иначе — принимается.

Критерий Кендэла

Другую ранговую меру связи ввел в 1938 г. М. Дж. Кендэл. Будем говорить, что пары (ξ_i, η_i) и (ξ_j, η_j) *согласованы* $(1 \leqslant i < j \leqslant n)$, если $\xi_i < \xi_j$ и $\eta_i < \eta_j$ или $\xi_i > \xi_j$ и $\eta_i > \eta_j$ (т. е. $\operatorname{sign}(\xi_j - \xi_i)\operatorname{sign}(\eta_j - \eta_i) = 1$). Пусть S — число согласованных пар, а R — число несогласованных пар. Тогда превышение согласованности над несогласованностью есть *)

$$T = S - R = \sum_{i<j} \operatorname{sign}(\xi_j - \xi_i)\operatorname{sign}(\eta_j - \eta_i).$$

Значения T изменяются от $-n(n-1)/2$ до $n(n-1)/2$. Например, $\max T = n(n-1)/2$ достигается при идеальном согласии порядка $\xi_1, \ldots, \xi_n$ и $\eta_1, \ldots, \eta_n$. Для измерения степени согласия Кендэл предложил коэффициент

$$\tau = \frac{T}{\max T} = \frac{2T}{n(n-1)} = \frac{2(S-R)}{n(n-1)} = 1 - \frac{4}{n(n-1)}\,R, \tag{23}$$

так как $S + R = n(n-1)/2$. Заметим, что величина R — это *количество инверсий* (см. пример 2 гл. 7), образованных величинами η_i, расположенными в порядке возрастания соответствующих ξ_i. Таким образом, коэффициент τ (линейно связанный с R) можно считать *мерой неупорядоченности* второй последовательности относительно первой.

Ввиду формулы (23) и асимптотической нормальности статистики R при справедливости гипотезы H_0 имеет место сходимость

$$\tau/\sqrt{\mathbf{D}\tau} \xrightarrow{d} Z \sim \mathcal{N}(0,1) \quad \text{при } n \to \infty,$$

где $\mathbf{D}\tau = 2(2n+5)/[9n(n-1)]$.

Обсудим *связь между коэффициентами* τ и ρ_S. Очевидно, статистика T представляется также в ранговой форме:

$$T = \sum_{i<j} \operatorname{sign}(R_j - R_i)\operatorname{sign}(S_j - S_i) = \sum_{i<j} \operatorname{sign}(T_j - T_i). \tag{24}$$

*) Предполагается, что среди ξ_i и среди η_i нет совпадений.

Аналогично, $R = \sum\limits_{i<j} I_{\{T_i>T_j\}}$. С учетом соотношения (23) получаем, что

$$\tau = 1 - \frac{4}{n^2-n} \sum_{i<j} I_{\{T_i>T_j\}}.$$

Согласно задаче 5 для коэффициента ρ_S верна похожая формула:

$$\rho_S = 1 - \frac{12}{n^3-n} \sum_{i<j} (j-i)\, I_{\{T_i>T_j\}}, \tag{25}$$

показывающая, что в случае ρ_S инверсиям придаются дополнительные веса $(j-i)$. Из-за этого возникает предположение, что ρ_S сильнее реагирует на несогласие ранжировок, чем τ. Однако М. Кендэл и А. Стьюарт в [35, с. 683] отмечают, что величины ρ_S и τ при справедливости гипотезы H_0 *сильно коррелированы*: коэффициент корреляции между ними равен $2\,(n+1)\,/\sqrt{2n(2n+5)}$. Он убывает от 1 при $n=2$ до 0,98 при $n=5$ и далее возрастает до 1 при $n\to\infty$.

Замечание 10. ***Обобщенный коэффициент корреляции*** [36]. Для удобства реализации на компьютере системы алгоритмов корреляционного анализа полезно вывести *обобщенную формулу* для вычисления разных парных корреляционных характеристик (таких, как τ, ρ_S и $\widehat{\rho}$, где

$$\widehat{\rho} = \sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y}) \Big/ \left[\sum_{i=1}^{n} (X_i - \overline{X})^2 \cdot \sum_{i=1}^{n} (Y_i - \overline{Y})^2 \right]^{1/2}$$

обозначает *обычный коэффициент корреляции* между выборками $\boldsymbol{X} = (X_1, \ldots, X_n)$ и $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$). С этой целью определим некоторое правило, в соответствии с которым каждой паре (X_i, X_j) компонент вектора $\boldsymbol{X}$ приписывается число («метка») $c_{ij} = c_{ij}(\boldsymbol{X})$, причем это правило будет обладать свойством *отрицательной симметричности*: $c_{ij} = -c_{ji}$, $c_{ii} = 0$. Тогда *обобщенный коэффициент корреляции* между $\boldsymbol{X}$ и $\boldsymbol{Y}$ определяется формулой

$$\widehat{r} = \sum_{i<j} c_{ij}(\boldsymbol{X})\, c_{ij}(\boldsymbol{Y}) \Big/ \left[\sum_{i<j} c_{ij}^2(\boldsymbol{X}) \cdot \sum_{i<j} c_{ij}^2(\boldsymbol{Y}) \right]^{1/2}.$$

Убедимся, что коэффициенты $\widehat{\rho}$, ρ_S и τ могут быть получены как **частные случаи** обобщенного коэффициента $\widehat{r}$ при соответствующем выборе правила приписывания числовых «меток» c_{ij}.

1) Установим для любых $X_1, \ldots, X_n$ и $Y_1, \ldots, Y_n$ справедливость тождества

$$\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y}) = \frac{1}{n} \sum_{i<j} (X_j - X_i)(Y_j - Y_i).$$

При $X_i = Y_i$ оно следует из *теоремы о межточечных расстояниях*, доказанной при решении задачи 5 гл. 16 ($m_i = 1$, $m = n$). В общем случае оно выводится из указанной теоремы с помощью представления $AB = [(A+B)/2]^2 - [(A-B)/2]^2$ (см. задачу 6).

Ввиду установленного тождества при выборе в качестве «меток» $c_{ij}(\boldsymbol{X}) = X_j - X_i$ коэффициент $\widehat{r}$ преобразуется в $\widehat{\rho}$.

2) Положим $c_{ij}(\boldsymbol{X}) = R_j - R_i$, где R_i — ранг X_i в выборке $\boldsymbol{X}$. С учетом предыдущих рассуждений и определения коэффициента Спирмена (20) видим, что $\widehat{r}$ в этом случае совпадает с ρ_S.

3) Пусть $c_{ij}(\boldsymbol{X}) = \operatorname{sign}(X_j - X_i) = \operatorname{sign}(R_j - R_i)$. Тогда делимым в формуле для $\widehat{r}$ служит определенная выше статистика T, а делитель равен $n(n-1)/2$. Принимая во внимание формулу (23), заключаем, что обобщенный коэффициент $\widehat{r}$ превращается в τ.

§ 8. МНОЖЕСТВЕННАЯ И ЧАСТНАЯ КОРРЕЛЯЦИИ

Может представлять интерес задача измерения статистической связи сразу между $k \geqslant 3$ выборками. С этой целью Кендэлом был предложен ранговый *коэффициент конкордации (согласованности)*

$$W = \frac{12}{k^2(n^3-n)} \sum_{i=1}^{n} \left(\sum_{j=1}^{k} R_{ij} - \frac{k(n+1)}{2} \right)^2,$$

где R_{ij} — ранг (от 1 до n) i-го элемента в j-й выборке (столбце).

Укажем **некоторые свойства** коэффициента W (см. [36, гл. 6]).

1) $0 \leqslant W \leqslant 1$, причем $W = 1$ тогда и только тогда, когда все k ранжировок совпадают. То, что W не принимает отрицательных значений, объясняется тем обстоятельством, что в отличие от случая парных связей для $k \geqslant 3$ выборок противоположность согласованности утрачивается: упорядочения могут полностью совпадать, но не могут полностью не совпадать.

2) Обозначим через $\overline{\rho}_S$ *среднее арифметическое коэффициентов Спирмена по всем* $k(k-1)/2$ *парам выборок*. Тогда

$$W = [(k-1)\overline{\rho}_S + 1]/k.$$

Таким образом, W и $\overline{\rho}_S$ линейно связаны. В частности, при $k = 2$ имеем $W = (\rho_S + 1)/2$, т. е. коэффициент конкордации W линейно зависит от коэффициента Спирмена ρ_S.

3) Сравнение с критерием Фридмана из § 2 гл. 17 (с точностью до замены обозначений $k \rightleftarrows n$) показывает, что при больших n статистика $k(n-1)W$ распределена приближенно по закону хи-квадрат с $(n-1)$ степенями свободы.

Вопрос 3. Будет ли значение $W = 0{,}09$ значимо велико на уровне 5% при $k = 20$ и $n = 15$? (Воспользуйтесь табл. Т3 критических значений χ^2-распределения.)

Перейдем теперь к обсуждению понятия *частной или «очищенной» корреляции*. Начнем с примера из [2, с. 64].

> «Даже если удалось установить тесную зависимость между двумя исследуемыми величинами, отсюда еще непосредственно не следует их *причинная взаимообусловленность*. Например, при анализе большого числа наблюдений, относящихся к отливке труб на сталелитейных заводах, была установлена положительная корреляционная

связь между временем плавки и процентом забракованных труб [3]. Дать какое-либо причинное истолкование этой стохастической связи было невозможно, а поэтому рекомендации ограничить продолжительность плавки для снижения процента забракованных труб выглядели малосостоятельными. Действительно, спустя несколько лет обнаружили, что большая продолжительность плавки всегда была связана с использованием сырья специального состава. Этот вид сырья приводил одновременно к длительному времени плавки и большому проценту брака, хотя оба этих фактора взаимно независимы.

Таким образом, высокий коэффициент корреляции между продолжительностью плавки и процентом забракованных труб полностью обусловливался влиянием третьего, не учтенного при исследовании фактора — характеристики качества сырья. Если же этот фактор был бы с самого начала учтен, то никакой значимой корреляционной связи между временем плавки и процентом забракованных труб мы бы не обнаружили. За счет подобных эффектов (одновременного влияния неучтенных факторов на исследуемые переменные) может искажаться и смысл истинной связи между переменными, т. е., например, подсчеты приводят к положительному значению парного коэффициента корреляции, в то время как истинная связь между ними имеет отрицательный смысл. Такую корреляцию между двумя переменными часто называют «ложной». Более детально подобные ситуации — обнаружение и исключение «общих причинных факторов», расчет «очищенных», или *частных*, коэффициентов корреляции и т. п. — исследуют методами многомерного корреляционного анализа.»

Определение. Частным коэффициентом корреляции между случайными величинами X и Y при исключении влияния случайной величины Z называется

$$\boldsymbol{\rho}(X,Y|Z) \equiv \rho_{XY|Z} = \frac{\boldsymbol{\rho}(X,Y) - \boldsymbol{\rho}(X,Z)\boldsymbol{\rho}(Y,Z)}{\sqrt{(1-\boldsymbol{\rho}^2(X,Z))(1-\boldsymbol{\rho}^2(Y,Z))}}\,.$$

К этой формуле приводит попытка исключить зависимость от Z, заменив X и Y такими случайными величинами

$$X' = X - aZ, \quad Y' = Y - bZ,$$

которые некоррелированы с Z: $\boldsymbol{\rho}(X',Z) = 0$ и $\boldsymbol{\rho}(Y',Z) = 0$. Тогда «оставшаяся» корреляция представляет собой обычную корреляцию между X' и Y'.

Доказательство. Допустим для простоты, что $\mathbf{M}X = \mathbf{M}Y = \mathbf{M}Z = 0$. Для краткости введем обозначения $\sigma_\xi = \sqrt{\mathbf{D}\xi}$ и $\rho_{\xi\eta} = \boldsymbol{\rho}(\xi,\eta)$. Константы a и b нужно выбрать так, чтобы имели место равенства

$$\mathbf{M}X'Z = \mathbf{M}XZ - a\mathbf{M}Z^2 = 0, \quad \mathbf{M}Y'Z = \mathbf{M}YZ - b\mathbf{M}Z^2 = 0.$$

Отсюда находим

$$a = \frac{\rho_{XZ}\,\sigma_X\,\sigma_Z}{\sigma_Z^2} = \rho_{XZ}\,\frac{\sigma_X}{\sigma_Z}\,, \quad b = \frac{\rho_{YZ}\,\sigma_Y\,\sigma_Z}{\sigma_Z^2} = \rho_{YZ}\,\frac{\sigma_Y}{\sigma_Z}\,. \tag{26}$$

Запишем обычный коэффициент корреляции между X' и Y':

$$\rho_{X'Y'} = \frac{\mathsf{M}(X - aZ)(Y - bZ)}{\sigma_{X-aZ}\,\sigma_{Y-bZ}}. \tag{27}$$

Числитель в формуле (27) можно представить в следующем виде:

$$\mathsf{M}XY - a\,\mathsf{M}YZ - b\,\mathsf{M}XZ + ab\,\mathsf{M}Z^2 =$$

$$= \rho_{XY}\,\sigma_X\,\sigma_Y - a\,\rho_{YZ}\,\sigma_Y\,\sigma_Z - b\,\rho_{XZ}\,\sigma_X\,\sigma_Z + ab\,\sigma_Z^2.$$

Заменив a и b в этом равенстве их значениями из соотношений (26), получим

$$\mathsf{M}(X - aZ)(Y - bZ) = (\rho_{XY} - \rho_{XZ}\,\rho_{YZ})\,\sigma_X\sigma_Y.$$

Точно также вычисляются дисперсии

$$\sigma^2_{X-aZ} = \mathsf{M}(X - aZ)^2 = (1 - \rho^2_{XZ})\,\sigma^2_X,$$

$$\sigma^2_{Y-bZ} = \mathsf{M}(Y - bZ)^2 = (1 - \rho^2_{YZ})\,\sigma^2_Y.$$

Подстановка всех этих выражений в формулу (27) приводит к равенству

$$\rho_{X'Y'} = \frac{\rho_{XY} - \rho_{XZ}\,\rho_{YZ}}{\sqrt{(1 - \rho^2_{XZ})(1 - \rho^2_{YZ})}} = \rho_{XY|Z}, \tag{28}$$

которое и требовалось установить. ■

Для получения оценки $\widehat{\rho}_{\boldsymbol{xy}|\boldsymbol{z}}$ [*)] для коэффициента $\rho_{XY|Z}$ надо заменить в соотношении (28) теоретические коэффициенты корреляции выборочными (см. определение (20)):

$$\widehat{\rho}_{\boldsymbol{xy}|\boldsymbol{z}} = \frac{\widehat{\rho}_{\boldsymbol{xy}} - \widehat{\rho}_{\boldsymbol{xz}}\widehat{\rho}_{\boldsymbol{yz}}}{\sqrt{(1 - \widehat{\rho}^2_{\boldsymbol{xz}})(1 - \widehat{\rho}^2_{\boldsymbol{yz}})}}. \tag{29}$$

К этой формуле можно прийти точно так же, как и выше, отталкиваясь от условия ортогональности реализации выборки $\boldsymbol{z}$ и линейных комбинаций $\boldsymbol{x}' = \boldsymbol{x} - a\boldsymbol{z}$ и $\boldsymbol{y}' = \boldsymbol{y} - b\boldsymbol{z}$. В этой интерпретации $\widehat{\rho}_{\boldsymbol{xy}|\boldsymbol{z}}$ представляет собой косинус угла между проекциями векторов $\boldsymbol{x}$ и $\boldsymbol{y}$ на подпространство в $\mathbb{R}^n$, ортогональное вектору $\boldsymbol{z}$ (рис. 26).

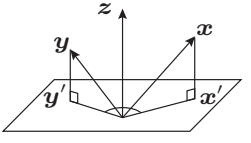


Рис. 26

Если дополнительно предположить, что $\boldsymbol{X}$, $\boldsymbol{Y}$ и $\boldsymbol{Z}$ — выборки из *независимых нормальных законов*, то (как доказано в [13, с. 370]) выборочный частный коэффициент корреляции $\widehat{\rho}_{\boldsymbol{XY}|\boldsymbol{Z}}$ будет распределен точно также, как и обычный выборочный коэффициент $\widehat{\rho}_{\boldsymbol{XY}}$, но для выборок размера не n, а $n-1$. Отсюда и из задачи 6 гл. 11 вытекает сходимость распределения случайной величины $\sqrt{n}\ \operatorname{arcth}\widehat{\rho}_{\boldsymbol{XY}|\boldsymbol{Z}}$ к закону $\mathcal{N}(0,1)$ при $n \to \infty$.

Пример 10 ([2, с. 85])**.** По итогам года у 37 однородных предприятий легкой промышленности были зарегистрированы следующие (среднемесячные) показатели их работы: $\boldsymbol{x}$ — значения характеристики качества ткани (в баллах), $\boldsymbol{y}$ — количества профилактических наладок автоматической линии, $\boldsymbol{z}$ — случаи обрывов нити.

[*)] Здесь неслучайный вектор $\boldsymbol{x} = (x_1, \ldots, x_n)$ обозначает реализацию выборки $\boldsymbol{X} = (X_1, \ldots, X_n)$, где X_i независимы и распределены так же, как и случайная величина X.

На основе этих данных были подсчитаны парные коэффициенты корреляции: $\widehat{\rho}_{\boldsymbol{xy}} = 0{,}105$, $\widehat{\rho}_{\boldsymbol{xz}} = 0{,}024$, $\widehat{\rho}_{\boldsymbol{yz}} = 0{,}996$. Проверка на статистическую значимость свидетельствует об отсутствии связи между качеством ткани, с одной стороны, и числом профилактических наладок и обрывов нити — с другой, что не согласуется с профессиональными представлениями технолога.

Однако расчет *частных* коэффициентов корреляции по формуле (29) дает значения $\widehat{\rho}_{\boldsymbol{xy|z}} = 0{,}908$ и $\widehat{\rho}_{\boldsymbol{xz|y}} = -0{,}907$, которые вполне соответствуют представлению о естественном характере связей между изучаемыми показателями.

В заключение отметим, что ранговый коэффициент Кендэла τ (в отличие от коэффициента Спирмена ρ_S) переносится на случай частной корреляции с помощью формулы, аналогичной формуле (29):

$$\tau_{\boldsymbol{xy|z}} = \frac{\tau_{\boldsymbol{xy}} - \tau_{\boldsymbol{xz}}\tau_{\boldsymbol{yz}}}{\sqrt{(1-\tau^2_{\boldsymbol{xz}})(1-\tau^2_{\boldsymbol{yz}})}}$$

(см. [36, гл. 8]). Критерии значимости для $\tau_{\boldsymbol{xy|z}}$ можно отыскать в журнальных статьях, указанных на с. 216 книги [86].

§ 9. ТАБЛИЦЫ СОПРЯЖЕННОСТИ

Рассмотрим задачу выявления статистической связи для сгруппированных (разбитых на категории) данных. Если ранее обсуждались случаи *количественных* (§§ 1–6) и *порядковых (ранговых)* (§§ 7–8) переменных, то теперь переменные — качественные и описываются номером группы, а данные представлены в виде *таблицы сопряженности (признаков)* $\|\nu_{ij}\|_{n\times m}$, в которой ν_{ij} — числа объектов с признаками i, j.*)

Опишем **три выборочные схемы**, приводящие к таблицам сопряженности ([2, с. 125]).

Схема I возникает в случае, когда строки $(\nu_{i1},\ldots,\nu_{im})$ таблицы данных $(i = 1,\ldots,n)$ можно рассматривать как независимые выборки из полиномиальных распределений (см. § 5 гл. 10 и доказательство теоремы 1 гл. 18) с вероятностями q_{ij} $(\sum\limits_{j=1}^{m} q_{ij} = 1)$ и заданными числами наблюдений $n_i = \sum\limits_{j=1}^{m} \nu_{ij}$. Такая организация данных обычно возникает, когда хотят сравнить между собой несколько одномерных распределений, представленных выборками заранее заданного размера. Наиболее важной для схемы I является *гипотеза однородности*

$$H_I\colon q_{ij} = q_{\cdot j}, \quad \text{где } q_{\cdot j} = \frac{1}{n}\sum_{i=1}^{n} q_{ij},$$

*) Таблицы с тремя и более входами анализируются в книгах [7], [47].

которая подробно обсуждалась ранее в § 3 гл. 18.

Схема II. Предполагается, что $(\nu_{11},\ldots,\nu_{nm})$ имеют полиномиальное распределение с вероятностями $(p_{11},\ldots,p_{nm})$ и фиксированным общим числом наблюдений $N=\sum\limits_{i,j}\nu_{ij}$. Таблица сопряженности в этом случае является обычной двумерной гистограммой для N наблюдений. Гипотезе H_I из схемы I в схеме II соответствует *гипотеза независимости*, состоящая в том, что совместное распределение есть произведение маргинальных (частных) распределений:

$$H_{II}\colon p_{ij}=r_is_j,\quad \text{где } r_i=\sum_{j=1}^{m}p_{ij},\quad s_j=\sum_{i=1}^{n}p_{ij}.$$

Схема III возникает, когда в схеме II общее число наблюдений рассматривается как случайная величина. Ее важным частным случаем является случай, когда все ν_{ij} независимы и распределены по закону Пуассона с параметрами λ_{ij}. Тогда их сумма N также имеет распределение Пуассона с параметром $c=\sum\limits_{i,j}\lambda_{ij}$ (см. задачу 3 гл. 10). Гипотезам H_I и H_{II} соответствует *гипотеза мультипликативности*

$$H_{III}\colon \lambda_{ij}=a_ib_j/c,\quad \text{где } a_i=\sum_{j=1}^{m}\lambda_{ij},\quad b_j=\sum_{i=1}^{n}\lambda_{ij}.$$

В качестве примера применения схемы III может быть рассмотрена задача, в которой ν_{ij} — число отказов (аварий) i-го вида на установках j-го типа в течение заданного времени наблюдения. Параметры λ_{ij} отражают ожидаемые количества отказов.

Можно доказать, что если в схеме III зафиксировать N, то она переходит в схему II с $p_{ij}=\lambda_{ij}/c$. При этом гипотеза H_{III} преобразуется в гипотезу H_{II}. Аналогично, если зафиксировать в схеме II суммы по строкам $n_i=\nu_{i1}+\ldots+\nu_{im}$, то схема II переходит в схему I с $q_{ij}=p_{ij}/r_i$, а гипотеза H_{II} — в гипотезу H_I.

Для проверки гипотез H_I—H_{III} применяется вариант **критерия хи-квадрат**, статистика которого имеет вид

$$X^2=N\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{(\nu_{ij}-n_im_j/N)^2}{n_im_j},\ \text{где } n_i=\sum_{j=1}^{m}\nu_{ij},\ m_j=\sum_{i=1}^{n}\nu_{ij}.$$

При справедливости проверяемой гипотезы при достаточно больших N статистика X^2 приближенно распределена по закону хи-квадрат с $(n-1)(m-1)$ степенями свободы.

Для схемы II это утверждение вытекает из теоремы Фишера (см. формулу (6) гл. 18). В этом случае имеется $(m+n-2)$ неизвестных параметров $r_1,\ldots,r_{n-1},s_1,\ldots,s_{m-1}$. Методом Лагранжа находим, что оценками максимального правдоподобия для них будут величины $\widehat{r}_i=n_i/N$ и $\widehat{s}_j=m_j/N$ (задача 7). При этом число степеней свободы предельного закона хи-квадрат равно $nm-1-(m+n-2)=(n-1)(m-1)$.

Пример 11 ([44, с. 481]). В приведенной ниже таблице представлены результаты социологического обследования о связи между доходом семей и количеством детей в них. Признак A означает количество детей и принимает значения 0, 1, 2, 3, $\geqslant 4$. Признак B указывает, какому из диапазонов (0–1), (1–2), (2–3), ($\geqslant 3$) (за единицу принято 1000 шведских крон) принадлежит доход семьи.

A \ B	0–1	1–2	2–3	$\geqslant 3$	n_i
0	2161	3577	2184	1636	**9558**
1	2755	5081	2222	1052	**11110**
2	936	1753	640	306	**3635**
3	225	419	96	38	**778**
$\geqslant 4$	39	98	31	14	**182**
m_j	**6116**	**10928**	**5173**	**3046**	**25263**

Значение статистики X^2 равно 568,6, что значимо больше критической границы 32,9 уровня 0,001 закона хи-квадрат с $(5-1)(4-1) = 12$ степенями свободы (см. табл. Т3). Поэтому гипотеза независимости признаков A и B отвергается.*)

ЗАДАЧИ

Помучишься — так научишься.

1. Пусть $\boldsymbol{A}$, $\boldsymbol{B}$ и $\boldsymbol{A}-\boldsymbol{B}$ — неотрицательно определенные матрицы. Докажите, что

а) $\operatorname{tr}\boldsymbol{A} \geqslant \operatorname{tr}\boldsymbol{B}$; $\operatorname{tr}\boldsymbol{A} = \operatorname{tr}\boldsymbol{B} \iff \boldsymbol{A} = \boldsymbol{B}$,

Указание. Используйте приведение $\boldsymbol{A}-\boldsymbol{B}$ к главным осям.

б) $\det\boldsymbol{A} \geqslant \det\boldsymbol{B}$; если $\det\boldsymbol{B} > 0$, то $\det\boldsymbol{A} = \det\boldsymbol{B} \iff \boldsymbol{A} = \boldsymbol{B}$.

Указание. Рассмотрите сначала случай $\boldsymbol{B} = \boldsymbol{E}$ и докажите, что тогда все собственные значения матрицы $\boldsymbol{A}-\boldsymbol{E}$ не меньше 1.

2*. Выведите соотношения (19).

3. Получите представление (22) рангового коэффициента Спирмена ρ_S.

Указание. Разложите $\sum\left[\left(i-\dfrac{n+1}{2}\right)-\left(T_i-\dfrac{n+1}{2}\right)\right]^2$.

4*. Докажите, что $\mathsf{D}\rho_S = 1/(n-1)$ при справедливости гипотезы H_0.

Указание. Положим $\xi_i = R_i - \dfrac{n+1}{2}$. В силу равенства $\sum\xi_i = 0$ имеем

$$0 = \mathsf{M}\,(\xi_1\textstyle\sum\xi_i) = \mathsf{M}\xi_1^2 + (n-1)\mathsf{M}\xi_1\xi_2.$$

5*. Проверьте справедливость для ρ_S представления (25).

6. Проверьте, что для любых $x_1,\ldots,x_n$ и $y_1,\ldots,y_n$ верно тождество

$$\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y}) = \frac{1}{n}\sum_{i<j}(x_j-x_i)(y_j-y_i).$$

*) Однако в [11, с. 412] отмечено, что более тонкий анализ этих данных указывает на очень слабую зависимость между A и B.

7. Используйте метод неопределенных множителей Лагранжа для нахождения оценок максимального правдоподобия неизвестных параметров $r_1,\ldots,r_{n-1},s_1,\ldots,s_{m-1}$ в схеме II из § 9.
Указание. См. похожую задачу 6 гл. 18.

РЕШЕНИЯ ЗАДАЧ

1. а) Так как матрица $\boldsymbol{A}-\boldsymbol{B}$ неотрицательно определена, все ее собственные значения λ_i неотрицательны (см. П10). Поэтому $\operatorname{tr}\boldsymbol{A}-\operatorname{tr}\boldsymbol{B}=\operatorname{tr}(\boldsymbol{A}-\boldsymbol{B})=\sum\lambda_i\geqslant 0$; если $\operatorname{tr}\boldsymbol{A}=\operatorname{tr}\boldsymbol{B}$, то $\operatorname{tr}(\boldsymbol{A}-\boldsymbol{B})=0$. Следовательно, все $\lambda_i=0$, $\boldsymbol{C}^T(\boldsymbol{A}-\boldsymbol{B})\boldsymbol{C}=\boldsymbol{0}$. В силу невырожденности $\boldsymbol{C}$ получаем, что $\boldsymbol{A}=\boldsymbol{B}$.
б) При $\det\boldsymbol{B}=0$ утверждение очевидно. Пусть $\det\boldsymbol{B}>0$, т. е. матрица $\boldsymbol{B}$ положительно определена. Сначала рассмотрим случай единичной матрицы: $\boldsymbol{B}=\boldsymbol{E}$. Пусть μ и $\boldsymbol{c}$ — собственное значение и соответствующий собственный вектор матрицы $\boldsymbol{A}$: $\boldsymbol{A}\boldsymbol{c}=\mu\boldsymbol{c}$. Тогда $(\boldsymbol{A}-\boldsymbol{E})\boldsymbol{c}=(\mu-1)\boldsymbol{c}$, т. е. $\lambda=\mu-1$ — собственное значение $\boldsymbol{A}-\boldsymbol{E}$. Но $\boldsymbol{A}-\boldsymbol{E}$ неотрицательно определена, поэтому $\lambda\geqslant 0\iff\mu\geqslant 1$. Отсюда $\det\boldsymbol{A}=\prod\mu_i\geqslant 1=\det\boldsymbol{E}$. Если $\det\boldsymbol{A}=1$, то каждое $\mu_i=1$, так что $\boldsymbol{A}=\boldsymbol{E}$ (см. вопрос 7 гл. 19).

Теперь рассмотрим общий случай положительно определенной матрицы $\boldsymbol{B}$. В этом случае матрица

$$\boldsymbol{D}=\boldsymbol{B}^{-1/2}(\boldsymbol{A}-\boldsymbol{B})\boldsymbol{B}^{-1/2}=\boldsymbol{B}^{-1/2}\boldsymbol{A}\boldsymbol{B}^{-1/2}-\boldsymbol{E}$$

будет неотрицательно определенной. Действительно,

$$\boldsymbol{x}^T\boldsymbol{D}\boldsymbol{x}=(\boldsymbol{x}^T\boldsymbol{B}^{-1/2})(\boldsymbol{A}-\boldsymbol{B})(\boldsymbol{B}^{-1/2}\boldsymbol{x})=\boldsymbol{y}^T(\boldsymbol{A}-\boldsymbol{B})\boldsymbol{y}\geqslant 0,$$

где $\boldsymbol{y}=\boldsymbol{B}^{-1/2}\boldsymbol{x}$. Следовательно, $\det(\boldsymbol{B}^{-1/2}\boldsymbol{A}\boldsymbol{B}^{-1/2})\geqslant 1$, т. е. $\det\boldsymbol{A}\det\boldsymbol{B}^{-1}\geqslant 1$, причем равенство имеет место тогда и только тогда, когда $\boldsymbol{B}^{-1/2}\boldsymbol{A}\boldsymbol{B}^{-1/2}=\boldsymbol{E}$, т. е. когда $\boldsymbol{A}=\boldsymbol{B}$.

Вопрос 4. Почему условие $\det\boldsymbol{B}>0$ необходимо?

2. Вычислим частную производную по x_{kr} функции

$$G=C_1F_1=\sum_{i<j}(\delta_{ij}-d_{ij})^2/d_{ij},\quad\text{где }\delta_{ij}=\left[\sum_{l=1}^{m}(x_{il}-x_{jl})^2\right]^{1/2}.$$

Если $i\neq k$ или $j\neq k$, то $\partial[(\delta_{ij}-d_{ij})^2/d_{ij}]/\partial x_{kr}=0$. Пусть сначала $i=k$, а $j=i+1,\ldots,n$ (рис. 27). Тогда

$$a_j=\frac{\partial[(\delta_{ij}-d_{ij})^2/d_{ij}]}{\partial x_{ir}}=\frac{2(\delta_{ij}-d_{ij})}{d_{ij}}\,\frac{\partial\delta_{ij}}{\partial x_{ir}}.$$

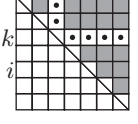


Рис. 27

Подсчитаем отдельно $\partial\delta_{ij}/\partial x_{ir}$:

$$\frac{\partial\delta_{ij}}{\partial x_{ir}}=\frac{1}{2\delta_{ij}}\,\frac{\partial}{\partial x_{ir}}\left[\sum_{l=1}^{m}(x_{il}-x_{jl})^2\right]=\frac{x_{ir}-x_{jr}}{\delta_{ij}}.$$

Подставляя этот результат в предыдущую формулу, получим

$$a_j=2(\delta_{ij}-d_{ij})(x_{ir}-x_{jr})/[\delta_{ij}d_{ij}].$$

Аналогично, в случае $j = k$ и $i = 1, \ldots, j-1$ имеем

$$b_i = \frac{\partial[(\delta_{ij} - d_{ij})^2/d_{ij}]}{\partial x_{jr}} = -2(\delta_{ij} - d_{ij})(x_{ir} - x_{jr})/[\delta_{ij} d_{ij}].$$

Наконец, сложим ненулевые слагаемые и поменяем индексы:

$$\frac{\partial G}{\partial x_{il}} = \sum_{j=i+1}^{n} a_j + \sum_{i=1}^{j-1} b_i = 2 \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{\delta_{ij} - d_{ij}}{\delta_{ij} d_{ij}} \, (x_{il} - x_{jl}).$$

3. Для краткости введем обозначения: $A = \sum (i - T_i)^2$, $B = \sum \left(i - \frac{n+1}{2} \right)^2$, $C = \sum \left(i - \frac{n+1}{2} \right) \left(T_i - \frac{n+1}{2} \right)$. Тогда

$$A = \sum \left[\left(i - \frac{n+1}{2} \right) - \left(T_i - \frac{n+1}{2} \right) \right]^2 = B - 2C + \sum \left(T_i - \frac{n+1}{2} \right)^2.$$

Поскольку последняя сумма также равна B, получаем равенство $A = 2B - 2C \Longleftrightarrow 1 - \frac{1}{2}\, A/B = C/B = \rho_S$.

4. Дополним обозначения из решения предыдущей задачи еще двумя: $\xi_i = R_i - \frac{n+1}{2}$ и $\eta_i = S_i - \frac{n+1}{2}$. При этом $B\rho_S = C = \sum \xi_i \eta_i$. Возводя в квадрат, получим соотношение

$$B^2 \rho_S^2 = \sum_{i,j} \xi_i \xi_j \eta_i \eta_j. \tag{30}$$

Ввиду того, что при справедливости гипотезы H_0 случайные величины ξ_i и η_i независимы, из формулы (30) следует равенство

$$B^2 \mathbf{M} \rho_S^2 = B^2 \mathbf{D} \rho_S = \sum_{i,j} (\mathbf{M} \xi_i \xi_j \cdot \mathbf{M} \eta_i \eta_j).$$

Эта сумма содержит n слагаемых с индексами $i = j$, и так как все индексы равноправны, то все эти слагаемые одинаковы. Точно так же равны друг другу остальные $n(n-1)$ слагаемых с индексами $i \neq j$. Поэтому

$$B^2 \mathbf{D} \rho_S = n \mathbf{M} \xi_1^2 \cdot \mathbf{M} \eta_1^2 + n(n-1) \mathbf{M} \xi_1 \xi_2 \cdot \mathbf{M} \eta_1 \eta_2. \tag{31}$$

В силу тождества $\sum \xi_i = 0$, имеем

$$0 = \mathbf{M} \left(\xi_1 \sum \xi_i \right) = \mathbf{M} \xi_1^2 + (n-1) \mathbf{M} \xi_1 \xi_2,$$

Следовательно, $\mathbf{M} \xi_1 \xi_2 = -\frac{1}{n-1}\, \mathbf{M} \xi_1^2$. Если взять математические ожидания от обеих частей равенства $\sum \xi_i^2 = B$, то найдем, что $\mathbf{M} \xi_1^2 = B/n$. Поэтому $\mathbf{M} \xi_1 \xi_2 = -B/[n(n-1)]$. Аналогичные формулы верны, конечно, и для $\mathbf{M} \eta_1^2$ и $\mathbf{M} \eta_1 \eta_2$. Подставив все в равенство (31), получим $B^2 \mathbf{D} \rho_S = B^2/n + B^2/[n(n-1)] = B^2/(n-1)$. Остается только сократить обе части на B^2.

5. Пусть, как и прежде, $A = \sum (i - T_i)^2$. Ввиду формулы (22), достаточно показать, что $A = 2D$, где

$$D = \sum_{i<j} (j - i)\, I_{\{T_i > T_j\}}.$$

Добавим к правой части и вычтем из нее $E = \sum\limits_{j<i} j\, I_{\{T_i > T_j\}}$:

$$\begin{aligned} D &= \sum_{i<j} j\, I_{\{T_i>T_j\}} - \sum_{i<j} i\, I_{\{T_i>T_j\}} + E - E = \\ &= \sum_{i,j} j\, I_{\{T_i>T_j\}} - \sum_{i<j} i \left[I_{\{T_i>T_j\}} + I_{\{T_j>T_i\}} \right] = \\ &= \sum_{j=1}^{n} j \sum_{i=1}^{n} I_{\{T_i>T_j\}} - \sum_{i=1}^{n-1} i \sum_{j=i+1}^{n} 1. \end{aligned}$$

Заметим, что $\sum\limits_{i=1}^{n} I_{\{T_i>T_j\}}$ равна числу тех T_i $(i = 1, \ldots, n)$, которые больше, чем T_j, т. е. равна $n - T_j$. Следовательно,

$$D = \sum_{j=1}^{n} j(n - T_j) - \sum_{i=1}^{n-1} i(n - i).$$

Заменяя в верхнем пределе второй суммы $(n-1)$ на n и приводя подобные члены — суммы, получаем

$$D = \sum_{i=1}^{n} i^2 - \sum_{j=1}^{n} jT_j.$$

Остается заметить, что $A = \sum (i - T_i)^2 = 2\sum i^2 - 2\sum iT_i$.

6. Введем обозначения $a_i = (x_i + y_i)/2$ и $b_i = (x_i - y_i)/2$. Используя тождество $AB = [(A+B)/2]^2 - [(A-B)/2]^2$, запишем для левой части представление

$$\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}) = \sum_{i=1}^{n} (a_i - \overline{a})^2 + \sum_{i=1}^{n} (b_i - \overline{b})^2.$$

Аналогично для правой части получаем равенство

$$\frac{1}{n} \sum_{i<j} (x_j - x_i)(y_j - y_i) = \frac{1}{n} \sum_{i<j} (a_j - a_i)^2 + \frac{1}{n} \sum_{i<j} (b_j - b_i)^2.$$

Завершает доказательство применение теоремы о межточечных расстояниях из решения задачи 5 гл. 16 для $m_i = 1$, $m = n$.

7. Запишем функцию правдоподобия сгруппированной выборки $L(\boldsymbol{\theta}) = c \prod\limits_{i,\,j} (r_i s_j)^{\nu_{ij}} = c \prod\limits_{i} r_i^{n_i} \prod\limits_{j} s_j^{m_j}$, где параметр $\boldsymbol{\theta}$ является вектором $\boldsymbol{\theta} = (r_1, \ldots, r_n, s_1, \ldots, s_m)$, и c от него не зависит. Таким образом, задача равносильна максимизации по $\boldsymbol{\theta}$ функции

$$f(\boldsymbol{\theta}) = \sum_{i} n_i \ln r_i + \sum_{j} m_j \ln s_j \quad \text{при выполнении условий}$$

$$g(\boldsymbol{\theta}) = 1 - \sum_{i} r_i = 0, \quad h(\boldsymbol{\theta}) = 1 - \sum_{j} s_j = 0.$$

Запишем систему уравнений для поиска экстремальных точек функции Лагранжа $F(\boldsymbol{\theta},\lambda,\mu) = f(\boldsymbol{\theta}) + \lambda\, g(\boldsymbol{\theta}) + \mu\, h(\boldsymbol{\theta})$:

$$\partial F/\partial r_i = n_i/r_i - \lambda = 0, \quad i = 1,\ldots,n;$$
$$\partial F/\partial s_j = m_j/s_j - \mu = 0, \quad j = 1,\ldots,m;$$
$$\partial F/\partial \lambda = g(\boldsymbol{\theta}) = 0, \quad \partial F/\partial \mu = h(\boldsymbol{\theta}) = 0.$$

Из n первых уравнений находим, что $r_i = n_i/\lambda$. Подставляя найденные r_i в уравнение $g(\boldsymbol{\theta}) = 0$, получим $\lambda = \sum n_i = N$, откуда $\widehat{r}_j = n_i/N$. Аналогично из оставшихся уравнений находим, что $\mu = \sum m_j = N$ и $\widehat{s}_j = m_j/N$.

ОТВЕТЫ НА ВОПРОСЫ

1. Евклидова норма матрицы при транспонировании не меняется: $|\boldsymbol{Q} - \boldsymbol{Y}\boldsymbol{P}|^2 = |\boldsymbol{Q}^T - \boldsymbol{P}^T\boldsymbol{Y}^T|^2$. Минимум достигается на $\widehat{\boldsymbol{Y}}^T = (\boldsymbol{P}\boldsymbol{P}^T)^{-1}\boldsymbol{P}\boldsymbol{Q}^T$. Остается только транспонировать обе части этого равенства с учетом свойств из П10.

2. Подставим $n - i + 1$ вместо T_i в правую часть формулы (21), получим, что второй сомножитель равен $-\left(i - \dfrac{n+1}{2}\right)$.

3. Поскольку $20 \cdot 14 \cdot 0{,}09 = 25{,}2$ больше, чем 5%-ное критическое значение 23,7 закона χ^2_{14}, то, несмотря на малость числа 0,09, гипотеза об отсутствии связи отвергается.

4. Необходимость условия $\det \boldsymbol{B} > 0$ показывает пример матриц $\boldsymbol{A} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ и $\boldsymbol{B} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$.

Глава 21

РЕГРЕССИЯ

К регрессионному анализу относятся задачи выявления искаженной случайным «шумом» функциональной зависимости интересующего исследователя показателя Y от измеряемых переменных $X_1,\ldots,X_m$. Данными служит таблица экспериментально полученных «зашумленных» значений Y на разных наборах $x_1,\ldots,x_m$. Основной целью обычно является как можно более точный прогноз (предсказание) Y на основе измеряемых (*предикторных*) переменных.

Регрессионный анализ по праву может быть назван основным методом современной математической статистики.

Н. Дрейлер, Г. Смит

Predict (*англ.*) — предсказывать.

§ 1. ПОДГОНКА ПРЯМОЙ

Термин «регрессия» ввел Ф. Гальтон в своей статье «Регрессия к середине в наследовании роста» (1885 г.), в которой он сравнивал средний рост детей Y со средним ростом их родителей X (на основе данных о 928 взрослых детях и 205 их родителях). Гальтон заметил, что рост детей у высоких (низких) родителей обычно также выше (ниже) среднего роста популяции $\mu \approx \overline{X} \approx \overline{Y}$, но при этом отклонение от μ у детей меньше, чем у родителей. Другими словами, экстремумы в следующем поколении сглаживаются, происходит возвращение назад (*регрессия*) к середине. По существу, Гальтон показал, что зависимость Y от X хорошо выражается уравнением $Y - \overline{Y} = (2/3)(X - \overline{X})$ (рис. 1).

Если кто-то способен предсказать, чем закончатся его исследования, то эта проблема не очень глубока и, можно сказать, практически не существует.

А. Шильд

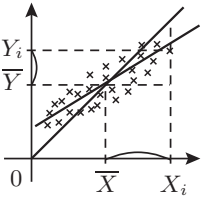


Рис. 1

Позднее регрессией стали называть любую функциональную зависимость между случайными величинами, даже в тех ситуациях, когда предикторные переменные являются неслучайными. В примечании переводчиков на с. 26 книги [23] высказано интересное мнение по поводу «живучести» термина «регрессия».

> «Можно предположить, что его удивительная устойчивость связана с переосмыслением значения. Постепенно исходная антропометрическая задача, занимавшая Гальтона, была забыта, а интерпретация вытеснилась благодаря ассоциативной связи с понятием «регресс», т. е. движение назад. Сначала берутся данные, а уж потом, задним числом, проводится их обработка. Такое понимание пришло на смену традиционной, еще средневековой, априорной модели, для которой данные были лишь инструментом подтверждения. Негативный оттенок, присущий понятию «регресс», думается и вызывает психологический дискомфорт, поскольку воспринимается одновременно с понятиями, описывающими такой прогрессивный метод, как регрессионный анализ.»

Проиллюстрируем основные идеи регрессии на примере подгонки прямой под «облако» экспериментальных точек (x_i,η_i), полученных в соответствии с моделью

$$\eta_i = a + bx_i + \varepsilon_i, \quad i = 1,\ldots,n. \tag{1}$$

Здесь коэффициенты прямой a и b — неизвестные параметры, x_i — (неслучайные) значения предиктора X (для простоты допустим, что $x_i \neq x_j$), ε_i — независимые и одинаково распределенные случайные ошибки, $\mathbf{M}\varepsilon_i = 0$. Для нахождения оценок коэффициентов a и b применим

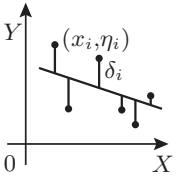


Рис. 2

Метод наименьших квадратов (МНК).

Естественным условием точности подгонки *пробной прямой* $y = \alpha + \beta x$ служит близость к нулю всех *остатков* $\delta_i(\alpha,\beta) = \eta_i - \alpha - \beta x_i$ (рис. 2). Общую меру близости к нулю можно выбирать по-разному (например, $\max|\delta_i|$ или $\sum|\delta_i|$), но наиболее простые формулы для оценок $\widehat{a}$ и $\widehat{b}$ получаются, если в качестве такой меры взять

$$F(\alpha,\beta) = \sum_{i=1}^{n} \delta_i^2 = \sum_{i=1}^{n} (\eta_i - \alpha - \beta x_i)^2. \tag{2}$$

Минимум $F(\alpha,\beta)$ достигается (см. задачу 1) в точке $(\widehat{a},\widehat{b})$, где

$$\widehat{b} = \sum_{i=1}^{n} (\eta_i - \overline{\eta})(x_i - \overline{x}) \Big/ \sum_{i=1}^{n} (x_i - \overline{x})^2\,, \quad \widehat{a} = \overline{\eta} - \widehat{b}\,\overline{x}. \tag{3}$$

А. М. Лежандр (1752–1833), французский математик.

Некоторые детали спора о приоритете между Лежандром и Гауссом приведены в примечании переводчиков к с. 32 книги [23].

Метод наименьших квадратов был впервые опубликован в 1805 г. Лежандром в работе, посвященной нахождению орбит комет. К. Гаусс утверждал, что использовал МНК еще до 1803 г.

Замечание 1. Прямая, подогнанная под «облако» точек МНК, вообще говоря, отличается от первой главной компоненты (см. § 1 гл. 20). Дело в том, что при построении главной компоненты переменные X и Y считаются равноправными, и минимизируется сумма квадратов длин отрезков, перпендикулярных компоненте, а не направленных вдоль оси Y, как в случае МНК (рис. 3).

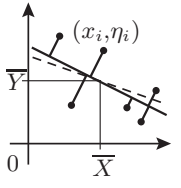


Рис. 3

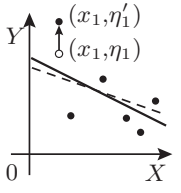


Рис. 4

Недостатком метода наименьших квадратов является излишняя *чувствительность* МНК-оценок к выделяющимся наблюдениям («выбросам»), возникающая вследствие зависимости меры F не от самих остатков δ_i, а от их квадратов. Стремление уменьшить остатки в точках «выбросов» может привести к значительному *смещению* оценок параметров. Так, если переместить вверх точку (x_1,η_1) на рис. 4, то, чтобы уменьшить δ_1^2, МНК-прямая довольно сильно повернется.

Одной из устойчивых к «выбросам» (*робастных*) альтернатив МНК может служить

Метод Тейла (см. [88, с. 215])

В этом методе оценки коэффициентов прямой задаются следующими формулами:

$$\tilde{b} = MED\{(\eta_j - \eta_i)/(x_j - x_i), \quad 1 \leqslant i < j \leqslant n\}, \tag{4}$$

$$\tilde{a} = MED\{\eta_i - \tilde{b}x_i, \quad i = 1, \ldots, n\}.$$

Причина робастности метода Тейла кроется в том, что одиночный «выброс» может исказить самое большее $(n-1)$ оценок $(\eta_j - \eta_i)/(x_j - x_i)$ коэффициента наклона b, в то время как медиана вычисляется по $n(n-1)/2$ оценкам.

Пример 1 ([86, с. 234])**.** Закон Хаббла в астрономии гласит: скорость удаления галактики прямо пропорциональна расстоянию до нее. В таблице ниже указаны расстояния Y (в миллионах световых лет) и скорости X (в сотнях миль в секунду) для 11 галактических созвездий (см. [66]).

Дистанции огромного размера...

Скалозуб в «Горе от ума» А. С. Грибоедова

Требуется подогнать прямую $Y = bX$ к этим данным.

Созвездие	X	Y	Y/X
Дева (Virgo)	22	7,5	0,341
Пегас (Pegasus)	68	24	0,353
Персей (Perseus)	108	32	0,296
Волосы Вероники (Coma Berenices)	137	47	0,343
Большая Медведица (Ursa Major No. 1)	255	93	0,365
Большая Медведица (Ursa Major No. 2)	700	260	0,371
Лев (Leo)	315	120	0,381
Северная Корона (Corona Borealis)	390	134	0,344
Близнецы (Gemini)	405	144	0,356
Волопас (Bootes)	685	245	0,358
Гидра (Hydra)	1100	380	0,345

Имеется $11 \times 10/2 = 55$ попарных наклонов, начиная с наименьшего 0,187 для Льва и Северной Короны. Медиана наклонов $\tilde{b} = 0{,}359$, оценка метода наименьших квадратов $\widehat{b} = 0{,}353 \approx \tilde{b}$.

Замечание 2. Интересно, что оценка $\tilde{b}$ в формуле (4) связана с ранговым коэффициентом Кендэла τ (см. § 7 гл. 20), причем эта связь аналогична связи МНК-оценки $\widehat{b}$ в формуле (3) с обычным выборочным коэффициентом корреляции $\widehat{\rho}$.

Действительно, занумеруем наблюдения так, что $x_1 < \ldots < x_n$. Если из η_i вычесть истинные значения bx_i, то $\eta_i - bx_i = a + \varepsilon_i$ $(i = 1, \ldots, n)$ образуют выборку (набор независимых и одинаково распределенных случайных величин). Не зная b, будем вычитать из η_i величины βx_i, где β меняется по нашему произволу. Чем ближе β к b, тем больше $\eta_i - \beta x_i$ будут похожи на выборку. В противном случае, они будут проявлять тенденцию к возрастанию (или убыванию) вместе с номером i (это зависит

от знака $b-\beta$). В этом легко убедиться, записав $\eta_i-\beta x_i$ в виде $\eta_i-\beta x_i=\eta_i-bx_i+x_i(b-\beta)=a+\varepsilon_i+x_i(b-\beta)$, из которого понятно, что в силу возрастания x_i с увеличением i у $\eta_i-\beta x_i$ появляется положительный (или отрицательный) «снос».

Тенденцию к изменению $\eta_i-\beta x_i$ с ростом i (или ее отсутствие) можно исследовать с помощью коэффициентов корреляции. Возьмем сначала *обычный выборочный коэффициент корреляции* $\widehat{\rho}(\beta)$ между рядами $(x_1,\ldots,x_n)$ и $(\eta_1-\beta x_1,\ldots,\eta_n-\beta x_n)$:

$$\widehat{\rho}(\beta)=\frac{\sum(x_i-\overline{x})[(\eta_i-\overline{\eta})-\beta(x_i-\overline{x})]}{\sqrt{\sum(x_i-\overline{x})^2}\sqrt{\sum[(\eta_i-\overline{\eta})-\beta(x_i-\overline{x})]^2}}\,.$$

Наименьшей зависимости $\eta_i-\beta x_i$ от x_i $(i=1,\ldots,n)$ соответствует значение $\widehat{\rho}=0$. По отношению к β это дает уравнение

$$\sum_{i=1}^{n}(x_i-\overline{x})(\eta_i-\overline{\eta})=\beta\sum_{i=1}^{n}(x_i-\overline{x})^2,$$

решением которого и является МНК-оценка $\widehat{b}$ из формулы (3).

В случае *коэффициента Кендэла* τ заменим x_i и $\eta_i-\beta x_i$ их рангами i и T_i соответственно. Тогда (см. формулу (24) гл. 20)

$$\tau=\frac{2}{n(n-1)}\sum_{i<j}\operatorname{sign}(T_j-T_i)=\frac{2}{n(n-1)}\sum_{i<j}\operatorname{sign}(\eta_j-\beta x_j-\eta_i+\beta x_i).$$

Правую часть можно переписать также в виде

$$\tau(\beta)=\frac{2}{n(n-1)}\sum_{i<j}\operatorname{sign}\left(\frac{\eta_j-\eta_i}{x_j-x_i}-\beta\right).$$

Рассуждения, аналогичные проведенным в комментарии 3 к критерию знаков из § 2 гл. 15, показывают, что решением уравнения $\tau(\beta)=0$ является оценка $\tilde{b}$, задаваемая формулой (4).

Какими свойствами обладают МНК-оценки? Как вытекает из задачи 2, в случае, когда вектор ошибок $\boldsymbol{\varepsilon}=(\varepsilon_1,\ldots,\varepsilon_n)$ *нормально распределен*, МНК совпадает с методом максимального правдоподобия (см. § 4 гл. 9) и, следовательно, является наиболее точным для больших выборок (асимптотически эффективным). Другие статистические свойства МНК-оценок обсуждаются в § 3.

§ 2. ЛИНЕЙНАЯ РЕГРЕССИОННАЯ МОДЕЛЬ

Предположим, что (с точностью до случайных ошибок) целевая переменная Y есть *линейная комбинация* $\theta_1X_1+\ldots+\theta_mX_m$ предикторных переменных $X_1,\ldots,X_m$ с неизвестными коэффициентами $\theta_1,\ldots,\theta_m$.

Измерения η_i переменной Y $(i=1,\ldots,n$, где $n\geqslant m)$, отвечающие заданным (не обязательно различным) значениям $x_{i1},\ldots,x_{im}$

предикторных переменных, имеют вид

$$\eta_i = \theta_1 x_{i1} + \ldots + \theta_m x_{im} + \varepsilon_i,$$

где ε_i — случайные ошибки.

Вводя для векторов и матриц обозначения $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)^T$, $\boldsymbol{X} = \|x_{il}\|_{n\times m}$, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)^T$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^T$, можно записать модель в матричной форме:

$$\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}. \tag{5}$$

Матрицу $\boldsymbol{X}$ называют *матрицей плана эксперимента*.

Будем предполагать, что для модели (5) выполняется допущение

Д1. *Столбцы* $\boldsymbol{x}_l = (x_{1l}, \ldots, x_{nl})^T$, $l = 1, \ldots, m$, *матрицы* $\boldsymbol{X}$ *линейно независимы. Иными словами, ввиду выполнения неравенства* $n \geqslant m$ *матрица* $\boldsymbol{X}$ *имеет ранг* m (см. П10).

Пример 2. *Подгонка полинома*. Рассмотрим $\boldsymbol{x}_1 = (1, \ldots, 1)^T$ и $\boldsymbol{x}_{l+1} = (u_1^l, \ldots, u_n^l)^T$, $l = 1, \ldots, m-1$. Здесь $u_1 < \ldots < u_n$ — так называемые «узлы», в которых вычисляются значения многочлена

$$p(u) = \theta_1 + \theta_2 u + \theta_3 u^2 + \ldots + \theta_m u^{m-1}$$

и «зашумляются» случайными ошибками ε_i (рис. 5 для $m = 4$). Матрица $\boldsymbol{X}$ имеет вид

$$\begin{pmatrix} 1 & u_1 & u_1^2 & \ldots & u_1^{m-1} \\ 1 & u_2 & u_2^2 & \ldots & u_2^{m-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & u_n & u_n^2 & \ldots & u_n^{m-1} \end{pmatrix}.$$

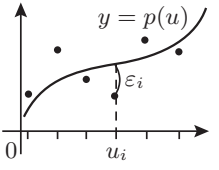


Рис. 5

Определитель подматрицы $\tilde{\boldsymbol{X}}$, образованной первыми m строками $\boldsymbol{X}$ ($m \leqslant n$), является известным из линейной алгебры *определителем Вандермонда*: $\det \tilde{\boldsymbol{X}} = \prod_{1\leqslant i<j\leqslant m}(u_j - u_i)$. Он отличен от нуля, поскольку «узлы» различны. Поэтому ранг подматрицы $\widetilde{\boldsymbol{X}}$ (и матрицы $\boldsymbol{X}$) равен m, а столбцы $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ линейно независимы.

В дальнейшем при изучении статистических свойств оценок параметров $\theta_1, \ldots, \theta_m$ (см. § 3) будем считать выполняющимся также допущение

Д2. *Случайные величины* $\varepsilon_1, \ldots, \varepsilon_n$ *одинаково распределены с* $\mathbf{M}\varepsilon_i = 0$, $\mathbf{D}\varepsilon_i = \sigma^2$ (параметр $0 < \sigma < \infty$ также неизвестен) *и некоррелированы:* $\mathbf{M}\varepsilon_i\varepsilon_j = 0$ *при* $i \neq j$.

Оценим параметры $\theta_1, \ldots, \theta_m$ **методом наименьших квадратов**, минимизируя по $\boldsymbol{\theta}$ функцию $F(\boldsymbol{\theta}) = \sum\limits_{i=1}^{n}(\eta_i - \theta_1 x_{i1} - \ldots - \theta_m x_{im})^2$. Точка ее минимума $\widehat{\boldsymbol{\theta}}$ называется *МНК-оценкой*, вектор $\widehat{\boldsymbol{\delta}}$ с

компонентами $\widehat{\delta}_i = \eta_i - \widehat{\theta}_1 x_{i1} - \ldots - \widehat{\theta}_m x_{im}$, $i = 1,\ldots,n$, — *вектором остатков*, а значение в точке минимума $RSS = F(\widehat{\boldsymbol{\theta}})$ — *остаточной суммой квадратов*.*)

Для нахождения оценки $\widehat{\boldsymbol{\theta}}$ интерпретируем задачу минимизации функции $F(\boldsymbol{\theta})$ в терминах пространства $\mathbb{R}^n$. Рассмотрим в $\mathbb{R}^n$ подпространство $L(\boldsymbol{X})$, порождаемое столбцами $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_m$ матрицы $\boldsymbol{X}$. Очевидно, что вектор $\boldsymbol{X\theta} = \theta_1\boldsymbol{x}_1 + \ldots + \theta_m\boldsymbol{x}_m$ пробегает $L(\boldsymbol{X})$, когда $\boldsymbol{\theta}$ пробегает $\mathbb{R}^m$. Поскольку

$$F(\boldsymbol{\theta}) = |\boldsymbol{\eta} - \boldsymbol{X\theta}|^2,$$

видим, что минимизация по $\boldsymbol{\theta}$ равносильна нахождению в подпространстве $L(\boldsymbol{X})$ вектора $\boldsymbol{X}\widehat{\boldsymbol{\theta}}$, наименее удаленного от $\boldsymbol{\eta}$. Как известно, таким вектором служит *ортогональная проекция* $\boldsymbol{\eta}$ на $L(\boldsymbol{X})$ (рис. 6). Следовательно, вектор остатков $\widehat{\boldsymbol{\delta}} = \boldsymbol{\eta} - \boldsymbol{X}\widehat{\boldsymbol{\theta}}$ должен быть ортогонален векторам $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_m$, порождающим подпространство $L(\boldsymbol{X})$, т. е.

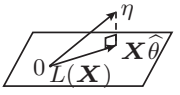


Рис. 6

$$\boldsymbol{X}^T(\boldsymbol{\eta} - \boldsymbol{X}\widehat{\boldsymbol{\theta}}) = \boldsymbol{0} \quad \text{или} \quad (\boldsymbol{X}^T\boldsymbol{X})\,\widehat{\boldsymbol{\theta}} = \boldsymbol{X}^T\boldsymbol{\eta}. \tag{6}$$

Соотношение (6) представляет собой систему линейных уравнений относительно $\widehat{\boldsymbol{\theta}}$ с *положительно определенной* (задача 3) матрицей $\boldsymbol{B} = \boldsymbol{X}^T\boldsymbol{X}$, которую называют *информационной*. Решить систему (6) можно, например, методом Холецкого (П10). Так как ввиду положительной определенности матрица $\boldsymbol{B}$ является невырожденнной, для МНК-оценки $\widehat{\boldsymbol{\theta}}$ получаем представление

$$\widehat{\boldsymbol{\theta}} = \boldsymbol{B}^{-1}\boldsymbol{X}^T\boldsymbol{\eta}. \tag{7}$$

Пример 3 ([2, с. 177], [27, с. 57])**.** На рис. 7 точками изображены результаты эксперимента по изучению зависимости между скоростью автомобиля V (в милях/час) и расстоянием Y (в футах), пройденным им после сигнала об остановке. Для каждого отдельного случая результат определяется в основном тремя факторами: скоростью V в момент подачи сигнала, временем реакции θ_1 водителя на этот сигнал и тормозами автомобиля. Автомобиль успеет проехать путь $\theta_1 V$ до момента включения водителем тормозов и еще $\theta_2 V^2$ после этого момента, поскольку согласно элементарным физическим законам теоретическое расстояние, пройденное до остановки с момента торможения, пропорционально квадрату скорости (убедитесь!).

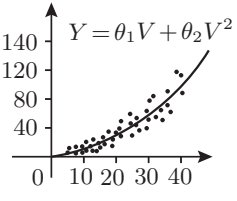


Рис. 7

Таким образом, в качестве модели годится $Y = \theta_1 V + \theta_2 V^2$. Для экспериментальных данных по формуле (7) были подсчитаны значения $\widehat{\theta}_1 = 0{,}76$ и $\widehat{\theta}_2 = 0{,}056$ (график параболы $Y = \widehat{\theta}_1 V + \widehat{\theta}_2 V^2$ приведен на рис. 7).

*) В обозначении использованы первые буквы соответствующего английского термина «Residual Sum of Squares».

Замечание 3. Удобно иметь дело с матрицей $\boldsymbol{X}$, столбцы которой *ортогональны* друг другу. В этом случае формула (7) упрощается: так как матрица $\boldsymbol{B}$ оказывается диагональной с элементами $|\boldsymbol{x}_l|^2$ $(l = 1, \ldots, m)$ на главной диагонали, из представления (7) получаем выражения

$$\widehat{\theta}_l = \boldsymbol{x}_l^T \boldsymbol{\eta} \,/ |\boldsymbol{x}_l|^2 \,, \quad l = 1, \ldots, m. \tag{8}$$

Если исходный набор линейно независимых векторов $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ таким свойством не обладает, то его можно ортогонализировать с помощью стандартной *процедуры Грама—Шмидта*, вводя новые векторы

$$\boldsymbol{x}'_l = \boldsymbol{x}_l - a_{(l-1)l}\boldsymbol{x}'_{l-1} - \ldots - a_{1l}\boldsymbol{x}'_1, \quad l = 1, \ldots, m, \tag{9}$$

где $a_{il} = \boldsymbol{x}_l^T \boldsymbol{x}'_i \,/ |\boldsymbol{x}'_i|^2$ (подробнее см. [43, с. 111]). Выражая $\boldsymbol{x}_l$ через $\boldsymbol{x}'_l, \ldots, \boldsymbol{x}'_1$, получим

$$\boldsymbol{x}_l = \boldsymbol{x}'_l + a_{(l-1)l}\boldsymbol{x}'_{l-1} + \ldots + a_{1l}\boldsymbol{x}'_1, \quad l = 1, \ldots, m. \tag{10}$$

Соотношения (10) можно записать в матричной форме: $\boldsymbol{X} = \boldsymbol{X}'\boldsymbol{A}$, где $\boldsymbol{A}$ — верхнетреугольная матрица с единицами на главной диагонали (она, очевидно, обратима). Подставляя это выражение в формулу (5), приходим к модели $\boldsymbol{\eta} = \boldsymbol{X}'\boldsymbol{\theta}' + \boldsymbol{\varepsilon}$, где $\boldsymbol{\theta}' = \boldsymbol{A}^{-1}\boldsymbol{\theta}$. Исходный вектор $\boldsymbol{\theta}$ восстанавливается по $\boldsymbol{\theta}'$ с помощью формулы

$$\boldsymbol{\theta} = \boldsymbol{A}\boldsymbol{\theta}'. \tag{11}$$

§ 3. СТАТИСТИЧЕСКИЕ СВОЙСТВА МНК-ОЦЕНОК

МНК-оценки параметров в линейной регрессионной модели (5) при выполнении допущений Д1 и Д2 обладают **рядом важных свойств**.

Утверждение 1.

1) Оценка $\widehat{\boldsymbol{\theta}}$, задаваемая формулой (7), является несмещенной, т. е. $\mathsf{M}\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}$.

2) *Матрицей ковариаций* $\mathbf{Cov}(\widehat{\boldsymbol{\theta}}) = \|\mathbf{cov}(\widehat{\theta}_k, \widehat{\theta}_l)\|_{m\times m}$ служит матрица $\sigma^2\boldsymbol{B}^{-1}$.

3) Для любого вектора $\boldsymbol{c} \in \mathbb{R}^m$ несмещенной оценкой для величины $\boldsymbol{c}^T\boldsymbol{\theta}$ служит $\boldsymbol{c}^T\widehat{\boldsymbol{\theta}}$, причем $\mathsf{D}\,\boldsymbol{c}^T\widehat{\boldsymbol{\theta}} = \sigma^2\boldsymbol{c}^T\boldsymbol{B}^{-1}\boldsymbol{c}$.

Доказательство. Воспользуемся формулами из П2

$$\mathsf{M}(\boldsymbol{A}\boldsymbol{\xi}) = \boldsymbol{A}\,\mathsf{M}\boldsymbol{\xi}, \quad \mathbf{Cov}(\boldsymbol{A}\boldsymbol{\xi}) = \boldsymbol{A}\,\mathbf{Cov}(\boldsymbol{\xi})\,\boldsymbol{A}^T, \tag{12}$$

где $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_m)^T$ — произвольный случайный вектор, для которого определены $\mathsf{M}\boldsymbol{\xi}$ и $\mathbf{Cov}(\boldsymbol{\xi})$, $\boldsymbol{A}$ — числовая $(k \times m)$-матрица.

Применим формулы (12) к (7) с учетом того, что для вектора ошибок $\boldsymbol{\varepsilon}$ в модели (5) выполняются равенства $\mathsf{M}\boldsymbol{\varepsilon} = \mathbf{0}$, $\mathbf{Cov}(\boldsymbol{\varepsilon}) = \sigma^2\boldsymbol{E}$:

$$\begin{aligned}
&\mathsf{M}\widehat{\boldsymbol{\theta}} = \boldsymbol{B}^{-1}\boldsymbol{X}^T\mathsf{M}\boldsymbol{\eta} = \boldsymbol{B}^{-1}\boldsymbol{X}^T\mathsf{M}(\boldsymbol{X}\boldsymbol{\theta}+\boldsymbol{\varepsilon}) = \\
&\quad = \boldsymbol{B}^{-1}\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{B}^{-1}\boldsymbol{X}^T\mathsf{M}\boldsymbol{\varepsilon} = \boldsymbol{\theta}, \\
&\mathbf{Cov}(\widehat{\boldsymbol{\theta}}) = \boldsymbol{B}^{-1}\boldsymbol{X}^T\mathbf{Cov}(\boldsymbol{\eta})\,\boldsymbol{X}\boldsymbol{B}^{-1} = \boldsymbol{B}^{-1}\boldsymbol{X}^T\mathbf{Cov}(\boldsymbol{\varepsilon})\,\boldsymbol{X}\boldsymbol{B}^{-1} = \\
&\quad = \boldsymbol{B}^{-1}\boldsymbol{X}^T\sigma^2\boldsymbol{E}\boldsymbol{X}\boldsymbol{B}^{-1} = \sigma^2\boldsymbol{B}^{-1}\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{B}^{-1} = \sigma^2\boldsymbol{B}^{-1}.
\end{aligned}$$

Свойство 3 сразу следует из свойства 2 и второй из формул (12). ■

Замечание 4. В случае, когда столбцы $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_m$ матрицы $\boldsymbol{X}$ ортогональны друг другу, из свойства 2 вытекает, что $\mathsf{D}\widehat{\theta}_l = \sigma^2/|\boldsymbol{x}_l|^2$, $l = 1,\ldots,m$. Оказывается, что в общем случае $\mathsf{D}\widehat{\theta}_l \geqslant \sigma^2/|\boldsymbol{x}_l|^2$, причем равенство при всех l достигается только для ортогональных столбцов (см. [71, с. 62]). Тем самым, так называемое *ортогональное планирование эксперимента*, при котором столбцы матрицы $\boldsymbol{X}$ выбираются ортогональными, является в приведенном выше смысле оптимальным.

Для получения более содержательных заключений о распределении оценки $\widehat{\boldsymbol{\theta}}$ предположим, что выполняется допущение

Д3. *Вектор $\boldsymbol{\varepsilon}$ имеет нормальное распределение* (см. П9).

Замечание 5. Если верны допущения Д1—Д2, то при любом векторе $\boldsymbol{c} \in \mathbb{R}^m$ оценка $\boldsymbol{c}^T\widehat{\boldsymbol{\theta}}$ имеет минимальную дисперсию в классе *линейных* (вида $\boldsymbol{d}^T\boldsymbol{\eta}$) *несмещенных оценок* для $\boldsymbol{c}^T\boldsymbol{\theta}$, а при справедливости также и допущения Д3 этот класс можно расширить до множества *произвольных несмещенных оценок* (см. [71, с. 54–55]).

Обозначим через Π_L *оператор проецирования (проектор)* на подпространство L в $\mathbb{R}^n$ (т. е. $\Pi_L\boldsymbol{x}$ — это проекция вектора $\boldsymbol{x}$ на L), а через $dim\,L$ — *размерность* L.

Dimension (*англ.*) — размерность.

В дальнейшем важную роль будет играть

Лемма 1. Пусть случайный вектор $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{E})$ (см. П9), а два подпространства L_1 и L_2 в $\mathbb{R}^n$ ортогональны между собой ($L_1 \perp L_2$). Тогда векторы $\Pi_{L_1}\boldsymbol{\xi}$ и $\Pi_{L_2}\boldsymbol{\xi}$ независимы и нормально распределены, причем случайная величина $\sigma^{-2}|\Pi_{L_k}\boldsymbol{\xi}|^2$ ($k = 1{,}2$) имеет хи-квадрат распределение с $n_k = dim\,L_k$ степенями свободы.

Доказательство. Так как проецирование — это линейное преобразование, проекция нормального вектора $\boldsymbol{\xi}$ также будет иметь многомерное нормальное распределение (П9).

Если $\boldsymbol{\zeta} = \boldsymbol{C}\boldsymbol{\xi}$, где $\boldsymbol{C}$ — *ортогональная* матрица, то согласно лемме 1 гл. 11 $\boldsymbol{\zeta} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{E})$. Это означает, что если в каком-нибудь ортонормированном базисе случайный вектор имеет независимые $\mathcal{N}(0,\sigma^2)$ компоненты, то этим свойством он обладает в любом таком базисе.

Для завершения доказательства леммы выберем в $\mathbb{R}^n$ ортонормированный базис $\boldsymbol{e}_1,\ldots,\boldsymbol{e}_n$, у которого первые n_1 векторов лежат в L_1, а следующие n_2 векторов — в L_2. Тогда

$$\Pi_{L_1}\boldsymbol{\xi} = \sum_{i=1}^{n_1} (\boldsymbol{e}_i^T\boldsymbol{\xi})\,\boldsymbol{e}_i, \quad \Pi_{L_2}\boldsymbol{\xi} = \sum_{i=n_1+1}^{n_1+n_2} (\boldsymbol{e}_i^T\boldsymbol{\xi})\,\boldsymbol{e}_i,$$

где коэффициенты $\boldsymbol{e}_i^T\boldsymbol{\xi}$ — это координаты $\boldsymbol{\xi}$ в базисе $\boldsymbol{e}_1,\ldots,\boldsymbol{e}_n$. ■

Для построения доверительных интервалов потребуется

Теорема 1. В случае выполнения допущений Д1—Д3 для линейной регрессионной модели (5) верны следующие утверждения.

1. Случайная величина $\sigma^{-2}RSS = \sigma^{-2}|\boldsymbol{\eta} - \boldsymbol{X}\widehat{\boldsymbol{\theta}}|^2$ имеет распределение хи-квадрат с $n-m$ степенями свободы и *не зависит* от оценки $\widehat{\boldsymbol{\theta}}$ и, поскольку $\mathsf{M}(\sigma^{-2}RSS) = n-m$ (см. вопрос 3 гл. 11), статистика $\widehat{\sigma}^2 = RSS/(n-m)$ несмещенно оценивает неизвестную дисперсию σ^2.
2. Для любого вектора $\boldsymbol{c} \in \mathbb{R}^m$ случайная величина

$$\boldsymbol{c}^T(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \Big/ \Big[\widehat{\sigma}\,\sqrt{\boldsymbol{c}^T\boldsymbol{B}^{-1}\boldsymbol{c}}\Big] \tag{13}$$

распределена по закону Стьюдента t_{n-m} с $(n-m)$ степенями свободы (определение t_k дано в примере 4 гл. 11).

Доказательство. Рассмотрим проекции вектора $\boldsymbol{\eta}$ на подпространство $L(\boldsymbol{X})$ и на его ортогональное дополнение $L^{\perp}(\boldsymbol{X})$ до $\mathbb{R}^n$. Из определения МНК-оценки $\widehat{\boldsymbol{\theta}}$ вытекает, что

$$\Pi_{L(\boldsymbol{X})}\boldsymbol{\eta} = \boldsymbol{X}\widehat{\boldsymbol{\theta}}, \quad \Pi_{L^{\perp}(\boldsymbol{X})}\boldsymbol{\eta} = \widehat{\boldsymbol{\delta}}, \tag{14}$$

где $\widehat{\boldsymbol{\delta}} = \boldsymbol{\eta} - \boldsymbol{X}\widehat{\boldsymbol{\theta}}$ — вектор остатков. Подставляя модель (5) в формулы (14) и учитывая, что $\boldsymbol{X}\widehat{\boldsymbol{\theta}} \in L(\boldsymbol{X})$, получаем

$$\boldsymbol{X}\boldsymbol{\theta} + \Pi_{L(\boldsymbol{X})}\boldsymbol{\varepsilon} = \boldsymbol{X}\widehat{\boldsymbol{\theta}}, \tag{15}$$

$$\Pi_{L^{\perp}(\boldsymbol{X})}\boldsymbol{\varepsilon} = \widehat{\boldsymbol{\delta}}. \tag{16}$$

Умножая обе части равенства (15) на матрицу $\boldsymbol{B}^{-1}\boldsymbol{X}^T$ слева, приходим к равенству

$$\boldsymbol{\theta} + \boldsymbol{B}^{-1}\boldsymbol{X}^T\Pi_{L(\boldsymbol{X})}\boldsymbol{\varepsilon} = \widehat{\boldsymbol{\theta}}, \tag{17}$$

так как $\boldsymbol{B}^{-1}\boldsymbol{X}^T\boldsymbol{X} = \boldsymbol{B}^{-1}\boldsymbol{B} = \boldsymbol{E}$. Из формул (16) и (17) на основании леммы 1 заключаем, что оценки $\widehat{\boldsymbol{\delta}}$ и $\widehat{\boldsymbol{\theta}}$ независимы. По-

скольку $RSS = |\widehat{\boldsymbol{\delta}}|^2$, в силу леммы о независимости из гл. 1 отсюда вытекает также независимость RSS и $\widehat{\boldsymbol{\theta}}$ ($\widehat{\sigma}$ и $\widehat{\boldsymbol{\theta}}$). С учетом леммы 1 и формулы (16) видим, что случайная величина $\sigma^{-2}RSS \sim \chi^2_{n-m}$ (так как $dim\, L^{\perp}(\boldsymbol{X}) = n - m$).

Перейдем к доказательству пункта 2. Из формулы (17) следует, что случайный вектор $\widehat{\boldsymbol{\theta}}$, будучи линейным преобразованием нормального вектора $\boldsymbol{\varepsilon}$, сам является нормальным. Поэтому для любого $\boldsymbol{c} \in \mathbb{R}^m$ случайная величина $\boldsymbol{c}^T\widehat{\boldsymbol{\theta}}$ нормально распределена. В силу утверждения 1

$$\boldsymbol{c}^T(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \Big/ \left[\sigma\sqrt{\boldsymbol{c}^T\boldsymbol{B}^{-1}\boldsymbol{c}}\right] \sim \mathcal{N}(0,1)$$

и, как выше установлено, не зависит от $\widehat{\sigma}$. Следовательно, случайная величина (13) имеет распределение Стьюдента t_{n-m}. ■

Вопрос 1. Является ли теорема 1 гл. 11 частным случаем доказанной теоремы?

На основе теоремы 1 построим методом 1 из § 3 гл. 11 доверительные интервалы для σ, компонент вектора $\boldsymbol{\theta}$ и значения целевой переменной Y в заданной точке $(x_1^0, \ldots, x_m^0)$.

Пусть z_p обозначает p-квантиль закона χ^2_{n-m} (см. § 3 гл. 7). Тогда в силу теоремы 1

$$\mathbf{P}\left(z_{\alpha/2} < \sigma^{-2}RSS < z_{1-\alpha/2}\right) = 1 - \alpha. \tag{18}$$

Разрешая неравенство в формуле (18) относительно σ, получаем для него доверительный интервал с коэффициентом доверия $1 - \alpha$ вида

$$\left(\sqrt{RSS/z_{1-\alpha/2}}, \sqrt{RSS/z_{\alpha/2}}\right). \tag{19}$$

Пусть y_p обозначает p-квантиль закона t_{n-m}. В силу симметрии закона Стьюдента верно равенство $y_{1-\alpha/2} = -y_{\alpha/2}$. Взяв в качестве вектора $\boldsymbol{c}$ вектор, у которого l-я компонента равна 1, а остальные равны 0, находим из (13) границы $(1-\alpha)$-доверительного интервала для θ_l:

$$\widehat{\theta}_l \pm y_{1-\alpha/2}\,\widehat{\sigma}\,\sqrt{d_{ll}}, \tag{20}$$

где d_{ll} — диагональный элемент матрицы $\boldsymbol{D} = \boldsymbol{B}^{-1}$.

Наконец, задавая $\boldsymbol{c} = (x_1^0, \ldots, x_m^0)^T$, для $\theta_1 x_1^0 + \ldots + \theta_m x_m^0$ получаем $(1 - \alpha)$-доверительный интервал с границами

$$\boldsymbol{c}^T\widehat{\boldsymbol{\theta}} \pm y_{1-\alpha/2}\,\widehat{\sigma}\,\sqrt{\boldsymbol{c}^T\boldsymbol{B}^{-1}\boldsymbol{c}}. \tag{21}$$

Пример 4 (упрощенный вариант из [63]). Данные в таблице ниже представляют собой «зашумленные» значения линейной зависимости $Y = a + bX$. Ошибки $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ моделировались при помощи датчика нормальных случайных чисел (см. § 5 гл. 4). В каждом из «узлов» $u_l = -1 - 1/10 + l/5$ ($l = 1, \ldots, 10$) ошибки добавлялись дважды. Поэтому матрица $\boldsymbol{X}$ состоит из двух столбцов

длины 20: $\boldsymbol{x}_1 = (1,\ldots,1)^T$ и $\boldsymbol{x}_2 = (u_1,u_1,u_2,u_2,\ldots,u_{10},u_{10})^T$. Кроме того, одно из значений $\eta_i = a + bx_i + \varepsilon_i$ было заменено на «выброс».

u_l	−0,9	−0,7	−0,5	−0,3	−0,1	0,1	0,3	0,5	0,7	0,9
η_l	0,94	0,93	0,75	0,79	0,40	0,44	0,42	0,17	0,23	−0,03
η_{2l}	1,03	0,96	0,84	0,65	0,64	0,44	0,21	0,21	0,46	0,22

Задача состоит в вычислении МНК-оценок неизвестных параметров a, b и σ, обнаружении «выброса» и построении 95%-ных доверительных интервалов для параметров и значения прямой в точке предполагаемого «выброса».

Прежде всего, обратим внимание на то, что, благодаря симметрии расположения «узлов» относительно нуля, столбцы $\boldsymbol{x}_1$ и $\boldsymbol{x}_2$ ортогональны. Поэтому для нахождения оценок $\widehat{a}$ и $\widehat{b}$ можно воспользоваться формулами (8). Получим $\widehat{a} \approx 0{,}535$, $\widehat{b} \approx -0{,}5$. На рис. 8 построена подогнанная прямая $y = \widehat{a} + \widehat{b}x$. Нетрудно подсчитать, что $RSS = \sum(\eta_i - \widehat{a} - \widehat{b}x_i)^2 \approx 0{,}23$. Отсюда $\widehat{\sigma} = \sqrt{RSS/(n-m)} \approx 0{,}11$ ($n = 20$, $m = 2$).

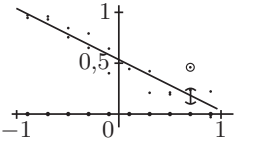


Рис. 8

Из табл. Т3 для χ^2_{18} находим квантили $z_{0,025} = 8{,}23$ и $z_{0,975} = 31{,}5$. Используя формулу (19), для параметра σ получаем доверительный интервал (0,08; 0,17).

Из табл. Т4 берем $y_{0,975} = 2{,}101$ для t_{18}. По формуле (20) вычисляем границы 95%-ных доверительных интервалов (0,48; 0,59) и (−0,59; −0,41) для a и b соответственно.

Наибольшую величину имеет остаток $\delta_{18} = 0{,}46 - 0{,}18 = 0{,}28$ (он больше, чем $2\widehat{\sigma} \approx 0{,}22$). С помощью формулы (21) получаем интервал (0,10; 0,27) для значения прямой в u_9 (см. рис. 8). Поскольку наблюдение 0,46 находится вблизи границы $0{,}27 + 0{,}22 = 0{,}49$, то оно, вероятно, является «выбросом».

Замечание 6. Ответ на вопрос об асимптотическом распределении МНК-оценки $\widehat{\boldsymbol{\theta}}$ при увеличении числа наблюдений дает следующая теорема.

Теорема 2. Потребуем, чтобы помимо выполнения допущений Д1—Д2 ошибки ε_i были *независимы*. Тогда, если матрица $n^{-1}\boldsymbol{B} = n^{-1}\boldsymbol{X}^T\boldsymbol{X}$ при $n \to \infty$ стремится к положительно определенной матрице $\boldsymbol{\Sigma}$, то

$$\sqrt{n}\,(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} \boldsymbol{\xi} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{\Sigma}^{-1}).$$

Выведем этот результат из теоремы 3, обобщающей теорему 1 из решения задачи 6 гл. 15.

Теорема 3. Пусть $\varepsilon_1,\varepsilon_2,\ldots$ — независимые и одинаково распределенные случайные величины, причем $\mathbf{M}\varepsilon_1 = 0$, $\mathbf{D}\varepsilon_1 = \sigma^2$, $0 < \sigma < \infty$. Пусть $\boldsymbol{A} = \|a_{il}\|_{n\times m}$, $\boldsymbol{\varepsilon} = (\varepsilon_1,\ldots,\varepsilon_n)^T$. Положим

$\boldsymbol{\xi}_n = n^{-1/2}\boldsymbol{A}^T\boldsymbol{\varepsilon}$. Если $n^{-1}\boldsymbol{A}^T\boldsymbol{A}$ стремится к положительно определенной матрице $\boldsymbol{\Lambda}$, то

$$\boldsymbol{\xi}_n \xrightarrow{d} \boldsymbol{\xi} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{\Lambda}) \quad \text{при } n \to \infty.$$

Доказательство теоремы 2. Подставим модель (5) в формулу (7):

$$\widehat{\boldsymbol{\theta}} = \boldsymbol{B}^{-1}\boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}) = \boldsymbol{\theta} + \boldsymbol{B}^{-1}\boldsymbol{X}^T\boldsymbol{\varepsilon},$$

и перепишем результат в виде

$$\sqrt{n}\,(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \sqrt{n}\,\boldsymbol{B}^{-1}\boldsymbol{X}^T\boldsymbol{\varepsilon} = n^{-1/2}\boldsymbol{A}^T\boldsymbol{\varepsilon},$$

где $\boldsymbol{A}^T = n\,\boldsymbol{B}^{-1}\boldsymbol{X}^T$. Поскольку

$$n^{-1}\boldsymbol{A}^T\boldsymbol{A} = n\,(\boldsymbol{B}^{-1}\boldsymbol{X}^T)(\boldsymbol{X}\boldsymbol{B}^{-1}) = n\boldsymbol{B}^{-1} \to \boldsymbol{\Lambda} = \boldsymbol{\Sigma}^{-1},$$

Доказательство теоремы 3 можно найти в [86, с. 308].

то остается воспользоваться теоремой 3. ■

§ 4. ОБЩАЯ ЛИНЕЙНАЯ ГИПОТЕЗА

Напомним, что в модели (5) неизвестный вектор $\mathsf{M}\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\theta}$ принадлежит *заданному* подпространству $L(\boldsymbol{X})$, порожденному столбцами матрицы $\boldsymbol{X}$. Более общим образом рассмотрим модель

$$\boldsymbol{\eta} = \boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \tag{22}$$

где (неслучайный) вектор $\boldsymbol{\gamma}$ неизвестен, но известно, что он лежит в заданном линейном подпространстве $L_0 \subset \mathbb{R}^n$ размерности $n_0 < n$, а случайный вектор $\boldsymbol{\varepsilon} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{E})$.*)

Проверяемая гипотеза H состоит в том, что $\boldsymbol{\gamma}$ лежит в некотором подпространстве $L_1 \subset L_0$, $dim\, L_1 = n_1 < n_0$.**)

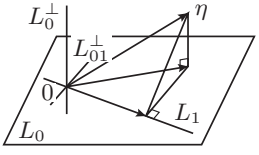


Рис. 9

Критерий для проверки гипотезы H естественно строить, основываясь на *сравнении расстояний* от $\boldsymbol{\eta}$ до подпространств L_0 и L_1 (рис. 9).

Введем обозначения $D_0 = |\boldsymbol{\eta} - \Pi_{L_0}\boldsymbol{\eta}|^2$, $D_1 = |\boldsymbol{\eta} - \Pi_{L_1}\boldsymbol{\eta}|^2$, где Π_{L_0} и Π_{L_1} — проекторы на подпространство L_0 и L_1 соответственно.

Так как векторы $\boldsymbol{\eta} - \Pi_{L_0}\boldsymbol{\eta}$ и $\Pi_{L_0}\boldsymbol{\eta} - \Pi_{L_1}\boldsymbol{\eta}$ ортогональны (см. рис. 9), то по «теореме Пифагора»

$$D_{01} = |\Pi_{L_0}\boldsymbol{\eta} - \Pi_{L_1}\boldsymbol{\eta}|^2 = D_1 - D_0. \tag{23}$$

Теорема 4. При выполнении гипотезы H статистика

$$R = \frac{D_{01}/(n_0 - n_1)}{D_0/(n - n_0)} \sim F_{n_0-n_1,\,n-n_0}, \tag{24}$$

*) Инвариантная (геометрическая) форма (22) модели (5) наглядна и удобна в рассуждениях. На практике же требуется та или иная параметризация модели, т. е. выбор векторов, порождающих пространство L_0.

**) Подчеркнем, что исходные предположения относительно модели (в частности, что $\boldsymbol{\gamma}$ принадлежит L_0) сомнению не подвергаются. Их иногда называют *основными предположениями*.

т. е. R распределена по закону Фишера—Снедекора с n_0-n_1 и $n-n_0$ степенями свободы (см. пример 1 гл. 14).

Отсюда получаем ***F*-критерий Фишера**, отвергающий гипотезу H на уровне значимости α, если наблюдаемое значение статистики R превосходит $(1-\alpha)$-квантиль закона $F_{n_0-n_1,n-n_0}$ (см. табл. Т5).

Доказательство теоремы 4. Согласно модели (22) и условию $\boldsymbol{\gamma} \in L_0$ запишем

$$\boldsymbol{\eta} - \Pi_{L_0}\boldsymbol{\eta} = \boldsymbol{\gamma} + \boldsymbol{\varepsilon} - \Pi_{L_0}(\boldsymbol{\gamma}+\boldsymbol{\varepsilon}) = \boldsymbol{\varepsilon} - \Pi_{L_0}\boldsymbol{\varepsilon},$$

т. е. вектор $\boldsymbol{\eta}-\Pi_{L_0}\boldsymbol{\eta}$ является проекцией $\boldsymbol{\varepsilon}$ на ортогональное дополнение $L_0^{\perp}$ к подпространству L_0 ($L_0 \oplus L_0^{\perp} = \mathbb{R}^n, dim\, L_0^{\perp} = n-n_0$). Поэтому согласно лемме 1 случайная величина $\sigma^{-2}D_0$ имеет хи-квадрат распределение с $n-n_0$ степенями свободы (независимо от того, верна или нет гипотеза $H\colon \boldsymbol{\gamma} \in L_1$).

В силу тех же условий (22) и $\boldsymbol{\gamma} \in L_0$ верно равенство $\Pi_{L_0}\boldsymbol{\eta} = \boldsymbol{\gamma} + \Pi_{L_0}\boldsymbol{\varepsilon}$. Если гипотеза H *справедлива*, то $\Pi_{L_1}\boldsymbol{\eta} = \boldsymbol{\gamma} + \Pi_{L_1}\boldsymbol{\varepsilon}$. Тогда из равенства (23) получаем: $D_{01} = |\Pi_{L_0}\boldsymbol{\varepsilon} - \Pi_{L_1}\boldsymbol{\varepsilon}|^2$. Очевидно, вектор $\Pi_{L_0}\boldsymbol{\varepsilon} - \Pi_{L_1}\boldsymbol{\varepsilon}$ является проекцией вектора $\boldsymbol{\varepsilon}$ на ортогональное дополнение $L_{01}^{\perp}$ подпространства L_1 до L_0 ($L_1 \oplus L_{01}^{\perp} = L_0, dim\, L_{01}^{\perp} = n_0 - n_1$). Так как $L_{01}^{\perp}$ и $L_0^{\perp}$ ортогональны (см. рис. 9), то в силу леммы 1 случайные величины D_{01} и D_0 независимы, причем $\sigma^{-2}D_{01} \sim \chi^2_{n_0-n_1}$. Следовательно, согласно определению распределения Фишера—Снедекора статистика R распределена по закону $F_{n_0-n_1,n-n_0}$. ■

Выведем из теоремы 4 несколько важных следствий.

1. Однофакторный дисперсионный анализ. Множество наблюдений η_{ij} представляет собой совокупность групп наблюдений Δ_j ($j=1,\ldots,k$) по $n_j > 1$ наблюдений в Δ_j. Всего $N = n_1 + \ldots + n_k$ наблюдений. Мы предполагаем, что

$$\eta_{ij} = \mu_j + \varepsilon_{ij}, \quad i = 1,\ldots,n_j, \tag{25}$$

где параметры $\mu_1,\ldots,\mu_k$ неизвестны, а все ошибки ε_{ij} независимы и одинаково распределены по закону $\mathcal{N}(0,\sigma^2)$. Тем самым, отклик η зависит от фактора с k уровнями. Для оценки этого влияния на каждом из уровней фактора проводится несколько независимых наблюдений.

Рассматривая η_{ij} как компоненты вектора $\boldsymbol{\eta}$ (записав их в порядке номеров групп Δ_j, $j=1,\ldots,k$), приведем модель (25) к

виду (5):

$$\begin{pmatrix}\eta_{11}\\ \dots\\ \eta_{n_1 1}\\ \eta_{12}\\ \dots\\ \eta_{n_2 2}\\ \dots\\ \eta_{1k}\\ \dots\\ \eta_{n_k k}\end{pmatrix} = \begin{pmatrix}1 & 0 & \dots & 0\\ \dots & \dots & \dots & \dots\\ 1 & 0 & \dots & 0\\ 0 & 1 & \dots & 0\\ \dots & \dots & \dots & \dots\\ 0 & 1 & \dots & 0\\ \dots & \dots & \dots & \dots\\ 0 & 0 & \dots & 1\\ \dots & \dots & \dots & \dots\\ 0 & 0 & \dots & 1\end{pmatrix} \begin{pmatrix}\mu_1\\ \mu_2\\ \dots\\ \mu_k\end{pmatrix} + \begin{pmatrix}\varepsilon_{11}\\ \dots\\ \varepsilon_{n_1 1}\\ \varepsilon_{12}\\ \dots\\ \varepsilon_{n_2 2}\\ \dots\\ \varepsilon_{1k}\\ \dots\\ \varepsilon_{n_k k}\end{pmatrix}. \tag{26}$$

Покажем, как с помощью F-критерия можно проверить *гипотезу однородности* $H'' : \mu_1 = \dots = \mu_k$ (см. пример 2 гл. 16). Ей соответствует подпространство L_1 в $\mathbb{R}^N$, порождаемое вектором $(1,\dots,1)^T$, $dim\, L_1 = 1$.

Ортогональным базисом подпространства $L_0 = L(\boldsymbol{X})$, очевидно, служат столбцы $\boldsymbol{x}_1,\dots,\boldsymbol{x}_k$ $(N \times k)$-матрицы $\boldsymbol{X}$ из формулы (26), $dim\, L_0 = k$. Найдем проекцию вектора $\boldsymbol{\eta}$ на L_0:

$$\Pi_{L_0}\boldsymbol{\eta} = \sum_{j=1}^{k} a_j \boldsymbol{x}_j, \quad \text{где } a_j = \boldsymbol{x}_j^T\boldsymbol{\eta}/|\boldsymbol{x}_j|^2 = \overline{\eta}_j = \frac{1}{n_j}\sum_{i=1}^{n_j}\eta_{ij}.$$

Величина $D_0 = |\boldsymbol{\eta} - \Pi_{L_0}\boldsymbol{\eta}|^2$ при этом представляется в виде

$$D_0 = \sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(\eta_{ij} - \overline{\eta}_j\right)^2. \tag{27}$$

Аналогично, нетрудно установить, что

$$D_1 = |\boldsymbol{\eta} - \Pi_{L_1}\boldsymbol{\eta}|^2 = \sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(\eta_{ij} - \overline{\eta}\right)^2, \quad \text{где } \overline{\eta} = \frac{1}{N}\sum_{j=1}^{k}\sum_{i=1}^{n_j}\eta_{ij}.$$

Сравнение с формулами (11) и (13) из примера 2 гл. 16 показывает, что $D_0 = V_{int}$, $D_1 = V_{tot}$. Из формул (12)–(13) гл. 16 следует, что $D_{01} = V_{out}$. Наконец, в силу теоремы 4 имеем

$$R = \frac{V_{out}/(k-1)}{V_{int}/(N-k)} = \frac{D_{01}/(k-1)}{D_0/(N-k)} \sim F_{k-1,N-k},$$

что было доказано непосредственно ранее в гл. 16.*)

2. Адекватность вида зависимости. Любую непрерывную зависимость между Y и X, где X принимает значения из некоторого отрезка, можно, согласно теореме Вейерштрасса, сколь угодно точно равномерно приблизить алгебраическим многочленом подходящей степени (см. § 2 гл. 5). Аналогично, периодическая непре-

*) Схему *двухфакторного дисперсионного анализа* из примера 1 гл. 17 также можно представить в виде линейной регрессионной модели (5) (см. [32, с. 205]).

рывная функция аппроксимируется тригонометрическими многочленами.

Обобщая пример 2, рассмотрим некоторую *систему базисных функций* $\{\varphi_1(x),\varphi_2(x),\ldots\}$ на $[-1,1]$, скажем,

$$\{1, x, x^2, \ldots\} \text{ или } \{1, \sin\pi x, \cos\pi x, \sin 2\pi x, \cos 2\pi x, \ldots\}.$$

Ограничиваясь начальным набором из m функций, приходим к исследованию зависимости вида

$$Y = \theta_1\varphi_1(X) + \ldots + \theta_m\varphi_m(X). \tag{28}$$

Выбирая на $[-1,1]$ «узлы» u_j $(j = 1,\ldots,k)$, произведем в каждом из них по $n_j > 1$ «зашумленных» измерений зависимости (28) (всего $N = n_1 + \ldots + n_k$ измерений). Для описания проведенного эксперимента используем модель

$$\eta_{ij} = \theta_1\varphi_1(u_j) + \ldots + \theta_m\varphi_m(u_j) + \varepsilon_{ij}, \quad i = 1,\ldots,n_j, \tag{29}$$

где случайные ошибки ε_{ij} независимы и одинаково распределены по закону $\mathcal{N}(0,\sigma^2)$.

Составим из величин η_{ij} вектор $\boldsymbol{\eta}$ в порядке возрастания j от 1 до k (см. левую часть (26)). Столбец с номером l $(l = 1,\ldots,m)$ $(N\times m)$-матрицы плана эксперимента $\boldsymbol{X}$ будет выглядеть следующим образом:

$$\boldsymbol{x}_l = (\underbrace{\varphi_l(u_1),\ldots,\varphi_l(u_1)}_{n_1}, \ldots, \underbrace{\varphi_l(u_k),\ldots,\varphi_l(u_k)}_{n_k}). \tag{30}$$

Допустим, что «узлы» $u_1,\ldots,u_k$ выбраны так, что для заданных базисных функций $\varphi_1(x),\ldots,\varphi_m(x)$ векторы $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_m$ *линейно независимы.*

Для проверки *гипотезы об адекватности* системы базисных функций возьмем в качестве L_1 подпространство $L(\boldsymbol{X})$, порожденное векторами $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_m$, $dim\, L_1 = m$. При этом, очевидно, $D_1 = RSS$.

В качестве подпространства L_0 используем подпространство, порожденное столбцами матрицы из правой части (26), $dim\, L_0 = k$. Величина D_0 для него задается формулой (27).

Вопрос 2. Почему $L_1 \subset L_0$ при $m<k$?

При справедливости гипотезы адекватности в силу теоремы 4

$$R = \frac{(RSS - D_0)/(k-m)}{D_0/(N-k)} \sim F_{k-m,N-k}. \tag{31}$$

Пример проверки гипотезы адекватности см. в задаче 4.

3. Понижение размерности модели. *Гипотеза H' допустимости понижения размерности* состоит в том, что в модели (5) компоненты $\theta_{m'+1},\ldots,\theta_m$ $(m' < m)$ параметрического вектора $\boldsymbol{\theta}$ можно считать равными нулю. Покажем, что она также является частным случаем общей линейной гипотезы.

Поскольку $\mathbf{M}\boldsymbol{\varepsilon} = \mathbf{0}$ (см. Д2 в § 2), то $\boldsymbol{\gamma} = \mathbf{M}\boldsymbol{\eta} \in L(\boldsymbol{X})$, где $L(\boldsymbol{X}) = L(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m)$ — подпространство, порожденное столбцами $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ матрицы $\boldsymbol{X}$. Возьмем $L(\boldsymbol{X})$ в качестве L_0, $dim\, L_0 = m$. Гипотеза H': $\theta_{m'+1} = \ldots = \theta_m = 0$ означает, что $\boldsymbol{\gamma}$ лежит в меньшем подпространстве $L_1 = L(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{m'})$, $dim\, L_1 = m' < m$.

Вычислим D_0 и D_{01}. Из определения МНК-оценки $\widehat{\boldsymbol{\theta}}$ следует (см. формулы (14)), что

$$\Pi_{L_0}\boldsymbol{\eta} = \boldsymbol{X}\widehat{\boldsymbol{\theta}} = \sum_{j=1}^{m} \widehat{\theta}_j\, \boldsymbol{x}_j. \tag{32}$$

Отсюда $D_0 = |\boldsymbol{\eta} - \Pi_{L_0}\boldsymbol{\eta}|^2 = |\Pi_{L_0^\perp}\boldsymbol{\eta}|^2 = |\widehat{\boldsymbol{\delta}}|^2 = RSS$.

Чтобы найти D_{01}, предположим, что векторы $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ *ортогональны.*[*)]

Обозначим через $L_{01}^\perp$ *ортогональное дополнение* L_1 *до* L_0: $L_1 \oplus L_{01}^\perp = L_0, dim\, L_{01}^\perp = m - m'$. Тогда $L_{01}^\perp = L(\boldsymbol{x}_{m'+1}, \ldots, \boldsymbol{x}_m)$ в силу ортогональности базисных векторов $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$. Из формулы (32) вытекает, что

$$\Pi_{L_1}\boldsymbol{\eta} = \Pi_{L_1}(\Pi_{L_0}\boldsymbol{\eta}) = \sum_{j=1}^{m'} \widehat{\theta}_j\, \boldsymbol{x}_j. \tag{33}$$

Подставляя равенства (32) и (33) в формулу (23), находим

$$D_{01} = \left| \sum_{j=m'+1}^{m} \widehat{\theta}_j\, \boldsymbol{x}_j \right|^2 = \sum_{j=m'+1}^{m} \widehat{\theta}_j^2\, |\boldsymbol{x}_j|^2. \tag{34}$$

Таким образом, при выполнении гипотезы H' (24) преобразуется в

$$R = \frac{D_{01}/(m-m')}{RSS/(n-m)} \sim F_{m-m',n-m}. \tag{35}$$

Примеры проверки гипотезы H' приведены в задачах 5 и 6.

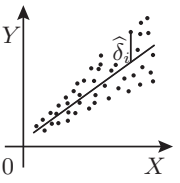
Замечание 7. Дополнительным преимуществом ортогональной модели является то, что в силу соотношений (8) при переходе к модели размерности $m' < m$ не надо пересчитывать оценки параметров $\theta_1, \ldots, \theta_{m'}$.

§ 5. ВЗВЕШЕННЫЙ МНК

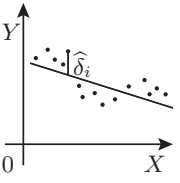
После оценивания параметров регрессионной модели (5) полезно изучить поведение реализации *вектора остатков* $\widehat{\boldsymbol{\delta}} = \boldsymbol{\eta} - \boldsymbol{X}\widehat{\boldsymbol{\theta}}$.

[*)] Если это условие не выполняется, перейдем к ортогональному базису в L_0 по формулам (9). С учетом того, что ортогонализационная матрица $\boldsymbol{A}$ является верхнетреугольной, в силу формулы (11) равенство $\theta'_{m'+1} = \ldots = \theta'_m = 0$ влечет $\theta_{m'+1} = \ldots = \theta_m = 0$. Так как $\boldsymbol{A}^{-1}$ — также верхнетреугольная, то верно и обратное. Поэтому гипотезу H' достаточно уметь проверять в ортогональной модели.

В случае хорошего соответствия данных и модели случайные величины $\widehat{\delta}_1,\ldots,\widehat{\delta}_n$ должны быть похожими на выборку из закона распределения ошибок ε_i. Однако нередко наблюдается увеличение разброса значений $\widehat{\delta}_i$ с ростом i (рис. 10, *а*) или чередование серий из положительных и отрицательных значений (рис. 10, *б*), возникающая из-за сильной коррелированности соседних остатков (см. § 5 гл. 15).*)



а)



б)

Рис. 10

В связи с этим рассмотрим два обобщения модели (5), возникающие в результате отказа от допущений одинаковой распределенности и некоррелированности ошибок ε_i.

1. Неравноточные наблюдения. Допустим, что в модели (5)

$$\mathbf{M}\boldsymbol{\varepsilon}=\mathbf{0},\quad \mathbf{Cov}(\boldsymbol{\varepsilon})=\|\mathbf{cov}(\varepsilon_i,\varepsilon_j)\|_{n\times n}=\sigma^2\boldsymbol{V}, \tag{36}$$

где $\boldsymbol{V}$ — известная *диагональная* матрица с элементами v_i на главной диагонали, σ — неизвестный параметр. Другими словами, ошибки некоррелированы и $\mathbf{D}\varepsilon_i=\sigma^2 v_i$.

Положим $w_i=1/v_i$ и, меняя масштаб с помощью преобразования $\tilde{\eta}_i=\sqrt{w_i}\,\eta_i$, получим *новую модель*:

$$\tilde{\boldsymbol{\eta}}=\tilde{\boldsymbol{X}}\boldsymbol{\theta}+\tilde{\boldsymbol{\varepsilon}}, \tag{37}$$

где $\tilde{x}_{ij}=\sqrt{w_i}\,x_{ij}$, $\tilde{\varepsilon}_i=\sqrt{w_i}\,\varepsilon_i$. Из формул (36) и (37) следует, что $\mathbf{M}\tilde{\varepsilon}_i=0$, $\mathbf{D}\tilde{\varepsilon}_i=\sigma^2$, $\mathbf{cov}(\tilde{\varepsilon}_i,\tilde{\varepsilon}_j)=0$ при $i\neq j$, т. е. для новой модели выполняется допущение Д2 из § 2.

МНК-оценки параметров $\theta_1,\ldots,\theta_m$ находятся (согласно § 2) путем минимизации функции

$$F(\boldsymbol{\theta})=\sum_{i=1}^{n}(\tilde{\eta}_i-\theta_1\tilde{x}_{i1}-\ldots-\theta_m\tilde{x}_{im})^2.$$

А для исходной модели имеем ($\tilde{\eta}_i=\sqrt{w_i}\,\eta_i$, $\tilde{x}_{ij}=\sqrt{w_i}\,x_{ij}$)

$$F(\boldsymbol{\theta})=\sum_{i=1}^{n}w_i(\eta_i-\theta_1x_{i1}-\ldots-\theta_mx_{im})^2. \tag{38}$$

Выходит, что теперь мы «взвешиваем» каждый член суммы, умножая его на $w_i=1/v_i$. Это придает больший вес наблюдениям, дисперсии ошибок которых меньше, что интуитивно представляется вполне разумным. Такая процедура называется **взвешенным методом наименьших квадратов**.

Нетрудно видеть, что матрица $\tilde{\boldsymbol{X}}$ из формулы (37) представляется в виде $\boldsymbol{V}^{-1/2}\boldsymbol{X}$.**) Аналогично, $\tilde{\boldsymbol{\eta}}=\boldsymbol{V}^{-1/2}\boldsymbol{\eta}$. Подставляя эти выражения в формулу (7), находим МНК-оценку

$$\widehat{\boldsymbol{\theta}}=(\tilde{\boldsymbol{X}}^T\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}^T\tilde{\boldsymbol{\eta}}=(\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{\eta}. \tag{39}$$

Почти все величайшие открытия в астрономии вытекают из рассмотрения того, что мы уже раньше назвали качественными или численными *остаточными феноменами*, иначе говоря, они вытекают из анализа той части числовых или качественных результатов наблюдения, которая «торчит» и остается необъясненной после выделения и учета всего того, что согласуется со строгим применением известных методов.

Дж. Гершель, «Основы астрономии», 1849 г.

*) Подробнее об исследовании остатков см. [23, с. 186].

**) Здесь $\boldsymbol{V}^{-1/2}$ — диагональная матрица с элементами $1/\sqrt{v_i}$ на диагонали.

Это и есть значение $\boldsymbol{\theta}$, минимизирующее сумму квадратов (38), которая в матричной записи имеет вид

$$F(\boldsymbol{\theta}) = (\boldsymbol{\eta} - \boldsymbol{X\theta})^T \boldsymbol{V}^{-1} (\boldsymbol{\eta} - \boldsymbol{X\theta}). \tag{40}$$

Согласно утверждению 1 из § 3, ковариационной матрицей случайного вектора $\widehat{\boldsymbol{\theta}}$ служит

$$\mathbf{Cov}(\widehat{\boldsymbol{\theta}}) = \sigma^2 (\tilde{\boldsymbol{X}}^T \tilde{\boldsymbol{X}})^{-1} = \sigma^2 (\boldsymbol{X}^T \boldsymbol{V}^{-1} \boldsymbol{X})^{-1}. \tag{41}$$

Разберем несколько приложений рассмотренной модели.

Пример 5. ***Взвешенное среднее.*** Если модель (5) содержит единственный неизвестный параметр θ ($m = 1$), то из формулы (39) несложно получить, что его МНК-оценка с учетом допущений (36) выглядит так:

$$\widehat{\theta} = \sum_{i=1}^{n} w_i x_i \eta_i \Big/ \sum_{i=1}^{n} w_i x_i^2. \tag{42}$$

В случае, когда все $x_i = 1$, она представляет собой просто *взвешенное среднее* наблюдений η_i с весами w_i.

Пример 6. ***Линейный закон — прямая, проходящая через начало координат*** ([75, с. 404])**.** Пусть Y — *потери тепла* в (стандартном) коттедже, а X обозначает *разность между внутренней и наружной температурой*. Рассмотрим линейную модель $Y = \theta X$, при этом дисперсия наблюдений $\eta_i = \theta x_i + \varepsilon_i$ ($i = 1, \ldots, n$) будет, скорее всего, возрастать с увеличением x_i. Простейший вид такой зависимости — это $\mathbf{D}\varepsilon_i = cx_i$, т. е. дисперсия пропорциональна разности температур. В этом случае из формулы (42) при $w_i = \sigma^2/(cx_i)$ находим, что $\widehat{\theta} = \sum \eta_i / \sum x_i = \overline{\eta}\,/\,\overline{x}$. Ввиду соотношений (41) $\mathbf{D}\widehat{\theta} = \sigma^2 / \sum w_i x_i^2 = c / \sum x_i$. Если же допустить, что $\mathbf{D}\varepsilon_i = cx_i^2$, то приходим к оценке $\widehat{\theta} = \dfrac{1}{n} \sum (\eta_i / x_i)$.

Вопрос 3. Чему равна дисперсия $\mathbf{D}\widehat{\theta}$ в этом случае?

Пример 7. ***Повторные наблюдения.*** Пусть η_i — среднее из k_i наблюдений, причем все они имеют одинаковое математическое ожидание $\sum\limits_{j=1}^{m} x_{ij}\theta_j$ и общую дисперсию σ^2, $i = 1, \ldots, n$. Если все наблюдения *независимы*, то $\mathbf{D}\eta_i = \sigma^2/k_i$, так что в формуле (36) надо положить $v_i = 1/k_i$.

2. Коррелированные наблюдения. Рассмотрим обобщение модели (36), допуская, что наблюдения могут не только иметь разную точность, но и быть зависимыми между собой. Для этого матрицу $\boldsymbol{V}$ в формуле (36) будем считать известной *положительно определенной* (см. П10) матрицей общего вида. Ковариации ошибок связаны с ее элементами v_{ij} равенствами $\mathbf{cov}(\varepsilon_i, \varepsilon_j) = \sigma^2 v_{ij}$.

Для такой матрицы существует единственная положительно определенная матрица $\boldsymbol{V}^{1/2}$, называемая *квадратным корнем* из матрицы $\boldsymbol{V}$, такая, что $\boldsymbol{V}^{1/2}\boldsymbol{V}^{1/2} = \boldsymbol{V}$ (см. П10).

Умножая обе части модели (5) на обратную к $\boldsymbol{V}^{1/2}$ матрицу $\boldsymbol{V}^{-1/2}$, приходим к *новой модели* $\tilde{\boldsymbol{\eta}} = \tilde{\boldsymbol{X}}\boldsymbol{\theta} + \tilde{\boldsymbol{\varepsilon}}$, где $\tilde{\boldsymbol{\eta}} = \boldsymbol{V}^{-1/2}\boldsymbol{\eta}$, $\tilde{\boldsymbol{X}} = \boldsymbol{V}^{-1/2}\boldsymbol{X}$, $\tilde{\boldsymbol{\varepsilon}} = \boldsymbol{V}^{-1/2}\boldsymbol{\varepsilon}$. В силу формул (12) и симметричности матрицы $\boldsymbol{V}^{-1/2}$ имеем:

$$\mathsf{M}\tilde{\boldsymbol{\varepsilon}} = \boldsymbol{V}^{-1/2}\,\mathsf{M}\boldsymbol{\varepsilon} = \boldsymbol{0},$$

$$\mathbf{Cov}(\tilde{\boldsymbol{\varepsilon}}) = \boldsymbol{V}^{-1/2}\,\mathbf{Cov}(\boldsymbol{\varepsilon})\,\boldsymbol{V}^{-1/2} = \sigma^2\,\boldsymbol{V}^{-1/2}\boldsymbol{V}\,\boldsymbol{V}^{-1/2} =$$
$$= \sigma^2\,\boldsymbol{V}^{-1/2}(\boldsymbol{V}^{1/2}\boldsymbol{V}^{1/2})\boldsymbol{V}^{-1/2} = \sigma^2\boldsymbol{E},$$

т. е. видим, что для новой модели опять выполняется допущение Д2 из § 2. Нетрудно убедиться, что формулы (39)–(41) остаются без изменений (благодаря симметричности $\boldsymbol{V}^{-1/2}$) и в случае любой положительно определенной матрицы $\boldsymbol{V}$.

Пример 8. ***Оценка параметров сдвига и масштаба с помощью МНК.*** Пусть элементы выборки $X_1,\ldots,X_n$ имеют функцию распределения $F((x-\mu)/\sigma)$, где $F(u)$ — известная функция распределения, $\mu \in \mathbb{R}$ и $\sigma > 0$ — неизвестные параметры сдвига и масштаба соответственно. Рассмотрим *стандартизованные* случайные величины $U_i = (X_i - \mu)/\sigma$, $i = 1,\ldots,n$. Тогда $U_1,\ldots,U_n$ — выборка из закона с функцией распределения $F(u)$. При этом порядковые статистики (см. § 2 гл. 5) двух выборок связаны формулами

$$U_{(i)} = (X_{(i)} - \mu)/\sigma, \quad i = 1,\ldots,n, \tag{43}$$

и так как функция распределения $F(u)$ случайной величины U_i известна, то в принципе любые вероятностные характеристики порядковых статистик $U_{(i)}$ могут быть рассчитаны. Положим

$$\mathsf{M}U_{(i)} = \alpha_i, \quad \mathbf{cov}\left(U_{(i)}, U_{(j)}\right) = v_{ij}.$$

Учитывая формулы (43), запишем соотношения

$$X_{(i)} = \mu + \sigma U_{(i)} = \mu + \sigma\alpha_i + \sigma(U_{(i)} - \alpha_i), \quad i = 1,\ldots,n. \tag{44}$$

Вводя обозначения $\eta_i = X_{(i)}$ и $\varepsilon_i = \sigma(U_{(i)} - \alpha_i)$, представим (44) в виде обобщенной линейной регрессионной модели:

$$\eta_i = \mu + \sigma\alpha_i + \varepsilon_i, \quad i = 1,\ldots,n,$$

где $\mathsf{M}\varepsilon_i = 0$, $\mathbf{cov}\,(\varepsilon_i, \varepsilon_j) = \sigma^2 v_{ij}$, или в матричной форме:

$$\begin{pmatrix}\eta_1\\ \vdots\\ \eta_n\end{pmatrix} = \begin{pmatrix}1 & \alpha_1\\ \vdots & \vdots\\ 1 & \alpha_n\end{pmatrix}\begin{pmatrix}\mu\\ \sigma\end{pmatrix} + \begin{pmatrix}\varepsilon_1\\ \vdots\\ \varepsilon_n\end{pmatrix}. \tag{45}$$

Согласно замечанию 5, МНК-оценки параметров μ и σ в равенстве (45), вычисляемые по формуле (39), имеют минимальную

дисперсию в классе линейных несмещенных оценок, построенных по порядковым статистикам $X_{(1)},\dots,X_{(n)}$. Приведем некоторые результаты вычислений для равномерного и показательного законов (подробности см. в [38, с. 89]).

Для U_i, *равномерно распределенных* на отрезке $[0,1]$, в задаче 1 гл. 5 были подсчитаны $\alpha_i = \mathbf{M}U_{(i)} = i/(n+1)$. В [38, с. 93] найдены

$$v_{ij} = \mathbf{M}U_{(i)}U_{(j)} - \alpha_i\alpha_j = \frac{i(n-j+1)}{(n+1)^2(n+2)}, \quad i \leqslant j.$$

Там же проверяется, что $\boldsymbol{V}^{-1}$ имеет элементы $(n+1)(n+2)p_{ij}$, где

$$p_{ii} = 2,\ p_{i,i-1} = p_{i-1,i} = -1,\ p_{ij} = 0 \text{ при } |j-i| > 1,$$

а МНК-оценки параметров масштаба и сдвига выглядят так:

$$\widehat{\sigma} = \frac{n+1}{n-1}\,(X_{(n)} - X_{(1)}), \quad \widehat{\mu} = \frac{X_{(1)}+X_{(n)}}{2} - \frac{\widehat{\sigma}}{2} = \frac{nX_{(1)} - X_{(n)}}{n-1}.$$

Они (с точностью до репараметризации модели) совпадают с оценками метода спейсингов из задачи 4 гл. 9.

Для *показательно распределенных* U_i ($F(u) = (1-e^{-u})I_{\{u\geqslant 0\}}$) при решении задачи 5 гл. 4 было установлено, что

$$Z_i = (n-i+1)(U_{(i)} - U_{(i-1)}), \quad U_{(0)} = 0, \quad i = 1,\dots,n,$$

независимы и распределены так же, как и U_i. Поскольку $\mathbf{M}Z_i = 1$,

$$\alpha_i = \mathbf{M}\sum_{k=1}^{i}\frac{Z_i}{n-k+1} = \sum_{k=1}^{i}\frac{\mathbf{M}Z_i}{n-k+1} = \sum_{k=1}^{i}\frac{1}{n-k+1},$$

$$v_{ij} = \sum_{k=1}^{i}\sum_{l=1}^{j}\frac{\mathbf{cov}\,(Z_i,Z_j)}{(n-k+1)(n-l+1)} = \sum_{k=1}^{\min(i,j)}\frac{1}{(n-k+1)^2}.$$

В [38, с. 96] проверяется, что матрица $\boldsymbol{V}^{-1}$ имеет элементы q_{ij}, где

$q_{ii} = (n-i+1)^2 + (n-i)^2$, $q_{i,i-1} = q_{i-1,i} = -(n-i+1)^2$

($q_{ij} = 0$ при $|j-i| > 1$), а МНК-оценки параметров σ и μ таковы:

$$\widehat{\sigma} = n(\overline{X} - X_{(1)})/(n-1), \quad \widehat{\mu} = \overline{X} - \widehat{\sigma} = (nX_{(1)} - \overline{X})/(n-1).$$

В случае *нормального закона* для моментов порядковых статистик $U_{(i)}$ нет простых формул. Известно, что $\widehat{\mu} = \overline{X}$, однако для вычисления $\widehat{\sigma}$ приходится пользоваться таблицами (см. [70]).

§ 6. ПАРАДОКСЫ РЕГРЕССИИ

Существуют три вида лжи: ложь, наглая ложь и статистика.

Марк Твен

Есть несколько **типичных ошибок** («тонких мест»), которые следует иметь в виду, применяя регрессионный анализ. Сами по себе они достаточно очевидны. Тем не менее, о них часто забывают при работе с реальными данными и в результате приходят к неверным выводам.

1. Неоднородность данных. Существенно исказить вид регрессионной зависимости могут не только выделяющиеся наблюдения («выбросы») по оси отклика Y, но и аномальные значения предиктора X.

Пример 9 ([2, с. 64], [54]). На рис. 11, *а* представлены данные о *числе телевизионных точек* Y (в десят. тыс.), установленных в 1953 г. в девяти городах США (Денвере, Сан-Антонио, Канзас-Сити, Сиэтле, Цинцинати, Буффало, Нью-Орлеане, Милуоки, Хьюстоне) и о *численности населения* X (в десят. тыс.) этих городов.

Выборочный коэффициент корреляции между наборами $x_1,\ldots,x_9$ и $y_1,\ldots,y_9$ $\widehat{\rho}=0{,}403$, что при $n=9$ свидетельствует о весьма малой степени линейной связи между X и Y.

Если же к этим данным добавить соответствующие сведения о Нью-Йорке ($x_{10}=802$, $y_{10}=345$), то пересчитанный для $n=10$ коэффициент $\widehat{\rho}=0{,}995$. На рис. 11, *б* изображены в более мелком масштабе точки с координатами (x_i,y_i), $i=1,\ldots,10$, и проведена регрессионная прямая, по сути, соединяющая центр масс первых девяти точек $(\overline{x},\overline{y})$ с (x_{10},y_{10}).

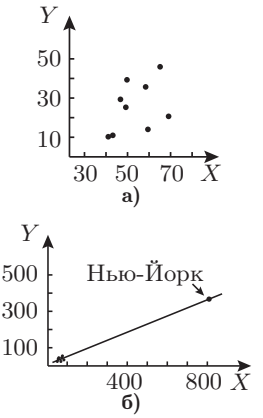


Рис. 11

2. Коррелированность предикторов. В случае, когда регрессионная модель включает много предикторов, некоторые из них могут оказаться приблизительно *линейно связанными* между собой. Обсудим связанные с этим проблемы.

При подгонке полинома на отрезке $[0,1]$ (см. пример 2) уже для предикторов $X_3=U^2$ и $X_4=U^3$, измеряемых в «узлах» $u_i=i/20$, $i=0{,}1,\ldots,20$, выборочный коэффициент корреляции $\widehat{\rho}$ между соответствующими столбцами матрицы $\boldsymbol{X}$ составляет 0,986 (рис. 12).

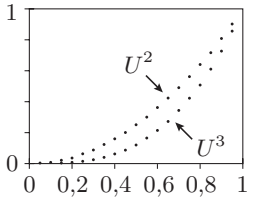


Рис. 12

Сильная коррелированность предикторов опасна тем, что приводит к *неустойчивости* МНК-оценок, вычисляемых по формуле (7), к малым возмущениям наблюдений η_i. Дело в том, что в этом случае столбцы матрицы $\boldsymbol{X}$ оказываются практически линейно зависимыми, вследствие чего матрица $\boldsymbol{B}=\boldsymbol{X}^T\boldsymbol{X}$ становится почти вырожденной, а задача поиска решения линейной системы (6) — *плохо обусловленной* (см. [6, с. 131]).

Интерполяция — частный случай регрессии при $n=m$.

Рассмотрим подробнее влияние возмущений на *интерполяционный полином* степени $n-1$, проходящий через точки плоскости с координатами (u_i,η_i), $i=1,\ldots,n$. В *форме Лагранжа* он выглядит так:

$$p(u)=\sum_{i=1}^{n}\eta_i L_i(u),\quad \text{где } L_i(u)=\prod_{\substack{j=1\\ j\neq i}}^{n}\frac{u-u_j}{u_i-u_j}\,.$$

То, что $p(u)$ интерполирует заданные точки, вытекает из равенств $L_i(u_i)=1$ и $L_i(u_j)=0$ при $j\neq i$.

На рис. 13, *а*, заимствованном из [53, с. 32], построен полином степени 8. На рис. 13, *б* продемонстрированы последствия добавления новой (десятой) точки: график даже не поместился в окне рисунка.

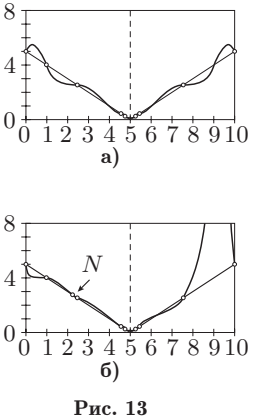


Рис. 13

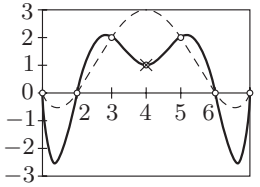


Рис. 14

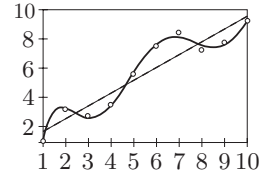


Рис. 15

Другой пример: даже для полинома степени 6 удаление точки с координатами (4,1) на рис. 14 приводит к значительным изменениям — возникновению осцилляций.

В регрессионных задачах полиномы высоких степеней имеют тенденцию *сглаживать ошибки* наблюдений, вместо того, чтобы выражать истинный вид зависимости отклика от предиктора. На рис. 15 десять точек, разбросанных вблизи прямой, сглажены для иллюстрации МНК-полиномом степени 6. Если зависимость на самом деле достаточно сложная, то для ее обнаружения могут оказаться полезными методы непараметрической регрессии, обсуждаемые в гл. 22.

Кроме вычислительной неустойчивости, коррелированность предикторов приводит к *затруднениям в интерпретации* результатов расчетов. Например, при исследовании зависимости *веса* Z студентов двух групп от их *роста* X и *размера обуви* Y методом наименьших квадратов (после предварительного выравнивания масштабов данных) в первой группе было получено регрессионное уравнение

$$Z - \overline{Z} = 0{,}9\,(X - \overline{X}) + 0{,}1\,(Y - \overline{Y}),$$

а для второй группы зависимость оказалась сильно отличающейся:

$$Z - \overline{Z} = 0{,}2\,(X - \overline{X}) + 0{,}8\,(Y - \overline{Y}).$$

Как объяснить существенное различие коэффициентов этих двух моделей?

На практике, подсчитав МНК-оценки $\widehat{\theta}_1, \ldots, \widehat{\theta}_m$ в линейной регрессионной модели, исследователь в первую очередь обращает внимание на предикторы с *самыми большими* (по абсолютной величине) $\widehat{\theta}_j$, так как именно их изменение сильнее всего сказывается на отклике. Исследователь нередко хочет не только точно предсказывать отклик для произвольных значений предикторов, но и желает, задавая эти значения, управлять откликом, надеясь, что регрессия отражает *причинно-следственную связь* между откликом и предикторами. Понятно, что «нажимать» надо на самые эффективные «рычаги».

В приведенном выше примере для первой группы важнейшим предиктором оказался X, а для второй — Y. Дело здесь в том, что X и Y сильно коррелируют друг с другом, вследствие чего общий «весовой» коэффициент при $(X - \overline{X}) + (Y - \overline{Y})$ случайным образом распределился между слагаемыми.

К счастью, рассмотренное затруднение нетрудно преодолеть. Достаточно проверить предикторы на наличие тесных линейных связей и каждую обнаруженную группу заменить в модели на ее единственного представителя.

3. Неадекватность модели. Когда простейшая линейная зависимость $Y = \theta_1 X_1 + \ldots + \theta_m X_m$ неадекватно описывает данные,

T	*1865*	*1866*	*1867*	*1868*	*1869*	*1870*	*1871*	*1872*	*1873*	*1874*	*1875*	*1876*
Y	9,10	9,66	10,06	10,71	11,95	12,26	12,85	14,84	15,12	13,92	14,12	13,96
T	*1877*	*1878*	*1879*	*1880*	*1881*	*1882*	*1883*	*1884*	*1885*	*1886*	*1887*	*1888*
Y	14,19	14,54	14,41	18,58	19,82	21,56	21,76	20,46	19,84	20,81	22,82	24,03
T	*1889*	*1890*	*1891*	*1892*	*1893*	*1894*	*1895*	*1896*	*1897*	*1898*	*1899*	*1900*
Y	25,88	27,87	26,17	26,92	25,26	26,03	29,37	31,29	33,46	36,46	40,87	41,35
T	*1901*	*1902*	*1903*	*1904*	*1905*	*1906*	*1907*	*1908*	*1909*	*1910*		
Y	41,14	44,73	46,82	46,22	54,79	59,66	61,30	48,80	60,60	66,20		

Рис. 16

для построения более сложной модели обычно пытаются отдельно изучить влияние каждого предиктора X_j на отклик Y. Для этого сглаживают двумерное «облако» точек при помощи некоторой нелинейной функции. Список наиболее часто используемых *монотонных* функций содержится в следующей таблице.

Поведение отклика	Уравнение	Усл. на b	x'	y'
Очень быстрый рост *)	$y = e^{a+bx}$	$b > 0$	x	$\ln y$
Быстрый (степенной) рост	$y = e^{a+b\ln x}$	$b > 1$	$\ln x$	$\ln y$
Медленный рост	$y = e^{a+b\ln x}$	$0 < b < 1$	$\ln x$	$\ln y$
Очень медленный рост	$y = a + b\ln x$	$b > 0$	$\ln x$	y
Медленная стабилизация	$y = a + b/x$	$b \neq 0$	$1/x$	y
Быстрая стабилизация	$y = a + be^{-x}$	$b \neq 0$	e^{-x}	y
Кривая S-образной формы	$y = 1/(a + be^{-x})$	$b > 0$	e^{-x}	$1/y$

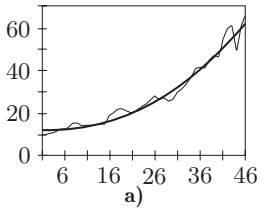
*) В [51, с. 46] содержится любопытная классификация углов из книги по альпинизму (изданной около 1900 г.): «Перпендикулярно — 60°, мой дорогой сэр, абсолютно перпендикулярно — 65°, нависающе — 70°».

[Последняя функция называется *логистической кривой*. При $a > 0$, $b > 0$ она возрастает, имеет две горизонтальные асимптоты $y = 0$, $y = 1/a$ и перегиб в точке с координатами $(\ln(b/a), 1/(2a))$.]

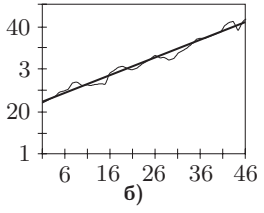
Переход к новым переменным x' и y' (см. таблицу) сводит задачу к подгонке прямой $y' = a + bx'$ из § 1.

Пример 10. В таблице на рис. 16 для каждого из годов с 1865 по 1910 (T — *номер года*) указано *количество чугуна* Y (в млн. тонн), которое выплавлялось за год во всем мире. Постараемся подогнать регрессионную кривую к этим данным.

Положим $X = T - 1864$. На рис. 17, *а* изображена кривая (ломаная), соединяющая точки плоскости с координатами (X_i, Y_i), $i = 1, \ldots, n$, где $n = 46$. Ван дер Варден (см. [13, с. 179]) пишет: «Эта кривая поднимается вверх значительно быстрее, чем прямая линия или квадратная парабола». Не соглашаясь с этим мнением, попытаемся сгладить кривую с помощью параболы. На рис. 17, *а* приведен график подогнанный параболы с помощью МНК. При этом остаточная сумма квадратов $RSS = 364{,}7$ и (согласно теореме 1) оценка стандартного отклонения ошибок $\widehat{\sigma} = \sqrt{RSS/(n-m)} = 2{,}91$, где $m = 3$.



а)



б)

Рис. 17

Поскольку заметна некоторая тенденция усиления колебаний кривой относительно параболы с ростом X, применим, следуя Ван дер Вардену, преобразование $Y' = \ln Y$. На рис. 17, *б* построена кривая, соединяющая точки с координатами (X_i, Y_i'), $i = 1, \ldots, n$. Она хорошо согласуется с подогнанной МНК-прямой, имеющей уравнение $y = \widehat{a} + \widehat{b}x$, где $\widehat{a} = 2{,}203$, $\widehat{b} = 0{,}0413$ (оценки коэффициентов прямой вычисляются по формуле (3)).

Однако, если попытаться использовать последнюю подгонку для предсказания (прогноза) Y_i на основе формулы $\tilde{Y}_i = \exp(\widehat{a} + \widehat{b}X_i)$, то получим значение $RSS = 381{,}8$, которое несколько больше, чем вычисленное ранее при подгонке параболы. Справедливости ради отметим, что *сумма модулей остатков* для параболы, наоборот, немного больше: $95{,}6 > 93{,}6$. В целом, можно считать подгонки примерно одинаковыми по точности.

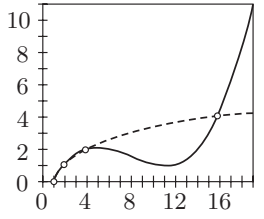


Рис. 18

Следует иметь в виду, что в случае ошибки при выборе типа сглаживающей кривой результаты *экстраполяции*[*)] могут оказаться совершенно неудовлетворительными. Это наглядно демонстрирует рис. 18, на котором зависимость $y = \log_2 x$ аппроксимируется полиномом степени 3, интерполирующим точки с координатами (1,0), (2,1), (4,2) и (16,4).

При построении модели нужно максимально учитывать всю имеющуюся информацию о *качественном поведении* регрессионной кривой: монотонность, выход на асимптоту и т. п. В идеале, желательно опираться на законы (физики, химии, экономики), лежащие в основе зависимости (как в примере 3). Следующий пример показывает, что в случае формальной подгонки кривой, взятой из некоторого класса функций, *необходима перепроверка*.

Пример 11 ([2, с. 177] по данным А. Я. Боярского)**.** Рассмотрим в качестве отклика Z *вес коровы*, а в качестве предикторов — *окружность ее туловища* X и *расстояние от хвоста до холки* Y. Сравнительному анализу были подвергнуты три регрессионные модели:

(a) *линейная*: $Z = \theta_1 + \theta_2 X + \theta_3 Y$,
(b) *степенная*: $Z = \theta_1' X^{\theta_2'} Y^{\theta_3'}$,
(c) *учитывающая содержательный смысл задачи*: $Z = \theta_0 X^2 Y$.

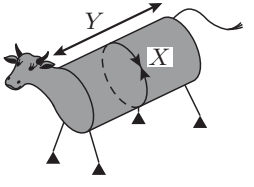


Рис. 19

Происхождение последней модели легко объяснить. Для этого следует представить себе приближенно тушу коровы в форме цилиндра с длиной образующей, равной Y, и радиусом основания, равным $X/(2\pi)$ (рис. 19). Если ρ — средняя плотность, то вес $Z = \rho\,\pi[X/(2\pi)]^2 Y = \theta_0 X^2 Y$ с точностью до головы и ног («рогов и копыт»).

[*)] То есть восстановления значений отклика по значениям предиктора, расположенным *вне* обследованного диапазона.

Вначале по всем ($n = 20$) имеющимся наблюдениям для каждой из моделей была вычислена МНК-оценка $\widehat{\boldsymbol{\theta}}$ векторного параметра $\boldsymbol{\theta}$ и оценка $\widehat{\sigma} = \sqrt{RSS/(n-m)}$ стандартного отклонения ошибок, где RSS — остаточная сумма квадратов (см. § 2), m — размерность вектора $\boldsymbol{\theta}$. Результаты расчетов приведены в левой стороне следующей таблицы из [2, с. 179].

Модель	По всем наблюдениям		По части наблюдений	
	$\widehat{\boldsymbol{\theta}}$	$\widehat{\sigma}$	$\widehat{\boldsymbol{\theta}}_{\text{тяж}}$	$\widehat{\sigma}_{\text{лег}}$
1	$\widehat{\theta}_1 = -984{,}7$ $\widehat{\theta}_2 = 4{,}73$ $\widehat{\theta}_3 = 4{,}70$	25,9	$\widehat{\theta}_1 = 453{,}2$ $\widehat{\theta}_2 = 0{,}62$ $\widehat{\theta}_3 = -0{,}22$	81
2	$\widehat{\theta}'_1 = 0{,}0011$ $\widehat{\theta}'_2 = 1{,}556$ $\widehat{\theta}'_3 = 1{,}018$	24,5	$\widehat{\theta}'_1 = 266{,}4$ $\widehat{\theta}'_2 = 0{,}203$ $\widehat{\theta}'_3 = -0{,}072$	79
3	$\widehat{\theta}_0 = 1{,}13 \cdot 10^{-4}$	26,6	$\widehat{\theta}_0 = 1{,}11 \cdot 10^{-4}$	28

Из них как-будто следует, что формальные модели 1 и 2 оказались несколько точней содержательной модели 3. Однако, это лишь кажущееся благополучие, что сразу выявляется при перепроверке путем разбиения данных на обучающую и контрольную подвыборки.

В качестве *обучающей* была взята подвыборка из 10 самых тяжелых коров, а в качестве *контрольной* — из 10 оставшихся легких коров. На основе обучающей подвыборки для каждой из моделей была заново подсчитана МНК-оценка $\widehat{\boldsymbol{\theta}}_{\text{тяж}}$ (см. правую сторону таблицы). Видим, что модели 1 и 2 не выдержали испытание на *устойчивость коэффициентов* ($\widehat{\theta}_3$ даже поменял знак).

Кроме того, при попытке предсказания веса легких коров с помощью модели с такими коэффициентами, оценка стандартного отклонения ошибок $\widehat{\sigma}_{\text{лег}}$ увеличилась для формальных моделей более чем в 3 раза!

Этот пример убедительно демонстрирует, что не следует переусложнять модель, ориентируясь на минимизацию $\widehat{\sigma}$: за счет трех управляемых параметров («рычагов») удалось «подогнать» формальные модели к данным лучше, чем однопараметрическую содержательную.

4. Скрытый фактор. Желание истолковывать регрессионную связь как причинно-следственную может приводить к парадоксам, подобным приведенным в двух следующих примерах (см. также близкое к этой теме понятие частной или «очищенной» корреляции из § 8 гл. 20).

Пример 12 (см. [57]). Во время второй мировой войны англичане исследовали зависимость *точности бомбометания* Z от ряда факторов, в число которых входили *высота бомбардировщика* H, *скорость ветра* V, *количество истребителей противника* X. Как

и ожидалось, Z увеличивалась при уменьшении H и V. Однако (что поначалу представлялось необъяснимым), точность бомбометания Z возрастала также и при увеличении X.

Дальнейший анализ позволил понять причину этого парадокса. Дело оказалось в том, что первоначально в модель не был включен такой важный фактор, как Y — *облачность*. Он сильно влияет и на Z (уменьшая точность), и на X (бессмысленно высылать истребители, если ничего не видно). Сильные отрицательные причинно-следственная связи в парах (Y,Z) и (X,Y) привели к появлению положительного коэффициента при X в линейной регрессионной модели для Z.

Пример 13 ([4, с. 171]). Если найти корреляцию между *ежегодным количеством родившихся в Голландии детей* Z и *количеством прилетевших аистов* X, то она окажется довольно значительной [37]. Можно ли на основе этого статистического результата заключить, что детей приносят аисты?

Рассмотрим проблему на содержательном уровне. Аисты появляются там, где им удобно вить гнезда; излюбленным же местом их гнездовья являются высокие дымовые трубы, какие строят в голландских сельских домах. По традиции новая семья строит себе новый дом — появляются новые трубы и, естественно, рождаются дети. Таким образом, и увеличение числа гнезд аистов, и увеличение числа детей являются следствиями одной причины Y — *образования новых семей*.

ЗАДАЧИ

Отложил на осень, а там и забросил.

1. Докажите, что функция $F(\alpha,\beta)$, определяемая формулой (2), имеет глобальный минимум в точке $(\widehat{a},\widehat{b})$ (см. (3)).

2. Проверьте, что если в модели (5) вектор ошибок $\boldsymbol{\varepsilon}$ нормально распределен (см. П9), то МНК-оценки являются оценками максимального правдоподобия (см. § 4 гл. 9).

3*. Убедитесь, что при выполнении допущения Д1 из § 2 матрица $\boldsymbol{B} = \boldsymbol{X}^T\boldsymbol{X}$ в модели (5) положительно определена.

4. Установите адекватность *линейной* $Y = a + bX$ и неадекватность *тригонометрической* $Y = a + b\sin(\pi X)$ зависимостей для данных из примера 4.

5. Проверьте для модели $Y = a + bX$ гипотезу $H'\colon b = 0$ на основе данных из следующей таблицы (подобной таблице из примера 4).

u_l	−0,9	−0,7	−0,5	−0,3	−0,1	0,1	0,3	0,5	0,7	0,9
η_l	0,55	0,53	0,72	0,51	0,45	0,45	0,47	0,54	0,51	0,33
η_{2l}	0,49	0,67	0,49	0,81	0,55	0,67	0,42	0,33	0,53	0,47

X	Y	X	Y	X	Y	X	Y
9,75	19	7,20	61	9,00	68	10,20	75
9,00	40	7,95	62	7,80	69	6,00	76
9,60	42	8,85	62	10,05	69	8,85	77
9,75	42	8,25	65	10,50	70	9,00	80
11,25	47	8,85	65	9,15	71	9,75	82
9,45	49	9,75	65	9,45	71	10,65	82
11,25	50	8,85	66	9,45	71	13,20	82
9,00	54	9,15	66	9,45	72	7,95	83
7,95	56	10,20	66	8,10	73	7,95	86
12,00	56	9,15	67	8,85	74	9,15	88
8,10	57	7,95	68	9,60	74	9,75	88
10,20	57	8,85	68	6,45	75	9,00	94
8,55	58			9,75	75		

Рис. 20

6. На рис. 20 представлены 50 пар наблюдений из исследования докторов Л. Матера и М. Уилсона (см. [23, с. 87]). Рассматривались переменные: X — *длина «линии жизни»* на левой руке в сантиметрах (с точностью до ближайших 0,15 см), Y — *продолжительность жизни человека* (округленная до ближайшего целого года). Верно ли, что Y и X связаны линейной регрессионной зависимостью?

Указание. Для вычисления оценок коэффициентов используйте то, что $\sum x_i = 459{,}9$, $\sum x_i^2 = 4308{,}57$, $\sum \eta_i = 3333$, $\sum x_i\eta_i = 30549{,}75$.

РЕШЕНИЯ ЗАДАЧ

1. Частные производные функции $F(\alpha,\beta)$ таковы:

$$\frac{\partial F}{\partial \alpha} = -2\sum(\eta_i - \alpha - \beta x_i), \quad \frac{\partial F}{\partial \beta} = -2\sum x_i(\eta_i - \alpha - \beta x_i).$$

Приравнивая их нулю, получим систему из двух линейных относительно α и β уравнений[*)]

$$\begin{cases} \alpha n + \beta \sum x_i = \sum \eta_i, \\ \alpha \sum x_i + \beta \sum x_i^2 = \sum x_i\eta_i. \end{cases} \tag{46}$$

Выражая α из первого уравнения и подставляя это выражение во второе, находим для решения β представление

$$\frac{n\sum x_i\eta_i - (\sum x_i)(\sum \eta_i)}{n\sum x_i^2 - (\sum x_i)^2}. \tag{47}$$

[*)] Убедитесь, что она является частным случаем системы (6).

После деления числителя и знаменателя на n^2 получаем в числителе выборочную ковариацию между $\boldsymbol{x}$ и $\boldsymbol{\eta}$, а в знаменателе — выборочную дисперсию набора $\boldsymbol{x}$ (см. формулу (1) гл. 6). Следовательно, дробь (47) совпадает с оценкой $\widehat{b}$ из (3).

Чтобы найти α, удовлетворяющее (46), надо подставить $\widehat{b}$ вместо β в первое из уравнений системы. Разделив обе части на n, приходим к оценке $\widehat{a}$ из (3).

Проверим, что найденная точка является точкой (сначала — локального, затем — глобального) минимума функции F. Для этого вычислим частные производные второго порядка:

$$\frac{\partial^2 F}{\partial \alpha^2} = 2n, \quad \frac{\partial^2 F}{\partial \alpha \partial \beta} = 2\sum x_i, \quad \frac{\partial^2 F}{\partial \beta^2} = 2\sum x_i^2.$$

Если не все x_i одинаковы*), то

$$\frac{\partial^2 F}{\partial \alpha^2} = 2n > 0, \quad \frac{\partial^2 F}{\partial \alpha^2}\,\frac{\partial^2 F}{\partial \beta^2} - \left(\frac{\partial^2 F}{\partial \alpha \partial \beta}\right)^2 = 4n\sum (x_i - \overline{x})^2 > 0,$$

т. е. выполнены достаточные условия наличия в точке $(\widehat{a},\widehat{b})$ локального минимума (см. [42, с. 335]).

Так как F дважды непрерывно дифференцируема (в данном случае вторые производные — константы), а *гессиан* $\nabla^2 F$ во всех точках плоскости есть неотрицательно определенная матрица, то F — выпуклая функция. Следовательно, минимум является глобальным (см. [55, с. 24]).

2. Согласно П9 для нормального вектора $\boldsymbol{\varepsilon}$ допущение Д2 влечет независимость компонент и их одинаковую распределенность по закону $\mathcal{N}(0,\sigma^2)$. Поэтому независимы и случайные величины $\eta_i = \mu_i + \varepsilon_i \sim \mathcal{N}(\mu_i,\sigma^2)$, где $\mu_i = x_{i1}\theta_1 + \ldots + x_{im}\theta_m$. Положим $\boldsymbol{y} = (y_1,\ldots,y_n)^T$. Плотность случайного вектора $\boldsymbol{\eta} = (\eta_1,\ldots,\eta_n)^T$ имеет вид

$$p_{\boldsymbol{\eta}}(\boldsymbol{y}) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}}\,\exp\left\{-\frac{(y_i-\mu_i)^2}{2\sigma^2}\right\} = c\cdot\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu_i)^2\right\}.$$

Максимизация этой функции по $\theta_1,\ldots,\theta_m$ при фиксированных $y_1,\ldots,y_n$ равносильна минимизации суммы квадратов остатков в аргументе экспоненты справа.

3. *Первое решение.* Поскольку для любого $\boldsymbol{y} \in \mathbb{R}^m$

$$\boldsymbol{y}^T\boldsymbol{B}\boldsymbol{y} = \boldsymbol{y}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{y} = (\boldsymbol{X}\boldsymbol{y})^T\boldsymbol{X}\boldsymbol{y} = |\boldsymbol{X}\boldsymbol{y}|^2 \geqslant 0,$$

матрица $\boldsymbol{B}$ неотрицательно определена (она, очевидно, симметрична: $(\boldsymbol{X}^T\boldsymbol{X})^T = \boldsymbol{X}^T(\boldsymbol{X}^T)^T = \boldsymbol{X}^T\boldsymbol{X}$). Из линейной алгебры известно свойство (см. [8, с. 17]): ранг $\boldsymbol{X}^T\boldsymbol{X}$ совпадает

*) В противном случае все точки (x_i,η_i) располагаются на одной вертикальной прямой, и по ним, очевидно, невозможно оценить коэффициент b угла наклона прямой.

с рангом матрицы $\boldsymbol{X}$. Ввиду Д1 ранг $\boldsymbol{X}$ равен m. Поэтому $\boldsymbol{B}$ невырождена и, следовательно, положительно определена.

Второе решение. В его основе лежит идея замены базиса из столбцов $\boldsymbol{X}$ на ортогональный (см. замечание 3). Воспользуемся тем, что существует невырожденная матрица $\boldsymbol{A}$ такая, что $\boldsymbol{X} = \boldsymbol{X}'\boldsymbol{A}$, причем столбцы матрицы $\boldsymbol{X}'$ ортогональны между собой (см. (9), (10)). Тогда

$$\boldsymbol{B} = \boldsymbol{X}^T\boldsymbol{X} = (\boldsymbol{X}'\boldsymbol{A})^T\boldsymbol{X}'\boldsymbol{A} = \boldsymbol{A}^T(\boldsymbol{X}')^T\boldsymbol{X}'\boldsymbol{A} = \boldsymbol{A}^T\boldsymbol{B}'\boldsymbol{A}.$$

Здесь $\boldsymbol{B}' = (\boldsymbol{X}')^T\boldsymbol{X}'$ — диагональная матрица с элементами $|\boldsymbol{x}'_j|^2$, где $\boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_m$ — столбцы матрицы $\boldsymbol{X}'$. Очевидно, $\boldsymbol{B}'$ положительно определена, т. е. $\boldsymbol{y}^T\boldsymbol{B}'\boldsymbol{y} > 0$ для любого $\boldsymbol{y} \in \mathbb{R}^m$, $\boldsymbol{y} \neq \boldsymbol{0}$. Полагая $\boldsymbol{z} = \boldsymbol{A}^{-1}\boldsymbol{y}$, получаем

$$\boldsymbol{z}^T\boldsymbol{B}\boldsymbol{z} = \boldsymbol{y}^T(\boldsymbol{A}^{-1})^T(\boldsymbol{A}^T\boldsymbol{B}'\boldsymbol{A})\boldsymbol{A}^{-1}y = \boldsymbol{y}^T\boldsymbol{B}'\boldsymbol{y} > 0$$

при любом $\boldsymbol{z} \in \mathbb{R}^m$, $\boldsymbol{z} \neq \boldsymbol{0}$, что и требовалось установить.

4. Для линейной зависимости в примере 4 уже были найдены оценки $\widehat{a} = 0{,}53$, $\widehat{b} = -0{,}50$ и подсчитана величина $RSS = 0{,}23$. Для данных из таблицы примера 4 вычисляем $k = 10$ полусумм ординат точек с одинаковыми абсциссами и по формуле (27) находим значение D_0. Подставляя его в (31), получаем $R = 1{,}00$. С помощью линейной интерполяции по аргументу $1/k_1$ табл. Т5 вычисляем квантиль уровня 95% F-распределения с $k_1 = k - m = 10 - 2 = 8$ и $k_2 = n - k = 20 - 10 = 10$ степенями свободы. Она приближенно равна 3,07. Так как $1{,}00 < 3{,}07$, то гипотеза об адекватности линейной модели не отвергается.

Аналогичные расчеты для тригонометрической модели (также ортогональной) дают следующие результаты: $\widehat{a} = 0{,}53$, $\widehat{b} = -0{,}34$, $RSS = 0{,}70$ (более чем в 3 раза превосходит RSS линейной модели), $R = 5{,}76$. Поскольку $5{,}76 > 3{,}07$, на уровне значимости 5% гипотеза адекватности отвергается.

На рис. 21 приведены для сравнения графики подогнанных методом наименьших квадратов прямой и синусоиды. Для последней заметно, в частности, значительное рассогласование с данными вблизи концов отрезка $[-1, 1]$.

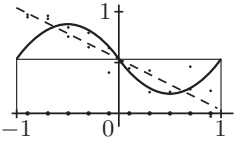


Рис. 21

5. Точно так же, как и в примере 4, вычисляем оценки $\widehat{a} = 0{,}52$, $\widehat{b} = -0{,}09$ и $RSS = 0{,}21$. В силу ортогональности модели для подсчета D_{01} можно воспользоваться формулой (34). Для $m' = 1$ она дает $D_{01} = \widehat{b}^2\,|\boldsymbol{x}_2|^2$. Подставляя значение D_{01} в (35), находим, что $R = 4{,}77$. С помощью табл. Т5 определяем квантиль уровня 95% F-распределения с $k_1 = m - m' = 2 - 1 = 1$ и $k_2 = n - m = 20 - 2 = 18$ степенями свободы. Она равна 4,41. Так как $4{,}77 < 4{,}41$, то на уровне значимости 5% понизить размерность модели не удается.

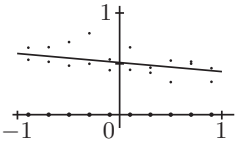


Рис. 22

На рис. 22 представлены данные задачи и подогнанная к ним прямая. По-видимому, все-таки существует тенденция к уменьшению ординат точек с ростом их абсцисс, т. е. истинная линейная зависимость имеет небольшой отрицательный наклон.

6. По формуле (47) вычисляем $\widehat{b} = -1{,}367$. Из (3) получаем $\widehat{a} = -79{,}23$. На основе данных таблицы, приведенной на рис. 20, подсчитываем (с помощью программы Excel) значение $D_0 = RSS = \sum(\eta_i - \widehat{\eta}_i)^2 = 9608{,}7$.

Так как столбцы матрицы $\boldsymbol{X}$ не ортогональны, для вычисления величины D_{01} будем использовать ее определение (23): $D_{01} = D_1 - D_0$. Остается только найти D_1. Для этого заметим, что проекцией вектора $\boldsymbol{\eta}$ на одномерное подпространство L_1, порождаемое вектором $(1,\ldots,1)^T$, очевидно, служит вектор $(\overline{\eta},\ldots,\overline{\eta})^T$, где, как обычно, $\overline{\eta}$ обозначает среднее арифметическое η_i (сравните с однофакторным дисперсионным анализом из § 4). Отсюда

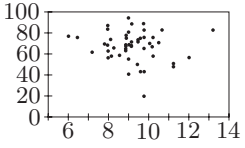


Рис. 23

$$D_1 = |\boldsymbol{\eta} - \Pi_{L_1}\boldsymbol{\eta}|^2 = \sum_{i=1}^{n} (\eta_i - \overline{\eta})^2 .$$

Подсчитав с помощью Excel значение $D_1 = 9755{,}2$, из формулы (35) при $n = 50$, $m = 2$ и $m' = 1$ имеем $R = 0{,}732$. Интерполяцией таблицы T5 по аргументу $1/k_2$ ($k_1 = 1$, $k_2 = 48$) вычисляем квантиль уровня 95%. Она примерно равна 4,04. Так как $0{,}732 < 4{,}04$, то на уровне значимости 5% коэффициент наклона прямой можно считать равным нулю. Иными словами, для рассматриваемых данных нет значимой зависимости продолжительности жизни Y от длины «линии жизни» X (что наглядно демонстрирует рис. 23).

ОТВЕТЫ НА ВОПРОСЫ

1. Да, является при $m = 1$ и $\boldsymbol{x}_1 = (1,\ldots,1)^T$.

2. В L_0 входят такие векторы из $\mathbb{R}^N$ ($N = n_1 + \ldots + n_k$), у которых первые n_1 координат одинаковы, следующие n_2 координат одинаковы и т. д. Столбцы $\boldsymbol{x}_l$ ($l = 1,\ldots,m$) матрицы $\boldsymbol{X}$ (см. (30)), очевидно, лежат в L_0. Поэтому порождаемое ими подпространство также содержится в L_0: $L_1 = L(\boldsymbol{X}) \subseteq L_0$. Если $dim\, L_1 = m < k = dim\, L_0$, то имеет место строгое включение $L_1 \subset L_0$.

3. Так как $w_i = \sigma^2/(cx_i^2)$, то $\mathbf{D}\widehat{\theta} = \sigma^2/\sum w_i x_i^2 = c/n$.

Часть VI
ОБОБЩЕНИЯ И ДОПОЛНЕНИЯ

В эту часть книги включен материал, дополняющий содержание предыдущих глав. В гл. 22 рассматриваются ядерные оценки плотности и функции регрессии (дополнение к гл. 21). Ранговые методы из гл. 14 и 15 обобщаются на случай многомерных данных в гл. 23. Глава 24 посвящена анализу двухвыборочной модели масштаба. В гл. 25 обсуждаются свойства так называемых L-, M- и R-оценок параметра сдвига (углубление материала гл. 8). Наконец, в гл. 26 некоторые ранее установленные результаты выводятся из теоремы о сходимости дифференцируемых функционалов от эмпирической функции распределения к функционалу от броуновского моста.

> Сократ постоянно указывал своим ученикам на то, что при правильно поставленном образовании в каждой науке надо доходить только до известного предела, который не следует преступать. По геометрии, говорил он, достаточно знать настолько, чтобы при случае быть в силах правильно измерить кусок земли, который продаешь или покупаешь, или чтоб разделить на части наследство, или чтоб суметь распределить работу рабочим. Но он не одобрял увлечения большими трудностями в этой науке, и хотя сам лично знал их, но говорил что они могут занять всю жизнь человека и отвлечь его от других полезных наук, тогда как они ни к чему не нужны.
>
> *Л. Н. Толстой*

Глава 22

ЯДЕРНОЕ СГЛАЖИВАНИЕ

§ 1. ОЦЕНИВАНИЕ ПЛОТНОСТИ

Из § 1 гл. 9 известно, что несмещенной и состоятельной оценкой неизвестной функции распределения $F(x)$ случайной величины X является *эмпирическая функция распределения*

$$\widehat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_{(i)} \leqslant x\}}, \tag{1}$$

которая возрастает скачками величины $\frac{1}{n}$ в точках $X_{(i)}$, где $X_{(1)} \leqslant \ldots \leqslant X_{(n)}$ — упорядоченные элементы выборки (график функции $\widehat{F}_n(x)$ изображен на рис. 1 гл. 9).

А как оценить *плотность* $p(x) = F'(x)$ (если она существует)? Так как $\widehat{F}_n(x)$ — кусочно-постоянная функция, то ее производная равна 0 всюду, за исключением точек скачков $\widehat{F}_n(x)$, и не годится в качестве оценки для $p(x)$.

Основная идея состоит в предварительном *сглаживании* эмпирического распределения за счет его свертки (см. П3) с распределениями, имеющими плотности и сходящимися к сосредоточенному в точке 0 распределению. Точнее: рассмотрим случайную величину $Z_n = X + h_n Y$, где случайная величина Y имеет (известную) плотность $q(y)$ и не зависит от X, а числа $h_n > 0$ и $h_n \to 0$ при $n \to \infty$.

Согласно следствию из формулы преобразования плотности (П8), плотностью случайной величины $h_n Y$ служит $\frac{1}{h_n}\, q(\frac{y}{h_n})$. С учетом формулы свертки (П3) выразим плотность $r_n(z)$ случайной величины Z_n:

$$r_n(z) = \frac{1}{h_n}\int q\left(\frac{z-x}{h_n}\right) F(dx) = \frac{1}{h_n}\int q\left(\frac{z-x}{h_n}\right) p(x)\,dx. \tag{2}$$

При $h_n \to 0$ случайные величины Z_n будут сходиться к X по вероятности (П5), а плотности $r_n(z)$ — к плотности $p(z)$ для почти

всех z при выполнении, скажем, условия 1 из приведенной ниже теоремы 1.*)

Заменив $F(x)$ в формуле (2) на эмпирическую функцию распределения $\widehat{F}_n(x)$, получим для $p(z)$ оценку

$$\widehat{p}_n(z) = \frac{1}{h_n}\int q\left(\frac{z-x}{h_n}\right)\widehat{F}_n(dx) = \frac{1}{nh_n}\sum_{i=1}^{n} q\left(\frac{z-X_i}{h_n}\right), \tag{3}$$

которую обычно называют *оценкой Розенблатта—Парзена* (Rosenblatt, 1956; Parzen, 1962). Какими статистическими свойствами она обладает?

Теорема 1. Пусть выполнены следующие условия:

1) плотность $q(y)$ непрерывна и ограничена, причем

$$\alpha = \int q^2(y)\,dy < \infty;$$

2) $h_n \to 0$ при $n \to \infty$ так, что $nh_n \to \infty$. Тогда

$$\widehat{p}_n(z) = r_n(z) + \xi_n(z)\Big/\sqrt{nh_n}\,, \tag{4}$$

где $r_n(z) \to p(z)$ при почти всех z, а случайные величины $\xi_n(z)$ асимптотически нормальны: $\xi_n(z) \xrightarrow{d} \xi(z) \sim \mathcal{N}(0, \alpha p(z))$.

Доказательство этой теоремы можно найти в учебнике [11, с. 58].

Естественно, возникает вопрос об *оптимальном выборе* $q(y)$ и скорости стремления к нулю последовательности h_n при $n \to \infty$. Ответ на него зависит от *свойств гладкости* оцениваемой плотности $p(x)$. Ограничимся классом плотностей, для которых верно допущение

Д1. *Носителем***) $p(x)$ *является конечный интервал, на котором* $p(x)$ *дважды непрерывно дифференцируема с условием*

$$\gamma = \int [p''(x)]^2\,dx < \infty.$$

Предположим также, что для $q(y)$ справедливы условия

Д2. $\int yq(y)\,dy = 0$ *(это всегда верно для четной функции* q*) и*

$$\beta = \int y^2 q(y)\,dy < \infty.$$

Совершив замену $y = (z-x)/h_n$ во втором интеграле в равенствах (2), применим формулу Тейлора:

$$\begin{aligned} r_n(z) &= \int q(y)p(z-yh_n)\,dy = \\ &= \int q(y)\left[p(z) - yh_n p'(z) + \frac{1}{2}\,y^2h_n^2p''(z) + o(y^2h_n^2)\right]dy = \\ &= p(z) + \frac{1}{2}\,h_n^2 p''(z)\int y^2q(y)\,dy + o(h_n^2). \end{aligned}$$

*) Достаточно ограниченности плотности $q(y)$ и выполнения при некоторых $c > 0$ и $\varepsilon > 0$ неравенства $q(y) \leqslant c/|y|^{1+\varepsilon}$ (см. [21, с. 18]).

**) То есть множеством $\{x\colon p(x) > 0\}$.

Подставляя результат в представление (4), видим, что

$$\widehat{p}_n(z) - p(z) = \frac{1}{2}\,\beta h_n^2 p''(z) + \xi_n(z)/\sqrt{nh_n} + o(h_n^2),$$

$$\mathbf{M}\,[\widehat{p}_n(z) - p(z)]^2 = \left[\frac{1}{2}\,\beta h_n^2 p''(z)\right]^2 + \alpha p(z)/(nh_n) + o(h_n^4),$$

поскольку из формул (3) и (2) имеем $\mathbf{M}\widehat{p}_n(z) = r_n(z)$, т. е. $\mathbf{M}\xi_n(z) = 0$.

Математическое ожидание во второй из формул представляет собой *квадратичный риск* оценки $\widehat{p}_n(z)$ (см. § 3 гл. 6).*) Однако риск зависит от z через неизвестные значения $p(z)$ и $p''(z)$. Чтобы избавиться от этого, усредним риск по всем z с помощью интеграла

$$\int \mathbf{M}\,[\widehat{p}_n(z) - p(z)]^2\,dz,$$

главная часть которого (после отбрасывания $o(h_n^4)$) равна

$$J(h_n) = \frac{1}{4}\,\beta^2\gamma h_n^4 + \alpha/(nh_n). \tag{5}$$

При $h_n \to 0$ первое слагаемое в сумме (5), отвечающее за смещение оценки $\widehat{p}_n(z)$, убывает. Второе слагаемое, отражающее дисперсию («разброс») оценки $\widehat{p}_n(z)$, наоборот, возрастает. Минимум функции $J(h_n)$ достигается при $h_n^* = [\alpha/(\beta^2\gamma)]^{1/5}\,n^{-1/5}$ (убедитесь!).

Подставляя h_n^* в формулу (5), находим минимальное значение

$$J(h_n^*) = \left[\frac{5}{4}\,(\alpha^2\beta)^{2/5}\gamma^{1/5}\right] n^{-4/5}. \tag{6}$$

Таким образом, порядок малости оптимального h_n^* есть $n^{-1/5}$. При этом скорость сходимости $\widehat{p}_n(z)$ к $p(z)$, определяемая величиной $\sqrt{J(h_n^*)}$, составляет лишь $n^{-2/5}$ в отличие от скорости $n^{-1/2}$, которая имеет место для сходимости $\widehat{F}_n(x)$ к $F(x)$ (см. § 2 гл. 12). Это можно объяснить тем, что в оценке значения $p(z)$ при фиксированном z принимают участие лишь X_i, находящиеся в некоторой уменьшающейся окрестности точки z, а не вся выборка.

Теперь выберем *оптимальную плотность* $q(y)$. Заметим, что $J(h_n^*)$ в формуле (6) зависит от $q(y)$ через $\alpha^2\beta$. Эта величина инвариантна относительно преобразования $\tilde{Y} = \sigma Y$, $\tilde{q}(y) = \frac{1}{\sigma}\,q(\frac{y}{\sigma})$, где $\sigma > 0$. Поэтому, не ограничивая общности, можно считать, что $\beta = \int y^2 q(y)\,dy = 1$. Тогда (ввиду Д2) приходим к задаче минимизации $\alpha = \int q^2(y)\,dy$ при условиях $\int q(y)\,dy = \int y^2 q(y)\,dy = 1$, $\int yq(y)\,dy = 0$. Ее решением ([86, с. 86]) является функция

$$\frac{3\sqrt{5}}{100}\,(5 - y^2)\,I_{\{|x| \leqslant \sqrt{5}\}},$$

*) *Абсолютный риск* $\mathbf{M}\,|\widehat{p}_n(z) - p(z)|$ изучается в монографии [21].

ранее встречавшаяся в теореме 4 гл. 8. При выборе $\beta = 1/5$ *оптимальная плотность* q^* имеет более простой вид:

$$q^*(y) = \frac{3}{4}\,(1-y^2)\,I_{\{|y|\leqslant 1\}}.$$

В статистической литературе ее обычно именуют *ядром Епанечникова* (Epanechnikov, 1969), хотя оно было еще раньше введено Бартлеттом (Bartlett, 1963).

Ядро — это непрерывная ограниченная четная функция с единичным интегралом $\int q(y)\,dy = 1$ (не обязательно неотрицательная).

Насколько важно использовать именно такое ядро? Конкуренцию ему могут составить ряд ядер, перечисленных в следующей таблице (см. также рис. 1):

N	Ядро	$q(y)$	$E(q)$
1	Епанечникова	$(3/4)(1-y^2)\,I_{\{\lvert y\rvert\leqslant 1\}}$	1
2	Квартическое	$(15/16)(1-y^2)^2 I_{\{\lvert y\rvert\leqslant 1\}}$	0,995
3	Треугольное	$(1-\lvert y\rvert)\,I_{\{\lvert y\rvert\leqslant 1\}}$	0,989
4	Гаусса	$(2\pi)^{-1/2}\exp\{-y^2/2\}$	0,961
5	Прямоугольное	$(1/2)\,I_{\{\lvert y\rvert\leqslant 1\}}$	0,943

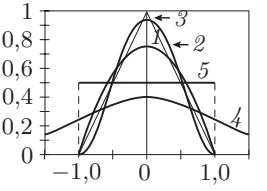


Рис. 1

Квартическое ядро, в отличие от ядра Епанечникова, дифференцируемо в точках -1 и 1. В случае *треугольного ядра* можно быстро производить пересчет оценки плотности $\widehat{p}_n(z)$ при увеличении переменной z с заданным шагом Δz. Это важно при интерактивном построении графиков $\widehat{p}_n(z)$ для «окон сглаживания» h_n разной ширины (см. пример 1 ниже). *Ядро Гаусса* (или *нормальное*) бесконечно дифференцируемо на всей оси. Однако оценка $\widehat{p}_n(z)$ вычисляется медленно по причине многократного подсчета значений экспонент. Достоинством *прямоугольного ядра* является его простой вид. Формально эту функцию нельзя называть ядром из-за разрывов в точках -1 и 1, но ее можно рассматривать как предел последовательности ядер, имеющих форму трапеции.

Ядра — чистый изумруд...

А. С. Пушкин, «Сказка о царе Салтане»

В последнем столбце таблицы указаны *эффективности*

$$E(q) = [\alpha(q)^2\beta(q)]^{-2/5}[\alpha(q^*)^2\beta(q^*)]^{2/5}$$

ядер по отношению к оптимальному ядру Епанечникова q^* (см. формулу (6)). На основе этой информации приходим к заключению, что выбор ядра *мало влияет* на величину среднеквадратичной ошибки оценки $\widehat{p}_n(z)$.

Отметим, что если неизвестная плотность $p(x)$ имеет непрерывные производные более высокого *четного порядка* $m > 2$, то можно добиться и большей, чем $n^{-2/5}$, скорости сходимости за счет использования ядер, принимающих отрицательные значения. При выполнении условий $\int q(y)\,dy = 1$, $\int y^k q(y)\,dy = 0$ при $1 \leqslant k < m$, $\int y^m q(y)\,dy = c \neq 0$ прежние рассуждения приводят к задаче

минимизации функционала

$$\left[\int q^2(y)\,dy\right]^m \left|\int y^m q(y)\,dy\right|,$$

решением которой на отрезке $[-1,1]$ является некоторый полином степени m (см. [85, с. 148]). В частности, для $m=4$ (при $c=-1/21$) получается ядро вида $(15/32)(7y^4-10y^2+3)\,I_{\{|y|\leqslant 1\}}$ (рис. 2). Скорость сходимости соответствующей ядерной оценки равна $n^{-4/9}$.

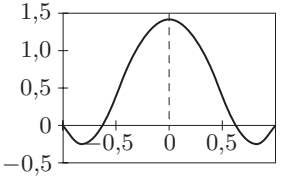
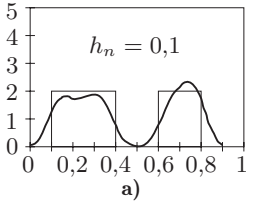


Рис. 2

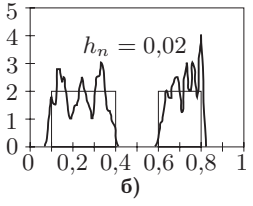
В общем случае погрешность оптимальной оценки имеет порядок малости $n^{-m/(2m+1)}$ (см. [11, с. 60]), который приближается к $n^{-1/2}$ при увеличении m. Это объясняется тем, что для более гладких плотностей к оцениванию значения $p(x)$ привлекаются элементы выборки, лежащие во все более широких окрестностях точки x.

Недостатком таких оценок является возможность появления у них отрицательных значений, что противоречит тому, что $p(x)\geqslant 0$.

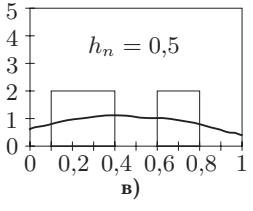
При практическом оценивании плотности $p(x)$ по заданной реализации выборки важен не столько выбор вида ядра, сколько правильное определение *ширины «окна сглаживания»* h_n.



а)



б)



в)

Рис. 3

Пример 1. С помощью таблицы случайных чисел Т1 была моделирована выборка размера $n=100$ из распределения с плотностью $p(z)=2\left(I_{[0,1;0,4]}+I_{[0,6;0,8]}\right)$. На рис. 3 приведены графики оценок $\widehat{p}_n(z)$, вычисленных на основе треугольного ядра для следующих значений h_n: 0,1; 0,02; 0,5.

Выбор слишком малого h_n приводит к быстро меняющейся, неустойчивой оценке, так как $\widehat{p}_n(z)$ опирается только на небольшое количество наблюдений из узкой окрестности точки z (велико второе слагаемое в формуле (5), отвечающее за дисперсию оценки).

С другой стороны, слишком большое значение h_n влечет чрезмерное сглаживание оцениваемой плотности, что не позволяет выявить ее характерные особенности (такие, например, как наличие нескольких максимумов). Оценка $\widehat{p}_n(z)$ в этом случае оказывается сильно смещенной.

§ 2. НЕПАРАМЕТРИЧЕСКАЯ РЕГРЕССИЯ

Изучаемая в этом параграфе задача заключается в сглаживании данных $\{(x_i,y_i)\}_{i=1}^n$ при помощи некоторой непрерывной кривой $y=m(x)$. Данные рассматриваются как реализация выборки, описываемой одной из двух моделей.

Модель со случайным планом эксперимента

Пары (X_i,Y_i) считаются независимыми и одинаково распределенными. *Кривая регрессии* определяется как (см. П7)

$$m(x)=\mathbf{M}(Y|X=x).$$

Это определение корректно, если $\mathbf{M}|Y| < \infty$. Если у вектора (X,Y) существует плотность $p(x,y)$, то

$$m(x) = \int y\, p(x,y)\, dy \Big/ p(x),$$

где $p(x) = \int p(x,y)\, dy$ — *маргинальная плотность* случайной величины X.

Например, пусть $(X,Y) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ (см. П9), где $\boldsymbol{\mu} = (\mu_1,\mu_2)$, $\boldsymbol{\Sigma} = \|\sigma_{ij}\|_{2\times 2}$. Несложно подсчитать ([2, с. 167]), что в этом случае

$$m(x) = \mu_2 + \frac{\sigma_{12}}{\sigma_{22}}\,(x - \mu_1),$$

т. е. кривая регрессии линейна.

В случае *управляемой неслучайной* предикторной (предсказывающей) переменной X применяется

Модель с фиксированным планом эксперимента

В этой модели

$$Y_i = m(x_i) + \varepsilon_i, \quad i = 1,\ldots,n,$$

где x_i — *«узлы»* (задаваемые значения X, в которых производятся измерения отклика Y), ε_i — независимые и одинаково распределенные *случайные ошибки*, $\mathbf{M}\varepsilon_i = 0$, $\mathbf{D}\varepsilon_i = \sigma^2$.

Замечание 1. Если допустимы многократные наблюдения в фиксированной точке $X = x$, то можно оценить $m(x)$ с помощью среднего арифметического соответствующих Y_i.

Однако, часто этому препятствуют финансовые ограничения: слишком дорого проводить более одного измерения для каждого фиксированного уровня x предиктора X.

Возможно также, что условия эксперимента нельзя воспроизвести из-за разрушения объекта (например, при изучении безопасности водителя при столкновении автомобилей) и т. п.

Если есть уверенность, что m — гладкая кривая, наблюдения X_i вблизи точки x должны содержать информацию о значении m в точке x. Таким образом, представляется возможным использовать нечто вроде локального усреднения близких к x данных для формирования оценки $m(x)$.

Р. Ойбанк

В основе процедуры оценивания кривой регрессии $m(x)$ лежит идея *локального усреднения.*

Оценка определяется как *взвешенное среднее*

$$\widehat{m}(x) = \sum_{i=1}^{n} w_i(x) Y_i \Big/ \sum_{i=1}^{n} w_i(x), \tag{7}$$

где *веса* $w_i(x)$ велики для X_i, близких к точке x, и малы для остальных X_i.

Замечание 2. В соответствии с примером 5 гл. 21 величина $\widehat{m}(x)$ при каждом x является *оценкой взвешенных наименьших квадратов*, так как она служит решением задачи минимизации функции

$$\sum_{i=1}^{n} w_i(x)(Y_i-\theta)^2 = \sum_{i=1}^{n} w_i(x)(Y_i-\widehat{m}(x))^2. \tag{8}$$

Один из наиболее простых способов определения весов $w_i(x)$ опирается на использование некоторого ядра $q(y)$:

$$w_i(x) = q_h(x-X_i), \quad \text{где } q_h(y) = h^{-1}q(y/h). \tag{9}$$

Здесь $h = h_n$ обозначает ширину «окна сглаживания». Для многомерной предикторной переменной $\boldsymbol{X}_i = (X_{i1},\ldots,X_{im})$ можно взять

$$w_i(x_1,\ldots,x_m) = \prod_{j=1}^{m} q_h(x_j - X_{ij}).$$

Задавать веса можно и по-другому. Для одномерной предикторной переменной Гассер и Мюллер в 1979 г. (см. [85, с. 328]) предложили

$$\tilde{w}_i(x) = \int_{X_{(i-1)}}^{X_{(i)}} q_h(x-y)\,dy, \tag{10}$$

где $X_{(1)} \leqslant X_{(2)} \leqslant \ldots \leqslant X_{(n)}$ — упорядоченные по возрастанию X_i, $X_{(0)} = -\infty$, $X_{(n+1)} = +\infty$. Заметим, что $\sum\limits_{i=1}^{n+1} \tilde{w}_i(x) = 1$.

В случае весов $w_i(x)$ из формулы (9) правая часть равенства (7) приобретает вид

$$\widehat{m}(x) = \sum_{i=1}^{n} q_h(x-X_i)\,Y_i \Big/ \sum_{i=1}^{n} q_h(x-X_i). \tag{11}$$

Выражение (11) определяет так называемую *оценку Надарая—Ватсона* (Nadaraya, 1964; Watson, 1964). С учетом формулы (3)

$$\widehat{m}(x) = \frac{1}{n}\sum_{i=1}^{n} q_h(x-X_i)\,Y_i \Big/ \widehat{p}_n(x),$$

где $\widehat{p}_n(x)$ — ядерная оценка Розенблатта—Парзена маргинальной плотности $p(x)$ переменной X.

В следующей теореме приведены условия, обеспечивающие состоятельность оценки Надарая—Ватсона $\widehat{m}(x)$.

Теорема 2. Пусть для модели со случайным планом эксперимента и одномерной предикторной переменной

1) $\int |q(y)|\,dy < \infty$,

2) $yq(y) \to 0$ при $|y| \to \infty$,
3) $\mathbf{M}Y^2 < \infty$,
4) $n \to \infty$, $h_n \to 0$, $nh_n \to \infty$.

Тогда $\widehat{m}(x) \xrightarrow{\mathsf{P}} m(x)$ в любой точке непрерывности функций $m(x)$, $p(x)$ и $\sigma^2(x) = \mathbf{D}(Y|X = x)$ такой, что $p(x) > 0$.

Доказательство этой теоремы приведено в [85, с. 51].

Сформулированный результат показывает, что оценка $\widehat{m}(x)$ из формулы (11) сходится по вероятности к $m(x)$ — истинной кривой регрессии. Естественно поставить вопрос, какова скорость этой сходимости. Одной из возможных мер точности оценки является *квадратичный риск* $R(x,h) = \mathbf{M}\,[\widehat{m}(x) - m(x)]^2$.

Теорема 3 (Гассер и Мюллер, 1984). Рассмотрим модель с фиксированным планом эксперимента[*)] для одномерной предикторной переменной X. Без ограничения общности можно считать, что «узлы» $x_i \in [0,1]$. Предположим, что

1) используется ядро $q(y)$ с
$$\alpha = \int q^2(y)\,dy < \infty \quad \text{и} \quad \beta = \int y^2 q(y)\,dy < \infty;$$
2) $q(y)$ имеет носитель $(-1,1)$, причем $q(-1) = q(1) = 0$;
3) веса $\tilde{w}_i(x)$ задаются формулой (10);
4) $m(x)$ дважды непрерывно дифференцируема;
5) $\max\limits_{1 \leqslant i \leqslant n+1} (x_{(i)} - x_{(i-1)}) = O(1/n) \qquad (x_{(0)} = 0,\ x_{(n+1)} = 1)$;
6) $\mathbf{D}\varepsilon_i = \sigma^2$, $i = 1, \ldots, n$;
7) $n \to \infty$, $h_n \to 0$, $nh_n \to \infty$.

Тогда
$$R(x,h_n) = \left[\frac{1}{4}\,\beta^2\,[m''(x)]^2\,h_n^4 + \alpha\sigma^2/(nh_n)\right](1 + o(1)).$$

Видим, что так же, как и в формуле (5), квадратичный риск состоит из двух частей: квадрата смещения и дисперсии, где смещение (как функция от h_n) возрастает, а дисперсия — убывает. В. Хардле в [85, с. 41] пишет: «Это качественное соображение раскрывает **сущность задачи сглаживания**: *баланс между дисперсией и квадратом смещения*».

Пример 2. ***Обнаружение локального максимума кривой регрессии*** **([85, с. 92]).** На рис. 4 представлены моделированные данные. Использовалась модель с фиксированным планом эксперимента: $n = 50$, $x_i = i/n$, $i = 1, \ldots, n$. Точки порождены «зашумлением» с помощью взятой из таблицы Т1 реализации $\varepsilon_i \sim \mathcal{N}(0,1)$ кривой регрессии
$$m(x) = 1 - x + \exp\left\{-200\,(x - 1/2)^2\right\}$$

[*)] Для случайного плана эксперимента скорость сходимости та же, но с другими константами (см. [85, с. 102]).

(т. е. на линейную функцию $y = 1 - x$ накладывается нормальный «горб» с центром в $1/2$ и шириной примерно равной $2\sigma = 1/10$).

Из-за большой интенсивности «шума» локальный максимум $m(x)$ практически не различим. Однако, он хорошо заметен на графике (пунктир) оценки $\widehat{m}(x)$ из (11), для которой использовалось гладкое квартическое ядро $q(y) = (15/16)(1 - y^2)^2 I_{\{|y| \leqslant 1\}}$ и был определен *методом пропуска* (см. ниже) оптимальный размер «окна сглаживания» $h = 0{,}09$.

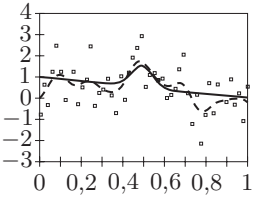


Рис. 4

Методы ядерного сглаживания могут быть также применены и для оценивания *производных кривой регрессии*. Так, первую производную $m'(x)$ можно оценить, используя веса

$$w_i^*(x) = q'(z_i) \sum_{j=1}^{n} q(z_j) \; - \; q(z_i) \sum_{j=1}^{n} q'(z_j), \quad \text{где } z_i = (x - X_i)/h,$$

получаемые дифференцированием правой части равенства (11) по x.

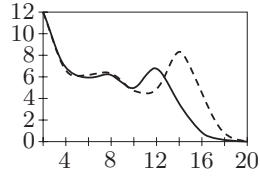


Рис. 5

Пример 3. ***Зависимость скорости роста от возраста*** **(см. [85, с. 46]).** На рис. 5 представлены ядерные оценки кривых *средней скорости* V (в см/год) *роста* H (т. е. $V = H'$) в зависимости от *возраста* T (в годах) для мальчиков (штрихованная линия) и для девочек (сплошная линия).

У девочек наблюдается локальный максимум скорости роста около 12 лет. У мальчиков аналогичный пик еще более выражен, но происходит примерно на 2 года позже.

Применение метода непараметрической регрессии позволило обнаружить и у мальчиков, и у девочек дополнительный локальный максимум V, так называемый *средний скачок роста*, в возрасте около 8 лет. Другие подходы, основанные на априорной фиксации параметрических моделей, приводят к значительным трудностям в обнаружении этого второго пика.

Так же, как и при оценивании плотности, решающую роль играет не вид функции $q(y)$, а правильный выбор ширины «окна сглаживания» h. Часто в этом помогает

Метод пропуска или кросс-проверка.

Минимизируется по h функция $S(h) = \sum\limits_{i=1}^{n} [Y_i - \widehat{m}_i(X_i)]^2$, где $\widehat{m}_i(X_i) = \sum\limits_{j \neq i} w_j(X_i)\, Y_j \Big/ \sum\limits_{j \neq i} w_j(X_i)$ представляет собой оценку для $m(X_i)$ по выборке, в которой i-е наблюдение пропущено. Здесь $S(h)$ — сумма квадратов ошибок предсказания Y_i (*невязок*) на основе выборки без (X_i, Y_i).

Для данных из примера 2 был подсчитан ряд значений $S(h)$:

h	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,50
$S(h)$	51,0	46,6	47,3	47,6	47,5	47,2	47,7	48,3	48,7	49,1

Заметим, что $S(h)$ имеет два локальных минимума: при $h \approx 0{,}1$ и $h \approx 0{,}3$. Уточним оптимальное значение h, проводя вычисления по сетке с меньшим шагом:

h	0,05	0,06	0,07	0,08	0,09	0,10	0,11	0,12	0,13	0,14	0,15
$S(h)$	51,0	48,7	47,4	46,7	46,4	46,6	46,7	46,7	46,8	47,1	47,3

Таким образом, с точностью до 0,01 функция $S(h)$ имеет глобальный минимум при $h = 0{,}09$.

Альтернативным методом, применяемым в случае, когда $X_1, \ldots, X_n$ располагаются очень неравномерно (*«где пусто, а где густо»*), является модификация ядерной оценки, у которой h зависит от x. В качестве такого $h(x)$ часто берут расстояние от точки x до ее *k-го (в смысле близости к x) соседа* X_i (см. [85, с. 54]). Значение параметра k ($1 \leqslant k \leqslant n$) задается исследователем.

В заключение главы обсудим проблему уменьшения влияния выделяющихся наблюдений («выбросов») на ядерные оценки кривой регрессии. К сожалению, локальное усреднение, осуществляемое этими оценками, имеет неограниченную возможность реагировать на «выбросы».

Например, график оценки $\widehat{m}(x)$ (пунктир) в правой половине рис. 4 особенно извилист из-за необходимости компенсировать квадраты остатков в точках с большими значениями ошибок ε_i (таких, как $\varepsilon_{35} = 1{,}73$ и $\varepsilon_{39} = -2{,}36$).

Повышение устойчивости сглаживания может быть достигнуто за счет уменьшения весов наблюдений с большими невязками. Кливленд (Cleveland W. S., 1979) предложил следующий простой алгоритм, реализующий эту идею (см. [85, с. 209]). Процедура начинается с вычисления пробной оценки, итеративно пересчитываются веса и несколько раз производится повторное сглаживание.

Алгоритм «LOWESS» (модификация для m-мерных предикторных переменных $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{im})$)

В оригинале LOcally WEighted Scatter plot Smoothing (*англ.*) — локально взвешенное сглаживание диаграммы рассеяния.

1. Задается k ($1 \leqslant k \leqslant n$). Для каждого $\boldsymbol{X}_i$ вычисляется расстояние $h(\boldsymbol{X}_i)$ до k-го соседа в $\mathbb{R}^m$ и по формуле (9) определяется соответствующий вес $w_i(\boldsymbol{X}_i)$. Веса всех наблюдений запоминаются в массиве.

2. В окрестности каждой точки $\boldsymbol{X}_i$ строится локальная линейная аппроксимация *поверхности регрессии* $m(\boldsymbol{X})$ на основе взвешенного метода наименьших квадратов (см. § 5 гл. 21) путем минимизации по $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$ функции

$$\sum_{i=1}^{n} w_i(\boldsymbol{X}_i)\,(Y_i - \theta_1 X_{i1} - \ldots - \theta_m X_{im})^2\,.$$

3. Используя оценки невязок $\widehat{\varepsilon}_i$, вычисляется выборочная медиана $\widehat{\sigma} = MED\{|\widehat{\varepsilon}_i|,\ i = 1, \ldots, n\}$ (оценка масштаба) и определяются *коэффициенты устойчивости* $c_i = q(\widehat{\varepsilon}_i/(6\widehat{\sigma}))$, где $q(y) = (15/16)(1 - y^2)^2 I_{\{|y| \leqslant 1\}}$ — квартическое ядро.

4. Строится локальная линейная аппроксимация, как в пункте 2, но с весами $c_i w_i(\boldsymbol{X}_i)$.

5. Пункты 3 и 4 повторяются до стабилизации величин c_i.

Пример 4. На рис. 6 из [85, с. 210] показано применение алгоритма «LOWESS» к моделированным данным. Точки генерировались в соответствии с моделью $Y_i = (1/50)x_i + \varepsilon_i$ $(n = 50)$, $x_i = i$, $\varepsilon_i \sim \mathcal{N}(0, 1)$. *Кривой регрессии* является прямая с малым коэффициентом наклона, причем дисперсия ошибок $\mathbf{D}\varepsilon_i$ настолько велика, что систематический прирост Y при увеличении X (так называемый *тренд*) почти не заметен из-за интенсивного «шума» (если оставить на рис. 6 только «крестики»).

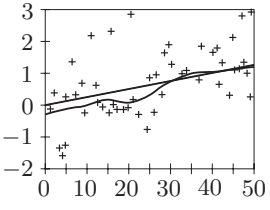


Рис. 6

Для $k = 25$, $m = 2$ и $\boldsymbol{X}_i = (1, x_i)$ на рисунке представлен также результат сглаживания после двух итераций алгоритма. Построенная оценка уверенно игнорирует «выбросы», расположенные вблизи границ рисунка, и довольно точно воспроизводит теоретический линейный тренд.

Глава 23

МНОГОМЕРНЫЕ МОДЕЛИ СДВИГА

§ 1. СТРАТЕГИЯ ПОСТРОЕНИЯ КРИТЕРИЕВ

В наше время накоплено огромное количество знаний, достойных изучения. Скоро наши способности будут слишком слабы, а жизнь — слишком коротка, чтобы усвоить хотя бы одну только наиболее полезную часть этих знаний. К нашим услугам полное изобилие богатств, но, восприняв их, мы должны снова отбрасывать многое, как бесполезный хлам. Было бы лучше иногда не обременять себя им.

И. Кант

Кто не понимает ничего, кроме химии, тот и ее понимает недостаточно.

Г. Лихтенберг

В этой главе одно- и двухвыборочные ранговые критерии из гл. 14 и гл. 15 обобщаются на многомерный случай.*) Если в многомерной модели участвуют p-мерные векторы, то нас будет интересовать некоторая векторная p-мерная статистика $\boldsymbol{S} = (S_1, \ldots, S_p)$, у которой каждая компонента S_j при нулевой гипотезе *центрирована*: $\mathbf{M}S_j = 0,\ j = 1, \ldots, p$.

При построении критериев станем придерживаться следующего **плана**:

1. установим, что в случае справедливости нулевой гипотезы при $n \to \infty$ распределение статистики $n^{-1/2}\boldsymbol{S}$ стремится к p-мерному нормальному закону $\mathcal{N}(\mathbf{0}, \boldsymbol{V})$ с *невырожденной* матрицей ковариаций $\boldsymbol{V}$ (см. П9);
2. найдем состоятельную оценку $\widehat{\boldsymbol{V}}$ для $\boldsymbol{V}$ (см. § 2 гл. 6);
3. исходя из этого, видим, что *статистика критерия*

$$S^* = n^{-1}\boldsymbol{S}^T\widehat{\boldsymbol{V}}^{-1}\boldsymbol{S} = \boldsymbol{S}^T(n\widehat{\boldsymbol{V}})^{-1}\boldsymbol{S}$$

распределена в пределе по закону χ^2_p (см. П5, П9).

§ 2. ОДНОВЫБОРОЧНАЯ МОДЕЛЬ

Пусть данные представляют собой реализацию n независимых и одинаково распределенных p-мерных случайных векторов $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, каждый из которых обладает p-мерной функцией распределения $F(x_1 - \theta_1, \ldots, x_p - \theta_p)$, где $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$ обозначает *неизвестный вектор сдвига*.

Данные можно представить в виде таблицы размера $n \times p$:

$$\begin{matrix} X_{11} & X_{12} & \ldots & X_{1p} \\ \vdots & \vdots & & \vdots \\ X_{n1} & X_{n2} & \ldots & X_{np} \end{matrix}$$

*) Изложение, в основном, следует гл. 6 монографии [86].

Здесь i-й строкой служит вектор $\boldsymbol{X}_i$, j-й столбец представляет собой выборку размера n, взятую по j-й компоненте векторов. (Такая таблица обычно возникает при записи p измерений по каждому из n объектов.)

Предположим, что выполняются **следующие допущения**.

Д1. *Распределение F имеет плотность $p(x_1,\ldots,x_p)$.*

Обозначим через $F_1(x),\ldots,F_p(x)$ маргинальные (частные) функции распределения, а через F_{jk} $(j,k=1,\ldots,p)$ — двумерные частные функции распределения.

Д2. *Для каждой из $F_j(x)$, $j=1,\ldots,p$, единственным решением уравнения $F_j(x)=1/2$ является* 0.

Другими словами, допущение Д2 означает, что 0 является единственной медианой распределений $F_j(x)$ (см. § 2 гл. 7).

Мы хотим проверить гипотезу

$$H_0\colon \boldsymbol{\theta}=\mathbf{0} \quad \text{против альтернативы } H_1\colon \boldsymbol{\theta}\neq\mathbf{0},$$

где $\mathbf{0}=(0,\ldots,0)$.[*)] Без дополнительных предположений о виде распределения F для построения критерия подходящей представляется статистика $\boldsymbol{S}$, компонентами которой служат *знаковые статистики* (см. § 2 гл. 15). Мы будем использовать (более удобную) симметричную форму знаковой статистики, имеющую нулевое математическое ожидание при условии справедливости гипотезы H_0:

$$S_j=\sum_{i=1}^{n}\operatorname{sign}X_{ij}=\Sigma_j^+-\Sigma_j^-,\quad j=1,\ldots,p,$$

где $\Sigma_j^+=\sum\limits_{i=1}^{n}I_{\{X_{ij}>0\}}$, $\Sigma_j^-=\sum\limits_{i=1}^{n}I_{\{X_{ij}<0\}}$, $\operatorname{sign}x$ — *знак числа x*, I_A — *индикатор события A*. (Здесь предполагается, что нет наблюдений, равных 0, что верно с вероятностью 1.) Поскольку $\Sigma_j^++\Sigma_j^-=n$, получаем $S_j=2\,\Sigma_j^+-n$, т. е. статистика S_j и статистика знаков из § 2 гл. 15, равная Σ_j^+, *линейно связаны*.

Многомерное обобщение критерия знаков

Чтобы сформулировать теорему, на которой основывается обобщение, потребуются некоторые дополнительные обозначения. Так, пусть $\boldsymbol{X}=(X_1,\ldots,X_p)$ — случайный вектор с функцией распределения F. Для $j,k=1,\ldots,p$ положим

$$p_{jk}^{++}=\mathbf{P}(X_j>0,X_k>0),\quad p_{jk}^{+-}=\mathbf{P}(X_j>0,X_k<0),$$
$$p_{jk}^{-+}=\mathbf{P}(X_j<0,X_k>0),\quad p_{jk}^{--}=\mathbf{P}(X_j<0,X_k<0).$$

[*)] Для заданного вектора $\boldsymbol{\theta}'$ гипотезу $H_0'\colon \boldsymbol{\theta}=\boldsymbol{\theta}'$ можно преобразовать в H_0, вычитая $\boldsymbol{\theta}'$ из векторов наблюдений.

Теорема 1. При справедливости гипотезы H_0 и выполнении допущений Д1, Д2 имеет место сходимость

$n^{-1/2}\boldsymbol{S} \xrightarrow{d} \boldsymbol{Z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{V})$ при $n \to \infty$,
где $\boldsymbol{V} = \|v_{jk}\|_{p\times p}$, $v_{jj} = 1$,
$v_{jk} = p_{jk}^{++} + p_{jk}^{--} - p_{jk}^{+-} - p_{jk}^{-+}$ в случае $k \neq j$.

Доказательство. Ввиду независимости X_{ij}, $i = 1, \ldots, n$, согласно свойствам математического ожидания и дисперсии (П2)

$$\mathsf{M}S_j = \sum_{i=1}^{n} \mathsf{M}\operatorname{sign} X_{ij} = n\,\mathsf{M}\operatorname{sign} X_j = n\,(\mathsf{P}(X_j > 0) - \mathsf{P}(X_j < 0)) = 0,$$

$$\mathsf{D}S_j = \sum_{i=1}^{n} \mathsf{D}\operatorname{sign} X_{ij} = n\,\mathsf{D}\operatorname{sign} X_j = n\,\mathsf{M}(\operatorname{sign} X_j)^2 = n,$$

откуда $v_{jj} = 1$. При $k \neq j$ находим, что

$$\mathsf{cov}\,(S_j, S_k) = \sum_{i=1}^{n} \mathsf{M}(\operatorname{sign} X_{ij} \operatorname{sign} X_{ik}) = n\,\mathsf{M}(\operatorname{sign} X_j \operatorname{sign} X_k) = nv_{jk}.$$

Теперь установим асимптотическую нормальность. Для произвольного вектора $\boldsymbol{\lambda} \in \mathbb{R}^p$ распределение случайных величин

$$\xi_n = n^{-1/2}\boldsymbol{\lambda}^T\boldsymbol{S} = n^{-1/2}\sum_{i=1}^{n} Y_i, \quad \text{где } Y_i = \sum_{i=1}^{p} \lambda_j \operatorname{sign} X_{ij},$$

Y_i — случайные величины с $\mathsf{M}Y_i = 0$ и $\mathsf{D}Y_i = \boldsymbol{\lambda}^T\boldsymbol{V}\boldsymbol{\lambda}$, сходится при $n \to \infty$ к закону $\mathcal{N}(0, \boldsymbol{\lambda}^T\boldsymbol{V}\boldsymbol{\lambda})$ в силу центральной предельной теоремы (П6). Согласно теореме непрерывности (П9) при всех t характеристические функции $\mathsf{M}\exp\{it\xi_n\} \to \exp\{-\boldsymbol{\lambda}^T\boldsymbol{V}\boldsymbol{\lambda}t^2/2\}$. Остается применить обратное утверждение теоремы непрерывности, положив $\boldsymbol{\lambda}' = t\boldsymbol{\lambda}$. ■

Для построения критерия нам нужна состоятельная оценка матрицы $\boldsymbol{V}$. Поскольку $\mathsf{M}(\operatorname{sign} X_j \operatorname{sign} X_k) = v_{jk}$, в силу закона больших чисел (П6) имеет место сходимость

$$\widehat{v}_{jk} = \frac{1}{n}\sum_{i=1}^{n} \operatorname{sign} X_{ij} \operatorname{sign} X_{ik} \xrightarrow{\mathsf{P}} v_{jk}. \tag{1}$$

Определим матрицу $\widehat{\boldsymbol{V}}$, полагая $\widehat{v}_{jj} = 1$, беря $\widehat{v}_{jk}$ из формулы (1). При нулевой гипотезе $\widehat{\boldsymbol{V}}$ — состоятельная оценка $\boldsymbol{V}$, а статистика

$$S^* = \boldsymbol{S}^T(n\widehat{\boldsymbol{V}})^{-1}\boldsymbol{S} \tag{2}$$

асимптотически распределена по закону χ^2_p при условии невырожденности матрицы $\boldsymbol{V}$.

*) Используя равенство маргинальной медианы нулю, нетрудно показать (проверьте!), что $v_{jk} = 4p_{jk}^{++} - 1$.

При $p = 2$ статистику S^* можно записать в виде (убедитесь!)

$$S^* = \frac{(N^{++} - N^{--})^2}{N^{++} + N^{--}} + \frac{(N^{+-} - N^{-+})^2}{N^{+-} + N^{-+}}, \tag{3}$$

где $N^{++} = \sum\limits_{i=1}^{n} I_{\{X_{i1}>0,\, X_{i2}>0\}}$, $N^{--} = \sum\limits_{i=1}^{n} I_{\{X_{i1}<0,\, X_{i2}<0\}}$ и т. п.

Эти величины удобно представлять в форме таблицы.

	X_2	
X_1	< 0	> 0
< 0	N^{--}	N^{-+}
> 0	N^{+-}	N^{++}

Приводимый ниже пример показывает, что многомерный критерий имеет преимущество перед наивной стратегией применения одномерных критериев по каждой из компонент.

Пример 1. ***Тест проверки способностей*** **[86, с. 300].** Был проведен выборочный опрос $n = 100$ американских студентов с использованием университетского психологического теста проверки способностей *Scholastic Aptitude Test*. Данные представляют собой пары измерений (V_i,Q_i), $\ldots,n$, где V_i — сумма баллов, полученных i-м студентом по *вербальной* (verbal) части теста, а Q_i — по *количественной* (quantitative) части теста. В таблице приведены числа измерений, которые ниже (меньше) и выше (больше) национального среднего по двум компонентам. Пусть $\boldsymbol{\theta}' = (\theta'_V,\theta'_Q)$ — вектор национальных средних. Проверим на уровне значимости $\alpha = 0{,}05$ гипотезу $H'_0\colon \boldsymbol{\theta} = \boldsymbol{\theta}'$ против альтернативы $H'_1\colon \boldsymbol{\theta} \neq \boldsymbol{\theta}'$.

	Q	
V	Ниже	Выше
Ниже	34	22
Выше	8	36

В соответствии с формулой (3) имеем $S^* = 2^2/70 + 14^2/30 \approx 6{,}59$. Эта величина превосходит 0,95-квантиль закона χ^2_2, равную 5,99 (см. табл. Т3). Следовательно, гипотеза H'_0 отвергается на уровне 5%.

С другой стороны, отдельно по компоненте V оказалось 56 измерений ниже θ'_V и 44 — выше θ'_V. Нормальное приближение для критерия знаков с поправкой на непрерывность (см. § 2 гл. 15 и табл. Т2) дает для таких данных приближенный фактический уровень значимости $\alpha_0 = 0{,}136$ (для односторонней альтернативы).

Аналогично, для компоненты Q имеем 42 измерения ниже θ'_Q и 58 — выше θ'_Q. Для таких данных $\alpha_0 = 0{,}069 > 0{,}05$.

Многомерное обобщение критерия знаковых рангов (см. § 3 гл. 15)

Для получения этого критерия потребуется ввести дополнительное допущение о *центральной симметричности плотности* $p(x_1,\ldots,x_p)$.

Д3. *Плотность* $p(x_1,\ldots,x_p)$ *при любых* $x_1,\ldots,x_p$ *удовлетворяет условию* $p(x_1,\ldots,x_p)=p(-x_1,\ldots,-x_p)$.

Из Д3, очевидно, следует, что каждая из частных плотностей $p_j(x)$, $j=1,\ldots,p$, является четной функцией.

На рис. 1, *а* изображена область, ограниченная характерной (типичной) линией уровня некоторой (взятой для примера) двумерной плотности, удовлетворяющей условию Д3, а на рис. 1, *б* — аналогичная область для плотности, не удовлетворяющей Д3.

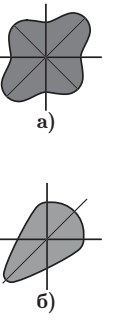


Рис. 1

Рассмотрим вектор (центрированных) знаковых ранговых статистик Уилкоксона $\boldsymbol{T}=(T_1,\ldots,T_p)$ с компонентами

$$T_j=\frac{1}{n+1}\sum_{i=1}^{n}R_{ij}\,\mathrm{sign}(X_{ij}),\quad j=1,\ldots,p, \tag{4}$$

где R_{ij} — ранг $|X_{ij}|$ относительно $|X_{1j}|,\ldots,|X_{nj}|$. Следующая теорема дает нам предельное распределение вектора $\boldsymbol{T}$ при условии справедливости гипотезы H_0.

Теорема 2. При справедливости гипотезы H_0 и выполнении условия Д3 имеет место сходимость

$$n^{-1/2}\,\boldsymbol{T}\xrightarrow{d}\boldsymbol{Z}\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{V})\ \text{ при } n\to\infty,$$

где $\boldsymbol{V}=\|v_{jk}\|_{p\times p}$, $v_{jj}=1/3$,

$$v_{jk}=4\int\limits_{-\infty}^{\infty}F_j(x)\,F_k(y)\,dF_{jk}(x,y)-1\ \text{ в случае } k\neq j.$$

Доказательство теоремы 2 можно найти в [86, с. 291].

Для построения критерия требуется состоятельная оценка матрицы $\boldsymbol{V}$. Учитывая, что $\mathrm{sign}(x)=2\,I_{\{x>0\}}-1$ при $x\neq 0$ и что сумма $\sum\limits_{i=1}^{n}R_{ij}=n(n+1)/2$, запишем правую часть равенства (4) в виде

$$T_j=\frac{2}{n+1}\sum_{i=1}^{n}R_{ij}I_{\{X_{ij}>0\}}\ -\ n/2\equiv\frac{2}{n+1}\ T_j^{+}\ -\ n/2.$$

Статистика T_j^{+} изучалась в § 3 гл. 15. В частности, при решении задачи 6 гл. 15 было установлено, что ее дисперсия $\mathbf{D}T_j^{+}=n(n+1)(2n+1)/24$. На основании этого возьмем

$$\widehat{v}_{jj}=\frac{1}{n}\ \mathbf{D}T_j=\frac{1}{n}\ \frac{4}{(n+1)^2}\ \mathbf{D}T_j^{+}=\frac{2n+1}{6(n+1)}\,. \tag{5}$$

В [86, с. 291] объясняется, что в качестве состоятельной оценки для v_{jk} при $j \neq k$ годится

$$\widehat{v}_{jk} = \frac{1}{n(n+1)^2} \sum_{i=1}^{n} R_{ij} \operatorname{sign}(X_{ij})\, R_{ik} \operatorname{sign}(X_{ik}). \tag{6}$$

С учетом соотношений (5) и (6) положим $\widehat{\boldsymbol{V}} = \|\widehat{v}_{jk}\|$. Тогда статистика

$$T^* = \boldsymbol{T}^T (n\widehat{\boldsymbol{V}})^{-1} \boldsymbol{T} \tag{7}$$

в случае справедливости гипотезы H_0 распределена в пределе по закону χ^2_p при условии невырожденности матрицы $\boldsymbol{V}$.

Пример 2. ***Влияние высоты проживания на величину артериального давления.*** В [86, с. 293] приведена таблица значений давления крови у $n = 15$ перуанских индейцев в возрасте 21 год, родившихся высоко над уровнем моря, родители которых также родились в высокогорье. Обозначим через X величину *систолического* давления, а через Y — величину *диастолического* давления.

На рис. 2 изображена диаграмма рассеяния точек (x_i,y_i), $i = 1,\ldots,n$. Если не обращать внимание на два «выброса», то можно предположить, что распределение случайного вектора (X,Y) *центрально симметрично* относительно $\boldsymbol{\theta} = (\theta_1,\theta_2)$, где θ_1 (θ_2) — центр частного распределения систолического (диастолического) давления. Проверим гипотезу $H_0'\colon \boldsymbol{\theta} = \boldsymbol{\theta}'$ против альтернативы $H_1'\colon \boldsymbol{\theta} \neq \boldsymbol{\theta}'$, где $\boldsymbol{\theta}' = (120,80)$ — стандарт для мужчин в возрасте 21 год в США.

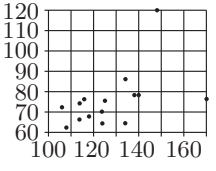


Рис. 2

Введем обозначения: $X_i' = X_i - 120$, $X_i'' = \operatorname{rank}(|X_i'|)\operatorname{sign}(X_i')$, $Y_i' = Y_i - 80$, $Y_i'' = \operatorname{rank}(|Y_i'|)\operatorname{sign}(Y_i')$. В таблице на рис. 3 приведены значения этих переменных.

По формуле (4) вычисляем $T_1 = 2{,}69$ и $T_2 = -4{,}81$. Из (5) имеем $\widehat{v}_{11} = \widehat{v}_{22} = 31/96 \approx 0{,}323$. На основе последней строки таблицы и формулы (6) находим $\widehat{v}_{12} = \widehat{v}_{21} = 0{,}096$. Используя функцию обращения матрицы из программы Excel, получаем

$$(n\widehat{\boldsymbol{V}})^{-1} = \begin{pmatrix} 0{,}226 & -0{,}067 \\ -0{,}067 & 0{,}226 \end{pmatrix}.$$

По формуле (7) подсчитываем значение $T^* = 8{,}605$. Так как оно превосходит 0,95-квантиль закона χ^2_2, равную 5,99 (см. табл. Т3),

X	170	125	148	140	106	108	124	134	116	114	118	138	134	124	114
X'	50	5	28	20	-14	-12	4	14	-4	-6	-2	18	14	4	-6
X''	15	5	14	13	-10	-8	3	10	-3	-6,5	-8	12	10	3	-6,5
Y	76	75	120	78	72	62	70	64	76	74	68	78	86	64	66
Y'	-4	-5	40	-2	-8	-18	-10	-16	-4	-6	-12	-2	6	-16	-14
Y''	-3,5	-5	15	-1,5	-8	-14	-9	-12,5	-3,5	-6,5	-10	-1,5	6,5	-12,5	-11

Рис. 3

то на приближенном*) уровне значимости 5% следует отклонить гипотезу H_0', т. е. давление крови перуанских индейцев *значимо отличается* от давления их американских сверстников.

Точечная оценка $\widehat{\boldsymbol{\theta}}$ для вектора $\boldsymbol{\theta}$ — пара медиан средних Уолша (см. § 3 гл. 8). Вычисления дают $\widehat{\boldsymbol{\theta}} = (126, 73)$.

Замечание 1. *Достоинствами* обоих рассмотренных выше критериев являются их устойчивость к «выбросам» и инвариантность относительно преобразований масштаба координатных осей.

К недостаткам относятся асимптотический характер (для конечной выборки они дают не точный, а приближенный уровень значимости) и отсутствие инвариантности относительно вращения осей.

При дополнительном предположении о *многомерной нормальности* распределения наблюдений

Д4. $\boldsymbol{X}_i \sim \mathcal{N}(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ *с невырожденной матрицей ковариаций* $\boldsymbol{\Sigma}$ (в этом случае распределение имеет плотность, см. приложение П9)

для проверки гипотезы H_0 используется **критерий Хотеллинга**, статистикой которого служит

$$G = [(n-p)/p]\, \overline{\boldsymbol{X}}^{\,T} \widehat{\boldsymbol{\Sigma}}^{-1} \overline{\boldsymbol{X}}, \tag{8}$$

где $\overline{\boldsymbol{X}} = (\overline{X}_1, \ldots, \overline{X}_p)$, $\widehat{\boldsymbol{\Sigma}} = \dfrac{1}{n} \sum\limits_{i=1}^{n} (\boldsymbol{X}_i - \overline{\boldsymbol{X}})(\boldsymbol{X}_i - \overline{\boldsymbol{X}})^T$ — выборочная ковариационная матрица.

С помощью критерия Хотеллинга можно отклонить гипотезу H_0 на уровне α, если наблюдаемое значение g статистики G превосходит $(1-\alpha)$-квантиль распределения Фишера—Снедекора с p и $(n-p)$ степенями свободы (см. табл. Т5).**)

Этот критерий не имеет указанных в замечании 1 недостатков. Более того, его статистика инвариантна относительно любых *невырожденных линейных* преобразований координат.

Действительно, пусть $\boldsymbol{Y}_i = \boldsymbol{C}\boldsymbol{X}_i$. В силу свойств ковариации и матричных операций (см. П2 и П10) запишем

$$\overline{\boldsymbol{Y}}^{\,T} \widehat{\boldsymbol{\Sigma}}_Y^{-1} \overline{\boldsymbol{Y}} = (\boldsymbol{C}\overline{\boldsymbol{X}})^T (\boldsymbol{C}\widehat{\boldsymbol{\Sigma}}_X \boldsymbol{C}^T)^{-1} (\boldsymbol{C}\overline{\boldsymbol{X}}) =$$
$$= \overline{\boldsymbol{X}}^{\,T} \boldsymbol{C}^T \left[(\boldsymbol{C}^T)^{-1} \widehat{\boldsymbol{\Sigma}}_X^{-1} \boldsymbol{C}^{-1} \right] \boldsymbol{C}\overline{\boldsymbol{X}} = \overline{\boldsymbol{X}}^{\,T} \widehat{\boldsymbol{\Sigma}}_X^{-1} \overline{\boldsymbol{X}}.$$

Однако, критерий Хотеллинга *не обладает устойчивостью к «выбросам»* (*робастностью*) ввиду того, что (как показывает следующий пример) неробастны статистики $\overline{\boldsymbol{X}}$ и $\widehat{\boldsymbol{\Sigma}}$.

*) Критерий базируется на *асимптотической* нормальности статистики $\boldsymbol{T}$, а применяется к конечной выборке размера $n = 15$.

**) О критерии Хотеллинга рассказывается в книге Андерсон Т. *Введение в многомерный статистический анализ*, М.: Физматгиз, 1963. Также см. монографию [8, с. 85].

Пример 3. Применим критерий Хотеллинга к данным примера 2. С помощью Excel вычисляем $\overline{\boldsymbol{X}} = (127{,}5;\ 75{,}3)$ и

$$\widehat{\boldsymbol{\Sigma}} = \begin{pmatrix} 269{,}2 & 106{,}5 \\ 106{,}5 & 182{,}7 \end{pmatrix}.$$

Для $\boldsymbol{\theta}' = (120{,}80)$ находим, что при $n = 15$ и $p = 2$ статистика

$$G' = [(n-p)/p]\,(\overline{\boldsymbol{X}} - \boldsymbol{\theta}')^T\,\widehat{\boldsymbol{\Sigma}}^{-1}\,(\overline{\boldsymbol{X}} - \boldsymbol{\theta}')$$

имеет значение 4,12. Согласно табл. Т5 0,95-квантиль распределения Фишера—Снедекора с 2 и 13 степенями свободы равна 3,81. Так как $4{,}12 > 3{,}81$, гипотеза $H_0'\colon \boldsymbol{\theta} = \boldsymbol{\theta}'$ отвергается на уровне 5%.

Исключим теперь два выделяющихся наблюдения — точки с координатами (170,76) и (148,120) (см. рис. 2). Для оставшихся данных в количестве $n_0 = 13$ вектор средних $\overline{\boldsymbol{X}}_0 = (122{,}7;\ 71{,}8)$ (как и следовало ожидать, средние значения уменьшились). В свою очередь выборочная ковариационная матрица оставшихся данных

$$\widehat{\boldsymbol{\Sigma}}_0 = \begin{pmatrix} 116{,}2 & 33{,}2 \\ 33{,}2 & 44{,}6 \end{pmatrix}.$$

Она совсем не похожа на прежнюю $\widehat{\boldsymbol{\Sigma}}$. Новое значение статистики $G_0' = 13{,}0$ превосходит даже 0,995-квантиль F-распределения с 2 и 11 степенями свободы, равную 8,91 (см. [10, с. 202]).

§ 3. ДВУХВЫБОРОЧНАЯ МОДЕЛЬ

Сравним две многомерные независимые выборки размеров n и m. Пусть $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ и $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_m$ — выборки из p-мерных распределений с функцией распределения $F(x_1, \ldots, x_p)$ и $F(x_1 - \Delta_1, \ldots, x_p - \Delta_p)$ соответственно. Выборки представляются в виде двух таблиц размеров $n \times p$ и $m \times p$, которые удобно объединить в *общую матрицу данных* $\boldsymbol{D}$ размера $N \times p$, где $N = n + m$:

$$\boldsymbol{D} = \begin{pmatrix} X_{11} & \ldots & X_{1p} \\ \vdots & & \vdots \\ X_{n1} & \ldots & X_{np} \\ Y_{11} & \ldots & Y_{1p} \\ \vdots & & \vdots \\ Y_{m1} & \ldots & Y_{mp} \end{pmatrix}.$$

Предположим, что верно допущение Д1 из § 2, а также выполнено условие

Д5. *Выборки $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ и $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_m$ независимы между собой (другими словами, строки матрицы $\boldsymbol{D}$ независимы).*

Отметим, что допущения о симметрии закона F *не делается*. Параметр $\boldsymbol{\Delta} = (\Delta_1, \ldots, \Delta_p)$ задает величину сдвига распределения $\boldsymbol{Y}$ относительно распределения $\boldsymbol{X}$. Мы хотим проверить гипотезу

$H_0\colon \boldsymbol{\Delta} = \mathbf{0}$ против альтернативы $H_1\colon \boldsymbol{\Delta} \neq \mathbf{0}$.

Рассматриваемый ниже критерий известен как

Многомерное обобщение критерия ранговых сумм Уилкоксона—Манна—Уитни
(см. § 5 гл. 14)

Пусть $\boldsymbol{W} = (W_1, \ldots, W_p)$, где

$$W_j = \frac{1}{N+1} \sum_{l=1}^{m} R_{lj} \; - \; m/2, \quad j = 1, \ldots, p, \tag{9}$$

где $N = n + m$, R_{lj} — ранг Y_{lj} относительно элементов j-го столбца матрицы $\boldsymbol{D}$. (Для удобства вычислений удобно преобразовать $\boldsymbol{D}$ в матрицу $\boldsymbol{R}$ из *рангов по столбцам*.) Ниже мы покажем, что W_j — это центрированная сумма рангов: $\mathsf{M}W_j = 0$ при H_0.

Следующая теорема дает предельное распределение статистики $N^{-1/2}\,\boldsymbol{W}$ при справедливости нулевой гипотезы.

Теорема 3. Пусть верна гипотеза H_0 и выполнены допущения Д1, Д5. Предположим, что $n, m \to \infty$ и $n/N \to \gamma$, $0 < \gamma < 1$. Тогда

$$N^{-1/2}\,\boldsymbol{W} \xrightarrow{d} \boldsymbol{Z} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{V}),$$

где $\boldsymbol{V} = \|v_{jk}\|_{p\times p}$, $v_{jj} = \gamma(1-\gamma)/12$,

$$v_{jk} = \gamma(1-\gamma) \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_j(x)\, F_k(y)\, dF_{jk}(x,y) \; - \; 1/4 \right] \quad \text{при } k \neq j.$$

Доказательство теоремы 3 можно найти в [86, с. 297].

Покажем, что $\mathsf{M}W_j = 0$ при справеливости гипотезы H_0. В соответствии с обозначениями из § 5 гл. 14 формулу (9), определяющую статистику W_j, можно записать в виде

$$W_j = V_j/(N+1) - m/2, \tag{10}$$

где $V_j = \sum\limits_{l=1}^{m} R_{lj}$. Согласно формулам (5) и (6) гл. 14 имеем

$$\mathsf{M}V_j = \mathsf{M}U_j + \frac{m(m+1)}{2} = \frac{nm}{2} + \frac{m(m+1)}{2} = \frac{m(N+1)}{2}\,. \tag{11}$$

Здесь $U_j = \sum\limits_{i=1}^{n} \sum\limits_{l=1}^{m} I_{\{X_{ij} < Y_{lj}\}}$ (см. формулу (4) гл. 14). Остается взять математическое ожидание обеих частей равенства (10) и подставить туда соотношение (11).

Для построения критерия (см. § 1) нужна состоятельная оценка $\widehat{\boldsymbol{V}}$ для асимптотической матрицы ковариаций $\boldsymbol{V}$. В силу (10) и формул (5)–(6) гл. 14

$$\mathsf{D}W_j = \frac{\mathsf{D}V_j}{(N+1)^2} = \frac{\mathsf{D}U_j}{(N+1)^2} = \frac{nm(n+m+1)}{12(N+1)^2} = \frac{nm}{12(N+1)}\,.$$

На основании этого возьмем

$$\widehat{v}_{jj} = \mathsf{D}\left(N^{-1/2}\,W_j\right) = \frac{nm}{12N(N+1)}\,. \tag{12}$$

Далее, если в формулу, определяющую v_{jk} (см. утверждение теоремы 3), вместо $F_j(x)$, $F_k(y)$ и $F_{jk}(x,y)$ подставить одномерные и двумерную эмпирические функции распределения, то получим оценку

$$\widehat{v}_{jk} = \frac{nm}{N^3}\left[\frac{1}{(N+1)^2}\sum_{t=1}^{N} R_{tj}R_{tk} \;-\; N/4\right]. \tag{13}$$

(Таким образом, потребуется вычислять скалярные произведения столбцов матрицы рангов $\boldsymbol{R}$.) Используя формулы (12) и (13) для определения элементов матрицы $\widehat{\boldsymbol{V}}$, получим, что при справедливости гипотезы H_0 *статистика критерия*

$$W^* = \boldsymbol{W}^T(N\widehat{\boldsymbol{V}})^{-1}\boldsymbol{W} \tag{14}$$

распределена в пределе по закону χ^2_p при условии невырожденности матрицы $\boldsymbol{V}$.

Пример 4. ***Обнаружение фальсификации данных.*** Исследователю-медику требовалось провести статистическую обработку данных из области диагностики пациентов, страдающих мигренями. При этом таблица данных состояла их двух частей: часть I составляли наблюдения, полученные самим исследователем, часть II — заимствованные данные.

У исследователя имелись основания сомневаться в добросовестности сторонней информации. Возникали подозрения, что данные были получены следующим образом: реально были записаны характеристики головной боли лишь небольшой группы больных, затем они была многократно дублированы и немного искажены («зашумлены»), чтобы не было заметно периодичности. (Забегая вперед, скажем, что проводимый ниже статистический анализ подтверждает наличие фальсификации.)

Для обнаружения различия между частями I и II таблицы было проведено исследование, состоящее из **нескольких этапов**.

1. Отбор больных и признаков. Из обеих частей таблицы были отобраны (с помощью фильтра в программе Excel) пациенты с диагнозом *«Эпизодическая головная боль напряжения»*: в группе I оказалось $n = 41$ таких больных, в группе II — $m = 58$ больных.

Среди ряда характеристик головной боли (таких, как *средняя интенсивность боли*, *характер боли*, *ее локализация* и т. д.) для анализа были взяты два признака: X — *средняя длительность одного приступа (часы)* и Y — *среднее частота приступов в месяц (дни)*. Соответствующие данные приведены в таблице на рис. 4.

Группа I		Группа I		Группа II		Группа II		Группа II	
X	Y	X	Y	X	Y	X	Y	X	Y
7	10	3,5	2	3,5	8	3	10	3	8
1,5	4	2,5	8	3,5	12	4	2	2,5	8
2,5	10	2,5	10	3	12	3	0	4	10
7	11	5	10	3	10	3	0	3,5	10
1,5	8	2,5	8	3,5	10	3,5	0	4	8
2,5	5,5	2,5	7	3,5	8	3	4	3,5	10
5	8	3,5	8	3,5	12	4	4	3,5	8
3,5	8	1,5	6	4,5	11	3,5	2	3	8
2	1	3,5	7	5	8	4,5	4	4,5	10
2	10	3,5	10	5	8	4	2	3	10
5	8	1	4	5,5	10	3,5	8	4	10
4	4	2,5	4	6,5	10	3	2	3	10
5	8	3,5	5,5	5	10	4	0	3	8
3,5	10	2,5	3	5	10	4	0	3,5	8
11	8	2,5	8	5	10	2,5	2	4	8
7	4	12	12	3,5	8	3	8	4	8
4,5	10	12	3	4	10	4	8	4	8
7	4	9	4,5	3	8	4	2		
3	10	5	10	4	8	2,5	2		
4	6	2,5	6	3,5	10	5	7		

Рис. 4

2. Визуальный анализ. На основе указанных данных была построена диаграмма рассеяния на плоскости (X,Y) (рис. 5). Парам (X_i,Y_i), относящимся к группе I, соответствуют *черные точки*, а парам (X_i,Y_i), относящимся к группе II — *серые квадраты*.

Рассматривая диаграмму, приходим к **ряду выводов**.

а) Центр «облака» серых квадратов смещен вверх по оси Y относительно центра «облака» черных точек.

б) Очевидно, что у серых квадратов разброс по обеим координатам намного меньше, чем у черных точек.

в) «Облако» серых квадратов имеет *округлую форму*, в то время как «облако» черных точек совсем не похоже на реализацию выборки из двумерного нормального закона. Обычно такая реализация (после исключения нескольких «выбросов») имеет вид эллипса, густо заполненного точками в центральной части и реже — по краям. В нашем же случае скорее наблюдается уменьшение плотности точек (расширение «облака») при увеличении значений обоих признаков, т. е. «облако» черных точек имеет *форму «веера»* (см. рис. 1, *б*).

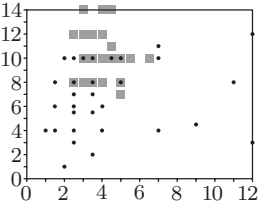


Рис. 5

г) Количество совпадений в группе II намного больше, чем в группе I: для $m = 58$ представителей группы II на диаграмме присутствуют только 21 серый квадрат ($21/58 \approx 0{,}36$), а для $n = 41$ представителя группы I — 31 черный кружок ($31/41 \approx 0{,}76$).

3. Статистическая проверка однородности. Вектор выборочных средних $(\overline{X}, \overline{Y})$ в группе I имеет значение $(4{,}2; 7{,}1)$, в группе II — $(3{,}7; 9{,}8)$. Средние *заметно отличаются*, особенно по второй координате. Стандартные отклонения в группе I равны $(2{,}7; 2{,}7)$, в группе II — $(0{,}8; 1{,}7)$. Видим, что разброс в группе II намного меньше, чем в группе I.

Ввиду пункта в) приведенных выше выводов для исследуемых данных неуместно применять критерии, опирающиеся на предположение о нормальности распределения наблюдений (скажем, критерий Хотеллинга). Более надежным подходом является использование многомерного обобщения рангового критерия Уилкоксона—Манна—Уитни. Вычисления дают $N = 99$, $\boldsymbol{W} = (1{,}305; 6{,}475)$, $v_{11} = v_{22} = 0{,}02$, $v_{12} = v_{21} = 0{,}0025$,

$$(N\widehat{\boldsymbol{V}})^{-1} = \begin{pmatrix} 0{,}513 & -0{,}064 \\ -0{,}064 & 0{,}513 \end{pmatrix}.$$

Наконец, $W^* = 21{,}28$, что превосходит 0,999-квантиль закона χ^2_2, равную 13,8. Таким образом, группы I и II нельзя считать однородными на уровне значимости 0,001.

Замечание 2. Если применить одномерный критерий Уилкоксона—Манна—Уитни к каждому из признаков отдельно, то будем иметь следующие результаты: для признака X нормированная статистика критерия (см. формулу (7) гл. 14) принимает значение 0,93, а для признака Y — значение 4,6.

Первое из них (согласно таблице Т2) незначимо на уровне 5%. Другими словами, критерий Уилкоксона—Манна—Уитни не улавливает различия между группами I и II по первому признаку. Это можно объяснить тем, что при переходе от самих наблюдений к их рангам нивелируется очевидное визуальное отличие на рис. 5 черных точек группы I с большими значениями признака X от серых квадратов группы II.

В свою очередь, для признака Y картина иная. Статистика значима велика даже на уровне 0,001, что подтверждает наличие систематического сдвига вверх по координате Y серых квадратов группы II относительно черных точек группы I. По степени разброса группы будут сравниваться в следующей главе.

Глава 24

ДВУХВЫБОРОЧНАЯ ЗАДАЧА О МАСШТАБЕ

В этой главе рассматриваются ранговые (и поэтому — устойчивые к «выбросам») альтернативы F-критерию, не предполагающие нормальности распределения наблюдений.[**)]

Данные представляют собой две независимые случайные выборки, по одной из двух генеральных совокупностей. На основе выборок мы хотим выяснить, есть ли различие в мерах рассеяния (масштаба) этих совокупностей.

Многие статистики развивали своего рода «логику», следы которой можно найти во многих прикладных книгах по статистике и суть которой можно свести к следующему: прежде чем применять двухвыборочный t-критерий, нужно сначала применить критерий нормальности к каждой выборке (например, критерий согласия хи-квадрат). Если результат не даст большой значимости, то можно считать, что распределения нормальны, и применять F-критерий равенства дисперсий. Если же последний также не выйдет за уровень значимости, то можно предполагать равенство дисперсий и позволить себе применить t-критерий. В этом рецепте нет ни слова о возможном отсутствии независимости, (...) ни слова о катастрофической необастности F-критерия для дисперсий, ни слова (или едва-едва что-нибудь) о том, что делать, если один из предварительных критериев превзойдет уровень значимости.[*)]

Из книги [84, с. 81]

§ 1. МЕДИАНЫ ИЗВЕСТНЫ ИЛИ РАВНЫ

Статистическая модель

Имеется $N = n + m$ наблюдений $X_1, \ldots, X_n$, $Y_1, \ldots, Y_m$:

$$X_i = \mu + \sigma_1 \varepsilon_i,\ i = 1, \ldots, n; \quad Y_j = \mu + \sigma_2 \varepsilon'_j,\ j = 1, \ldots, m;$$

где μ — неизвестная общая медиана распределения наблюдений, $\sigma_1 > 0$ и $\sigma_2 > 0$ — неизвестные параметры масштаба, ε_i и ε'_j — случайные ошибки.

Допущения

Д1. *Все ошибки* $\{\varepsilon_i, \varepsilon'_j\}$ *независимы.*

Д2. *Все* $\{\varepsilon_i, \varepsilon'_j\}$ *имеют одинаковое непрерывное (неизвестное) распределение, медиана* (см. § 2 гл. 7) *которого равна* 0.

Для проверки гипотезы H_0: $\sigma_1 = \sigma_2$ применяется

Свободный от распределения ранговый критерий Ансари—Брэдли

1. Упорядочить N наблюдений от меньшего к большему.
2. Наименьшему и наибольшему из наблюдений присвоить ранг 1, второму и предпоследнему по величине присвоить

[*)] Критерий Стьюдента (t-критерий) и F-критерий для дисперсий приведены в примере 1 гл. 14. Критерий хи-квадрат для проверки нормальности рассматривался в примере 3 гл. 18. Проблема зависимости наблюдений обсуждалась в § 4 гл. 15, необастность F-критерия — в замечании 2 гл. 16 ($k = 2$).

[**)] Изложение, в основном, следует главе 5 монографии [88].

ранг 2 и т. д. Если N четно, то расположение рангов будет $1,2,\ldots,N/2,N/2,\ldots,2,1$; если же N нечетно, то расположение рангов будет $1,2,\ldots,(N-1)/2,(N+1)/2,(N-1)/2,\ldots,2,1$.

3. Обозначив через S_j ранг Y_j в указанной ранжировке, вычислить значение *статистики критерия Ансари—Брэдли*

$$Z=\sum_{j=1}^{m} S_j,$$

т. е. сумму рангов, относящихся к $Y_1,\ldots,Y_m$.

Поясним, почему слишком большие (или слишком малые) значения Z противоречат H_0. Пусть, например, $\sigma_1<\sigma_2$. Тогда значения Y будут проявлять склонность к бóльшему рассеянию (разбросу), чем значения X. Поэтому при ранжировке (по схеме 2-го шага метода) наблюдения Y_j, в большинстве своем, будут иметь малые ранги, а X_i — большие. В результате величина Z окажется малой. (Примером крайнего случая служит такое распределение представителей X и Y в упорядоченной объединенной выборке: $YYXXXYY$.)

Малые выборки. Критические значения статистики Z для $n,m\leqslant 10$ приведены в табл. $A.6$ книги [88].

Большие выборки. Если гипотеза H_0 верна, то статистика

$$Z^*=(Z-\mathbf{M}Z)\Big/\sqrt{\mathbf{D}Z} \tag{1}$$

асимптотически (при $\min\{n,m\}\to\infty$) распределена по стандартному нормальному закону $\mathcal{N}(0,1)$. Математическое ожидания и дисперсия в (1) задаются формулами

$$\mathbf{M}Z=\frac{m(N+2)}{4},\qquad \mathbf{D}Z=\frac{nm(N+2)(N-2)}{48\,(N-1)},\quad \text{если N — четное;}$$

$$\mathbf{M}Z=\frac{m(N+1)^2}{4N},\quad \mathbf{D}Z=\frac{nm(N+1)(N^2+3)}{48\,N^2},\quad \text{если N — нечетное.}$$

Обозначим через $x_{1-\alpha}$ квантиль уровня $(1-\alpha)$ закона $\mathcal{N}(0,1)$ (§ 3 гл. 7), а через z^* — наблюдаемое значение статистики Z^*.

Односторонний критерий приближенного уровня α для проверки гипотезы H_0 против альтернативы $H_1\colon \sigma_1<\sigma_2$ таков: если $z^*\leqslant -x_{1-\alpha}$, то отвергаем гипотезу H_0, иначе — принимаем.

Односторонний критерий приближенного уровня α для проверки гипотезы H_0 против альтернативы $H_2\colon \sigma_1>\sigma_2$ таков: если $z^*\geqslant x_{1-\alpha}$, то отвергаем гипотезу H_0, иначе — принимаем.

Совпадения. Если среди N наблюдений есть одинаковые, то для подсчета Z надо использовать средние ранги. Для больших выборок следует также заменить $\mathbf{D}Z$ в формуле (1) на величину

$$nm\left[16\sum_{k=1}^{g} l_k\overline{R}_k^2-N(N+2)^2\right]\Big/\left[16\,N(N-1)\right],\qquad \text{если } N \text{ — четное;}$$

$$nm\left[16N\sum_{k=1}^{g} l_k\overline{R}_k^2 - (N+1)^4\right] \Big/ \left[16\,N^2(N-1)\right],\ \text{если } N\text{ — нечетное.}$$

Здесь g — число групп совпадений среди N наблюдений, l_k — количество элементов в k-й группе, $\overline{R}_k$ — средний ранг наблюдений в k-й группе. Не совпадающие с другими наблюдения рассматриваются как группы размера 1.

Пример 1. Продолжим анализ данных из примера 4 гл. 23. Напомним, что критерий ранговых сумм Уилкоксона—Манна—Уитни не позволил обнаружить сдвига по компоненте X наблюдений в группе I относительно наблюдений в группе II. Поэтому можно предположить, что *выборки имеют одинаковую медиану*. Исходя из этого допущения, попробуем обнаружить различие в рассеянии выборок с помощью критерия Ансари—Брэдли.

Для вычислений удобно использовать такие операции программы Excel, как «Сортировка» и «Фильтр». В результате их применения (для $n = 41$ и $m = 58$) получим следующие значения: $z = 1784{,}2$ $\mathbf{M}Z = 1464{,}6$ и $\mathbf{D}Z = 4955{,}7$ (с учетом совпадений — 4623,7), $z^* = 4{,}7$. Эта величина значимо велика даже на уровне $\alpha = 0{,}001$ (см. табл. Т2). Таким образом, приходим к заключению, что выборки нельзя считать однородными, поскольку у наблюдений в группе I рассеяние по компоненте X *значимо больше*.

Комментарии

1. При справедливости гипотезы H_0 вероятность каждого из C_N^m возможных размещений Y-ов по N местам одинакова. Благодаря этому можно найти распределение Z при нулевой гипотезе.

Например, для $n = 2$ и $m = 2$ имеется $C_4^2 = 6$ размещений: $XXYY$, $XYXY$, $XYYX$, $YXXY$, $YXYX$, $YYXX$. Набор соответствующих значений Z таков: $(3,3,4,2,3,3)$. Отсюда, скажем, находим, что $\mathbf{P}(Z \geqslant 4) = 1/6$ при H_0.

2. Во многих двухвыборочных ситуациях интересно выявить различие совокупностей в сдвиге или разбросе. Одно из решений — применение критериев, предназначенных для более общих альтернатив (например, критерия Смирнова или критерия Розенблатта из гл. 14). Другой путь — совместное применение двух критериев: одного для сдвига (например, критерия Уилкоксона—Манна—Уитни из § 5 гл. 14), другого — для масштаба (скажем, критерия Ансари—Брэдли). Р. Рандлес и Р. Хогг в 1971 г. показали, что при выполнении H_0 статистики последних двух критериев *асимптотически независимы*. Отсюда следует, что если мы применяем критерий Уилкоксона—Манна—Уитни на уровне α_1 и критерий Ансари—Брэдли на уровне α_2, то вероятность отклонить гипотезу об идентичности совокупностей X и Y *хотя бы одним* из критериев, будет приближенно равна $\alpha_1 + \alpha_2 - \alpha_1\alpha_2$.

3. Критерий Ансари—Брэдли можно применять исключительно в ситуации, когда совокупности X и Y различаются *лишь* в мерах рассеяния, в частности, *не имеют сдвига* друг относительно друга.

Пусть, например, $n = 3$, $m = 4$ и все наблюдения из совокупности X оказались меньше любого наблюдения из совокупности Y. Тогда Z имеет значение 10 вне зависимости от степени разброса каждой из выборок.

В то же время требование равенства медиан не обязательно для классического F-критерия, основанного на отношении выборочных дисперсий. В следующем параграфе рассматривается (похожий на

ранговый) *критерий Мозеса*, для которого равенство медиан также не требуется.

§ 2. МЕДИАНЫ НЕИЗВЕСТНЫ И НЕРАВНЫ

Статистическая модель

Имеется две группы наблюдений: $X_1,\dots,X_n$ и $Y_1,\dots,Y_m$:

$X_i = \mu_1 + \sigma_1\varepsilon_i,\ i=1,\dots,n;\quad Y_j = \mu_2 + \sigma_2\varepsilon'_j,\ j=1,\dots,m;$

где μ_1 и μ_2 — неизвестные медианы совокупностей (мешающие параметры), σ_1 и σ_2 — неизвестные (интересующие нас) параметры масштаба, ε_i и ε'_j — случайные ошибки.

Допущения те же, что и в § 1.

Для проверки гипотезы $H_0\colon \sigma_1 = \sigma_2$ используется

Свободный от распределения похожий на ранговый критерий Мозеса.

1. Задать целое $k \geqslant 2$ и *случайным образом* разбить каждую из групп наблюдений на $n' = [n/k]$ и $m' = [m/k]$ подгрупп размера k соответственно.*) Все не вошедшие в подгруппы наблюдения отбросить.
2. Обозначить через $X_{1r},\dots,X_{kr}$ k наблюдений, попавших в r-ю подгруппу для X, $r=1,\dots,n'$, а через $Y_{1s},\dots,Y_{ks}$ — k наблюдений, попавших в s-ю подгруппу для Y, $s=1,\dots,m'$.
3. Подсчитать $C_1,\dots,C_{n'}$ по формулам
$$C_r = \sum_{t=1}^{k} X_{tr}^2 - \frac{1}{k}\left(\sum_{t=1}^{k} X_{tr}\right)^2,\quad r=1,\dots,n'.$$
4. Подсчитать $D_1,\dots,D_{m'}$ по формулам
$$D_s = \sum_{t=1}^{k} Y_{ts}^2 - \frac{1}{k}\left(\sum_{t=1}^{k} Y_{ts}\right)^2,\quad s=1,\dots,m'.$$
5. Применить к $C_1,\dots,C_{n'}$ и $D_1,\dots,D_{m'}$ критерий ранговых сумм Уилкоксона—Манна—Уитни (см. § 5 гл. 14).

Если H_0 не выполняется, например, $\sigma_1 < \sigma_2$, то величины D_s имеют тенденцию превосходить величины C_r. Это приводит к увеличению рангов D_s и, как следствие, к большим значениям статистики критерия ранговых сумм.

Оценка отношения коэффициентов масштаба. В случае невыполнения H_0 представляет интерес величина $\gamma = \sigma_2/\sigma_1$. Для ее оценивания Г. Шорак в 1969 г. (см. [88, с. 116]) предложил следующую процедуру.

*) Здесь $[\cdot]$ обозначает целую часть числа.

Группа I				
3,5	2,5	7	7	1,5
3,5	2,5	2	1,5	5
5	4	2,5	3,5	5
2,5	7	2,5	5	11
2,5	4,5	2,5	1,5	3,5
3,5	7	3,5	1	3,5
2,5	3	2,5	12	2,5
9	4	5	2,5	12
C_1	C_2	C_3	C_4	C_5
33,5	23,0	20,7	97,5	106,0

Группа II						
5	3,5	4,5	3,5	3	3	3,5
4	3,5	4	5	3,5	5	3,5
4	4	3	4,5	5,5	3	3,5
4	4	3	4	5	4	6,5
3	3,5	3	3	3	5	3
4,5	4	2,5	3	3,5	4	3,5
3	4	4	2,5	3	3,5	2,5
3	4	3	3,5	4	3,5	4
D_1	D_2	D_3	D_4	D_5	D_6	D_7
4,0	0,5	3,4	4,9	6,5	4,4	10,0

Рис. 1

1. Вычислить $n'm'$ отношений D_s/C_r, $r = 1, \ldots, n'$, $s = 1, \ldots, m'$.
2. Обозначить через $Q_{(1)} \leqslant \ldots \leqslant Q_{(n'm')}$ упорядоченные значения отношений D_s/C_r.
3. Взять в качестве оценки

$$\widehat{\gamma} = \begin{cases} Q_{(l)}^{1/2} & \text{при } n'm' = 2l+1, \\ (Q_{(l)}Q_{(l+1)})^{1/4} & \text{при } n'm' = 2l. \end{cases}$$

Пример 2. Применим критерий Мозеса к данным из примера 4 гл. 23. Сравним рассеяния измерений в группах I и II по компоненте X *без предположения о равенстве медиан совокупностей* (использовавшегося в примере 1 ранее).

Если задать $k = 8$, то в первой выборке размера $n = 41$ потребуется отбросить всего одно наблюдение ($n' = 5$), а во второй выборке размера $m = 58$ — два наблюдения ($m' = 7$). Для выбора номеров отбрасываемых наблюдений используем таблицу равномерных случайных чисел Т1. Будем считывать ее по строкам, начиная с первой. Получим числа 10, 09, 73, 25,.... В соответствии с ними отбросим в первой выборке элемент с номером 10, а во второй — с номерами 9 и 25 (так как $73 > m = 58$, то это число игнорируем).

Далее, разобьем оставшиеся наблюдения группы I *случайным образом* на подгруппы размера 8. Опишем, как это удобно сделать. Продолжим считывание табл. Т1 со второй строки: 37, 54, ... Обращая внимание только на первую цифру, отнесем первый элемент выборки группы I к подгруппе с номером 3, второй — к подгруппе с номером 5 и т. д. (ноль и цифры, большие $n' = 5$ игнорируются). При этом полезно отмечать в соседнем столбце текущий размер пополняемой подгруппы и вносить в момент заполнения подгруппы (когда размер станет равным $k = 8$) ее номер в *список номеров уже заполненных подгрупп* (скажем, записываемый на листе бумаги). Когда подгруппа заполнена, ее номер, встретившийся в таблице, также игнорируется.

Аналогично разобьем на $m' = 7$ подгрупп наблюдения группы II. Результаты разбиения приведены в двух таблицах на рис. 1.

Также на рис. 1 указаны соответствующие C_r и D_s, $r = 1,\ldots,5$, $s = 1,\ldots,7$. Видим, что все C_r больше любого из D_s. Согласно таблице критических значений статистики критерия Уилкоксона—Манна—Уитни из [10, с. 358] вероятность этого при справедливости гипотезы H_0 для выборок размеров $r = 5$ и $s = 7$ менее 0,001. Таким образом, приходим к тому же выводу, что и в примере 1: рассеяние в группе I по компоненте X *значимо больше*, чем в группе II.

Путем написания простой программы для процедуры Шорака на Visual Basic в Excel было подсчитано значение оценки для «отношения разбросов» $1/\gamma = \sigma_1/\sigma_2$, равное 3,112.

Комментарии

1. Г. Шорак рекомендует выбирать размер подгрупп k как можно больше (но не более 10), однако не делая n' и m' настолько малыми, что распределение соответствующей статистики критерия ранговых сумм Уилкоксона—Манна—Уитни не сможет обеспечить разумный уровень значимости. Размер подгрупп k надо выбирать, учитывая только размеры выборок n и m. М. Холлендер и Д. Вулф (см. [88, с. 114]) пишут: «Было бы неверным так подбирать различные значения k, чтобы получились значимые решения. Такой подход делает незаконными все выводы!»

2. У критерия Мозеса есть **недостаток**: два человека, обрабатывая одни и те же данные, могут получить различные выводы при одном и том же уровне значимости α. Эта возможность возникает из-за случайности процесса разбиения на подгруппы. Иными словами, повторное применение критерия Мозеса к тем же данным, но с другим разбиением на подгруппы, может привести в противоположным выводам. В связи с этим подчеркнем, что при использовании критерия допускается лишь одна чисто случайная группировка. Совершенно недопустимо манипулировать группировками для достижения значимого результата!

3. М. Холлендер в 1968 г. показал, что при выполнении гипотезы H_0 и *симметричности распределения* статистика критерия Мозеса и статистика критерия ранговых сумм Уилкоксона—Манна—Уитни некоррелированы и асимптотически независимы. Это влечет те же заключения, что и в комментарии 2 из § 1.

4. Известны и другие критерии для проверки гипотезы H_0, применимые *для выборок большого размера*. Так, в [88, с. 117] приводится критерий Миллера, основанный на методе «складного ножа» (jackknife). Он обладает более высокой эффективностью сравнительно к критерием Мозеса, но свободен от распределения лишь асимптотически.

Глава 25

КЛАССЫ ОЦЕНОК

§ 1. L-ОЦЕНКИ

Пусть расцветают сто цветов.

Китайская мудрость

Рассмотрим вариационный ряд $X_{(1)} \leqslant X_{(2)} \leqslant \ldots \leqslant X_{(n)}$ (см. § 4 гл. 4). *L-оценками* называются линейные комбинации $\sum w_i X_{(i)}$ порядковых статистик, где w_i — некоторые веса. Эти оценки обобщают усеченное среднее (см. § 2 гл. 8)

$$\overline{X}_\alpha = \frac{1}{n-2k}\left(X_{(k+1)} + \ldots + X_{(n-k)}\right), \text{ где } 0 < \alpha < \frac{1}{2},\ k = [\alpha n]^{*)},$$

у которого k крайним с каждой стороны наблюдениям приданы нулевые веса, а оставшимся $(n-2k)$ центральным наблюдениям — одинаковые $w_i = 1/(n-2k)$. Чтобы изучить поведение L-оценок при больших n, удобно задавать веса w_{in} $(i = 1, \ldots, n;\ n = 1,2,\ldots)$ в виде

$$w_{in} = \frac{1}{n}\,\lambda\left(\frac{i}{n+1}\right), \tag{1}$$

где $\lambda(t)$ — некоторая функция, определенная на отрезке $[0,1]$. Например, усеченному среднему $\overline{X}_\alpha$ отвечает

$$\lambda(t) = \frac{1}{1-2\alpha}\; I_{\{\alpha<t<1-\alpha\}}. \tag{2}$$

Естественно рассмотреть более гладкое распределение весов. Приведем условия (из [50, с. 328]), обеспечивающие асимптотическую нормальность L-оценок.

Теорема 1. Пусть случайные величины X_i независимы и одинаково распределены на интервале (a,b), $-\infty \leqslant a < b \leqslant \infty$, согласно распределению F, у которого

1) существует положительная при $a < x < b$ плотность $p(x)$,
2) $\mathbf{M}X_1^2 < \infty$.

Потребуем также, чтобы выполнялось условие

3) $\lambda(t)$ — ограниченная функция, непрерывная почти всюду (относительно лебеговой меры) с $\int\limits_0^1 \lambda(t)\,dt = 1$.

*) Здесь $[\cdot]$ обозначает целую часть числа.

Положим

$$\mu(F,\lambda)=\int\limits_0^1 \lambda(t)\,F^{-1}(t)\,dt,\ \sigma^2(F,\lambda)=\int\limits_0^1 G^2(t)\,dt-\left(\int\limits_0^1 G(t)\,dt\right)^2,$$

где $G(t)$ — любая функция с $G'(t)=\lambda(t)\,/p\,(F^{-1}(t))$.

Тогда при условии $\sigma(F,\lambda)>0$ для статистики

$$L_n=\frac{1}{n}\sum_{i=1}^{n}\lambda\left(\frac{i}{n+1}\right)X_{(i)} \tag{3}$$

справедлива сходимость

$$\sqrt{n}\,(L_n-\mu(F,\lambda))\xrightarrow{d}\xi\sim\mathcal{N}(0,\sigma^2(F,\lambda))\ \text{ при } n\to\infty.$$

Следствие. Пусть элементы выборки X_i распределены согласно закону $F(x-\theta)$, где F симметрична относительно 0, а функция λ — относительно $1/2$. Тогда легко проверить, что $\mu(F,\lambda)=\theta$, и поэтому $\sqrt{n}\,(L_n-\theta)\xrightarrow{d}\xi\sim\mathcal{N}(0,\sigma^2(F,\lambda))$.

Для заданного распределения F наиболее эффективная L-оценка параметра сдвига θ получается путем минимизации дисперсии $\sigma^2(F,\lambda)$ по λ.

Теорема 2. Допустим, что плотность $p(x)=F'(x)$ имеет две производные $p'(x)$, $p''(x)$ почти всюду и что $p(x)\to 0$ при $x\to\pm\infty$. Положим

$$\gamma(x)=-p'(x)/p(x). \tag{4}$$

Тогда асимптотическая дисперсия $\sigma^2(F,\lambda)$ минимальна при $\lambda=\lambda^*$, которая *пропорциональна* функции $\gamma'\,(F^{-1}(t))$ (см. условие 3 теоремы 1). В этом случае

$$\sigma^2(F,\lambda^*)=1/I(F),\quad \text{где } I(F)=\int\limits_{-\infty}^{\infty}\left[p'(x)\right]^2/p(x)\,dx \tag{5}$$

обозначает информацию Фишера, введенную ранее в § 3 гл. 9.

Другими словами, такая L-оценка является асимптотически эффективной (как и оценка максимального правдоподобия).

Доказательство этой теоремы приведено в [50, с. 331].

Пример 1. Нетрудно проверить, что для *стандартного нормального закона* ($F=\Phi$) функция $\gamma(x)=x$. Отсюда оптимальная $\lambda^*\equiv 1$. Соответствующая L_n совпадает с ОМП $\overline{X}$. Информация Фишера $I(\Phi)=1$.

В случае *логистического распределения* с $F(x)=1/(1+e^{-x})$ находим, что $\gamma(x)=2F(x)-1=\operatorname{th}(x/2)$. Оптимальная $\lambda^*(t)=6\,t\,(1-t)$, причем $I(F)=1/3$.

Для *закона Коши* с плотностью $p(x) = 1/[\pi(1+x^2)]$ получается весьма своеобразная оптимальная функция

$$\lambda^*(t) = 2\cos(2\pi t)[\cos(2\pi t) - 1],$$

которая отрицательна при $\left|t - \frac{1}{2}\right| > 1/4$ (рис. 1).*) Здесь $I(F) = 1/2$.

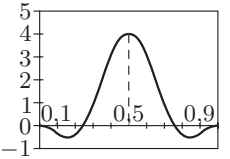


Рис. 1

В случае *«наименее благоприятного распределения» Хьюбера*, определяемого в примере 2 ниже, оптимальной L-оценкой является α-усеченное среднее $\overline{X}_\alpha$ ($\alpha = F(-b)$) с весовой функцией, задаваемой формулой (2). Таблица значений $I(F)$ для разных b содержится в [89, с. 92].

Замечание 1. Помимо эффективности, важную роль играет устойчивость оценки по отношению к выделяющимся наблюдениям («выбросам»). Если значения весовой функции $\lambda(t)$ вблизи концов отрезка $[0,1]$ малы по сравнению со значениями в середине, то L-оценка относительно нечувствительна к «выбросам». Тем не менее, для того, чтобы оценка имела положительную толерантность (см. § 4 гл. 8) необходимо, чтобы носителем функции $\lambda(t)$ был интервал $(\alpha, 1-\alpha)$ при некотором $0 < \alpha < 1/2$.

§ 2. M-ОЦЕНКИ

Усеченное среднее $\overline{X}_\alpha$ служит компромиссом между средним $\overline{X}$ ($\alpha = 0$) и выборочной медианой MED ($\alpha \to 1/2$). Другой компромисс подсказан тем фактом, что среднее и медиана суть величины, минимизирующие, соответственно, $\sum(X_i - \theta)^2$ и $\sum|X_i - \theta|$ (см. задачу 1 гл. 1 и задачу 4 гл. 7). П. Хьюбер предложил минимизировать *меру «разброса»*

$$\sum_{i=1}^{n} \rho_b(X_i - \theta), \tag{6}$$

где функция $\rho_b(x)$ при некотором $b > 0$ задается формулой

$$\rho_b(x) = \begin{cases} x^2/2, & \text{если } |x| \leqslant b, \\ b|x| - b^2/2, & \text{если } |x| \geqslant b. \end{cases} \tag{7}$$

Она пропорциональна x^2 при $|x| \leqslant b$, а вне отрезка $[-b, b]$ заменяет ветви параболы на две прямые линии. Эти куски графика согласованы так, что $\rho_b(x)$ и ее производная $\psi_b(x) = \rho_b'(x)$ непрерывны (рис. 2). При увеличении b функция ρ_b будет совпадать с $x^2/2$ на все большей части ее определения, так что оценка Хьюбера будет сближаться с $\overline{X}$. В свою очередь, при уменьшении b происходит сближение с выборочной медианой. (В качестве компромисса иногда предлагается значение $b = 1{,}5$.)

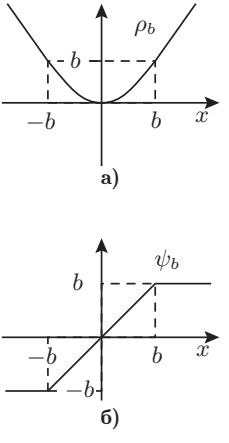


Рис. 2

*) Для неотрицательности $\lambda^*(t)$ необходимо и достаточно, чтобы $\gamma'(x) \geqslant 0$ при всех x. Это эквивалентно выпуклости функции $-\log p(x)$. Такие распределения называются *сильно унимодальными*.

Подобно тому, как усеченные средние $\overline{X}_\alpha$ $(0 < \alpha < 1/2)$ образуют однопараметрическое подмножество класса L-оценок, оценки Хьюбера $(0 < b < \infty)$ образуют однопараметрическое подмножество класса так называемых M-оценок.

M-оценка M_n параметра сдвига θ определяется для некоторой функции $\rho(x)$ как точка минимума по θ функции

$$\sum_{i=1}^{n} \rho(X_i - \theta). \tag{8}$$

Если ρ — выпуклая и четная, то можно показать, что значения θ, минимизирующие сумму (8), образуют некоторый отрезок, а если ρ строго выпукла (П4), то минимизирующее значение единственно (см. [50, с. 57]).

Если функция ρ имеет производную $\psi = \rho'$, то оценка M_n является решением уравнения

$$\sum_{i=1}^{n} \psi(X_i - \theta) = 0. \tag{9}$$

Пусть элементы выборки X_i имеют функцию распределения $F(x - \theta)$, где F обладает *четной* плотностью $p(x)$. Тогда при слабых предположениях относительно ψ и F

$$\sqrt{n}\,(M_n - \theta) \xrightarrow{d} \zeta \sim \mathcal{N}(0, \sigma^2(F,\psi)) \quad \text{при } n \to \infty, \tag{10}$$

Доказательство (10) можно найти в монографии [89, с. 58], где приведено детальное изложение теории M-оценок не только для модели сдвига, но и для более общих параметрических моделей.

где *асимптотическая дисперсия* $\sigma^2(F,\psi)$ задается формулой

$$\sigma^2(F,\psi) = \int \psi^2(x)\,p(x)\,dx \Big/ \left[\int \psi'(x)\,p(x)\,dx\right]^2. \tag{11}$$

Если $\rho(x) = -\ln p(x)$, то $\psi = \rho' = \gamma(x)$ (см. формулу (4)). В этом случае минимизация суммы (8) эквивалентна максимизации произведения $\prod p(X_i - \theta)$, т. е. M-оценка параметра сдвига совпадает с оценкой максимального правдоподобия (ОМП) (см. § 4 гл. 9). Таким образом, M-оценки *обобщают ОМП* (ей они обязаны своим названием). Известно, что при выполнении условий регулярности (см. § 3 гл. 9) ОМП является асимптотически эффективной оценкой с $\sigma^2(F,\gamma) = 1/I(F)$, где $I(F)$ определена в (5).

Пример 2. Для *закона* $\mathcal{N}(0,1)$ функция $\gamma(x) = x$, поэтому оптимальная оценка M_n совпадает с $\overline{X}$. Нулевая толерантность этой оценки (см. § 4 гл. 8) связана с неограниченностью функции $\gamma(x)$.

Для *логистического распределения* с $F(x) = 1/(1 + e^{-x})$ имеем $\gamma(x) = \operatorname{th}(x/2)$. В силу строгой монотонности и ограниченности $\gamma(x)$ (см. [89, с. 61]) толерантность решения уравнения (9) равна $1/2$.

В случае *закона Лапласа* с плотностью $p(x) = \dfrac{1}{2}\, e^{-|x|}$ весовая функция $\rho(x) = |x|$ не имеет производной в нуле. Полагая $\gamma(0) = 0$,

получаем $\gamma(x) = \operatorname{sign} x$. Соответствующая ОМП — выборочная медиана *MED*, имеющая $\sigma^2(F,\gamma) = 1$ и обладающая асимптотической толерантностью $1/2$.

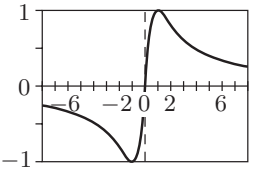


Рис. 3

Для *распределения Коши* с плотностью $p(x) = 1/[\pi(1 + x^2)]$ функция $\gamma(x) = 2x/(1 + x^2)$ оказывается немонотонной (рис. 3). Поэтому уравнение (9) может иметь несколько решений (см. вопрос 3 в гл. 9). Коллинз (1976) и Кларк (1984) (см. [84, с. 137]) показали, что если а) брать ближайшее к выборочной медиане решение или б) пользоваться методом Ньютона, выбирая в качестве начального приближения *MED* (см. теорему 4 гл. 9), то оценка будет асимптотически эффективной, причем от медианы будет унаследована толерантность $1/2$.

Наконец, рассмотрим случай *«наименее благоприятного распределения» Хьюбера* с плотностью $p(x)$, пропорциональной функции $\exp\{-\rho_b(x)\}$, где $\rho_b(x)$ задается формулой (7). Для этого закона $\gamma(x)$ совпадает с $\psi_b(x) = \rho_b'(x)$. Толерантность оценки Хьюбера также равна $1/2$.

Пример 3. ***Минимаксный подход Хьюбера.*** Вместо того, чтобы предполагать функцию распределения F известной, более реалистично допустить, что она известна лишь приближенно. Одной из таких моделей, предложенных Хьюбером, является представление F в виде смеси

$$F(x) = (1 - \varepsilon)\, G(x) + \varepsilon\, H(x), \tag{12}$$

где функция распределения $G(x)$ и $0 \leqslant \varepsilon \leqslant 1$ заданы, а $H(x)$ — это *произвольная* неизвестная функция распределения (предполагается только, что G и H симметричны относительно нуля (см. гл. 8)).

Обозначим через $\Omega(G,\varepsilon)$ *семейство распределений, удовлетворяющих* (12) *для фиксированных* G *и* ε. (Легко видеть, что любое распределение вида $(1 - \varepsilon')\, G(x) + \varepsilon' H(x)$, где $0 \leqslant \varepsilon' \leqslant \varepsilon$, также входит в $\Omega(G,\varepsilon)$.) В соответствии с **минимаксным подходом** представляет интерес нахождение M-оценки (функции ψ), на которой достигается

$$\min_{\psi} \sup_{F \in \Omega(G,\varepsilon)} \sigma^2(F,\psi). \tag{13}$$

В частном случае, когда G совпадает с функцией распределения стандартного нормального закона Φ с плотностью $\varphi(x) = (2\pi)^{-1/2} e^{-x^2/2}$, решением этой задачи (как доказано в [50, с. 335]) служит оценка Хьюбера с $\psi_b(x)$, у которой $b = b(\varepsilon)$ определяется из соотношения

$$1/(1 - \varepsilon) = 2\,\Phi(b) - 1 + 2\,\varphi(b)/b. \tag{14}$$

Некоторые значения $b(\varepsilon)$ приведены в таблице (взятой из [89, с. 92]):

ε	0	0,001	0,01	0,05	0,1	0,2	0,5	1
$b(\varepsilon)$	∞	2,630	1,945	1,399	1,140	0,862	0,436	0

В частности, видим, что 5%-му «загрязнению» отвечает $b \approx 1{,}4$.

По ходу доказательства в [50] устанавливается, что при $\psi = \psi_b(x)$ наибольшее значение $\sigma^2(F,\psi_b)$ достигается на распределении F с плотностью $p(x)$, пропорциональной $\exp\{-\rho_b(x)\}$, ввиду чего его и называют *«наименее благоприятным распределением» Хьюбера*.

Замечание 2 (см. [84, с. 149]). Дж. Тьюки в 1970 г. предложил для параметра сдвига θ так называемую *W-оценку* $\tilde\theta$, представляющую собой решение уравнения

$$\tilde\theta = \sum_{i=1}^{n} w(X_i - \tilde\theta)\, X_i \Big/ \sum_{i=1}^{n} w(X_i - \tilde\theta)\ , \tag{15}$$

где $w(u)$ — некоторая *весовая функция*. Правая часть (15) является *взвешенным средним* наблюдений X_i с весами

$$w_i = w(X_i - \tilde\theta), \quad i = 1,\ldots,n.$$

Для подсчета $\tilde\theta$ обычно прибегают к использованию итерационного алгоритма. Начиная с $\tilde\theta_0 = MED$, вычисляются величины

$$\tilde\theta_{j+1} = \sum_{i=1}^{n} X_i\, w(X_i - \tilde\theta_j) \Big/ \sum_{i=1}^{n} w(X_i - \tilde\theta_j)\ , \quad j = 0{,}1,\ldots, \tag{16}$$

до тех пор, пока не будет достигнута стабилизация $\tilde\theta_j$.[*)]

Отметим, что W-оценки являются на самом деле M-оценками специального вида, так как в силу (15) $\sum\limits_{i=1}^{n}(X_i - \tilde\theta)\, w(X_i - \tilde\theta) = 0$, и значит, $\sum\limits_{i=1}^{n} \psi_w(X_i - \tilde\theta) = 0$ с $\psi_w(u) = uw(u)$.

Из-за простоты вычисления популярны *одношаговые W-оценки* $\tilde\theta_1$ в (16), имеющие очень хорошие характеристики: они наследуют толерантность 1/2 выборочной медианы и обладают асимптотической дисперсией M-оценки с $\psi = \psi_w$, задаваемой формулой (11).

Замечание 3. Если для описания данных используется не модель сдвига, а модель сдвига-масштаба, то прежде чем оценивать сдвиг, надо оценить параметр масштаба с помощью какой-нибудь робастной оценки, например, *(нормированной) медианы абсолютных отклонений*

$$MAD = \frac{1}{\Phi^{-1}(3/4)}\ MED\{|X_i - MED|, i = 1,\ldots,n\},$$

[*)] Процедура *итерационного перевзвешивания* часто применяется в регрессионном анализе (см., например, алгоритм «LOWESS» в § 2 гл. 22).

где $\Phi^{-1}(x)$ — функция, обратная к функции распределения $\mathcal{N}(0,1)$, $1/\Phi^{-1}(3/4) \approx 1{,}483$ (она ранее встречалась в § 1 гл. 19). Толерантность MAD равна $1/2$.

Затем для некоторой ψ оценка параметра сдвига $\widehat{\theta}$ находится как решение уравнения

$$\sum_{i=1}^{n} \psi\left(\frac{X_i - \widehat{\theta}}{MAD}\right) = 0.$$

С другой стороны, можно также совместно оценить параметры сдвига и масштаба при помощи обобщения M-оценок на случай векторного параметра (см. [84, с. 269], [89, с. 141]).

§ 3. R-ОЦЕНКИ

Так называемые R-оценки ведут свое происхождение от ранговых критериев для проверки гипотез о значении параметра сдвига θ (см. гл. 14, 15).

Пусть $X_{(1)} < \ldots < X_{(n)}$ обозначают упорядоченные наблюдения, $d_1, \ldots, d_n$ — набор неотрицательных чисел. Рассмотрим $n(n+1)/2$ полусумм вида $(X_{(j)} + X_{(k)})/2$ при $1 \leqslant j \leqslant k \leqslant n$ (при $j = k$ это просто сами порядковые статистики $X_{(j)}$). Каждой такой полусумме припишем вес

$$w_{jk} = d_{n-(k-j)} \Big/ \sum_{i=1}^{n} i d_i. \tag{17}$$

Легко проверить (убедитесь!), что в сумме эти веса дают 1. Рассмотрим дискретное распределение, присваивающее вероятности w_{jk} значениям $(X_{(j)} + X_{(k)})/2$. По определению, *R-оценка* R_n есть медиана этого распределения.

> Отметим аналогию с L-оценками $L_n = \sum w_i X_{(i)}$, у которых веса $w_i \geqslant 0$, $\sum w_i = 1$. В этом случае L_n есть математическое ожидание распределения, приписывающего вероятности w_i значениям $X_{(i)}$.

Пример 4. Пусть $d_1 = \ldots = d_{n-1} = 0$, $d_n = 1$. Тогда $\sum i d_i = n$, $w_{jk} = 1/n$ при $j = k$ и равны нулю в противном случае. Дискретное распределение, по которому строится R-оценка, приписывает вероятность $1/n$ каждому из $X_{(1)}, \ldots, X_{(n)}$, т. е. R_n — это просто *выборочная медиана* наблюдений.

Возьмем, наоборот, $d_1 = 1$, $d_2 = \ldots = d_n = 0$. Тогда веса w_{jk} положительны, лишь когда $k - j = n - 1$, т. е. когда $j = 1$ и $k = n$. Таким образом, полусумме $(X_{(1)} + X_{(n)})/2$ присваивается вероятность 1, и R_n представляет собой *середину размаха выборки*.

Когда $d_1 = \ldots = d_n = 1$, все величины $(X_{(j)} + X_{(k)})/2$ имеют одинаковый вес. Медиана R_n этих полусумм появилась в § 3 гл. 8 под названием *медианы средних Уолша* W.

Напротив, *оценка Гальтона*, определенная в задаче 3 гл. 8 как

$$\widehat{\theta} = MED\left\{ \left(X_{(i)} + X_{(n-i+1)}\right)/2,\ i = 1,\ldots,[(n+1)/2]\right\}$$

(здесь $[\cdot]$ обозначает *целую часть* числа), *не входит в число R-оценок*. Дело в том, что полусуммам вида $(X_{(i)}+X_{(n-i+1)})/2$ приписаны равные веса, а всем остальным — нулевые, но согласно (17) полусуммы, находящиеся на одной диагонали $k-j=\text{const}$ (рис. 4), должны иметь одинаковый вес.

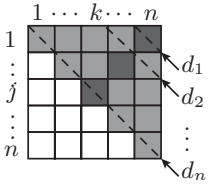


Рис. 4

Для того чтобы получить асимптотическое распределение для R_n, зададим $d_i = d_{in}$ с помощью функции $K(t)$, определенной на (0,1), неубывающей и удовлетворяющей соотношению $K(1-t) = -K(t)$ (т. е. график $K(t)$ центрально симметричен относительно точки плоскости с координатами $(1/2, 0)$), полагая

$$d_{in} = K\left(\frac{i+1}{2n+1}\right) - K\left(\frac{i}{2n+1}\right).$$

Для модели сдвига $X_i \sim F(x-\theta)$, где распределение F предполагается симметричным с (четной) плотностью $p(x)$, при некоторых условиях регулярности на K и F имеет место сходимость

$$\sqrt{n}\,(R_n - \theta) \xrightarrow{d} \xi \sim \mathcal{N}(0, \sigma^2(F,K)),$$

$$\text{где } \sigma^2(F,K) = \int\limits_0^1 K^2(t)\,dt \Big/ \left[\int\limits_{-\infty}^{\infty} K'[F(x)]\,p^2(x)\,dx\right]^2.$$

Как и в случае L-оценок и M-оценок, для заданной функции распределения F существует асимптотически эффективная оценка R-оценка. Она соответствует функции

$$K(t) = \gamma\left(F^{-1}(t)\right), \quad \text{где } \gamma \text{ задается формулой (4).}$$

Пример 5. Для *логистического распределения* в примере 1 было отмечено, что $\gamma(x) = 2F(x) - 1$, так что оптимальная функция $K(t) = 2t - 1$, и, значит, $d_{in} = 2/(2n+1)$ для всех $i = 1,\ldots,n$. Так как R-оценка в силу соотношения (17) не меняется, когда все d_{in} умножаются на одну и ту же положительную константу, то получающаяся в результате асимптотически эффективная R-оценка есть медиана средних Уолша W. Согласно задаче 4 гл. 8 толерантность $\tau_W = 1 - \sqrt{2}/2 \approx 0{,}293$.

В случае *закона Коши* оптимальная $K(t) = -\sin(2\pi t)$.

Для *стандартного нормального закона* $\gamma(x) = x$ и, следовательно, $K(t) = \Phi^{-1}(t)$, где Φ^{-1} — обратная к функции распределения закона $\mathcal{N}(0,1)$. Соответствующая ей R-оценка называется *оценкой с нормальными весами (или метками)* и обозначается через N.

Оценка N из примера 5 интересна тем, что на классе Ω_s гладких симметричных распределений (см. определение в § 1 гл. 8) она *доминирует по эффективности над оценкой* $\overline{X}$ (см. [86, с. 121]).

Теорема 3. Если X_i имеют функцию распределения $F(x-\theta)$, то

$e_{N,\overline{X}}(F) \geqslant 1$ для всех $F \in \Omega_s$,

где неравенство строгое, за исключением случая $F = \Phi$.

Тем самым N улучшает нижнюю границу медианы средних Уолша W: $e_{W,\overline{X}}(F) \geqslant 108/125 \approx 0{,}864$ для всех $F \in \Omega_s$, приведенную в теореме 4 гл. 8.

В [84, с. 146] указана толерантность $\tau_N = 2\,\Phi(-\sqrt{\ln 4}) \approx 0{,}239$. Однако П. Хьюбер в [89, с. 77] высказывает мнение: «Не следует рекомендовать N для практического использования — показатели ее количественной робастности[*)] возрастают очень быстро при отклонении от нормальной модели, и оценка N очень быстро становится хуже, чем, например, оценка W».[**)]

В заключение отметим любопытную **взаимосвязь** между M-, L- и R-оценками. Для заданной $F \in \Omega_s$ и нечетной ψ обозначим через λ_ψ функцию, пропорциональную $\psi'(F^{-1}(t))$, и положим $K_\psi = \psi(F^{-1}(t))$. Тогда

$$\sigma^2_M(F,\psi) = \sigma^2_L(F,\lambda_\psi) = \sigma^2_R(F,K_\psi).$$

Другими словами, для всякой M-оценки найдутся *асимптотически эквивалентные* L-оценка и R-оценка.

В частности, при $\psi = \gamma$ согласно примерам 1, 2 и 5 имеем

а) для *нормального закона* с $F = \Phi$
   M: $\psi(x) = x$, выборочное среднее $\overline{X}$,
   L: $\lambda_\psi(t) \equiv 1$, выборочное среднее $\overline{X}$,
   R: $K_\psi(t) = \Phi^{-1}(t)$, оценка с нормальными весами N;

б) для *логистического распределения* с $F = 1/(1+e^{-x})$
   M: $\psi(x) = \operatorname{th}(x/2)$,
   L: $\lambda_\psi(t) = 6\,t\,(1-t)$,
   R: $K_\psi(t) = 2t-1$, медиана средних Уолша W.

К рассмотренной взаимосвязи имеет отношение следующий результат о применении к классу L-оценок минимаксного подхода Хьюбера (см. пример 3).

Сначала покажем, что M-оценке Хьюбера с $\psi = \psi_b$ (независимо от вида F) соответствует в классе L-оценок усеченное среднее $\overline{X}_\alpha$, где $\alpha = F(-b)$. Действительно, $\psi_b'(x) = I_{\{|x|<b\}}$. Тогда функция λ_{ψ_b} пропорциональна

$$\psi_b'\left(F^{-1}(t)\right) = 1, \quad \text{если } |F^{-1}(t)| < b,$$

[*)] Приведенные в [89, с. 21].

[**)] Быстрый рост показателей связан с неограниченностью $K(t)$.

и равна нулю в противном случае. Так как неравенство $|F^{-1}(t)| < b$ эквивалентно неравенству $F(-b) < t < F(b)$, то $\lambda_{\psi_b}(t)$ постоянна на интервале $(\alpha, 1-\alpha)$ и равна 0 вне этого интервала.

Отсюда ввиду (13) возникает вопрос: не является ли L-оценка, минимизирующая для $G = \Phi$

$$\sup_{F\in\Omega(G,\varepsilon)} \sigma_L^2(F,\lambda),$$

усеченным средним $\overline{X}_\alpha$ с $\alpha = \Phi(-b)$, соответствующим оценке Хьюбера с константой b, определяемой по ε из соотношения (14)? Л. Джекел в 1971 г. доказал, что это действительно так (см. [50, с. 339]).

§ 4. ФУНКЦИЯ ВЛИЯНИЯ

Обозначим через $\Omega^{\pm} = \{F\}$ множество функций, представимых в виде $F = H_1 - H_2$, где H_1 и H_2 — некоторые ограниченные неубывающие функции. Это множество включает, в частности, произвольные функции распределения ($H_2 \equiv 0$). Мы будем рассматривать функционалы T, заданные на $\Omega^{\pm}$, т. е. отображения из $\Omega^{\pm}$ в $\mathbb{R}$.

Нас будут интересовать оценки вида $T(\widehat{F}_n)$, где $\widehat{F}_n(x)$ — эмпирическая функция распределения (см. § 1 гл. 9). Например, для функции распределения F естественными оценками *математического ожидания* $T_1(F) = \int x\, dF(x)$ и *медианы* $T_2(F) = F^{-1}(1/2)$ являются *выборочное среднее* $T_1(\widehat{F}_n) = \int x\, d\widehat{F}_n(x) = n^{-1}\sum X_i = \overline{X}$ и *выборочная медиана* $T_2(\widehat{F}_n) = \widehat{F}_n^{-1}(1/2) = MED$ (см. § 2 гл. 7).

Чтобы понять насколько такие оценки устойчивы к «выбросам», Ф. Хампель (см. [84]) предложил изучить влияние на функционал «атома» единичной массы, помещенного в точку с «большой» координатой y (функция распределения $\delta_y(x) = I_{\{x\geqslant y\}}$). Степень реакции на подобное «возмущение» выражается характеристикой

$$A(y) = \lim_{t\to 0} \frac{T((1-t)\,F + t\,\delta_y) - T(F)}{t}\,, \tag{18}$$

называемой *функцией влияния*. Если функция $A(y)$ неограничена при $y \to \pm\infty$, то оценка быстро смещается в сторону «выброса», и, следовательно, неробастна (см. § 4 гл. 8).

Вычислим $A(y)$ для определенного выше функционала T_1. Пусть $F_t = (1-t)\,F + t\,\delta_y$. Тогда

$$T_1(F_t) = \int x\, dF_t(x) = \int x\, dF(x) + t\left[\int x\, d\delta_y(x) - \int x\, dF(x)\right].$$

Отсюда $A_1(y) = \dfrac{d}{dt}\, T_1(F_t) = \int x\, d\delta_y(x) - \int x\, dF(x) = y - T_1(F)$. Таким образом, математическое ожидание T_1 имеет неограниченную линейную функцию влияния.

Дифференцируя тождество $1/2 = F_t(F_t^{-1}(1/2))$, нетрудно установить (проверьте!), что функцией влияния функционала T_2 служит $A_2(y) = [1/2 - \delta_y(x_{1/2})]/p(x_{1/2})$, где $p(x)$ — плотность распределения F, $x_{1/2} = F^{-1}(1/2)$ — медиана F. В частности, при $x_{1/2} = 0$ имеем $A_2(y) = [2\,p(0)]^{-1} \operatorname{sign} y$. Следовательно, $MED = T_2(\widehat{F}_n)$ робастна при условии $p(x_{1/2}) > 0$.

Покажем, как функция влияния связана с асимптотической дисперсией оценки $T(\widehat{F}_n)$. При некоторых условиях на функционал T для произвольной функции $G \in \Omega^{\pm}$ выполняется следующее равенство, обобщающее (18):

$$\lim_{t\to 0} \frac{T((1-t)\,F + t\,G) - T(F)}{t} = \int A(x)\,dG(x). \tag{19}$$

Из него при $G = F$ вытекает, что

$$\int A(x)\,dF(x) = 0. \tag{20}$$

Поэтому в формулу (19) вместо dG можно подставить $d(G - F)$.

Далее, если функция $\widetilde{F}$ «близка» к F, то имеет смысл воспользоваться разложением функционала T в «точке» F в ряд Тейлора до члена первого порядка:

$$T(\widetilde{F}) = T(F) + \int A(x)\,d(\widetilde{F} - F)(x) + \dots . \tag{21}$$

В силу *теоремы Гливенко* (см. [19, с. 201]) при $n \to \infty$ эмпирическая функция $\widehat{F}_n$ стремится к F с вероятностью 1. Взяв ее в качестве $\widetilde{F}$ в формуле (21) и учитывая равенство (20), получим

$$\sqrt{n}\,(T(\widehat{F}_n) - T(F)) \approx \sqrt{n} \int A(x)\,d\widehat{F}_n(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} A(X_i). \tag{22}$$

Применяя к правой части (22) центральную предельную теорему (П6) (с точностью до обоснования законности отбрасывания остаточного члена из (21)), выводим сходимость

$$\sqrt{n}\,(T(\widehat{F}_n) - T(F)) \xrightarrow{d} \xi \sim \mathcal{N}\left(0,\, \int A^2(x)\,dF(x)\right).$$

Отсюда, в частности, немедленно находим асимптотическую дисперсию выборочной медианы MED (см. теорему 1 гл. 7).

В заключение приведем результат Ф. Хампеля (1968 г.), связывающий *функцию влияния и M-оценки* (см. [84, с. 151, 173]): для нормального закона ($F = \Phi$) решением задачи минимизации задаваемой формулой (11) асимптотической дисперсии $\sigma^2(F,\psi)$ M-оценки при условии ограниченности ее функции влияния: $\sup|A(y)| < c$, служит оценка Хьюбера с $\psi_b(x)$, где b определяется по c из соотношения $c = b/(2\,\Phi(b) - 1)$.

Глава 26

БРОУНОВСКИЙ МОСТ

Физическое явление, описанное Робертом Броуном, состояло в сложном и беспорядочном движении взвешенных в жидкости частиц цветочной пыльцы. За многие годы, прошедшие со времени этого описания*), броуновское движение превратилось в объект изучения как чистой, так и прикладной математики. Даже теперь продолжают обнаруживать многие его важные свойства, и, без сомнения, остаются еще не открытыми новые и полезные аспекты этого процесса.

Т. Хида, [87, с. 8]

§ 1. БРОУНОВСКОЕ ДВИЖЕНИЕ

В § 6 гл. 14 была введена модель симметричного случайного блуждания. Допустим, что для заданного значения n выполнено следующее преобразование графика траектории блуждания (рис. 1): масштаб по оси абсцисс сжат в n раз, а по оси ординат — в $\sqrt{n}$ раз. При этом кусок графика на $[0, n]$ преобразуется в (случайную) ломаную, заданную на $[0, 1]$. Обозначим ее через $W_n(t)$. Известно (см. [12, с. 350]), что в пределе при $n \to \infty$ на $[0, 1]$ возникает случайный процесс $W(t)$, называемый *стандартным броуновским движением или винеровским процессом***). Сходимость $W_n(t)$ к $W(t)$ можно понимать как *сходимость конечномерных распределений* процессов: для любых $0 \leqslant t_1 < t_2 < \ldots < t_k \leqslant 1$ последовательность $(W_n(t_1), \ldots, W_n(t_k)) \xrightarrow{d} (W(t_1), \ldots, W(t_k))$ (см. П5).

Процесс $W(t)$ характеризуется **следующими свойствами**:

1) $W(0) = 0$,
2) приращения $W(t_2) - W(t_1), \ldots, W(t_k) - W(t_{k-1})$ независимы,
3) $W(t) - W(s) \sim \mathcal{N}(0, t - s)$ при $0 \leqslant s < t$,
4) траектории $W(t)$ непрерывны с вероятностью 1.

Однако траектории процесса $W(t)$ *нигде не дифференцируемы* с вероятностью 1 (доказательство приведено, скажем, в [87, с. 65]). Этот факт соответствует физической природе процесса, а математически объясняется разной скоростью сжатия в горизонтальном и

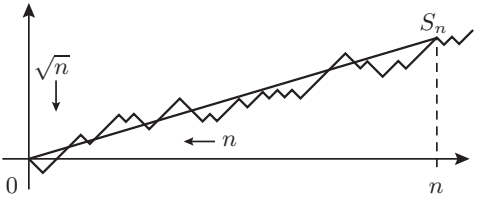


Рис. 1

*) Наблюдения проводились в 1827 г.

**) Обозначение $W(t)$ введено в честь Н. Винера, внесшего значительный вклад в построение теории броуновского движения.

вертикальном направлениях ломаной случайного блуждания, вследствие чего каждый локальный максимум (минимум) траектории становится с ростом n все более «острым».

Из свойств 1 и 3 вытекает, что $\mathbf{M}W(t) = 0$ при всех t. Вычислим *ковариационную функцию* $R(s,t) = \mathbf{M}W(s)W(t)$. При $s < t$, используя свойства 2 и 3, находим, что

$$R(s,t) = \mathbf{M}[W(s) - W(0)][W(t) - W(s)] + \mathbf{M}W^2(s) = \\ = \mathbf{M}[W(s) - 0] \cdot \mathbf{M}[W(t) - W(s)] + s = s.$$

В общем случае, очевидно, $R(s,t) = \min\{s,t\}$.

§ 2. ЭМПИРИЧЕСКИЙ ПРОЦЕСС

Построим на основе $W(t)$ новый случайный процесс.

Определение. *Броуновским мостом* называется случайный процесс $B(t) = W(t) - tW(1)$ при $0 \leqslant t \leqslant 1$ (рис. 2).

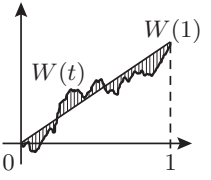


Рис. 2

Название объясняется тем, что оба конца каждой траектории $B(t)$ закреплены: $B(0) = B(1) = 0$. Нетрудно проверить, что

$$\mathbf{cov}\,(B(s), B(t)) = \min\{s,t\} - st. \tag{1}$$

Броуновский мост $B(t)$ возникает в качестве предельного процесса при описанном выше преобразовании случайного блуждания, из которого вычитается прямая, соединяющая начало координат с точкой (n,S_n) (см. рис. 1).

Другой путь, приводящий в $B(t)$, состоит в изучении условного распределения траекторий блуждания, возвращающихся в нуль в момент n (см. [39, с. 211]). На языке процесса $W(t)$ это соответствует переходу к пределу при $\varepsilon \to 0$ на множестве броуновских траекторий, у которых $|W(1)| < \varepsilon$.

Еще один подход, приводящий к появлению броуновского моста, который мы обсудим подробно, основан на сходимости к $B(t)$ так называемого *эмпирического процесса*

$$B_n(t) = \sqrt{n}\,(\widehat{F}_n(t) - t), \quad t \in [0,1],$$

где $\widehat{F}_n(t)$ — эмпирическая функция распределения (см. § 1 гл. 9), построенная по выборке $\eta_1, \ldots, \eta_n$ из равномерного распределения на отрезке $[0,1]$ (рис. 3). Можно доказать (см. [39, с. 235]), что конечномерные распределения $B_n(t)$ сходятся при $n \to \infty$ к распределениям $B(t)$. Однако нам потребуется более сильное утверждение о сходимости непрерывных функционалов*) от $B_n(t)$.

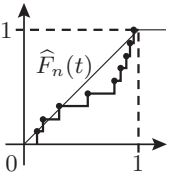


Рис. 3

*) Функционал — это отображение из пространства функций в $\mathbb{R}$.

Рассмотрим для примера статистику $U_n = \sup\limits_{0\leqslant t\leqslant 1} B_n(t)$. Верно ли, что $U_n \xrightarrow{d} U = \sup\limits_{0\leqslant t\leqslant 1} B(t)$ при $n\to\infty$? Нетривиальность ответа здесь связана с невозможностью представления U как функции от значений $B_n(t)$ в конечном числе точек.

К тому же траектории предельного процесса $B(t)$ непрерывны с вероятностью 1, в то время как траектории $B_n(t)$ имеют разрывы в точках скачков функции $\widehat{F}_n(t)$.

Для преодоления указанных трудностей рассмотрим на $[0,1]$ два пространства функций: $C[0,1]$ — *множество всех непрерывных функций*, $D[0,1]$ — *множество функций, непрерывных справа (в точке* 1 *слева) и имеющих лишь конечное число скачков* (см. [11, с. 40]). Так как $C[0,1]\subset D[0,1]$, то $D[0,1]$ (вместе с σ-алгеброй $\mathcal{A}_D$ (П1), порожденной *цилиндрическими множествами*[*)]) можно считать выборочным пространством процессов $B_n(t)$ и $B(t)$. Положим $\|y\| = \sup\limits_{0\leqslant t\leqslant 1} |y(t)|$.

Теорема 1 (*функциональная предельная теорема для эмпирического процесса*). Пусть G — функционал, определенный на $D[0,1]$ и непрерывный в равномерной метрике в каждой точке $y\in D[0,1]$: для любых $y_n, y\in D[0,1]$ условие $\|y_n - y\|\to 0$ при $n\to\infty$ влечет $G(y_n)\to G(y)$. Тогда $G(B_n)\xrightarrow{d} G(B)$.

Эта теорема вытекает из результата, доказанного в [11, с. 423].

Очевидно, рассмотренный выше для примера функционал U удовлетворяет условиям теоремы 1. При этом распределение случайной величины U можно найти в явном виде (см. задачу 6 гл. 14):

$$\mathbf{P}(U\leqslant x) = 1 - e^{-2x^2} \quad \text{при } x\geqslant 0.$$

Далее в § 3 обсуждается ряд следствий более сильного утверждения, справедливого для дифференцируемых функционалов от эмпирической функции распределения.

§ 3. ДИФФЕРЕНЦИРУЕМЫЕ ФУНКЦИОНАЛЫ

Интересно, что такие, на первый взгляд, не связанные между собой **статистические результаты**, как

а) предельные распределения статистики Колмогорова и статистики омега-квадрат (§ 2 гл. 12),
б) асимптотическая нормальность выборочных квантилей (теорема 2 гл. 7),
в) критерий хи-квадрат (§ 1 гл. 18),

[*)] То есть множествами функций $y(t)$, удовлетворяющих условиям вида $\{y(t_1)\in A_1,\ldots,y(t_k)\in A_k\}$, где $A_1,\ldots,A_k$ — произвольные борелевские множества (см. П2) на действительной прямой.

на самом деле вытекают из приведенной ниже теоремы 2. Для ее формулировки потребуется следующее

Определение. Функционал G, заданный на $D[0,1]$, называется *непрерывно дифференцируемым порядка k в точке $y_0 \in D[0,1]$*, если существует такой функционал $g(y_0,v)$, что для любой функции $v \in C[0,1]$ и произвольной последовательности $v_h \in D[0,1]$, $\|v_h - v\| \to 0$ при действительном $h \to 0$, выполняются соотношения

$$\frac{G(y_0 + hv_h) - G(y_0)}{h^k} \to g(y_0,v), \quad g(y_0,v_h) \to g(y_0,v).$$

Второе из соотношений означает непрерывность в равномерной метрике в точках из $C[0,1]$ функционала $g(y_0,v)$, который можно называть *производной порядка k от G в y_0 по направлению v*.

Пример 1. Рассмотрим функционал $G(y) = \|y - y_0\|$. Он непрерывно дифференцируем по любому направлению, так как $G(y_0) = 0$,

$$g(y_0,v) = \lim_{h\to 0} \frac{G(y_0 + hv_h)}{h} = \lim_{h\to 0} \|v_h\| = \|v\|.$$

Пример 2. Для любого натурального k функционал

$$G(F) = \int_0^1 |F - F_0|^k \, dF,$$

где F и F_0 — функции распределения случайных величин, принимающих значения только из $[0,1]$, является непрерывно дифференцируемым k-го порядка по любому направлению, поскольку

$$g(F_0,v) = \lim_{h\to 0} \frac{G(F_0 + hv_h)}{h^k} = \lim_{h\to 0} \int_0^1 |v_h|^k \, dF = \int_0^1 |v|^k \, dF.$$

Пример 3. Покажем, что для произвольного $p \in (0,1)$ функционал $G(F) = F^{-1}(p) = \inf\{t\colon F(t) \geqslant p\}$ непрерывно дифференцируем в точке $F_0(t) = t$, где $t \in [0,1]$, причем $g(F_0,v) = -v(p)$.

Доказательство. Введем обозначение

$$t^* = G(F_0 + hv_h) = \inf\{t\colon t + hv_h(t) \geqslant p\}. \tag{2}$$

Так как функционал G является непрерывным в равномерной метрике в точке F_0, мы можем положить $t^* = p+\delta$, где $\delta = \delta(h) \to 0$ при $h \to 0$. Из $\|v_h - v\| \to 0$ и $v \in C[0,1]$ вытекает, что при $h \to 0$

$$R(h) = |v_h(p+\delta) - v(p)| \leqslant |v_h(p+\delta) - v(p+\delta)| + |v(p+\delta) - v(p)| \to 0. \tag{3}$$

В силу (2) и (3) имеем неравенство

$$p \leqslant F_0(t^*) + hv_h(t^*) = p + \delta + hv_h(p+\delta) = p + \delta + h[v(p) + \theta R(h)],$$

где $|\theta| \leqslant 1$. Аналогично можно написать обратное неравенство:

$$p > F_0(t^*-) + hv_h(t^*-) = p + \delta + hv_h(p+\delta-) = p + \delta + h[v(p) + \theta_1 R_1(h)],$$

где $|\theta_1| \leqslant 1$, $R_1(h) = |v_h(p+\delta-) - v(p)| \to 0$ при $h \to 0$. Поэтому

$$[G(F_0 + hv_h) - G(F_0)]/h = (p + \delta - p)/h = \delta/h \to -v(p). \qquad \blacksquare$$

Пример 4. Рассмотрим разбиение отрезка $[0,1]$ точками $0 = t_0 < t_1 < \ldots < t_N = 1$. Тогда для $F_0(t) = t$ функционал

$$G(F) = \sum_{j=1}^{N} \frac{(\Delta_j F - \Delta_j F_0)^2}{\Delta_j F_0},$$

где $\Delta_j F = F(t_j) - F(t_{j-1})$, $\Delta_j F_0 = t_j - t_{j-1}$, будет непрерывно дифференцируем 2-го порядка, поскольку для него

$$g(F_0,v) = \frac{G(F_0 + hv)}{h^2} = \sum_{j=1}^{N} \frac{(\Delta_j v)^2}{\Delta_j F_0}.$$

Сформулируем предельную теорему для статистик, которые можно представить в виде дифференцируемого функционала от эмпирической функции распределения $\widehat{F}_n(t)$, построенной по выборке $\eta_1, \ldots, \eta_n$ из равномерного распределения на $[0,1]$.*)

Теорема 2. Если $F_0(t) = t$, $t \in [0,1]$, и функционал $G(F)$ непрерывно дифференцируем порядка k в F_0, то при $n \to \infty$

$$n^{k/2}\,[G(\widehat{F}_n) - G(F_0)] \xrightarrow{d} g(F_0,B),$$

где B есть броуновский мост.

Доказательство см. в [11, с. 50].

Следствие 1. Рассмотрим статистику критерия Колмогорова $D_n = \sup\limits_{0 \leqslant t \leqslant 1} |\widehat{F}_n(t) - t|$. Ввиду примера 1 из теоремы 2, примененной к функционалу $D_n(\widehat{F}_n)$, вытекает, что $\sqrt{n}\,D_n \xrightarrow{d} \|B\|$. А. Н. Колмогоровым для функции распределения случайной величины $\|B\|$ получено представление в виде ряда:

$$K(x) = \mathbf{P}(\|B\| \leqslant x) = 1 + 2\sum_{k=1}^{\infty} (-1)^k e^{-2k^2x^2} \quad \text{при } x > 0.$$

Распределение статистики $\sup\limits_{-\infty \leqslant x \leqslant \infty} |\widehat{F}_n(x) - F(x)|$ согласно замечанию 1 гл. 12 оказывается одним и тем же для выборки $X_1, \ldots, X_n$ из закона с произвольной непрерывной функции распределения $F(x)$.

Замечание. Дополним объяснение из решения задачи 2 гл. 12. Пусть $0 < \mu = \mathbf{M}X_i$, $\sigma^2 = \mathbf{D}X_i$, $S_i = X_1 + \ldots + X_i$, $S_i^0 = S_i - i\mu$. Рассмотрим

$$\xi_n = \sqrt{n} \max_{1 \leqslant i \leqslant n} \left| \frac{S_i}{S_n} - \frac{i}{n} \right| = \sqrt{n} \max_{1 \leqslant i \leqslant n} \left| \frac{i\mu + S_i^0}{n\mu + S_n^0} - \frac{i\mu}{n\mu} \right| =$$

$$= \frac{1}{\sqrt{n}} \max_{1 \leqslant i \leqslant n} \left| \frac{nS_i^0 - iS_n^0}{n\mu + S_n^0} \right| = \sigma \left| \frac{n}{n\mu + S_n^0} \right| \cdot \max_{1 \leqslant i \leqslant n} \left| \frac{S_i^0}{\sigma\sqrt{n}} - \frac{i}{n} \frac{S_n^0}{\sigma\sqrt{n}} \right|.$$

*) Существуют статистики, которые не представляются в таком виде, скажем, $\eta_{(n)} = \max\{\eta_1, \ldots, \eta_n\}$ или $T = \sum\limits_{i=1}^{n-1} \eta_i \eta_{i+1}$.

Здесь первый сомножитель сходится по вероятности к σ/μ в силу закона больших чисел, а второй — по распределению к $\sup\limits_{0\leqslant t\leqslant 1}|W^0(t)| \sim K(x)$. В частности, для равномерного распределения на отрезке $[0,1]$ имеем $\mu = 1/2$, $\sigma = 1/\sqrt{12}$, $\sigma/\mu = 1/\sqrt{3} \approx 0{,}6$.

Следствие 2. Распределение статистики Крамéра—Мизеса $\omega_n^2 = \int\limits_{-\infty}^{\infty} [\widehat{F}_n(x) - F(x)]^2\, dF(x)$ также не зависит от функции распределения F в случае ее непрерывности. Для $F_0 = t$, $t \in [0,1]$, функционал $\omega_n^2(\widehat{F}_n) = \int\limits_0^1 [\widehat{F}_n(t) - t]^2\, dt$ согласно примеру 2 непрерывно дифференцируем 2-го порядка. Поэтому $n\omega_n^2 \xrightarrow{d} \int\limits_0^1 B^2(t)\, dt$. Для предельного распределения так же, как и для $K(x)$, известно разложение в ряд и составлены таблицы (см. [10, с. 83, 348]).

Следствие 3. Пусть $\eta_{([pn]+1)}$ обозначает выборочную p-квантиль $(0 < p < 1)$ (см. § 3 гл. 7) для выборки $\eta_1, \ldots, \eta_n$ из равномерно распределенных на отрезке $[0,1]$. Тогда $\eta_{([pn]+1)} = \widehat{F}_n^{-1}(p)$. Ввиду примера 3 из теоремы 2 следует, что $\sqrt{n}\,(\eta_{([pn]+1)} - p) \xrightarrow{d} -B(p)$ при $n \to \infty$.

Случайная величина $B(p) = W(p) - p\,W(1)$, будучи линейной комбинацией компонент нормального вектора $(W(p), W(1))$, сама распределена нормально (согласно П9). Далее, $\mathbf{M}B(p) = \mathbf{M}W(p) - p\,\mathbf{M}W(1) = 0$. Из равенства (1) вытекает, что дисперсия $\mathbf{D}B(p) = \mathbf{cov}\,(B(p), B(p)) = p(1-p)$, т. е. $B(p) \sim \mathcal{N}(0, p(1-p))$. Очевидно, что случайной величины $-B(p)$ распределена так же. Тем самым установлена асимптотическая нормальность выборочных квантилей для выборки из равномерного распределения на отрезке $[0,1]$.

Для того, чтобы обобщить этот результат на случай распределения с заданной плотностью $p(x)$ (т. е. доказать теорему 2 гл. 7), остается лишь применить лемму 1 гл. 7.

Следствие 4. Для $F_0 = t$ положим $p_j = \Delta_j F_0 = t_j - t_{j-1}$, $j = 1, \ldots, N$ (см. пример 4). Тогда статистика хи-квадрат

$$X_n^2 = G(\widehat{F}_n) = \sum_{j=1}^{N} (\Delta_j \widehat{F}_n - p_j)^2/p_j,$$

где $n\Delta_j \widehat{F}_n$ представляет собой число η_i в промежутке $[t_{j-1}, t_j)$. Согласно примеру 4 при $n \to \infty$

$$X_n^2 \xrightarrow{d} \zeta = \sum_{j=1}^{N} (\Delta_j B)^2/p_j.$$

Покажем, что $\zeta \sim \chi^2_{N-1}$. Для этого рассмотрим случайный вектор $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_N)$, где $\xi_j = \Delta_j B$. Так как $\Delta_j B = \Delta_j W - p_j W(1)$, то $\boldsymbol{\xi}$ — нормальный вектор. Вычислим его ковариационную матрицу $\boldsymbol{\Sigma} = \|\sigma_{jk}\|_{N\times N}$. Запишем:

$$\Delta_j B = \Delta_j W - p_j \sum_{k=1}^{N} \Delta_k W = \sum_{k=1}^{N} a_{jk}\, \Delta_k W, \quad j = 1, \ldots, N, \tag{4}$$

где $a_{jk} = \delta_{jk} - p_j$.*) В силу свойств 2 и 3 броуновского движения

$$\mathbf{M}(\Delta_j W)(\Delta_k W) = \delta_{jk} p_j. \tag{5}$$

С учетом соотношений (4), (5) и равенства $p_1 + \ldots + p_N = 1$ нетрудно вывести (проверьте!), что

$$\sigma_{jk} = \mathbf{M}(\Delta_j B)(\Delta_k B) = \sum_{l=1}^{N} a_{jl} a_{kl} p_l = \delta_{jk} p_j - p_j p_k.$$

Таким образом, ковариационная матрица $\boldsymbol{\Sigma}$ совпадает (с точностью до обозначений) с матрицей, задаваемой формулой (3) гл. 18. Дальнейшие рассуждения повторяют приведенные во второй части доказательства теоремы 1 гл. 18.

Нажми, водитель, тормоз, наконец,
Ты нас тиранил три часа подряд,
Слезайте, граждане, приехали, конец...

Ю. Визбор, «Охотный ряд»

*) Здесь $\delta_{jk} = 1$ при $j = k$ и $\delta_{jk} = 0$ при $j \neq k$ (*символ Кронекера*).

ПРИЛОЖЕНИЕ. НЕКОТОРЫЕ СВЕДЕНИЯ ИЗ ТЕОРИИ ВЕРОЯТНОСТЕЙ И ЛИНЕЙНОЙ АЛГЕБРЫ

РАЗДЕЛ 1

АКСИОМАТИКА ТЕОРИИ ВЕРОЯТНОСТЕЙ

Целью придумавших систему было, очевидно, скрыть, что в этих значках содержится какой-то смысл.

А. Конан Дойл, «Пляшущие человечки»

Вероятностной моделью называется тройка $(\Omega, \mathcal{A}, \mathsf{P})$, где

- $\Omega = \{\omega\}$ — *множество элементарных событий* ω;
- $\mathcal{A}$ — такой набор подмножеств Ω, что
  1) $\Omega \in \mathcal{A}$,
  2) если $A \in \mathcal{A}$, то и его *дополнение* $\overline{A} = \Omega \setminus A \in \mathcal{A}$,
  3) если $A_1, A_2, \ldots$ принадлежат $\mathcal{A}$, то и их *объединение* $\bigcup A_k \in \mathcal{A}$.

  Совокупность подмножеств Ω, обладающая свойствами 1 – 3, называется *σ-алгеброй*, а сами подмножества — *событиями*;
- P — *вероятность*, т. е. такая функция от событий, что
  1) $\mathsf{P}(A) \geqslant 0$ для любого $A \in \mathcal{A}$,
  2) $\mathsf{P}(\Omega) = 1$,
  3) для непересекающихся событий $A_1, A_2, \ldots$ $\mathsf{P}(\bigcup A_k) = \sum\limits_k \mathsf{P}(A_k)$.

Свойство непрерывности.
Для вложенных событий $A_1 \supseteq A_2 \supseteq \ldots$ $\lim\limits_{k\to\infty} \mathsf{P}(A_k) = \mathsf{P}(\bigcap A_k)$.

Определение. *Функция $\xi(\omega)$ называется случайной величиной, если для произвольного числа x множество $\{\omega\colon \xi(\omega) \leqslant x\} \in \mathcal{A}$ (в таком случае говорят, что ξ измерима относительно σ-алгебры $\mathcal{A}$).*

Набор $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_k)$ случайных величин называется *случайным вектором.*

РАЗДЕЛ 2

МАТЕМАТИЧЕСКОЕ ОЖИДАНИЕ И ДИСПЕРСИЯ

Определение. *Для случайной величины $\xi \geqslant 0$*

$$\mathsf{M}\xi = \int\limits_{\Omega} \xi(\omega)\, \mathsf{P}(d\omega) = \lim_{n\to\infty} \left[\sum_{i=1}^{n2^n} \frac{i-1}{2^n}\, \mathsf{P}\left(\frac{i-1}{2^n} \leqslant \xi < \frac{i}{2^n} \right) + n\, \mathsf{P}\left(\xi \geqslant n\right) \right].$$

Эта формула означает, что если заменить ξ дискретной случайной величиной $\xi^{(n)}$ таким образом, что $\xi^{(n)} = (i-1)\,2^{-n}$ на множестве $\{\omega\colon (i-1)\,2^{-n} \leqslant \xi < i\,2^{-n}\}$, то, вычисляя $\mathsf{M}\xi^{(n)}$ и устремляя $n \to \infty$, приходим к пределу, который и называется $\mathsf{M}\xi$.

Определение. *Для произвольной случайной величины* ξ

$$\mathsf{M}\xi = \mathsf{M}\xi^+ - \mathsf{M}\xi^-, \quad \text{где } \xi^+ = \max\{\xi, 0\}, \quad \xi^- = -\min\{\xi, 0\},$$

и хотя бы одно из математических ожиданий конечно. В случае, когда $\mathsf{M}\xi^+ = \mathsf{M}\xi^- = +\infty$, $\mathsf{M}\xi$ *(интеграл Лебега) не существует.*

Определение. *Борелевской* σ*-алгеброй* $\mathcal{B}$ *называется наименьшая* σ*-алгебра, содержащая всевозможные интервалы на прямой.*

Определение. *Функция* $\varphi(x)$ *— борелевская, если она измерима относительно* $\mathcal{B}$ *(в частности, такова любая непрерывная функция).*

Теорема о замене переменных. *Для борелевской функции* φ

Справа в формуле стоит *интеграл Стильтьеса.*

$$\mathsf{M}\varphi(\xi) = \int\limits_{\Omega} \varphi(\xi(\omega))\,\mathsf{P}(d\omega) = \int\limits_{-\infty}^{+\infty} \varphi(x)\,F_\xi(dx),$$

где $F_\xi(x) = \mathsf{P}(\xi \leqslant x)$ *— функция распределения случайной величины* ξ.

Следствие. *Момент* k*-го порядка* $\mathsf{M}\xi^k = \int\limits_{-\infty}^{+\infty} x^k F_\xi(dx)$.

Свойства математического ожидания

1) $\mathsf{M}(\xi + c) = \mathsf{M}\xi + c, \quad \mathsf{M}(c\,\xi) = c\,\mathsf{M}\xi$.
2) $\mathsf{M}(\xi + \eta) = \mathsf{M}\xi + \mathsf{M}\eta$, если нет неопределенности вида $\infty - \infty$.
3) Если $\xi \leqslant \eta$, то $\mathsf{M}\xi \leqslant \mathsf{M}\eta$.
4) $|\mathsf{M}\xi| \leqslant \mathsf{M}|\xi|$.
5) $\mathsf{M}(\xi\eta) = \mathsf{M}\xi \cdot \mathsf{M}\eta$, когда ξ и η независимы, а $\mathsf{M}\xi$ и $\mathsf{M}\eta$ конечны.

Для случайных величин ξ и η таких, что $\mathsf{M}\xi^2 < \infty$ и $\mathsf{M}\eta^2 < \infty$ определены

- *дисперсия* $\mathsf{D}\xi = \mathsf{M}(\xi - \mathsf{M}\xi)^2 = \mathsf{M}\xi^2 - (\mathsf{M}\xi)^2$,
- *ковариация* $\mathsf{cov}\,(\xi, \eta) = \mathsf{M}(\xi - \mathsf{M}\xi)(\eta - \mathsf{M}\eta) = \mathsf{M}(\xi\eta) - \mathsf{M}\xi\,\mathsf{M}\eta$,
- *коэффициент корреляции* $\boldsymbol{\rho}(\xi,\eta) = \mathsf{cov}\,(\xi, \eta)\,/\sqrt{\mathsf{D}\xi\,\mathsf{D}\eta}$.

Свойства дисперсии и ковариации

1) $\mathsf{D}(\xi + c) = \mathsf{D}\xi, \quad \mathsf{D}(c\,\xi) = c^2\mathsf{D}\xi$.
2) Если ξ и η независимы, то $\mathsf{D}(\xi + \eta) = \mathsf{D}\xi + \mathsf{D}\eta$.
3) Для случайных величин $\xi_1, \ldots, \xi_n$ с конечными $\mathsf{M}\xi_i^2$

$$\mathsf{D}\left(\sum_{i=1}^{n} \xi_i\right) = \sum_{i=1}^{n} \mathsf{D}\xi_i + 2\sum_{i<j} \mathsf{cov}\,(\xi_i, \xi_j)\,.$$

4) $\mathsf{cov}\,(a\xi + b, c\eta + d) = ac\,\mathsf{cov}\,(\xi, \eta)$ и $\boldsymbol{\rho}(a\xi + b, c\eta + d) = \boldsymbol{\rho}(\xi,\eta)$ при $ac > 0$.
5) Пусть случайный вектор $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_k)$ имеет *ковариационную матрицу* $\boldsymbol{\Sigma} = \|\mathsf{cov}\,(\xi_i, \xi_j)\,\|_{k\times k}$, $\boldsymbol{B}$ — произвольная числовая $(m \times k)$-матрица (см. раздел 10). Тогда $\boldsymbol{\eta} = \boldsymbol{B}\boldsymbol{\xi}$ имеет ковариационную матрицу $\|\mathsf{cov}\,(\eta_i, \eta_j)\,\|_{m\times m} = \boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{B}^T$.

РАЗДЕЛ 3

ФОРМУЛА СВЕРТКИ

Рассмотрим *независимые* случайные величины ξ и η, т. е. такие, что для любых чисел x и y верно равенство

$\mathsf{P}(\xi \leqslant x, \eta \leqslant y) = \mathsf{P}(\xi \leqslant x)\,\mathsf{P}(\eta \leqslant y)$.

Если они распределены дискретно, то в силу независимости

$$\mathsf{P}(\xi+\eta = x_i) = \sum_j \mathsf{P}(\xi = x_i - y_j, \eta = y_j) = \sum_j \mathsf{P}(\xi = x_i - y_j)\,\mathsf{P}(\eta = y_j).$$

Если же у них есть плотности $p_\xi(x)$ и $p_\eta(x)$, то плотность суммы

$$p_{\xi+\eta}(x) = \int\limits_{-\infty}^{+\infty} p_\xi(x-y)\,p_\eta(y)\,dy.$$

В общем случае функция распределения суммы независимых случайных величин ξ и η вычисляется по формуле

$$F_{\xi+\eta}(x) = \int\limits_{-\infty}^{+\infty} F_\xi(x-y)\,F_\eta(dy).$$

Интеграл справа называется *сверткой* $F_\xi(x)$ и $F_\eta(x)$ (обозначается $F_\xi * F_\eta$).

РАЗДЕЛ 4

ВЕРОЯТНОСТНЫЕ НЕРАВЕНСТВА

1) Для случайной величины $\xi \geqslant 0$ и любого $\varepsilon > 0$ верно *неравенство Чебышёва*

   $$\mathsf{P}(\xi \geqslant \varepsilon) \leqslant \frac{\mathsf{M}\xi}{\varepsilon}.$$

   Следствие. *Взяв* $\eta = (\xi - \mathsf{M}\xi)^2$, *получим* $\mathsf{P}(|\xi - \mathsf{M}\xi| \geqslant \varepsilon) \leqslant \dfrac{\mathsf{D}\xi}{\varepsilon^2}$.

2) $\mathsf{M}|\xi\eta| \leqslant \sqrt{\mathsf{M}\xi^2\mathsf{M}\eta^2}$ (*неравенство Коши—Буняковского*).

   Следствие. *Применяя его к случайным величинам* $\xi' = \xi - \mathsf{M}\xi$ *и* $\eta' = \eta - \mathsf{M}\eta$ *и используя свойство* 4 *математического ожидания, получаем, что для коэффициента корреляции верно неравенство* $|\boldsymbol{\rho}(\xi,\eta)| \leqslant 1$.

3) Если $\varphi(x)$ — *выпуклая вниз* функция и $\mathsf{M}\xi$, $\mathsf{M}\varphi(\xi)$ конечны, то

   $\varphi(\mathsf{M}\xi) \leqslant \mathsf{M}\varphi(\xi)$ (*неравенство Иенсена*).

   Для *строго выпуклых*[*)] φ в случае $\xi \neq \text{const}$ неравенство строгое.

   Замечание. Для бернуллиевской случайной величины ξ неравенство Иенсена означает, что график $y = \varphi(x)$ на $[0,1]$ лежит под хордой, соединяющей точки с координатами $(0,\varphi(0))$ и $(1,\varphi(1))$.

[*)] $\varphi(px+qy) < p\varphi(x) + q\varphi(y)$ для любых $x \neq y$, $0 < p < 1$, $q = 1-p$.

РАЗДЕЛ 5

СХОДИМОСТЬ СЛУЧАЙНЫХ ВЕЛИЧИН И ВЕКТОРОВ

Определение. *Случайные величины* $\xi_1, \xi_2, \ldots$ *сходятся при* $n \to \infty$ *к случайной величине* ξ

1) *почти наверное* ($\xi_n \xrightarrow{\text{п. н.}} \xi$), *если* $\mathsf{P}\{\omega\colon \xi_n(\omega) \to \xi(\omega)\} = 1$;
2) *в среднем квадратическом* ($\xi_n \xrightarrow{\text{с. к.}} \xi$), *если* $\mathsf{M}(\xi_n - \xi)^2 \to 0$;
3) *по вероятности* ($\xi_n \xrightarrow{\mathsf{P}} \xi$), *если* $\forall \varepsilon > 0 \quad \mathsf{P}(|\xi_n - \xi| > \varepsilon) \to 0$;
4) *по распределению* ($\xi_n \xrightarrow{d} \xi$), *если функции распределения* $F_{\xi_n}(x)$ *сходятся к* $F_\xi(x)$ *в точках непрерывности последней.*

Буква «d» происходит от distribution (*англ.*) — распределение.

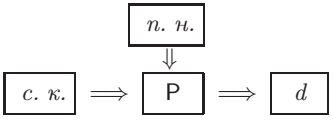
Пример $\xi_n = 1/n \xrightarrow{d} \xi = 0$ показывает, что функции F_{ξ_n} могут не сходиться к F_ξ в точках разрыва последней: $F_{\xi_n}(0) = 0 \neq F_\xi(0) = 1$.

Случайные векторы $\boldsymbol{\xi}_n$ сходятся к вектору $\boldsymbol{\xi}$ в смысле $\xrightarrow{\text{п. н.}}$, $\xrightarrow{\text{с. к.}}$ или $\xrightarrow{\mathsf{P}}$, если в соответствующем смысле сходятся все их компоненты.

Определение. *Говорят, что* $\boldsymbol{\xi}_n \xrightarrow{d} \boldsymbol{\xi}$, *если* $\mathsf{M}\varphi(\boldsymbol{\xi}_n) \to \mathsf{M}\varphi(\boldsymbol{\xi})$ *для любой непрерывной и ограниченной функции* φ *на* $\mathbb{R}^k$.

Критерием $\boldsymbol{\xi}_n \xrightarrow{d} \boldsymbol{\xi}$ является сходимость линейных комбинаций компонент $\boldsymbol{\lambda}^T\boldsymbol{\xi}_n \xrightarrow{d} \boldsymbol{\lambda}^T\boldsymbol{\xi}$ для всевозможных $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_k)$.

Диаграмма зависимости видов сходимости



Теорема о монотонной сходимости. *Пусть* $\xi_1 \leqslant \xi_2 \leqslant \ldots$ *и* $\mathsf{M}\xi_1 > -\infty$. *Если* $\xi_n \xrightarrow{\text{п. н.}} \xi$ *при* $n \to \infty$, *то* $\mathsf{M}\xi_n \uparrow \mathsf{M}\xi$.

Теорема Лебега о мажорируемой сходимости. *Пусть* $\xi_n \xrightarrow{\mathsf{P}} \xi$ *при* $n \to \infty$. *Если найдется такая случайная величина* η, *что* $|\xi_n| \leqslant \eta$ *и* $\mathsf{M}\eta < \infty$, *то*

$$\mathsf{M}|\xi| < \infty, \quad \mathsf{M}\xi_n \to \mathsf{M}\xi \quad \text{и} \quad \mathsf{M}|\xi_n - \xi| \to 0 \quad \text{при } n \to \infty.$$

Свойства сходимости

Пусть a, b, c — константы, α_n, β_n, ξ_n, ξ — случайные величины. Тогда

1) $\alpha_n \xrightarrow{\mathsf{P}} a$, $\beta_n \xrightarrow{\mathsf{P}} b$, $\xi_n \xrightarrow{d} \xi \implies \alpha_n + \beta_n\xi_n \xrightarrow{d} a + b\xi$;
2) $\xi_n \xrightarrow{d} c \implies \xi_n \xrightarrow{\mathsf{P}} c$.
3) Для любых случайных векторов $\boldsymbol{\xi}_n$, $\boldsymbol{\xi}$ и *непрерывной* функции φ на $\mathbb{R}^k$ $\boldsymbol{\xi}_n \xrightarrow{\mathsf{P}} \boldsymbol{\xi} \implies \varphi(\boldsymbol{\xi}_n) \xrightarrow{\mathsf{P}} \varphi(\boldsymbol{\xi})$ и $\boldsymbol{\xi}_n \xrightarrow{d} \boldsymbol{\xi} \implies \varphi(\boldsymbol{\xi}_n) \xrightarrow{d} \varphi(\boldsymbol{\xi})$.

РАЗДЕЛ 6

ПРЕДЕЛЬНЫЕ ТЕОРЕМЫ

Пусть $X_1, X_2, \ldots$ — независимые одинаково распределенные случайные величины. Положим $S_n = X_1 + \ldots + X_n$.

Усиленный закон больших чисел. *Если существует* $\mu = \mathbf{M}X_1$ *(не обязательно конечное), то* $S_n/n \xrightarrow{\text{п. н.}} \mu$ *при* $n \to \infty$.

Из приведенной выше диаграммы зависимости видов сходимости следует сходимость $S_n/n \xrightarrow{\mathsf{P}} \mu$ (*закон больших чисел*).

Центральная предельная теорема. *Если* $0 < \sigma^2 = \mathbf{D}X_1 < \infty$, *то*

$$S_n^* = \frac{S_n - \mathbf{M}S_n}{\sqrt{\mathbf{D}S_n}} = \frac{S_n - \mu n}{\sigma\sqrt{n}} \xrightarrow{d} Z \quad \text{при } n \to \infty,$$

где Z — *стандартная нормальная случайная величина с функцией распределения*

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2}\, du.$$

Обозначим через $F_n(x) = \mathsf{P}(S_n^* \leqslant x)$. Центральная предельная теорема равносильна сходимости $F_n(x) \to \Phi(x)$ при всех x. На самом деле можно показать, что эта сходимость является *равномерной*:

$$\sup_x |F_n(x) - \Phi(x)| \to 0 \quad \text{при } n \to \infty.$$

Следующая теорема позволяет оценить скорость сходимости.

Теорема Берри—Эссеена. *Пусть* $\mathbf{M}|X_1|^3 < \infty$. *Тогда*

$$\sup_x |F_n(x) - \Phi(x)| \leqslant \frac{C\,\mathbf{M}|X_1 - \mu|^3}{\sigma^3\sqrt{n}} \quad \text{при всех } n,$$

где C — *постоянная.*[*)] (Доказательство см. в [12, с. 420].)

Приведем два обобщения центральной предельной теоремы на случаи неодинаково распределенных и зависимых слагаемых.

Теорема Линдеберга. *Пусть* $X_1, X_2, \ldots$ — *независимые случайные величины,* $\mu_i = \mathbf{M}X_i$, $0 < \sigma_i^2 = \mathbf{D}X_i < \infty$, $a_n = \mathbf{M}S_n = \sum\limits_{i=1}^{n} \mu_i$, $b_n^2 = \mathbf{D}S_n = \sum\limits_{i=1}^{n} \sigma_i^2$, $C_{i,n} = \{\omega\colon |X_i(\omega) - \mu_i| > \varepsilon b_n\}$. *Тогда, если для любого* $\varepsilon > 0$

$$\lim_{n\to\infty} \frac{1}{b_n^2} \sum_{i=1}^{n} \mathbf{M}\left[(X_i - \mu_i)^2 I_{C_{i,n}}\right] = 0,$$

то $(S_n - a_n)/b_n \xrightarrow{d} Z \sim \mathcal{N}(0,1)$ *при* $n \to \infty$ (см. [90, с. 351]).

[*)] Берри утверждал, что $C \leqslant 1{,}88$, но его вычисления содержали ошибку. Эссеен показал, что $C \leqslant 7{,}59$. В дальнейшем C удалось уменьшить до следующих значений: 2,9 (Эссеен, 1956); 2,05 (Уоллес, 1958); 0,9051 (В. М. Золотарев, 1966); 0,7655 (И. С. Шиганов, 1982). Известно также, что $C \geqslant 1/\sqrt{2\pi} \approx 0{,}3989$. (см. [90, с. 402]).

Последовательность $X_1,X_2,\ldots$ называется *стационарной*, если для любого n совместное распределение $(X_{1+i},\ldots,X_{n+i})$ одно и то же при всех i. Последовательность называется *m-зависимой*, если $(X_1,\ldots,X_n)$ и $(X_j,X_{j+1},\ldots)$ независимы при $j > n+m$.*)

Следующая теорема — частный случай результата В. Хефдинга и Х. Робинса (1948 г.).

Теорема Хефдинга и Робинса. *Пусть случайные величины $X_1,X_2,\ldots$ образуют стационарную m-зависимую последовательность; $\mathsf{M}X_1 = \mu$, $\mathsf{M}|X_1|^3 < \infty$. Тогда при $n \to \infty$*

$$\frac{S_n - \mu n}{\sqrt{n}} \xrightarrow{d} \xi \sim \mathcal{N}(0,\sigma^2), \quad \text{где } \sigma^2 = \mathsf{D}X_1 + 2\sum_{j=1}^{m} \mathsf{cov}\,(X_1, X_{1+j})\,.$$

РАЗДЕЛ 7

УСЛОВНОЕ МАТЕМАТИЧЕСКОЕ ОЖИДАНИЕ

Определение. *Условной вероятностью события A при условии события B (предполагается, что $\mathsf{P}(B) > 0$) называется*

$$\mathsf{P}(A \,|\, B) = \mathsf{P}\,(A \bigcap B)\,/\mathsf{P}(B).$$

Формула полной вероятности. *Если события $A_1,A_2,\ldots$ не пересекаются и $\bigcup A_k = \Omega$, то для любого события B*

$$\mathsf{P}(B) = \sum_k \mathsf{P}\,(B \bigcap A_k) = \sum_k \mathsf{P}(B \,|\, A_k)\,\mathsf{P}(A_k).$$

Пусть случайные величины ξ и η имеют дискретные распределения.

Определение. *Условное математическое ожидание ξ при условии события $\{\omega\colon \eta(\omega) = y_j\}$ есть*

$$\mathsf{M}(\xi \,|\, \eta = y_j) = \sum_i x_i\,\mathsf{P}(\xi = x_i \,|\, \eta = y_j).$$

Определение. *Условным математическим ожиданием $\mathsf{M}(\xi \,|\, \eta)$ называется случайная величина, принимающая на множествах вида $\{\omega\colon \eta(\omega) = y_j\}$ значения $\mathsf{M}(\xi \,|\, \eta = y_j)$.*

Пусть случайные величины ξ и η имеют совместную плотность $p_{(\xi,\eta)}(x,y)$ (см. раздел 8).

Определение. *Условная плотность ξ при условии η есть*

$$p_{\xi|\eta}(x,y) = \begin{cases} p_{(\xi,\eta)}(x,y)/p_\eta(y), & \text{если } p_\eta(y) > 0, \\ 0, & \text{если } p_\eta(y) = 0. \end{cases}$$

*) Примером могут служить $\eta_i = \varphi(\xi_i,\ldots,\xi_{i+m-1})$, где ξ_i — независимые и одинаково распределенные случайные величины, φ — произвольная борелевская функция на $\mathbb{R}^m$ (например, линейная).

Определение. *Условное математическое ожидание ξ при условии события $\{\omega\colon \eta(\omega) = y\}$ задается формулой*

$$\mathsf{M}(\xi \,|\, \eta = y) = \int_{-\infty}^{\infty} x\, p_{\xi|\eta}(x,y)\, dx.$$

Определение. *Условное математическое ожидание $\mathsf{M}(\xi \,|\, \eta)$ — случайная величина, принимающая на множествах $\{\omega\colon \eta(\omega) = y\}$ значения $\mathsf{M}(\xi \,|\, \eta = y)$.*

Определение $\mathsf{M}(\xi \,|\, \eta)$ в общем случае приведено в [90, с. 231].

Свойства условного математического ожидания

1) $\mathsf{M}(\mathsf{M}(\xi \,|\, \eta)) = \mathsf{M}\xi$.
2) Если ξ не зависит от η, то $\mathsf{M}(\xi \,|\, \eta) = \mathsf{M}\xi$ (п. н.).
   Для любой борелевской функции $\varphi(x)$ имеют место равенства
3) $\mathsf{M}(\varphi(\eta)\xi \,|\, \eta) = \varphi(\eta)\mathsf{M}(\xi \,|\, \eta)$ (п. н.),
4) $\mathsf{M}[\mathsf{M}(\xi \,|\, \eta) \,|\, \varphi(\eta)] = \mathsf{M}(\xi \,|\, \varphi(\eta))$ (п. н.).

РАЗДЕЛ 8

ПРЕОБРАЗОВАНИЕ ПЛОТНОСТИ СЛУЧАЙНОГО ВЕКТОРА

Случайный вектор $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_k)$ (его распределение) имеет *плотность* $p_{\boldsymbol{\xi}}(\boldsymbol{x}) \geqslant 0$, где $\boldsymbol{x} = (x_1, \ldots, x_k)$, если для любого борелевского множества $B \in \mathbb{R}^k$

$$\mathsf{P}(\boldsymbol{\xi} \in B) = \int\limits_B p_{\boldsymbol{\xi}}(\boldsymbol{x})\, d\boldsymbol{x}.$$

В частности, *совместная функция распределения* компонент $\boldsymbol{\xi}$

$$F_{\boldsymbol{\xi}}(\boldsymbol{x}) = \mathsf{P}(\xi_1 \leqslant x_1, \ldots, \xi_n \leqslant x_k) = \int\limits_{-\infty}^{x_1} \cdots \int\limits_{-\infty}^{x_k} p_{\boldsymbol{\xi}}(\boldsymbol{x})\, dx_1 \ldots dx_k.$$

Плотность $p_{\boldsymbol{\xi}}(\boldsymbol{x})$ можно найти из равенств (верных почти всюду)

$$\frac{\partial^k F_{\boldsymbol{\xi}}(\boldsymbol{x})}{\partial x_1 \ldots \partial x_k} = p_{\boldsymbol{\xi}}(\boldsymbol{x}) = \lim_{\delta \downarrow 0} \frac{\mathsf{P}(x_1 \leqslant \xi_1 \leqslant x_1 + \delta, \ldots, x_k \leqslant \xi_k \leqslant x_k + \delta)}{\delta^k}.$$

Плотность подвектора $(\xi_1, \ldots, \xi_j)$, $1 \leqslant j < k$, вычисляется интегрированием $p_{\boldsymbol{\xi}}(\boldsymbol{x})$ по координатам $x_{j+1}, \ldots, x_k$:

$$p_{(\xi_1,\ldots,\xi_j)}(x_1, \ldots, x_j) = \int\limits_{-\infty}^{\infty} \cdots \int\limits_{-\infty}^{\infty} p_{\boldsymbol{\xi}}(\boldsymbol{x})\, dx_{j+1} \ldots dx_k.$$

Допустим, что соотношения $y_i = \varphi_i(x_1, \ldots, x_k)$, $i = 1, \ldots, k$, задают взаимно однозначное дифференцируемое отображение $\boldsymbol{y} = \boldsymbol{\varphi}(\boldsymbol{x})$ некоторой

области из $\mathbb{R}^k$ на другую. Обозначим через $\boldsymbol{\varphi}^{-1}$ обратное отображение,

$$J(\boldsymbol{y}) = \det \begin{pmatrix} \dfrac{\partial \varphi_1^{-1}}{\partial y_1} & \cdots & \dfrac{\partial \varphi_k^{-1}}{\partial y_1} \\ \cdots & \cdots & \cdots \\ \dfrac{\partial \varphi_1^{-1}}{\partial y_k} & \cdots & \dfrac{\partial \varphi_k^{-1}}{\partial y_k} \end{pmatrix} — \textit{ якобиан } \boldsymbol{\varphi}^{-1} \text{ (см. раздел 10).}$$

Формула преобразования плотности. *Плотность распределения случайного вектора* $\boldsymbol{\eta} = \boldsymbol{\varphi}(\boldsymbol{\xi})$ *вычисляется по формуле*

$$p_{\boldsymbol{\eta}}(\boldsymbol{y}) = |J(\boldsymbol{y})|\, p_{\boldsymbol{\xi}}(\boldsymbol{\varphi}^{-1}(\boldsymbol{y})).$$

Следствие. *Если* $\eta = a + b\xi$, *то* $p_\eta(y) = \dfrac{1}{|b|}\, p_\xi\left(\dfrac{y-a}{b}\right)$ *при* $b \neq 0$.

РАЗДЕЛ 9

ХАРАКТЕРИСТИЧЕСКИЕ ФУНКЦИИ И МНОГОМЕРНОЕ НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

В этом разделе i обозначает комплексное число $\sqrt{-1}$. Используемые термины матричного исчисления определены в разд. 10.

Определение. *Характеристической функцией* (х. ф.) *случайной величины* ξ *называется комплекснозначная функция действительного аргумента* $\psi_\xi(t) = \mathsf{M}e^{it\xi} = \mathsf{M}\cos(t\xi) + i\,\mathsf{M}\sin(t\xi)$.

Например, для $X \sim \mathcal{N}(\mu, \sigma^2)$ характеристической функцией служит $\psi_X(t) = \exp\left\{it\mu - \frac{1}{2}\sigma^2 t^2\right\}$ (см. [90, с. 296]).

По $\psi_\xi(t)$ функция распределения $F_\xi(x)$ определяется однозначно. В частности, если $\int\limits_{-\infty}^{\infty} |\psi_\xi(t)|\, dt < \infty$, то у случайной величины ξ есть плотность, вычисляемая по формуле *обратного преобразования Фурье*:

Ж. Фурье (1768–1830), французский физик и математик.

$$p_\xi(x) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} e^{-ixt} \psi_\xi(t)\, dt.$$

Свойства характеристических функций

1) $\psi_\xi(0) = 1, \quad |\psi_\xi(t)| \leqslant \mathsf{M}|e^{it\xi}| = 1.$
2) Если ξ и η независимы, то
$$\psi_{\xi+\eta}(t) = \mathsf{M}e^{it(\xi+\eta)} = \mathsf{M}(e^{it\xi} \cdot e^{it\eta}) = \\ = \mathsf{M}e^{it\xi} \cdot \mathsf{M}e^{it\eta} = \psi_\xi(t) \cdot \psi_\eta(t).$$
3) $\psi_{a+b\xi}(t) = \mathsf{M}e^{it(a+b\xi)} = e^{ita}\mathsf{M}e^{i(bt)\xi} = e^{ita}\psi_\xi(bt).$
4) Если $\mathsf{M}|\xi|^n < \infty$, то $\psi_\xi^{(n)}(0) = i^n \mathsf{M}\xi^n$. Обратно, если существует и конечна $\psi_\xi^{(2n)}(0)$, то $\mathsf{M}\xi^{2n} < \infty$.

Определение. *Характеристической функцией случайного вектора* $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_k)$ *называется функция от* $\boldsymbol{t} = (t_1, \ldots, t_k)$

$$\psi_{\boldsymbol{\xi}}(\boldsymbol{t}) = \mathsf{M} e^{i\boldsymbol{t}^T \boldsymbol{\xi}} = \mathsf{M} e^{i(t_1\xi_1 + \ldots + t_k\xi_k)}.$$

Теорема непрерывности. *Из сходимости* $\boldsymbol{\xi}_n \xrightarrow{d} \boldsymbol{\xi}$ *при* $n \to \infty$ (см. раздел 5) *вытекает, что* $\psi_{\boldsymbol{\xi}_n}(\boldsymbol{t}) \to \psi_{\boldsymbol{\xi}}(\boldsymbol{t})$ *при всех* $\boldsymbol{t} \in \mathbb{R}^k$.

Обратно: пусть $\psi_{\boldsymbol{\xi}_n}(\boldsymbol{t}) \to \psi(\boldsymbol{t})$ *при всех* $\boldsymbol{t}$, *где* ψ *непрерывна в* $\mathbf{0}$, *тогда* ψ *является характеристической функцией распределения, предельного для распределений случайных векторов* $\boldsymbol{\xi}_n$.

Определение. *Случайный вектор* $\boldsymbol{\xi}$ *называется нормальным с параметрами* $\boldsymbol{\mu}$ *и* $\boldsymbol{\Sigma}$ *(обозн.* $\boldsymbol{\xi} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$*), где* $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)$, *матрица* $\boldsymbol{\Sigma} = \|\sigma_{lj}\|_{k\times k}$ *неотрицательно определена, если*

$$\psi_{\boldsymbol{\xi}}(\boldsymbol{t}) = \exp\left\{i\boldsymbol{t}^T\boldsymbol{\mu} - \frac{1}{2}\,\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t}\right\} = \exp\left\{i\sum_{j=1}^{k} t_j\mu_j - \frac{1}{2}\sum_{l=1}^{k}\sum_{j=1}^{k}\sigma_{lj}t_lt_j\right\}.$$

Для такого вектора $\mathsf{M}\xi_j = \mu_j$ и $\mathsf{cov}\,(\xi_l, \xi_j) = \sigma_{lj}$.

Некоррелированность компонент нормального вектора ($\sigma_{lj} = 0$ при $l \neq j$) эквивалентна их независимости.

Вектор $\boldsymbol{\xi}$ является нормальным тогда и только тогда, когда для любого $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_k)$ линейная комбинация его компонент $\boldsymbol{\lambda}^T\boldsymbol{\xi}$ имеет нормальное распределение (см. [90, с. 322]).

Если $\boldsymbol{\xi} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ и $\boldsymbol{B}$ — произвольная числовая $(m \times k)$-матрица, то $\boldsymbol{\eta} = \boldsymbol{B}\boldsymbol{\xi} \sim \mathcal{N}\left(\boldsymbol{B}\boldsymbol{\mu}, \boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{B}^T\right)$.

Когда ранг матрицы $\boldsymbol{\Sigma}$ равен $r < k$, распределение $\boldsymbol{\xi}$ сосредоточено в r-мерной гиперплоскости $\boldsymbol{\mu} + \boldsymbol{\Sigma}\boldsymbol{x}$, где $\boldsymbol{x}$ пробегает $\mathbb{R}^k$.

В случае невырожденности $\boldsymbol{\Sigma}$ (т. е., когда $r = k$) нормальный вектор $\boldsymbol{\xi}$ имеет положительную на всем $\mathbb{R}^k$ плотность

$$p_{\boldsymbol{\xi}}(\boldsymbol{x}) = (2\pi)^{-k/2}\,(\det\boldsymbol{\Sigma})^{-1/2}\exp\left\{-\frac{1}{2}\,(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\},$$

где $\det\boldsymbol{\Sigma}$ и $\boldsymbol{\Sigma}^{-1}$ суть соответственно определитель и обратная матрица матрицы ковариаций $\boldsymbol{\Sigma}$.

В частности, для $k = 2$, $\mu_1 = \mu_2 = 0$, $\sigma_{11} = \sigma_{22} = 1$, $\sigma_{12} = \sigma_{21} = \rho$

$$p_{\boldsymbol{\xi}}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left\{-\frac{x^2 - 2\rho\,xy + y^2}{2(1-\rho^2)}\right\},$$

где ρ — коэффициент корреляции между ξ_1 и ξ_2 (здесь $|\rho| < 1$).

С помощью этих свойств можно показать, что нормальности компонент *недостаточно* для того, чтобы вектор был нормальным. Действительно, рассмотрим случайный вектор

$$(\xi',\eta') = \begin{cases}(\xi,|\eta|), & \text{если } \xi \geqslant 0,\\ (\xi,-|\eta|), & \text{если } \xi < 0,\end{cases}$$

где $\xi \sim \mathcal{N}(0,1)$ и $\eta \sim \mathcal{N}(0,1)$ независимы. Нетрудно проверить, что каждая из компонент ξ' и η' будет стандартной нормальной. При этом плотность вектора (ξ',η') равна удвоенной плотности (ξ,η) в I и III координатных углах и равна 0 во II и IV координатных углах.

Если $\boldsymbol{\xi} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ и матрица $\boldsymbol{\Sigma}$ невырождена, то $\boldsymbol{\xi} = \boldsymbol{\mu} + \boldsymbol{A\zeta}$, где $\boldsymbol{A} = \boldsymbol{\Sigma}^{1/2}$ и $\boldsymbol{\zeta}$ имеет независимые $\mathcal{N}(0,1)$ компоненты.

Из этого представления немедленно вытекает, что для невырожденной $\boldsymbol{\Sigma} = \|\sigma_{ij}\|_{k\times k}$ квадратичная форма $(\boldsymbol{\xi} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\xi} - \boldsymbol{\mu})$ имеет распределение хи-квадрат с k степенями свободы.

РАЗДЕЛ 10

ЭЛЕМЕНТЫ МАТРИЧНОГО ИСЧИСЛЕНИЯ

Определение. *Прямоугольная таблица действительных чисел* a_{ij} $(i = 1,\ldots,m;\ j = 1,\ldots,n)$, *имеющая* m *строк и* n *столбцов, называется* $(m \times n)$-*матрицей* (обозначается $\boldsymbol{A} = \|a_{ij}\|_{m\times n}$):

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix}.$$

Определение. *Матрица, чьи столбцы являются строками матрицы* $\boldsymbol{A}$ *(при сохранении их порядка), называется транспонированной по отношению к матрице* $\boldsymbol{A}$ *и обозначается* $\boldsymbol{A}^T$.

Во всех формулах, где участвуют матрицы, под *вектором* $\boldsymbol{x} = (x_1,\ldots,x_m)^T$ мы понимаем *вектор-столбец*, т. е. $(m \times 1)$-матрицу с элементами x_i.

Матрицу можно умножать на действительное число b путем умножения на b каждого ее элемента. Матрицы одинаковой размерности можно складывать поэлементно.

Определение. *Матрица, все элементы которой равны нулю, называется нулевой и обозначается* $\mathbf{0}$.

Определение. *Векторы* $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k$ *называются линейно зависимыми, если найдутся числа* $c_1,\ldots,c_k$, *не все равные* 0, *что*

$c_1\boldsymbol{x}_1 + \ldots + c_k\boldsymbol{x}_k = \mathbf{0}$ (здесь $\mathbf{0}$ — нулевой вектор).

Ранг матрицы $\boldsymbol{A}$ определяется как максимально возможное число линейно независимых строк (или, что эквивалентно, столбцов) $\boldsymbol{A}$.

Определение. *Произведением* $\boldsymbol{C} = \boldsymbol{AB}$ $(m \times l)$-*матрицы* $\boldsymbol{A}$ *на* $(l \times n)$-*матрицу* $\boldsymbol{B}$ *называется* $(m \times n)$-*матрица* $\boldsymbol{C}$ *с элементами*

$$c_{ij} = \sum_{k=1}^{l} a_{ik}b_{kj}, \quad i = 1,\ldots,m;\ j = 1,\ldots,n.$$

Свойства умножения матриц

1) $(\boldsymbol{AB})^T = \boldsymbol{B}^T\boldsymbol{A}^T$.
2) $(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC})$ (*ассоциативность*).
3) $(\boldsymbol{A} + \boldsymbol{B})\boldsymbol{C} = \boldsymbol{AC} + \boldsymbol{BC}$ (*дистрибутивность*).

Однако, в общем случае умножение *некоммутативно*: $\boldsymbol{AB} \neq \boldsymbol{BA}$.

Введем еще несколько понятий, имеющих смысл исключительно *для квадратных матриц.*

Определение. *Верхнетреугольной (нижнетреугольной) называют квадратную матрицу, у которой все элементы a_{ij} под (над) главной диагональю a_{11},a_{22},$\ldots$,a_{nn} равны нулю.*

Определение. *Квадратная матрица, на главной диагонали которой стоят единицы, а остальные элементы — нули, называется единичной и обозначается $\boldsymbol{E}$.*

Определение. *Говорят, что $(n\times n)$-матрица $\boldsymbol{A}$ невырождена, если ее ранг равен n.*

В этом случае найдется $(n\times n)$-матрица $\boldsymbol{D}$ такая, что $\boldsymbol{DA}=\boldsymbol{E}$. Она называется *обратной* к $\boldsymbol{A}$ и обозначается $\boldsymbol{A}^{-1}$.

Когда обратные матрицы существуют, выполняются тождества

$$(\boldsymbol{A}^T)^{-1}=(\boldsymbol{A}^{-1})^T \quad \text{и} \quad (\boldsymbol{AB})^{-1}=\boldsymbol{B}^{-1}\boldsymbol{A}^{-1}.$$

Важнейшими характеристиками $(n\times n)$-матрицы $\boldsymbol{A}$ являются ее след $\operatorname{tr}\boldsymbol{A}$ и определитель $\det\boldsymbol{A}$.

Сокращения tr и det происходят от английских терминов trace (след) и determinant (определитель).

Определение. *Следом называется сумма элементов $\boldsymbol{A}$ на главной диагонали:* $\operatorname{tr}\boldsymbol{A}=a_{11}+\ldots+a_{nn}$.

Пусть $p=(i_1,i_2,\ldots,i_n)$ — это перестановка длины n, т. е. числа $1,2\ldots,n$, записанные в произвольном порядке.

Определение. *Определитель матрицы $\boldsymbol{A}$ задается формулой*

$$\det\boldsymbol{A}=\sum_p Z(p)\,a_{1i_1}a_{2i_2}\ldots a_{ni_n},\quad Z(p)=\prod_{1\leqslant k<l\leqslant n}\operatorname{sign}(i_l-i_k).$$

Здесь p пробегает все $n!$ перестановок, $\operatorname{sign}(x)$ *— знак числа x.*

В частности, $\det\boldsymbol{A}=a_{11}a_{22}-a_{12}a_{21}$ при $n=2$,

$$\det\boldsymbol{A}=a_{11}a_{22}a_{33}+a_{13}a_{21}a_{32}+a_{12}a_{23}a_{31}-a_{13}a_{22}a_{31}-a_{11}a_{23}a_{32}-a_{12}a_{21}a_{33}$$

при $n=3$ и $\det\boldsymbol{A}=a_{11}a_{22}\ldots a_{nn}$ для (верхне-) нижнетреугольной матрицы.

Условие $\det\boldsymbol{A}\neq 0$ равносильно невырожденности матрицы $\boldsymbol{A}$.

Иногда удается вычислить определитель с помощью следующей формулы, называемой *разложением по i-й строке*:

$$\det\boldsymbol{A}=\sum_{j=1}^{n}(-1)^{i+j}\,a_{ij}M_{ij},$$

где M_{ij} — определитель матрицы, получаемой из $\boldsymbol{A}$ вычеркиванием i-й строки и j-го столбца. Аналогично можно разложить определитель и по j-му столбцу.

Свойства следа и определителя

1) $\det \boldsymbol{A}^T = \det \boldsymbol{A}$.
2) $\operatorname{tr}(\boldsymbol{AB}) = \operatorname{tr}(\boldsymbol{BA})$ и $\det(\boldsymbol{AB}) = \det \boldsymbol{A} \cdot \det \boldsymbol{B}$.

Следствие 1. *Если $\boldsymbol{A}$ невырождена, то* $\det \boldsymbol{A}^{-1} = 1/\det \boldsymbol{A}$.

Следствие 2. *Пусть $\boldsymbol{D}$ невырождена и $\boldsymbol{B} = \boldsymbol{D}^{-1}\boldsymbol{AD}$ (такую $\boldsymbol{B}$ называют подобной $\boldsymbol{A}$). Тогда* $\operatorname{tr} \boldsymbol{B} = \operatorname{tr} \boldsymbol{A}$ *и* $\det \boldsymbol{B} = \det \boldsymbol{A}$.

Определение. *Квадратная матрица $\boldsymbol{A}$ называется симметричной, если* $\boldsymbol{A}^T = \boldsymbol{A}$.

Определение. *Матрица $\boldsymbol{C}$ ортогональна, когда* $\boldsymbol{C}^T\boldsymbol{C} = \boldsymbol{E} \iff \boldsymbol{C}^{-1} = \boldsymbol{C}^T$. (Из свойств определителя следует, что $|\det \boldsymbol{C}| = 1$.)

Теорема о приведении к главным осям. *Для любой симметричной $(n \times n)$-матрицы $\boldsymbol{A}$ существует ортогональная матрица $\boldsymbol{C}$ такая, что* $\boldsymbol{C}^T\boldsymbol{AC} = \boldsymbol{\Lambda}$, *где*

$$\boldsymbol{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix},$$

причем действительные числа $|\lambda_1| \geqslant |\lambda_2| \geqslant \dots \geqslant |\lambda_n|$ *являются корнями характеристического многочлена n-й степени от λ*

$$\det(\boldsymbol{A} - \lambda \boldsymbol{E}) = 0.$$

Эти корни называют *собственными значениями* матрицы $\boldsymbol{A}$, а задачу их нахождения — *проблемой собственных значений*. Каждому λ_i соответствует *собственный вектор* $\boldsymbol{e}_i$ (иначе называемый *главной осью*), для которого $\boldsymbol{Ae}_i = \lambda_i \boldsymbol{e}_i$ и $\boldsymbol{e}_i^T \boldsymbol{e}_i = 1$. Если $\lambda_i \neq \lambda_j$, то $\boldsymbol{e}_i$ и $\boldsymbol{e}_j$ *ортогональны*, т. е. $\boldsymbol{e}_i^T \boldsymbol{e}_j = 0$. Когда все λ_i различны, собственные оси $\boldsymbol{e}_i$ определяются однозначно с точностью до выбора направления (одновременной смены знака всех компонент вектора). В этом случае в качестве $\boldsymbol{e}_i$ можно взять столбцы матрицы $\boldsymbol{C}$.

Для симметричной матрицы $\boldsymbol{A}$ имеют место очевидные равенства

$$\operatorname{tr} \boldsymbol{A} = \sum_{i=1}^{n} \lambda_i, \quad \text{и} \quad \det \boldsymbol{A} = \prod_{i=1}^{n} \lambda_i.$$

Определение. *Симметричная матрица $\boldsymbol{A}$ называется неотрицательно определенной, если* $\boldsymbol{x}^T \boldsymbol{Ax} \geqslant 0$ *для любого вектора* $\boldsymbol{x} \in \mathbb{R}^n$.

Замечание. Ковариационная матрица $\boldsymbol{\Sigma}$ произвольного случайного вектора $\boldsymbol{\xi} = (\xi_1, \dots, \xi_n)$ обладает этим свойством. Действительно,

$$0 \leqslant \mathsf{M}\left[\sum_{i=1}^{n} x_i(\xi_i - \mathsf{M}\xi_i)\right]^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i\, x_j\, \mathsf{cov}\,(\xi_i, \xi_j) = \boldsymbol{x}^T \boldsymbol{\Sigma x}.$$

Неотрицательная определенность $\boldsymbol{A}$ равносильна неотрицательности всех собственных значений λ_i, $i = 1, \dots n$.

Для неотрицательно определенной матрицы $\boldsymbol{A}$ число положительных λ_i совпадает с рангом $\boldsymbol{A}$. Если ранг равен n, то матрица называется *положительно определенной.*

Положительная определенность $\boldsymbol{A}$ равносильна положительности всех *главных миноров*

$$D_1 = a_{11},\ D_2 = \det\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \ldots,\ D_n = \det\begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \ldots & \ldots & \ldots \\ a_{n1} & \ldots & a_{nn} \end{pmatrix}.$$

Теорема Холецкого (метод квадратных корней). *Для положительно определенной матрицы* $\boldsymbol{A}$ *возможно представление*

А.-Л. Холецкий (1875–1918), французский геодезист.

$$\boldsymbol{A} = \boldsymbol{U}^T\boldsymbol{U},$$

где $\boldsymbol{U}$ *— верхнетреугольная матрица, элементы которой последовательно вычисляются по формулам*

$$u_{11}^2 = a_{11},\ u_{1j} = a_{1j}/u_{11},\ j = 2,3,\ldots,n;\ \ldots;$$

$$u_{ii}^2 = a_{ii} - \sum_{k=1}^{i-1} u_{ki}^2,\ u_{ij} = \frac{1}{u_{ii}}\left(a_{ij} - \sum_{k=1}^{i-1} u_{ki}u_{kj}\right),\ j = i+1,\ldots,n;\ \ldots;$$

$$u_{nn}^2 = a_{nn} - \sum_{k=1}^{n-1} u_{kn}^2.$$

Доказательство приведено в [6, с. 159].

Для вычисления u_{ii} приходится извлекать квадратные корни. Это объясняет название метода.

Метод интересен тем, что позволяет решать систему линейных уравнений $\boldsymbol{Ax} = \boldsymbol{b}$ с положительно определенной матрицей $\boldsymbol{A}$ примерно вдвое быстрее метода Гаусса (после представления $\boldsymbol{A}$ в виде $\boldsymbol{U}^T\boldsymbol{U}$ требуется всего лишь провести дважды «обратную прогонку»: $\boldsymbol{U}^T\boldsymbol{y} = \boldsymbol{b}$ и $\boldsymbol{Ux} = \boldsymbol{y}$). При этом метод Холецкого численно устойчив.

Отметим, что в классе невырожденных квадратных матриц $\boldsymbol{X}$ уравнение $\boldsymbol{X}^T\boldsymbol{X} = \boldsymbol{A}$ имеет много решений: подходит любая матрица $\boldsymbol{X} = \boldsymbol{CU}$, где $\boldsymbol{C}$ — произвольная ортогональная матрица. Действительно,

$$\boldsymbol{X}^T\boldsymbol{X} = (\boldsymbol{CU})^T\boldsymbol{CU} = (\boldsymbol{U}^T\boldsymbol{C}^T)\,\boldsymbol{CU} = \boldsymbol{U}^T(\boldsymbol{C}^T\boldsymbol{C})\,\boldsymbol{U} = \boldsymbol{A}.$$

Среди этих решений только одна матрица *симметрична.* Она называется *квадратным корнем* из $\boldsymbol{A}$ и обозначается $\boldsymbol{A}^{1/2}$. Ее можно явно указать на основе теоремы о приведении $\boldsymbol{A}$ к главным осям: если $\boldsymbol{C}^T\boldsymbol{AC} = \boldsymbol{\Lambda}$, то $\boldsymbol{A}^{1/2} = \boldsymbol{C\Lambda}^{1/2}\boldsymbol{C}^T$. Матрица $\boldsymbol{A}^{1/2}$ имеет наибольший след среди всех решений $\boldsymbol{X}$ (см. [86, с. 22]).

Обобщенная проблема собственных значений. *Для неотрицательно определенной матрицы* $\boldsymbol{A}$ *и положительно определенной* $\boldsymbol{B}$ *найдется такая невырожденная матрица* $\boldsymbol{D}$*, что*

$$\boldsymbol{D}^T\boldsymbol{AD} = \boldsymbol{\Lambda} \quad \textit{и} \quad \boldsymbol{D}^T\boldsymbol{BD} = \boldsymbol{E}.$$

Здесь $\boldsymbol{\Lambda}$ *— диагональная матрица с неотрицательными элементами* $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_n$ *на главной диагонали, которые могут быть получены как решения характеристического уравнения*

$$\det(\boldsymbol{A} - \lambda\boldsymbol{B}) = 0 \quad \textit{или} \quad \det(\boldsymbol{AB}^{-1} - \lambda\boldsymbol{E}) = 0.$$

Числа λ_i называются собственными значениями матриц $\boldsymbol{A}$ и $\boldsymbol{B}$.

Доказательство. Отметим, что матрица $\boldsymbol{AB}^{-1}$ не обязательно является симметричной. Однако $\boldsymbol{S} = \boldsymbol{B}^{-1/2}\boldsymbol{AB}^{-1/2}$, очевидно, симметрична (здесь $\boldsymbol{B}^{-1/2} \equiv (\boldsymbol{B}^{1/2})^{-1}$). Она также неотрицательно определена:

$$\boldsymbol{x}^T\boldsymbol{S}\boldsymbol{x} = \left(\boldsymbol{x}^T\boldsymbol{B}^{-1/2}\boldsymbol{A}^{1/2}\right)\left(\boldsymbol{A}^{1/2}\boldsymbol{B}^{-1/2}\boldsymbol{x}\right) = \left|\boldsymbol{A}^{1/2}\boldsymbol{B}^{-1/2}\boldsymbol{x}\right|^2 \geqslant 0.$$

При этом характеристическое уравнение

$$\det\left[\boldsymbol{B}^{1/2}\left(\boldsymbol{B}^{-1/2}\boldsymbol{AB}^{-1/2} - \lambda\boldsymbol{E}\right)\boldsymbol{B}^{1/2}\right] = 0 \iff \det\left(\boldsymbol{S} - \lambda\boldsymbol{E}\right) = 0.$$

Обозначим через $\boldsymbol{C}$ ортогональную матрицу преобразования, приводящего $\boldsymbol{S}$ к главным осям: $\boldsymbol{C}^T\boldsymbol{SC} = \boldsymbol{\Lambda}$. Тогда $\boldsymbol{D} = \boldsymbol{B}^{-1/2}\boldsymbol{C}$. Действительно,

$$\boldsymbol{D}^T\boldsymbol{AD} = \boldsymbol{C}^T\boldsymbol{B}^{-1/2}\boldsymbol{AB}^{-1/2}\boldsymbol{C} = \boldsymbol{C}^T\boldsymbol{SC} = \boldsymbol{\Lambda},$$
$$\boldsymbol{D}^T\boldsymbol{BD} = \boldsymbol{C}^T\boldsymbol{B}^{-1/2}\boldsymbol{B}^{1/2}\boldsymbol{B}^{1/2}\boldsymbol{B}^{-1/2}\boldsymbol{C} = \boldsymbol{E}.$$

Из доказанной теоремы легко вывести, что выполняются равенства

$$\operatorname{tr}(\boldsymbol{AB}^{-1}) = \sum_{i=1}^{n}\lambda_i \quad \text{и} \quad \det\left(\boldsymbol{AB}^{-1}\right) = \prod_{i=1}^{n}\lambda_i. \qquad \blacksquare$$

ТАБЛИЦЫ

Таблица 1

Случайные числа*)

Равномерно распределенные

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	10	09	73	25	33	76	52	01	35	86	34	67	35	48	76
2	37	54	20	48	05	64	89	47	42	96	24	80	52	40	37
3	08	42	26	89	53	19	64	50	93	03	23	20	90	25	60
4	99	01	90	25	29	09	37	67	07	15	38	31	13	11	65
5	12	80	79	99	70	80	15	73	61	47	64	03	23	66	53
6	66	06	57	47	17	34	07	27	68	50	36	69	73	61	70
7	31	06	01	08	05	45	57	18	24	06	35	30	34	26	14
8	85	26	97	76	02	02	05	16	56	92	68	66	57	48	18
9	63	57	33	21	35	05	32	54	70	48	90	55	35	75	48
10	73	79	64	57	53	03	52	96	47	78	35	80	83	42	82
11	98	52	01	77	67	14	90	56	86	07	22	10	94	05	58
12	11	80	50	54	31	39	80	82	77	32	50	72	56	82	48
13	83	45	29	96	34	06	28	89	80	83	13	74	67	00	78
14	88	68	54	02	00	86	50	75	84	01	36	76	66	79	51
15	99	59	46	73	48	87	51	76	49	69	91	82	60	89	28
16	65	48	11	76	74	17	46	85	09	50	58	04	77	69	74
17	80	12	43	56	35	17	72	70	80	15	45	31	82	23	74
18	74	35	09	98	17	77	40	27	72	14	43	23	60	02	10
19	69	91	62	68	03	66	25	22	91	48	36	93	68	72	03
20	09	89	32	05	05	14	22	56	85	14	46	42	75	67	88

Нормально распределенные

1	2	3	4	5	6	7	8	9	10
−1,75	−0,33	−1,26	0,32	1,53	0,35	−0,96	−0,06	0,42	−1,08
−0,29	0,09	1,70	−1,09	−0,44	−0,29	0,25	−0,54	−1,38	0,32
−0,93	0,13	0,63	0,90	1,41	−0,88	−0,10	0,23	0,13	0,37
−0,45	−0,24	0,07	1,03	1,73	−0,06	−1,49	−0,08	−2,36	−0,99
0,51	−0,88	0,49	−1,30	−0,27	0,76	−0,36	0,19	−1,08	0,53

*) Фрагменты таблиц из [10, с. 366, 372].

Таблица 2

Вероятности верхнего «хвоста» стандартного нормального распределения

$$y = 1 - \Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{x}^{+\infty} e^{-u^2/2}\, du$$

x	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,500	0,496	0,492	0,488	0,484	0,480	0,476	0,472	0,468	0,464
0,1	0,460	0,456	0,452	0,448	0,444	0,440	0,436	0,433	0,429	0,425
0,2	0,421	0,417	0,413	0,409	0,405	0,401	0,397	0,394	0,390	0,386
0,3	0,382	0,378	0,375	0,371	0,367	0,363	0,359	0,356	0,352	0,348
0,4	0,344	0,341	0,337	0,334	0,330	0,326	0,323	0,319	0,316	0,312
0,5	0,309	0,305	0,302	0,298	0,295	0,291	0,288	0,284	0,281	0,278
0,6	0,274	0,271	0,268	0,264	0,261	0,258	0,255	0,251	0,248	0,245
0,7	0,242	0,239	0,236	0,233	0,230	0,227	0,224	0,221	0,218	0,215
0,8	0,212	0,209	0,206	0,203	0,201	0,198	0,195	0,192	0,189	0,187
0,9	0,184	0,181	0,179	0,176	0,174	0,171	0,169	0,166	0,164	0,161
1,0	0,159	0,156	0,154	0,152	0,149	0,147	0,145	0,142	0,140	0,138
1,1	0,136	0,134	0,131	0,129	0,127	0,125	0,123	0,121	0,119	0,117
1,2	0,115	0,113	0,111	0,109	0,108	0,106	0,104	0,102	0,100	0,099
1,3	0,097	0,095	0,093	0,092	0,090	0,089	0,087	0,085	0,084	0,082
1,4	0,081	0,079	0,078	0,077	0,075	0,074	0,072	0,071	0,069	0,068
1,5	0,067	0,066	0,064	0,063	0,062	0,061	0,059	0,058	0,057	0,056
1,6	0,055	0,054	0,053	0,052	0,051	0,050	0,049	0,048	0,047	0,046
1,7	0,045	0,044	0,043	0,042	0,041	0,040	0,039	0,038	0,038	0,037
1,8	0,036	0,035	0,034	0,034	0,033	0,032	0,031	0,031	0,030	0,029
1,9	0,029	0,028	0,027	0,027	0,026	0,026	0,025	0,024	0,024	0,023
2,0	0,023	0,022	0,022	0,021	0,021	0,020	0,020	0,019	0,019	0,018
2,1	0,018	0,017	0,017	0,017	0,016	0,016	0,015	0,015	0,015	0,014
2,2	0,014	0,014	0,013	0,013	0,013	0,012	0,012	0,012	0,011	0,011
2,3	0,011	0,010	0,010	0,010	0,010	0,009	0,009	0,009	0,009	0,008
2,4	0,008	0,008	0,008	0,008	0,007	0,007	0,007	0,007	0,007	0,006
2,5	0,006	0,006	0,006	0,006	0,006	0,005	0,005	0,005	0,005	0,005
2,6	0,005	0,005	0,004	0,004	0,004	0,004	0,004	0,004	0,004	0,004
2,7	0,004	0,003	0,003	0,003	0,003	0,003	0,003	0,003	0,003	0,003
2,8	0,003	0,003	0,002	0,002	0,002	0,002	0,002	0,002	0,002	0,002
2,9	0,002	0,002	0,002	0,002	0,002	0,002	0,002	0,002	0,001	0,001
3,0	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001

Примечание. В силу симметрии распределения $\Phi(-x) = 1 - \Phi(x)$.

Таблица 3

Распределение хи-квадрат с k степенями свободы

$$p = \frac{1}{2^{k/2}\Gamma(k/2)} \int\limits_0^{z_p} z^{\frac{k}{2}-1} e^{-\frac{z}{2}}\, dz$$

k \ p	0,001	0,025	0,05	0,1	0,5	0,9	0,95	0,975	0,99	0,999
1	0,000	0,001	0,004	0,016	0,46	2,71	3,84	5,02	6,64	10,8
2	0,002	0,05	0,10	0,21	1,39	4,61	5,99	7,38	9,21	13,8
3	0,024	0,22	0,35	0,58	2,37	6,25	7,82	9,35	11,3	16,3
4	0,09	0,48	0,71	1,06	3,36	7,78	9,49	11,1	13,3	18,5
5	0,21	0,83	1,15	1,61	4,35	9,24	11,1	12,8	15,1	20,5
6	0,38	1,24	1,64	2,2	5,35	10,6	12,6	14,4	16,8	22,4
7	0,60	1,69	2,17	2,83	6,35	12,0	14,1	16,0	18,5	24,3
8	0,86	2,18	2,73	3,49	7,34	13,4	15,5	17,5	20,1	26,1
9	1,15	2,70	3,33	4,17	8,34	14,7	16,9	19,0	21,7	27,9
10	1,48	3,25	3,94	4,87	9,34	16,0	18,3	20,5	23,2	29,6
11	1,83	3,82	4,58	5,58	10,3	17,3	19,7	21,9	24,7	31,3
12	2,21	4,40	5,23	6,30	11,3	18,5	21,0	23,3	26,2	32,9
13	2,62	5,01	5,89	7,04	12,3	19,8	22,4	24,7	27,7	34,5
14	3,04	5,63	6,57	7,79	13,3	21,1	23,7	26,1	29,1	36,1
15	3,48	6,26	7,26	8,55	14,3	22,3	25,0	27,5	30,6	37,7
16	3,94	6,91	7,96	9,31	15,3	23,5	26,3	28,8	32,0	39,3
17	4,42	7,56	8,67	10,1	16,3	24,8	27,6	30,2	33,4	40,8
18	4,91	8,23	9,39	10,9	17,3	26,0	28,9	31,5	34,8	42,3
19	5,41	8,91	10,1	11,7	18,3	27,2	30,1	32,9	36,2	43,8
20	5,92	9,59	10,9	12,4	19,3	28,4	31,4	34,2	37,6	45,3
21	6,45	10,3	11,6	13,2	20,3	29,6	32,7	35,5	38,9	46,8
22	6,98	11,0	12,3	14,0	21,3	30,8	33,9	36,8	40,3	48,3
23	7,53	11,7	13,1	14,8	22,3	32,0	35,2	38,1	41,6	49,7
24	8,09	12,4	13,8	15,7	23,3	33,2	36,4	39,4	43,0	51,2
25	8,65	13,1	14,6	16,5	24,3	34,4	37,7	40,6	44,3	52,6
26	9,22	13,8	15,4	17,3	25,3	35,6	38,9	41,9	45,6	54,1
27	9,80	14,6	16,2	18,1	26,3	36,7	40,1	43,2	47,0	55,5
28	10,4	15,3	16,9	18,9	27,3	37,9	41,3	44,5	48,3	56,9
29	11,0	16,0	17,7	19,8	28,3	39,1	42,6	45,7	49,6	58,3
30	11,6	16,8	18,5	20,6	29,3	40,3	43,8	47,0	50,9	59,7

Таблица 4

Распределение Стьюдента с k степенями свободы

$$p = \frac{\Gamma((k+1)/2)}{\Gamma(k/2)\sqrt{\pi k}} \int\limits_{-\infty}^{y_p} \left(1 + \frac{y^2}{k}\right)^{-\frac{k+1}{2}} dy$$

k \ p	0,9	0,95	0,975	0,99	0,995	0,999
1	3,078	6,314	12,71	31,82	63,66	318,3
2	1,886	2,920	4,303	6,965	9,925	22,33
3	1,638	2,353	3,182	4,541	5,841	10,21
4	1,533	2,132	2,776	3,747	4,604	7,173
5	1,476	2,015	2,571	3,365	4,032	5,893
6	1,440	1,943	2,447	3,143	3,707	5,208
7	1,415	1,895	2,365	2,998	3,500	4,785
8	1,397	1,860	2,306	2,897	3,355	4,501
9	1,383	1,833	2,262	2,821	3,250	4,297
10	1,372	1,813	2,228	2,764	3,169	4,144
12	1,356	1,782	2,179	2,681	3,055	3,930
14	1,345	1,761	2,145	2,625	2,977	3,787
16	1,337	1,750	2,120	2,584	2,921	3,686
18	1,330	1,734	2,101	2,552	2,878	3,611
20	1,325	1,725	2,086	2,528	2,845	3,552
22	1,321	1,717	2,074	2,508	2,819	3,505
24	1,318	1,711	2,064	2,492	2,797	3,467
26	1,315	1,706	2,056	2,479	2,779	3,435
28	1,313	1,701	2,048	2,467	2,763	3,408
30	1,310	1,697	2,042	2,457	2,750	3,385
32	1,309	1,694	2,037	2,449	2,739	3,365
34	1,307	1,691	2,032	2,441	2,728	3,348
36	1,306	1,688	2,028	2,435	2,720	3,333
38	1,304	1,686	2,024	2,429	2,712	3,319
40	1,303	1,684	2,021	2,423	2,705	3,307
60	1,296	1,671	2,000	2,390	2,660	3,232
∞	1,282	1,645	1,960	2,326	2,576	3,090

Примечание. В силу симметрии распределения $y_{1-p} = -y_p$.

Таблица 5

Распределение Фишера—Снедекора с k_1 и k_2 степенями свободы

$$p = \frac{\Gamma((k_1+k_2)/2)}{\Gamma(k_1/2)\Gamma(k_2/2)} \left(\frac{k_1}{k_2}\right)^{k_1/2} \int_0^{v_p} v^{\frac{k_1}{2}-1} \left(1+\frac{k_1}{k_2}\,v\right)^{-(k_1+k_2)/2} dv$$

Для каждой пары (k_2,k_1) в таблице приведены значения трех квантилей v_p: при $p = 0{,}9$ (верхнее), $p = 0{,}95$ (среднее) и $p = 0{,}99$ (нижнее). Левые границы доверительных интервалов находятся из условия $v_{1-p}(k_1,k_2) = 1/v_p(k_2,k_1)$. Интерполяцию рекомендуется проводить по аргументам $1/k_1$ и $1/k_2$. (Более подробные таблицы см. в [8], [10].)

k_2 \ k_1	1	2	3	4	5	6	7	10	20	40	∞
	39,9	49,5	53,6	55,8	57,2	58,2	58,9	60,2	61,7	62,5	63,3
1	161	200	216	225	230	234	237	242	248	251	254
	4052	5000	5403	5625	5764	5859	5928	6056	6209	6287	6366
	8,53	9,00	9,16	9,24	9,29	9,33	9,35	9,39	9,44	9,47	9,49
2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,5	19,5
	98,5	99,0	99,2	99,2	99,3	99,3	99,4	99,4	99,4	99,5	99,5
	5,54	5,46	5,39	5,34	5,31	5,28	5,27	5,23	5,18	5,16	5,13
3	10,1	9,55	9,28	9,12	9,01	8,94	8,88	8,79	8,66	8,59	8,53
	34,1	30,8	29,5	28,7	28,2	27,9	27,7	27,2	26,7	26,4	26,1
	4,54	4,32	4,19	4,11	4,05	4,01	3,98	3,92	3,84	3,80	3,76
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	5,96	5,80	5,72	5,63
	21,2	18,0	16,7	16,0	15,5	15,2	15,0	14,5	14,0	13,7	13,5
	4,06	3,78	3,62	3,52	3,45	3,40	3,37	3,30	3,21	3,16	3,11
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,74	4,56	4,46	4,37
	16,3	13,3	12,1	11,4	11,0	10,7	10,5	10,1	9,55	9,29	9,02
	3,78	3,46	3,29	3,18	3,11	3,05	3,01	2,94	2,84	2,78	2,72
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,06	3,87	3,77	3,67
	13,7	10,9	9,78	9,15	8,75	8,47	8,26	7,87	7,40	7,14	6,88
	3,59	3,26	3,07	2,96	2,88	2,83	2,78	2,70	2,59	2,54	2,47
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,64	3,44	3,34	3,23
	12,2	9,55	8,45	7,85	7,46	7,19	6,99	6,62	6,16	5,91	5,65
	3,46	3,11	2,92	2,81	2,73	2,67	2,62	2,54	2,42	2,36	2,29
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,35	3,15	3,04	2,93
	11,3	8,65	7,59	7,01	6,63	6,37	6,18	5,81	5,36	5,12	4,86
	3,36	3,01	2,81	2,69	2,61	2,55	2,51	2,42	2,30	2,23	2,16
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,14	2,94	2,83	2,71
	10,6	8,02	6,99	6,42	6,06	5,80	5,61	5,26	4,81	4,57	4,31
	3,29	2,92	2,73	2,61	2,52	2,46	2,41	2,32	2,20	2,13	2,06
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	2,98	2,77	2,66	2,54
	10,0	7,56	6,55	5,99	5,64	5,39	5,20	4,85	4,41	4,17	3,91
	3,23	2,86	2,66	2,54	2,45	2,39	2,34	2,25	2,12	2,05	1,97
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,85	2,65	2,53	2,40
	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,54	4,10	3,86	3,60
	3,18	2,81	2,61	2,48	2,39	2,33	2,28	2,19	2,06	1,99	1,90
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,75	2,54	2,43	2,30
	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,30	3,86	3,62	3,36

k_2 \ k_1	1	2	3	4	5	6	7	10	20	40	∞
	3,14	2,76	2,56	2,43	2,35	2,28	2,23	2,14	2,01	1,93	1,85
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,67	2,46	2,34	2,21
	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,10	3,66	3,43	3,17
	3,10	2,73	2,52	2,39	2,31	2,24	2,19	2,10	1,96	1,89	1,80
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,60	2,39	2,27	2,13
	8,86	6,51	5,56	5,04	4,70	4,46	4,28	3,94	3,51	3,27	3,00
	3,07	2,70	2,49	2,36	2,27	2,21	2,16	2,06	1,92	1,85	1,76
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,54	2,33	2,20	2,07
	8,68	6,36	5,42	4,89	4,56	4,32	4,14	3,80	3,37	3,13	2,87
	2,97	2,59	2,38	2,25	2,16	2,09	2,04	1,94	1,79	1,71	1,61
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,35	2,12	1,99	1,84
	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,37	2,94	2,69	2,42
	2,92	2,53	2,32	2,18	2,09	2,02	1,97	1,87	1,72	1,63	1,52
25	4,24	3,38	2,99	2,76	2,60	2,49	2,40	2,24	2,01	1,87	1,71
	7,77	5,57	4,68	4,18	3,86	3,63	3,46	3,13	2,70	2,45	2,17
	2,88	2,49	2,28	2,14	2,05	1,98	1,93	1,82	1,67	1,57	1,46
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,16	1,93	1,79	1,62
	7,56	5,39	4,51	4,02	3,70	3,47	3,30	2,98	2,55	2,30	2,01
	2,84	2,44	2,23	2,09	2,00	1,93	1,87	1,76	1,61	1,51	1,38
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,08	1,84	1,69	1,51
	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,80	2,37	2,11	1,80
	2,79	2,39	2,18	2,04	1,95	1,87	1,82	1,71	1,54	1,44	1,29
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	1,99	1,75	1,59	1,39
	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,63	2,20	1,94	1,60
	2,75	2,35	2,13	1,99	1,90	1,82	1,77	1,65	1,48	1,37	1,19
120	3,92	3,07	2,68	2,45	2,29	2,18	2,09	1,91	1,66	1,50	1,25
	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,47	2,03	1,76	1,38
	2,71	2,30	2,08	1,94	1,85	1,77	1,72	1,60	1,42	1,30	1,00
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,83	1,57	1,39	1,00
	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,32	1,88	1,59	1,00

Таблица 7

Односторонние критические значения коэффициента Спирмена ρ_S
$p = \mathsf{P}(\rho_S \geqslant r_p)$, k — размер выборки

k \ p	0,1	0,05	0,025	0,01	0,005	0,001
4	1,000					
5	0,800	0,900	1,000			
6	0,657	0,829	0,886	0,943	1,000	
7	0,571	0,714	0,786	0,893	0,929	1,000
8	0,524	0,643	0,738	0,833	0,881	0,952
9	0,483	0,600	0,700	0,783	0,833	0,917
10	0,455	0,564	0,648	0,745	0,794	0,879
11	0,427	0,536	0,618	0,709	0,755	0,845
12	0,406	0,503	0,587	0,671	0,727	0,825
13	0,385	0,484	0,560	0,648	0,703	0,802
14	0,367	0,464	0,538	0,622	0,675	0,776
15	0,354	0,443	0,521	0,604	0,654	0,754
16	0,341	0,429	0,503	0,582	0,635	0,732
17	0,328	0,414	0,485	0,566	0,615	0,713
18	0,317	0,401	0,472	0,550	0,600	0,695
19	0,309	0,391	0,460	0,535	0,584	0,677
20	0,299	0,380	0,447	0,520	0,570	0,662
21	0,292	0,370	0,435	0,508	0,556	0,648
22	0,284	0,361	0,425	0,496	0,544	0,634
23	0,278	0,353	0,415	0,486	0,532	0,622
24	0,271	0,344	0,406	0,475	0,521	0,610
25	0,265	0,337	0,398	0,466	0,511	0,598
26	0,259	0,331	0,390	0,457	0,501	0,587
27	0,255	0,324	0,382	0,448	0,491	0,577
28	0,250	0,317	0,375	0,440	0,483	0,567
29	0,245	0,312	0,368	0,433	0,475	0,558
30	0,240	0,306	0,362	0,425	0,467	0,549
35	0,222	0,283	0,335	0,394	0,433	0,510
40	0,207	0,264	0,313	0,368	0,405	0,479
45	0,194	0,248	0,294	0,347	0,382	0,453
50	0,184	0,235	0,279	0,329	0,363	0,430
100	0,129	0,165	0,197	0,233	0,257	0,307

Примечание. В силу симметрии распределения $\mathsf{P}(|\rho_S| \geqslant r_p) = 2p$.

ЛИТЕРАТУРА

Классика — это то, что все хотели бы прочитать, но никто читать не хочет.

Марк Твен

1. *Айвазян С. А., Бухштабер В. М., Енюков И. С., Мешалкин Л. Д.* Прикладная статистика: Классификация и снижение размерности. — М.: Финансы и статистика, 1989
2. *Айвазян С. А., Енюков И. С., Мешалкин Л. Д.* Прикладная статистика: Исследование зависимостей. — М.: Финансы и статистика, 1985
3. *Айвазян С. А.* Статистическое исследование зависимостей. — М.: Металлургия, 1968
4. *Александров В. В., Алексеев А. И., Горский Н. Д.* Анализ данных на ЭВМ. — М.: Финансы и статистика, 1990
5. *Амелькин В. В.* Дифференциальные уравнения в приложениях. — М.: Наука, 1987
6. *Амосов А. А., Дубинский Ю. А., Копченова Н. В.* Вычислительные методы для инженеров. — М.: Высшая школа, 1994
7. *Аптон Г.* Анализ таблиц сопряженности. — М.: Финансы и статистика, 1982
8. *Аренс Х., Лейтер Ю.* Многомерный дисперсионный анализ. — М.: Финансы и статистика, 1985
9. *Березовский Б. А., Гнедин А. В.* Задача наилучшего выбора. — М.: Наука, 1984
10. *Большев Л. Н., Смирнов Н. В.* Таблицы математической статистики. — М.: Наука, 1983
11. *Боровков А. А.* Математическая статистика. — М.: Наука, 1984
12. *Боровков А. А.* Теория вероятностей. — М.: Наука, 1986
13. *Ван дер Варден Б. Л.* Математическая статистика. — М.: Издательство Иностранной литературы, 1960
14. *Вихман Э.* Квантовая физика. — М.: Наука, 1986
15. *Воинов В. Г., Никулин М. С.* Несмещенные оценки и их применения. — М.: Наука, 1989
16. *Волков Е. А.* Численные методы. — М.: Наука, 1987
17. *Галамбош Я.* Асимптотическая теория экстремальных порядковых статистик. — М.: Наука, 1984
18. *Гилл Ф., Мюррей У., Райт М.* Практическая оптимизация. — М.: Мир, 1985
19. *Гнеденко Б. В.* Курс теории вероятностей. — М.: Наука, 1988
20. *Двайт Г. Б.* Таблицы интегралов и другие математические формулы. — М.: Наука, 1983

21. *Деврой Л., Дьерфи Л.* Непараметрическое оценивание плотности. — М.: Мир, 1988
22. *Демидович Б. П.* Сборник задач и упражнений по математическому анализу. — М.: Наука, 1990
23. *Дрейпер Н., Смит Г.* Прикладной регрессионный анализ: В 2-х кн. Кн. 1. — М.: Финансы и статистика, 1986
24. *Дынкин Е. Б., Юшкевич А. А.* Теоремы и задачи о процессах Маркова. — М.: Наука, 1967
25. *Дэйвисон М.* Многомерное шкалирование. — М.: Финансы и статистика, 1988
26. *Дюран Н., Оделл П.* Кластерный анализ. — М.: Статистика, 1977
27. *Езекиэл М., Фокс К. А.* Методы анализа корреляций и регрессий. — М.: Статистика, 1966
28. *Емеличев В. А., Мельников О. И., Сарванов В. И., Тышкевич Р. И.* Лекции по теории графов. — М.: Наука, 1990
29. *Ермаков С. М., Михайлов Г. А.* Статистическое моделирование. — М.: Наука, 1982
30. *Жамбю М.* Иерархический кластер-анализ и соответствия. — М.: Финансы и статистика, 1988
31. *Журбенко И. Г.* Анализ стационарных и однородных случайных систем. — М.: Издательство МГУ, 1987
32. *Ивченко Г. И., Медведев Ю. И.* Математическая статистика. — М.: Высш. шк., 1984
33. *Карлин С.* Основы теории случайных процессов. — М.: Мир, 1971
34. *Картер А.* Структурные изменения в экономике США. — М.: Статистика, 1974
35. *Кендалл М. Дж., Стьюарт А.* Статистические выводы и связи. — М.: Наука, 1973
36. *Кендэл М.* Ранговые корреляции. — М.: Статистика, 1975
37. *Кимбл Г.* Как правильно пользоваться статистикой. — М.: Финансы и статистика, 1982
38. *Козлов М. В., Прохоров А. В.* Введение в математическую статистику. — М.: Издательство МГУ, 1987
39. *Козлов М. В.* Элементы теории вероятностей в примерах и задачах. — М.: Издательство МГУ, 1990
40. *Кокс Д., Льюис П.* Статистический анализ последовательностей событий. — М.: Мир, 1969
41. *Колмогоров А. Н., Фомин С. В.* Элементы теории функций и функционального анализа. — М.: Наука, 1989
42. *Корн Г., Корн Т.* Справочник по математике для научных работников и инженеров. — М.: Наука, 1984
43. *Кострикин А. И., Манин Ю. И.* Линейная алгебра и геометрия. — М.: Наука, 1986
44. *Крамер Г.* Математические методы статистики. — М., 1975
45. *Кудрявцев Л. Д.* Курс математического анализа, Т. 1. — М.: Высш. шк., 1988
46. *Кудрявцев Л. Д.* Курс математического анализа, Т. 2. — М.: Высш. шк., 1988
47. *Кульбак С.* Теория информации и статистика. — М.: Наука, 1967
48. *Ламперти Дж.* Вероятность. — М.: Наука, 1973

49. *Ланкастер П.* Теория матриц. — М.: Наука, 1982
50. *Леман Э.* Теория точечного оценивания. — М.: Наука, 1991
51. *Литлвуд Дж.* Математическая смесь. — М.: Наука, 1978
52. *Мандель И. Д.* Кластерный анализ. — М.: Финансы и статистика, 1988
53. *Математика и САПР, Кн. 1.* — М.: Мир, 1988
54. *Миллс Ф.* Статистические методы. — М.: Госстатиздат, 1958
55. *Мину М.* Математическое программирование. — М.: Наука, 1990
56. *Миркин Б. Г.* Анализ качественных признаков и структур. — М.: Статистика, 1980
57. *Мостеллер Ф., Тьюки Дж.* Анализ данных и регрессия. — М.: Финансы и статистика, 1982
58. *Мэйндоналд Дж.* Вычислительные алгоритмы в прикладной статистике. — М.: Финансы и статистика, 1988
59. *Никольский С. М.* Курс математического анализа, Т. 2. — М.: Наука, 1983
60. *Орлов А. И.* Некоторые вероятностные вопросы теории классификации. Прикладная статистика. — М.: Наука, 1983. — с. 166–179
61. *Питмен Э.* Основы теории статистических выводов. — М.: Мир, 1986
62. *Пойа Д.* Математическое открытие. — М.: Наука, 1976
63. *Практикум по математической статистике / Под редакцией Н. С. Бахвалова, Ю. А. Розанова, И. Г. Журбенко, А. В. Михалева.* — М.: Издательство МГУ, 1987
64. *Прасолов В. В.* Задачи по планиметрии, Ч. 2. — М.: Наука, 1991
65. *Прохоров Ю. В., Розанов Ю. А.* Теория вероятностей. — М.: Наука, 1987
66. *Рей Г.* Звезды. Новые очертания старых созвездий. — М.: Мир, 1969
67. *Рейф Ф.* Статистическая физика. — М.: Наука, 1986
68. *Розанов Ю. А.* Введение в теорию случайных процессов. — М.: Наука, 1982
69. *Розанов Ю. А.* Теория вероятностей, случайные процессы и математическая статистика. — М.: Наука, 1985
70. *Сархан А., Гринберг Б.* Введение в теорию порядковых статистик. — М.: Статистика, 1970
71. *Себер Дж.* Линейный регрессионный анализ. — М.: Мир, 1980
72. *Секей Г.* Парадоксы в теории вероятностей и математической статистике. — М.: Мир, 1990
73. *Сидоров Ю. В., Федорюк М. В., Шабунин М. И.* Лекции по теории функций комплексного переменного. — М.: Наука, 1982
74. *Соболь И. М., Статников Р. Б.* Выбор оптимальных параметров в задачах со многими критериями. — М.: Наука, 1981
75. *Справочник по прикладной статистике, Т. 1 / Под редакцией Э. Ллойда, У. Ледермана.* — М.: Финансы и статистика, 1989
76. *Справочник по прикладной статистике, Т. 2 / Под редакцией Э. Ллойда, У. Ледермана.* — М.: Финансы и статистика, 1989
77. *Сухарев А. Г., Тимохов А. В., Федоров В. В.* Курс методов оптимизации. — М.: Наука, 1986
78. *Терехина А. Ю.* Анализ данных методами многомерного шкалирования. — М.: Наука, 1986

79. *Тутубалин В. Н.* Теория вероятностей и случайных процессов. — М.: Издательство МГУ, 1992
80. *Тюрин Ю. Н., Макаров А. А.* Анализ данных на компьютере. — М.: ИНФРА-М, Финансы и статистика, 1995
81. *Феллер В.* Введение в теорию вероятностей и ее приложения, Т. 1. — М.: Мир, 1984
82. *Феллер В.* Введение в теорию вероятностей и ее приложения, Т. 2. — М.: Мир, 1984
83. *Хальд А.* Математическая статистика с техническими приложениями. — М.: ИЛ, 1956
84. *Хампель Ф., Рончетти Э., Рауссеу П., Штаэль В.* Робастность в статистике. Подход на основе функций влияния. — М.: Мир, 1989
85. *Хардле В.* Прикладная непараметрическая регрессия. — М.: Мир, 1993
86. *Хеттманспергер Т.* Статистические выводы, основанные на рангах. — М.: Финансы и статистика, 1987
87. *Хида Т.* Броуновское движение. — М.: Наука, 1987
88. *Холлендер М., Вулф Д.* Непараметрические методы статистики. — М.: Финансы и статистика, 1983
89. *Хъюбер П.* Робастность в статистике. — М.: Мир, 1984
90. *Ширяев А. Н.* Вероятность. — М.: Наука, 1989
91. *Яблонский С. В.* Введение в дискретную математику. — М.: Наука, 1986
92. *Яглом А. М., Яглом И. М.* Вероятность и информация. — М.: Физматгиз, 1960
93. *Cochran W. G., Cox G. M.* Experimental Designs, 2nd ed. — New York.: Wiley, 1957

ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

О книгах Жордана говорили, что если ему нужно было ввести четыре аналогичные или родственные величины (такие, как, например, a, b, c, d), то они у него получали обозначения a, M_3', ε_2, $\Pi_{1,2}''$.

Дж. Литлвуд, [51, с. 46]

П1,. . . ,П10 — ссылки, соответственно, на разделы 1,. . . ,10 приложения

$C_n^k = n!/(k!(n-k)!)$ — где $n! = 1 \cdot 2 \cdot \ldots \cdot (n-1) \cdot n$, $0! = 1$

$\xi \sim$ — случайная величина ξ распределена по закону (распределена так же, как)

$\mathcal{N}(0,1)$, $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ — стандартный (§ 2 гл. 3) и многомерный (П9) нормальные законы

χ_k^2 — распределение хи-квадрат с k степенями свободы (§ 2 гл. 11)

F_{k_1,k_2} — распределение Фишера — Снедекора с k_1 и k_2 степенями свободы (§ 5 гл. 14)

$\Gamma(\alpha,\lambda)$ — гамма-распределение с параметрами α и λ (§ 4 гл. 3)

$\Gamma(p)$, $B(r,s)$ — гамма- и бета-функции Эйлера (§ 4 гл. 3)

$\xrightarrow{п.\,н.}$, $\xrightarrow{с.\,к.}$, $\xrightarrow{\mathsf{P}}$, $\xrightarrow{d}$ — символы сходимости почти наверное, в среднем, по вероятности, по распределению (П5)

I_A — индикатор множества A (§1 гл. 1)

$\operatorname{sign} x$ — знак числа x (§5 гл. 4)

$\mathbf{P}(A)$ — вероятность события A (П1)

$\mathbf{P}(A \mid B)$ — условная вероятность A при условии B (П7)

$F_\xi(x)$, $p_\xi(x)$, $\psi_\xi(t)$ — функция распределения, плотность (§ 1 гл. 1) и характеристическая функция случайной величины ξ (П9)

$\mathbf{M}\xi$ — математическое ожидание случайной величины ξ (§ 2 гл. 1, П2)

$\mathbf{M}(\xi \mid \eta)$ — условное математическое ожидание ξ при условии η (П7)

$\mathbf{D}\xi$ — дисперсия случайной величины ξ (§ 2 гл. 1, П2)

$\mathbf{cov}\,(\xi,\eta)$ — ковариация случайной величины ξ и η (П2)

$\mathbb{R}^n$ — n-мерное действительное пространство

$\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)$ — n-мерный случайный вектор (набор случайных величин)

$F_{\boldsymbol{\xi}}(x_1, \ldots, x_n)$, $p_{\boldsymbol{\xi}}(x_1, \ldots, x_n)$ — функция распределения, плотность случайного вектора $\boldsymbol{\xi}$ (П8)

$\mathbf{Cov}(\boldsymbol{\xi})$ — матрица ковариаций случайного вектора $\boldsymbol{\xi}$ (П2)

$\overline{X} = (X_1 + \ldots + X_n)/n$ — выборочное среднее наблюдений

$X_{(1)} \leqslant X_{(2)} \leqslant \ldots \leqslant X_{(n)}$ — вариационный ряд выборки $X_1, \ldots, X_n$, элементы вариационного ряда — порядковые статистики (§ 4 гл. 4)

$\det \boldsymbol{A}$, $\operatorname{tr} \boldsymbol{A}$ — определитель и след матрицы $\boldsymbol{A}$ (П10)

$\boldsymbol{A}^T$, $\boldsymbol{A}^{-1}$, $\boldsymbol{A}^{1/2}$ — транспонированная, обратная матрицы и квадратный корень из матрицы $\boldsymbol{A}$ (П10)

$a_n = o(b_n)$ — $a_n/b_n \to 0$,

$a_n \sim b_n$ — $a_n/b_n \to 1$

$a_n = O(b_n)$ — последовательность a_n/b_n ограничена

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

ОГЛАВЛЕНИЕ

Минимальные системные требования определяются соответствующими требованиями программы Adobe Reader версии не ниже 11-й для платформ Windows, Mac OS, Android, iOS, Windows Phone и BlackBerry; экран 10"

Учебное электронное издание

Лагутин Михаил Борисович

НАГЛЯДНАЯ МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Учебное пособие

Ведущий редактор *М. Стригунова*
Художник *С. Инфантэ*
Оригинал-макет подготовлен *О. Лапко* в пакете LaTeX 2_ε

Подписано к использованию 19.03.15.
Формат 155×225 мм

Издательство «БИНОМ. Лаборатория знаний»
125167, Москва, проезд Аэропорта, д. 3
Телефон: (499) 157-5272
e-mail: info@pilotLZ.ru, http://www.pilotLZ.ru