

УДК 519.2(078.5)
ББК 22.171+22.172
Б 86

Бочаров П. П., Печинкин А. В. **Теория вероятностей. Математическая статистика.** — 2-е изд. — М.: ФИЗМАТЛИТ, 2005. — 296 с. — ISBN 5-9221-0633-3.

В первой части рассматриваются основные понятия теории вероятностей, при этом используются относительно простые математические конструкции, но, тем не менее, изложение ведется на основе аксиоматического построения, предложенного академиком А. Н. Колмогоровым. Во второй части излагаются основные понятия математической статистики. Рассматриваются наиболее часто встречающиеся задачи оценивания неизвестных параметров и проверки статистических гипотез и описываются основные методы их решения. Каждое приведенное положение иллюстрируется примерами. Излагаемый материал в целом соответствует государственному образовательному стандарту.

Студентам, аспирантам и преподавателям вузов, научным работникам различных специальностей и желающим получить первое представление о теории вероятностей и математической статистике.

ОГЛАВЛЕНИЕ

Предисловие	6
Введение	8

I. Теория вероятностей

Глава 1. Вероятностное пространство	15
1. Пространство элементарных исходов	15
2. События, действия над ними	16
3. σ -алгебра событий	21
4. Вероятность	25
Глава 2. Классическая и геометрическая вероятности	29
1. Классическая вероятность	29
2. Элементы комбинаторики в теории вероятностей	30
3. Геометрическая вероятность	36
Глава 3. Условная вероятность. Независимость событий. Формулы полной вероятности и Байеса	40
1. Условная вероятность	40
2. Формула умножения вероятностей	42
3. Независимость событий	44
4. Формула полной вероятности	47
5. Формула Байеса	48
Глава 4. Схема Бернулли	52
1. Формула Бернулли	52
2. Формула Пуассона	53

3. Формулы Муавра–Лапласа	54
4. Применение приближенных формул Пуассона и Муавра–Лапласа	57
5. Теорема Бернулли	62
6. Вычисление определенных интегралов методом Монте-Карло	63
7. Полиномиальная схема	67
 Г л а в а 5. Случайные величины и их распределения	69
1. Случайная величина.	69
2. Функция распределения случайной величины	71
3. Дискретные случайные величины	74
4. Непрерывные случайные величины	77
5. Функции от случайной величины	84
 Г л а в а 6. Многомерные случайные величины и их свойства	89
1. Многомерная случайная величина	89
2. Совместная функция распределения	90
3. Дискретные двумерные случайные величины	92
4. Непрерывные двумерные случайные величины	95
5. Условные распределения	101
6. Независимые случайные величины	105
7. Функции от многомерных случайных величин	108
 Г л а в а 7. Числовые характеристики случайных величин	114
1. Математическое ожидание случайной величины	114
2. Математическое ожидание функции от случайной величины. Свойства математического ожидания	117
3. Дисперсия. Моменты высших порядков	120
4. Ковариация и корреляция случайных величин	125
5. Условное математическое ожидание. Регрессия	129
6. Другие числовые характеристики случайных величин	133
 Г л а в а 8. Предельные теоремы теории вероятностей	137
1. Неравенство Чебышева. Закон больших чисел	138
2. Усиленный закон больших чисел. Закон повторного логарифма . .	140
3. Характеристическая функция	143
4. Центральная предельная теорема	150
 Список литературы	152

II. Математическая статистика

Г л а в а 1. Общие сведения	155
1. Задачи математической статистики	155
2. Основные понятия математической статистики	158
3. Простейшие статистические преобразования.	160
4. Основные распределения математической статистики.	169
 Г л а в а 2. Оценки неизвестных параметров	173
1. Статистические оценки и их свойства	173
2. Достаточные оценки.	183
3. Метод моментов	191
4. Метод максимального правдоподобия	193
5. Метод минимального расстояния	198
6. Метод номограмм	199
7. Доверительные интервалы	201
 Г л а в а 3. Проверка статистических гипотез	207
1. Статистическая гипотеза. Критерий	207
2. Простые гипотезы	212
3. Однопараметрические гипотезы. Равномерно наилучшие критерии	223
4. Многопараметрические гипотезы	232
5. Критерий согласия	238
6. Критерии однородности двух выборок	246
 Г л а в а 4. Некоторые задачи, связанные с нормальными выборками	252
1. Общая характеристика задач	252
2. Критерии согласия	253
3. Критерии равенства дисперсий	256
4. Выборочная корреляция	260
5. Общая линейная модель, метод наименьших квадратов	263
6. Регрессионный анализ	271
7. Дисперсионный анализ	278
8. Планирование эксперимента	285
 Список литературы	292
Приложение	293

Предисловие

Предлагаемое учебное пособие написано на основе курсов по теории вероятностей и математической статистике, читаемых авторами в течение ряда лет студентам самых различных специальностей, начиная от инженерных и кончая прикладной математикой. Для его изучения достаточно знание математики в объеме стандартного курса высшей математики для вузов.

Следует обратить внимание читателя на то, что хотя теория вероятностей и математическая статистика выступают в роли единой математической дисциплины, ее первая часть — теория вероятностей — существенно более проста для понимания. Гораздо сложнее вторая часть — математическая статистика, причем необходимым условием для ее изучения является хорошее знание теории вероятностей. Этим объясняется определенное различие в стиле изложения первой и второй частей. Материал первой части, ориентированный на детальную проработку основных понятий теории вероятностей, достаточно однороден, в то время как во второй части применен «многоуровневый» подход к изложению: доказательства и более глубокие сведения, которые можно опустить при первом знакомстве с математической статистикой, выделены более мелким шрифтом.

В пособие включено большое количество примеров, иллюстрирующих приведенные положения.

В книге принята раздельная нумерация параграфов, примеров, формул, рисунков и таблиц по главам. При ссылке на объект из другой главы добавляется номер главы (например, «см. пример 3 из гл. 2»). В части 2 («Математическая статистика») имеются ссылки на часть 1 («Теория вероятностей»), и тогда добавляется «часть 1» (например, «см. часть 1, гл. 2, параграф 3»).

В книге принятые следующие соглашения об использовании математических знаков равенства и приближенного равенства. Соотношение $f \equiv g$ означает тождественное равенство функций f и g при всех значениях аргумента. Запись $a \approx b$ означает асимптотическое равенство a и b при использовании предельных формул и эквивалентна математическому соотношению $a = b + o(1)$. Наконец, запись $a \cong b$ означает равенство a и b с точностью до ошибок округления.

В книге приводятся отдельные списки литературы для первой и второй частей. При этом принят следующий порядок: сначала по возрастанию сложности идут учебники, затем — дополнительная литература для тех читателей, кто желает более глубоко ознакомиться с тем

или иным разделом, и, наконец, также по возрастанию сложности приводятся задачники.

При написании настоящей книги были использованы материалы учебника Б. В. Гнеденко и задачника Л. Д. Мешалкина (см. [8] и [17], литература к части 1). Кроме того, хотя в приложении и приводятся таблицы значений распределения Пуассона, плотности нормального распределения и интеграла Лапласа, для решения примеров по математической статистике необходимо иметь статистические таблицы Л. Н. Больщева и Н. В. Смирнова ([1], литература к части 2).

Авторы благодарны всем лицам, оказавшим помощь при подготовке настоящего издания.

Бочаров П. П., Печинкин А. В.
2005 г.

Введение

Теория вероятностей и математическая статистика относится к числу прикладных математических дисциплин, поскольку она направлена на решение прикладных задач и возникла из чисто практических потребностей, а использует математические методы. В свою очередь двойное название дисциплины говорит о том, что в ней можно выделить два направления: теорию вероятностей и математическую статистику. Если попытаться кратко объяснить, чем занимаются каждое из направлений, то это будет выглядеть так: теория вероятностей производит пересчет заданных вероятностей «простых» событий в вероятности «сложных» событий, а математическая статистика по наблюденным данным восстанавливает вероятности событий или проверяет, правы ли мы в своих предположениях относительно этих вероятностей.

Постараемся подробнее остановиться на том, какие закономерности изучает теория вероятностей и математическая статистика и какое отношение это имеет к практике.

Начнем с первого направления — *теории вероятностей*. Уже само название наводит на мысль, что основная задача теории вероятностей — изучение численных закономерностей в опытах, результаты которых не могут быть предсказаны однозначно до проведения испытаний.

Для того чтобы лучше понять существование дела, выясним подробнее тот смысл, который вкладывается в понятия «вероятно», «маловероятно», «весьма вероятно».

Первое необходимое условие употребления этих понятий заключается, как уже говорилось, в невозможности предсказания исхода некоторого действия. Так, мы не можем предсказать, какой стороной упадет подброшенная монета, сколько очков выпадет на игральной кости, сколько времени проработает электрическая лампочка, какое вещество образуется в результате определенной химической реакции (если, конечно, результат не предопределен известной нам теорией), сколько частиц будет зарегистрировано счетчиком Гейгера–Мюллера за заданный промежуток времени, какой номер телефона у знакомого, какая будет погода 1 июня 2010 г. и т. д.

Однако погоду 1 июня 2010 г. мы не будем знать до этого дня, а после наступления 1 июня 2010 г. она будет полностью определена. Поэтому применение слов «маловероятно», «весьма вероятно» к погоде именно 1 июня 2010 г. некорректно из-за невозможности повторения испытания и погодой 1 июня 2010 г. до этого дня занимается метеорология, а после — история, но никак не теория вероятностей. Итак, второе необходимое условие — возможность повторения испытания

с первоначальным комплексом исходных данных, причем, хотя бы теоретически, бесконечное число раз.

Далее, выяснив номер телефона знакомого, мы сколько бы раз его ни спрашивали, новых цифр не добьемся. Аналогично, определив исход химической реакции, а затем проведя ее снова при тех же условиях, мы, естественно, нового вещества не получим. Отсюда вытекает третье необходимое условие — невозможность точного предсказания результатов не только первого испытания, но и каждого последующего.

Таким образом, в приведенных выше примерах высказывания «маловероятно», «весыма вероятно» в наиболее точном их смысле можно отнести только к подбрасыванию монеты, бросанию игральной кости, испытанию электрической лампочки и регистрации частиц счетчиком Гейгера.

Попробуем теперь формализовать понятие «вероятного», связывая с ним числовую характеристику. Очевидно, эпитет «маловероятно» мы приписываем событиям, доля появления которых в общем числе испытаний мала, и, наоборот, «весыма вероятно» — событиям, происходящим практически во всех испытаниях. Введем количественную оценку понятия «вероятного». Для этого рассмотрим частоту $f_A = N_A/N$ появления некоторого события A , где N_A — число появлений этого события, а N — общее число испытаний. Оказывается, как было установлено опытным путем, с ростом числа испытаний N для довольно широко класса явлений частота f_A стабилизируется, т.е. стремится к некоторому предельному значению P_A , которое естественно принять в качестве вероятности события A . К явлениям подобного рода относятся события, связанные с выпадением «герба» или «цифры» при подбрасывании монеты, с выпадением определенного числа очков при бросании игральной кости, с работой электрической лампочки в определенных границах времени, регистрацией частиц счетчиком и многие, многие другие. Такие события разумно назвать случайными. Поскольку в настоящем курсе другие события встречаться не будут, то прилагательное «случайное» мы будем для краткости опускать.

Эмпирический закон предельного постоянства частоты является той основной физической предпосылкой, которая необходима для практического применения методов теории вероятностей. Более того, хотя современная теория вероятностей и строится на аксиоматическом определении вероятности (см. ниже), при осмыслении полученных результатов мы рекомендуем всегда пользоваться частотной интерпретацией. Так, если после вычислений Вы получили, что вероятность некоторого события A равна 0,15, то это нужно трактовать следующим образом: при многократном повторении соответствующего опыта на каждые 100 испытаний приходится 15 появлений события A .

Однако попытка отождествить вероятность с частотой не выдерживает никакой более или менее существенной критики. Частота меняется от испытания к испытанию, а бесконечного числа испытаний, как известно, никто не проводил и вряд ли проведет в обозримом будущем.

Математика же привыкла иметь дело с точными, логически безупречными понятиями, и частотное определение ее никак не устраивает. Поэтому выдающимся математиком прошлого века Андреем Николаевичем Колмогоровым было предложено *аксиоматическое определение вероятности*, основанное на общем понятии меры. При этом аксиомы Колмогорова отражают три основных свойства частоты, перенесенных им на вероятность.

1. Частота появления случайного события неотрицательна — аксиома неотрицательности вероятности.
2. Частота появления достоверного события, т. е. события, происходящего в каждом испытании, равна единице — аксиома нормированности.
3. Если два события не могут одновременно произойти в одном и том же испытании (несовместны), то частота появления хотя бы одного из них совпадает с суммой частот появления каждого — аксиома сложения вероятностей.

Если теперь сопоставить каждому событию вероятность, т. е. число, удовлетворяющее трем вышеперечисленным аксиомам, то, оказывается, можно построить весьма содержательную теорию. При этом единственное добавление, необходимое для обеспечения математической строгости, заключается в замене аксиомы сложения вероятностей З расширенной аксиомой сложения вероятностей:

- 3'. Вероятность суммы равна сумме вероятностей не только для двух, но и для произвольного счетного (т. е. такого, которое можно пересчитать с помощью чисел натурального ряда) числа несовместных событий.

Аксиоматический подход позволяет не только описать хорошо известные явления, но и найти закономерности более общего типа. Так, вспомним пример с подбрасыванием монеты. Естественно предположить (и это подтверждается неоднократными опытами), что частоты выпадения «герба» и «цифры» одинаковы, и приписать каждому из этих двух событий одинаковую вероятность $1/2$. Однако при аксиоматическом построении теории вероятностей мы вправе приписать выпадению «герба» любую вероятность p , заключенную между нулем и единицей, а тогда выпадению «цифры» мы обязаны сопоставить вероятность $q = 1 - p$. Такое определение вероятности описывает случай несимметричной монеты. В свою очередь, последовательное подбрасывание несимметричной монеты (или аналогичный опыт) носит в теории вероятностей название последовательности независимых одинаковых испытаний или схемы Бернулли и является одной из наиболее часто применяемых на практике «базовых» моделей, позволяющих наглядно представить себе основные задачи этой теории.

Теперь мы можем привести примеры простейших задач, которые решает теория вероятностей. Так, считая, что при бросании игральной

кости падение ее на любую грань одинаково вероятно, нужно найти вероятность того, что выпадет четное число очков. Или, зная вероятность p выпадения «герба» при одном подбрасывании несимметричной монеты, необходимо определить вероятность выпадения ровно одного «герба» при двух подбрасываниях этой же монеты.

Вернемся к примерам, рассмотренным в самом начале введения. При первом прочтении мы отбросили высказывания о погоде, результате химической реакции, номере телефона знакомого. При более детальном рассмотрении оказывается, однако, что и к ним можно применить методы теории вероятностей. Заменяя, например, высказывание «погода 1 июня 2010 г. будет солнечная» высказыванием «погода 1 июня будет солнечная» (не указывая, какого именно года), мы уже вправе использовать вероятностные соображения. Аналогично, в результате химической реакции могут появиться побочные продукты, объем которых может оказаться случайным. Предоставим читателю самому придумать постановку задачи о номере телефона знакомого, для решения которой также можно было бы применить вероятностные методы.

Перейдем ко второму направлению — *математической статистике*. В теории вероятностей игнорируется вопрос: откуда берутся исходные вероятности? Они считаются заданными «извне» или определенными какой-либо теорией. Однако в реальной жизни при применении методов теории вероятностей постоянно приходится сталкиваться с ситуациями, когда эти вероятности нам неизвестны. Конечно, при отсутствии всякой информации давать какие-либо прогнозы — вещь весьма опасная. Поэтому подумаем, как можно выйти из создавшейся ситуации.

Обратимся к примеру. Пусть нам предлагают играть в «орлянку» неизвестной монетой, о которой мы не знаем, как часто выпадает «герб» и как часто — «цифра». Осторожный человек, прежде чем «бросаться головой в омут», всегда посмотрит, как это делают другие. Поэтому он сначала понаблюдает за ходом игры и только потом, оценив для себя шансы, будет ставить на тот или иной исход.

Именно задачами восстановления на основе предварительных наблюдений данных, недостающих для расчетов методами теории вероятностей, и занимается математическая статистика. Очевидно, что в силу случайности результатов наблюдений сделать достоверные выводы о параметрах того явления, которое мы далее собираемся исследовать, невозможно. Тем не менее у нас есть определенная зацепка: это тот же самый эмпирический закон предельного постоянства частоты (однако, если мы работаем в рамках аксиоматического подхода, мы должны этот закон доказать строго математически!). В соответствии с ним: чем дольше производить наблюдения, тем более точные выводы можно получить. Математическая статистика как раз и учит тому, как нужно обрабатывать наблюдения, чтобы «выжать» из них наиболее полную

информацию, и как оценить степень достоверности полученных выводов.

Вернемся к игре в «орлянку». Понаблюдав за игрой, мы можем задать себе такой вопрос: что можно сказать на основе полученных данных о вероятности выпадения «герба»? Или спросить себя: а действительно ли вероятности выпадения «герба» и «цифры» одинаковы, как это нам обещают? Именно эти два вопроса и определяют два основных направления в математической статистике. Первое направление связано с оценкой неизвестного параметра (в частности, вероятности выпадения «герба»), а второе — с проверкой определенных предположений, или гипотез (в нашем примере — гипотезы о равновероятности выпадения «герба» и «цифры»).

Правильность исходных предпосылок теории вероятностей и математической статистики, как и всякой другой прикладной теории, проверяется практикой. На сегодняшний день трудно найти такую область человеческих знаний, где в той или иной мере не применялись бы методы теории вероятностей и математической статистики. Сюда, наряду с естественными отраслями науки и техники, такими, как физика, химия, инженерия, можно отнести и, казалось бы, далекие от математики области: историю, медицину, генетику, социологию, лингвистику и т. д.

На базе теории вероятностей и математической статистики сформировались и выросли такие разделы, как теория случайных процессов, теория массового обслуживания, математическая теория надежности и т. д. В свою очередь теория вероятностей и математическая статистика опираются на другие математические дисциплины: функциональный анализ, алгебру, аналитическую геометрию, теорию функций комплексного переменного, дифференциальные уравнения и др.

В заключение скажем несколько слов об истории развития теории вероятностей и математической статистики. Возникновение теории вероятностей обычно относят к XVII в. и связывают с комбинаторными задачами азартных игр. Хотя азартные игры и нельзя рассматривать как серьезный объект для изучения, именно они привели к решению задач, не укладывавшихся в рамки существовавших тогда математических моделей, и способствовали появлению новых понятий и идей. Эти новые элементы можно встретить уже у Я. Бернулли, П. С. Лапласа, К. Ф. Гаусса и целого ряда других видных математиков того времени. Как самостоятельный раздел математики дисциплина «Теория вероятностей и математическая статистика» сложилась в конце XIX — начале XX веков, причем если говорить о математической статистике как об отдельном направлении, то ее бурное развитие началось лишь в прошлом столетии. Окончательное становление теории вероятностей и математической статистики, как уже говорилось, связано с именем А. Н. Колмогорова.

Часть I

ТЕОРИЯ ВЕРОЯТНОСТЕЙ

Глава 1

ВЕРОЯТНОСТНОЕ ПРОСТРАНСТВО

Взяв практически любую статью по теории вероятностей, мы увидим, что либо она начинается словами: «Пусть (Ω, \mathcal{B}, P) — вероятностное пространство», либо в одной из первых же фраз написано: «где (Ω, \mathcal{B}, P) — вероятностное пространство». Иногда добавляется: « Ω — пространство элементарных исходов, \mathcal{B} — σ -алгебра (сигма-алгебра) событий, P — вероятностная мера (или вероятность)». Естественно, у неискушенного читателя пропадает всякое желание читать эту статью дальше. Однако понятие вероятностного пространства является весьма естественным математическим обобщением хорошо известных физических понятий: исход опыта, случайное событие, вероятность события. В настоящей главе мы попытаемся, насколько это возможно, дать читателю, знакомому только с основами высшей математики, разъяснение этого основополагающего для теории вероятностей понятия.

1. Пространство элементарных исходов

Рассмотрим простейший вариант случайного испытания — подбрасывание монеты. Если отвлечься от чисто гипотетических возможностей — падения монеты на ребро или вообще исчезновения монеты, то возможны только два исхода: выпадение «герба» и выпадение «цифры». Эти два исхода в рамках данного опыта уже нельзя разбить на более мелкие составляющие, т. е. они являются в некотором роде «элементарными». При бросании игральной кости такими неделимыми исходами являются: выпадение одного очка, выпадение двух очков, ..., выпадение шести очков. Значит, мы имеем уже 6 элементарных исходов. Более сложный пример получим, если рассмотрим падение идеальной (т. е. не имеющей размера) частицы на плоскость. Тогда результат испытания представляет собой попадание частицы в определенную точку плоскости и его можно отождествить с двумерным вектором в некоторой системе координат на плоскости.

Аналогично, если проанализировать любое испытание со случайнym исходом, можно заметить, что его результат представляет собой один из множества допустимых исходов. Поскольку в математике принято абстрагироваться от несущественных деталей, то всегда можно рассматривать все возможные в данном опыте исходы как некоторое множество Ω , которое и носит название *пространства элементарных исходов* или *пространства элементарных событий*. Сами элементарные исходы будем обозначать строчной буквой ω , снабжая ее при необходимости индексами.

Пример 1. При подбрасывании монеты пространство элементарных исходов Ω состоит всего из 2 исходов: ω_1 — выпадение «герба» и ω_2 — выпадение «цифры». \square

Пример 2. При бросании игральной кости возможны 6 элементарных исходов: ω_1 — выпадение одного очка, ω_2 — выпадение 2 очков, ..., ω_6 — выпадение 6 очков. \square

Пример 3. При подбрасывании двух монет пространство элементарных исходов Ω содержит уже 4 исхода. Перечислим их: ω_1 — пара «герб»—«герб», ω_2 — «герб»—«цифра», ω_3 — «цифра»—«герб», ω_4 — «цифра»—«цифра». При подбрасывании трех монет возможны 8 элементарных исходов типа «герб»—«цифра»—«герб» и т. д. \square

Пример 4. При определении времени жизни элементарной частицы пространство элементарных исходов Ω представляет собой полуправую прямую $[0, \infty)$. \square

Следует отметить, что в практических исследованиях существует определенный произвол в описании пространства элементарных исходов Ω . Так, однократное подбрасывание монеты (пример 1) можно рассматривать как часть более сложного опыта, заключающегося в подбрасывании двух или более монет (пример 3). При определении времени жизни частицы (пример 4) можно также рассматривать типы получившихся после распада частиц и т. д. Очевидно, при решении практических задач разумно выбирать всегда наиболее простой вариант пространства элементарных исходов, необходимый для решения стоящей перед исследователем задачи.

2. События, действия над ними

Понятие «событие» лингвистически отличается от понятия «элементарное событие» только отсутствием прилагательного «элементарное». Естественно поэтому определить событие так же, как исход испытания, но только не обязательно неделимый.

Пример 5. При бросании игральной кости (см. пример 2) событиями являются: выпадение четного числа очков (это событие происходит тогда и только тогда, когда появляется один из элементарных исходов $\omega_2, \omega_4, \omega_6$); выпадение нечетного числа очков (элементарные исходы $\omega_1, \omega_3, \omega_5$); выпадение не менее двух очков (элементарные исходы $\omega_2, \omega_3, \omega_4, \omega_5, \omega_6$) и т. д. \square

Пример 6. При подбрасывании двух монет примерами событий будут: падение обеих монет на одну и ту же сторону (появлению этого события благоприятствуют элементарные исходы ω_1 и ω_4 из примера 3); падение монет на разные стороны (элементарные исходы ω_2 и ω_3); выпадение, по крайней мере, одного «герба» (элементарные исходы ω_1, ω_2 и ω_3) и т. п. \square

Пример 7. При определении времени безотказной работы электрической лампочки можно привести следующие примеры событий: безотказная работа лампочки до момента T ; отказ лампочки до момента T ; отказ лампочки между моментами T_1 и T_2 и т. д. Здесь так же, как и в примере 4, пространство элементарных исходов Ω представляет собой полуправую прямую $[0, \infty)$. Тогда первому событию соответствует множество точек на полуправой прямой $[T, \infty)$, второму — на интервале $[0, T)$, третьему — на интервале (T_1, T_2) . \square

Вспоминая, что в результате опыта может произойти один и только один элементарный исход ω из пространства элементарных исходов Ω , мы приходим к теоретико-множественному определению *события* как произвольного набора элементарных исходов или, иными словами, произвольного подмножества множества элементарных исходов Ω . События будем обозначать прописными латинскими буквами, снабженными при необходимости индексами: A, B, C_1, H_2 и т. д.

Заметим, что приведенное выше определение события *не всегда позволяет построить логически безупречную аксиоматику теории вероятностей*. Поэтому в следующем параграфе мы уточним понятие «событие». Сейчас же наша цель состоит в описании теоретико-множественных операций над событиями, и нам удобно отказаться от несущественных пока деталей.

Часто бывает полезным наглядное представление событий в виде так называемой *диаграммы Эйлера–Венна*. Будем изображать все пространство элементарных исходов прямоугольником (рис. 1). Тогда каждый элементарный исход ω соответствует точке внутри прямоугольника, а каждое событие A отождествимо с некоторой областью.

Само пространство элементарных исходов Ω представляет собой событие, состоящее из всех возможных исходов, т. е. происходящее всегда (при любом элементарном исходе ω), и носит название *достоверного события*. Таким образом, пространство элементарных исходов выступает в двух качествах: в качестве собственно множества всех элементарных исходов и в качестве достоверного события.

Для дальнейшего нам удобно ввести еще одно событие \emptyset , называемое невозможным. *Невозможное событие* не происходит никогда, т. е. не содержит ни одного элементарного исхода.

Пример 8. При бросании игральной кости событие «выпадение не менее одного очка» является достоверным (Ω), событие «выпадение более 6 очков» — невозможным (\emptyset). \square

Над событиями как над подмножествами фиксированного множества можно производить действия, которые мы сейчас опишем.

Пересечением (произведением) двух событий A и B называется событие C , происходящее тогда и только тогда, когда наступают одновременно оба события A и B , или, иными словами, состоящее из тех элементарных исходов, которые принадлежат и A , и B (рис. 2).

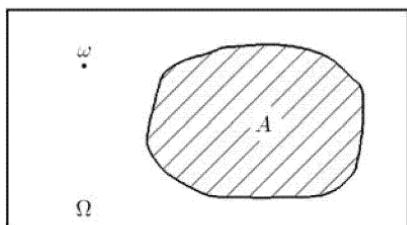


Рис. 1

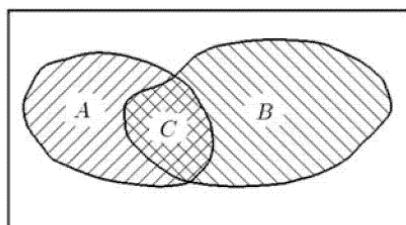


Рис. 2

Пересечение событий A и B записывается следующим образом:

$$C = A \cap B$$

или

$$C = A \cdot B = AB.$$

Аналогично определяется пересечение трех и более событий.

Пример 9. Событие A — при подбрасывании двух монет падение их одной стороной, событие B — выпадение хотя бы одного «герба». Пересечением событий A и B является событие C , состоящее в выпадении двух «гербов». \square

Пример 10. Событие A — выпадение четного числа очков при бросании игральной кости, событие B — выпадение не менее 3 очков. Пересечение A и B — событие C , состоящее в выпадении 4 или 6 очков. \square

События A и B называются *непересекающимися* или *несовместными*, если их пересечение является невозможным событием, т. е. $A \cap B = \emptyset$ (рис. 3).

Для трех и более событий понятие несовместности можно определить разными способами. Мы будем, в основном, пользоваться следующим понятием несовместности n событий, которое также называется *попарной несовместностью* событий: события A_1, \dots, A_n называются (попарно) несовместными, или (попарно) непересекающимися, если $A_i \cap A_j = \emptyset$ для любых i и j при $i \neq j$.

Пример 11. Событие A — выпадение четного числа очков при бросании игральной кости, событие B — выпадение нечетного числа очков. События A и B несовместны. \square

Нетрудно видеть, что справедливы следующие простейшие формулы для пересечения двух событий, одно из которых достоверно или невозможно:

$$A\Omega = A, \quad A\emptyset = \emptyset.$$

Объединением (суммой) двух событий A и B называется событие C , происходящее тогда и только тогда, когда наступает хотя бы одно из событий A или B , т. е. состоящее из тех элементарных исходов, которые принадлежат хотя бы одному из множеств A или B (рис. 4).

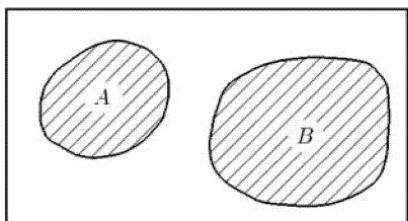


Рис. 3

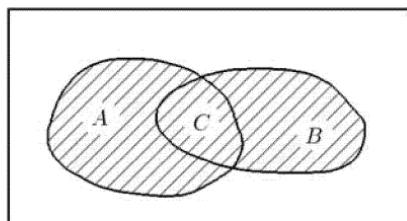


Рис. 4

Для объединения событий A и B применяется запись

$$C = A \cup B.$$

Пример 12. Событие A — выпадение 1 или 3 очков при бросании игральной кости, событие B — выпадение 3 или 5 очков. Объединением событий A и B является событие C , состоящее в выпадении нечетного числа очков. \square

Для объединения двух событий, одно из которых достоверно или невозможено, имеют место следующие формулы:

$$\Omega \cup A = \Omega, \quad \emptyset \cup A = A.$$

В том случае, когда события A и B несовместны, наряду со знаком « \cup » для их объединения употребляют знак « $+$ ». Обычно знак « $+$ » применяют тогда, когда заведомо известно, что A и B несовместны, и это особо хотят подчеркнуть. В частности, поскольку невозможное событие несовместно с любым событием A , то

$$\emptyset \cup A = \emptyset + A = A.$$

Аналогично определяется объединение трех и более событий. При этом знак « $+$ » используется в случае попарной несовместности входящих в объединение событий.

Разностью двух событий A и B называется событие C , происходящее тогда и только тогда, когда происходит событие A , но не происходит событие B , т. е. состоящее из тех элементарных исходов, которые принадлежат A , но не принадлежат B (рис. 5).

Разность событий A и B записывается в виде

$$C = A \setminus B.$$

Пример 13. Событие A — выпадение хотя бы одного «герба» при подбрасывании двух монет, событие B — падение обеих монет одной стороной. Разность C событий A и B представляет собой событие, заключающееся в выпадении ровно одного «герба». \square

Справедливы следующие формулы для разности двух событий, одно из которых достоверно или невозможно:

$$A \setminus \Omega = \emptyset, \quad A \setminus \emptyset = A, \quad \emptyset \setminus A = \emptyset.$$

Кроме того, если A и B несовместны ($AB = \emptyset$), то $A \setminus B = A$.

Симметрической разностью двух событий A и B (обозначается знаком Δ или \circ) называется событие C , представляющее собой объединение событий $A \setminus B$ и $B \setminus A$:

$$C = A \Delta B = (A \setminus B) \cup (B \setminus A).$$

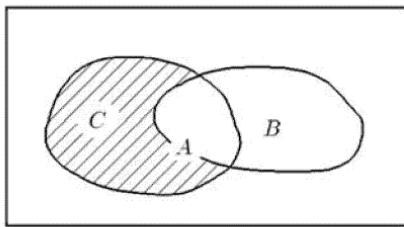


Рис. 5

Поскольку события $A \setminus B$ и $B \setminus A$ несовместны (рис. 6), симметрическую разность можно записать также в виде

$$A \Delta B = (A \setminus B) + (B \setminus A).$$

Нетрудно заметить, что симметрическая разность есть объединение событий A и B без их общей части:

$$A \Delta B = (A \cup B) \setminus (B \cap A).$$

Пример 14. Событие A — выпадение не менее 2 очков при бросании игральной кости, событие B — выпадение не более 4 очков. Симметрической разностью событий A и B является событие C , заключающееся в выпадении 1, 5 или 6 очков. \square

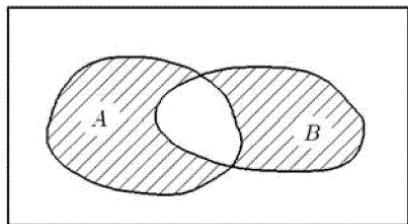


Рис. 6

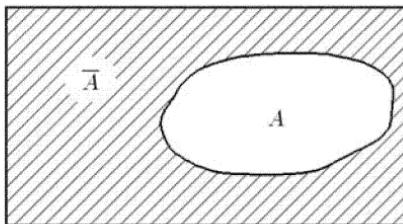


Рис. 7

Если A и B несовместны, то

$$A \Delta B = A + B.$$

Дополнением события A (обычно обозначается \bar{A}) называется событие, происходящее тогда и только тогда, когда не происходит событие A (рис. 7), или, иными словами,

$$\bar{A} = \Omega \setminus A.$$

Пример 15. Событие A — выпадение четного числа очков при бросании игральной кости. Дополнительное событие \bar{A} — выпадение нечетного числа очков. \square

Справедливы формулы:

$$\bar{\Omega} = \emptyset, \quad \emptyset = \Omega, \quad \bar{\bar{A}} = A.$$

Если некоторое событие записано в виде нескольких действий над различными событиями, то сначала вычисляются дополнения, затем выполняются умножения и, наконец, сложения и вычитания событий. Так, формула

$$C = A_1 \bar{A}_2 B_1 \cup A_3 \bar{B}_2 \setminus B_3$$

эквивалентна формуле

$$C = \{[A_1(\bar{A}_2)B_1] \cup [A_3(\bar{B}_2)]\} \setminus B_3.$$

Пользуясь диаграммой Эйлера–Венна, нетрудно показать справедливость следующих формул (*формулы де Моргана*):

$$A \cap B = \overline{\overline{A} \cup \overline{B}}, \quad A \cup B = \overline{\overline{A} \cap \overline{B}}.$$

Формулы де Моргана элементарно переносятся на произвольное число событий. В частности, для n событий A_1, \dots, A_n они имеют вид:

$$A \cap \dots \cap A_n = \overline{\overline{A}_1 \cup \dots \cup \overline{A}_n}, \quad A \cup \dots \cup A_n = \overline{\overline{A}_1 \cap \dots \cap \overline{A}_n}.$$

Следует отметить, что все действия над событиями можно получить с помощью только двух действий — объединения и дополнения (или пересечения и дополнения). Основанием для этого утверждения служат формулы де Моргана, а также соотношение

$$A \setminus B = A\overline{B}.$$

Кроме вышеперечисленных действий над событиями нам в дальнейшем понадобится понятие включения. Событие A принадлежит (содержится в, включается в) событию B (записывается $A \subset B$), если появление события A обязательно влечет за собой наступление события B (рис. 8) или, иными словами, каждый элементарный исход ω , принадлежащий A , обязательно принадлежит и B . Ясно, что включение $A \subset B$ эквивалентно выполнению равенства $AB = A$.

Используют и обратное понятие: событие B содержит (включает) событие A ($B \supset A$), если $A \subset B$.

Пример 16. Событие A — выпадение четного числа очков при бросании игральной кости, событие B — выпадение не менее 2 очков. Событие A принадлежит событию B , поскольку если выпало четное число очков (2, 4 или 6), то обязательно выпало не менее 2 очков. \square

Следующие включения очевидны:

$$\emptyset \subset A \subset \Omega.$$

Кроме того, если $A \subset B$, то

$$A \cup B = B, \quad A \setminus B = \emptyset, \quad A \Delta B = B \setminus A.$$

3. σ -алгебра событий

Итак, мы назвали событием *произвольное подмножество* пространства элементарных исходов Ω . Такое определение прекрасно работает, когда Ω конечно или даже счетно (т. е. его можно пересчитать с помощью чисел натурального ряда). Однако если Ω более чем счетно,

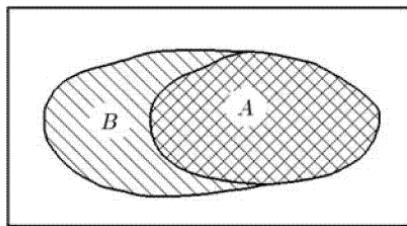


Рис. 8

то, вообще говоря, мы уже не сможем построить логически непротиворечивую теорию, называя событием произвольное подмножество Ω . Причина этого заключается в существовании так называемых неизмеримых множеств, что в свою очередь кроется в топологической структуре классических рассматриваемых пространств (прямой, плоскости, трехмерного пространства и т. д.). Поэтому приходится отказаться от, казалось бы, естественного желания назвать событием любое подмножество пространства элементарных исходов Ω и выделить среди всех подмножеств некоторый класс подмножеств \mathfrak{B} . Именно только подмножества из выделенного класса \mathfrak{B} и будут называться *событиями*. Интуитивно ясно, что описанные в предыдущем пункте теоретико-множественные операции над событиями не должны приводить к подмножествам, не являющимся событиями.

С точки зрения повседневной практики подмножества пространства элементарных исходов Ω , не являющиеся событиями, представляют собой чистую математическую абстракцию и в реальной жизни никогда не встречаются. Даже само доказательство их существования представляет весьма сложную задачу. Поэтому читателю, не желающему вдаваться в математические тонкости, мы рекомендуем пропустить параграф, посвященный σ -алгебре событий, и в дальнейшем под событием понимать произвольное подмножество элементарных исходов Ω , а под σ -алгеброй — систему всех этих подмножеств. Любознательному читателю мы предоставляем возможность познакомиться со строгим определением последнего понятия, излагаемым ниже.

Алгеброй событий \mathfrak{A} назовем непустую систему подмножеств Ω , удовлетворяющую следующим аксиомам:

- A1.** Если подмножество A принадлежит \mathfrak{A} (является событием), то дополнение \bar{A} также принадлежит \mathfrak{A} (является событием).
- A2.** Если подмножества A и B принадлежат \mathfrak{A} (являются событиями), то и объединение $A \cup B$ принадлежит \mathfrak{A} (является событием).

Как мы знаем, любую из рассмотренных нами операций над подмножествами можно получить с помощью только двух операций: дополнения и объединения. Поэтому пересечение и разность двух событий также будут событиями. Поскольку $\Omega = A \cup \bar{A}$ и $\emptyset = \bar{\Omega}$, то все пространство элементарных исходов Ω и пустое подмножество \emptyset обязательно являются событиями в любой алгебре событий. Очевидно также, что объединение и пересечение любого конечного числа событий снова будет событием. Иными словами, алгебру событий \mathfrak{A} можно определить как *систему подмножеств пространства элементарных*

исходов Ω , замкнутую¹⁾ относительно конечного числа теоретико-множественных операций.

Однако понятие алгебры событий также оказывается недостаточным для аксиоматического построения теории вероятностей в том случае, когда пространство элементарных исходов Ω не является конечным. Интересы общей теории меры требуют, чтобы аксиома А2 была заменена на более сильную, и мы приходим к новому определению:

σ -алгеброй событий \mathfrak{B} назовем систему подмножеств из Ω , удовлетворяющую аксиоме А1 и аксиоме

2'. Если подмножества $A_1, A_2, \dots, A_n, \dots$ принадлежат \mathfrak{B} (являются событиями), то и их (счетное) объединение $A_1 \cup A_2 \cup \dots \cup A_n \cup \dots$ также принадлежит \mathfrak{B} (является событием).

Основываясь на формулах де Моргана, нетрудно показать, что пересечение счетного числа событий снова будет событием. Таким образом, σ -алгебру событий \mathfrak{B} можно определить как *систему подмножеств пространства элементарных исходов Ω , замкнутую относительно счетного числа теоретико-множественных операций*.

Любая σ -алгебра событий является одновременно и алгеброй событий. Обратное, вообще говоря, не верно, т. е. существуют алгебры событий, не являющиеся σ -алгебрами. Однако если пространство элементарных исходов Ω конечно, то любая алгебра событий будет также и σ -алгеброй событий, т. е. в этом случае понятия алгебры событий и σ -алгебры событий эквивалентны.

σ -алгебра событий является второй компонентой вероятностного пространства $(\Omega, \mathfrak{B}, P)$.

Пример 17. Для любого пространства элементарных исходов Ω , содержащего хотя бы один исход, семейство подмножеств, состоящее всего из двух подмножеств Ω и \emptyset , является σ -алгеброй. Ясно, однако, что на такой σ -алгебре, состоящей всего из достоверного и невозможного событий, скольнибудь содержательную теорию построить невозможно, и мы ее в дальнейшем рассматривать не будем. \square

Пример 18. Пусть пространство элементарных исходов Ω содержит по крайней мере два исхода. Возьмем в Ω некоторое подмножество A , отличное от \emptyset и Ω . Тогда система из четырех подмножеств \emptyset, A, \bar{A} и Ω будет являться σ -алгеброй \mathfrak{B} . Поскольку в дальнейшем мы будем рассматривать только события, а других событий, кроме перечисленных четырех, σ -алгебра \mathfrak{B} не содержит, то естественно отождествить ее с σ -алгеброй, определенной на пространстве элементарных исходов Ω' , состоящем всего из двух элементарных исходов: $\omega'_1 = A$ и $\omega'_2 = \bar{A}$, и содержащей подмножества $\emptyset, \{\omega'_1\}, \{\omega'_2\}$ и Ω . Здесь мы имеем дело с тем принципом упрощения пространства элементарных исходов, о котором говорилось в параграфе 1.

¹⁾ Система подмножеств замкнута относительно операции, если в результате применения этой операции получаем подмножество, снова принадлежащее этой системе.

В качестве иллюстрации рассмотрим время работы электрической лампочки. Первоначально пространство элементарных исходов представляет собой полупрямую $[0, \infty)$. Однако если наблюдателю доступна только информация, произошел отказ за фиксированное время T (событие A) или нет (событие \bar{A}), то он фактически имеет дело с двумя элементарными исходами: $\omega'_1 = A$ и $\omega'_2 = \bar{A}$; соответствующая σ -алгебра состоит из четырех событий, описанных выше. В этом случае наблюдение за работой электрической лампочки с точки зрения числа возможных элементарных исходов ничем не отличается от наблюдения за подбрасыванием монеты. \square

Пример 19. Пусть пространство элементарных исходов Ω содержит конечное ($\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$) или счетное ($\Omega = \{\omega_1, \omega_2, \dots, \omega_n, \dots\}$) число исходов. Такое пространство называется дискретным. В качестве σ -алгебры \mathfrak{B} возьмем систему всех подмножеств. Именно эту σ -алгебру мы всегда будем рассматривать в дальнейшем в случае дискретного пространства элементарных исходов. Нетрудно видеть, что любое событие A можно отождествить с последовательностью из 0 и 1, причем 0 на i -м месте означает, что элементарный исход ω_i не принадлежит событию A . В частности, невозможному событию \emptyset соответствует последовательность 0, 0, ..., а достоверному $\Omega = 1, 1, \dots$ Ясно, что если число исходов n конечно, то σ -алгебра \mathfrak{B} содержит 2^n событий (на каждом из n мест последовательности может стоять одно из двух чисел: 0 или 1).

В случае дискретного Ω σ -алгебру \mathfrak{B} можно определить исходя и из других предпосылок. Для этого достаточно объявить *событиями* все *элементарные исходы* ω_i . Поскольку в случае дискретного Ω любое его подмножество будет содержать не более счетного числа элементарных исходов, то в соответствии с аксиомой А2' оно обязательно будет *событием*. Таким образом, σ -алгебру \mathfrak{B} можно трактовать как σ -алгебру, порожденную всеми элементарными исходами.

В частности, в случае конечного Ω σ -алгебра \mathfrak{B} порождается конечным числом элементарных исходов и поэтому совпадает с алгеброй \mathfrak{A} , порожденной всеми элементарными исходами ω_i . Однако в случае счетного Ω алгебра \mathfrak{A} , порожденная всеми элементарными исходами ω_i , уже не будет совпадать с σ -алгеброй \mathfrak{B} , поскольку она будет содержать только подмножества, состоящие из конечного числа элементарных исходов (как объединения событий в соответствии с аксиомой А2) или подмножества, состоящие из *всех* элементарных исходов за исключением их конечного числа (как дополнения к подмножествам первого типа в соответствии с аксиомой А1). \square

Пример 20. Пусть пространство элементарных исходов Ω представляет собой прямую $(-\infty, \infty)$. И здесь система всех подмножеств будет представлять собой σ -алгебру. Однако оказывается, что такая «максимальная» σ -алгебра в наиболее интересных случаях представляет собой негодный объект для дальнейших исследований. Дело в том, что введение понятия σ -алгебры является *вспомогательным процессом*, необходимым для дальнейшего определения собственно вероятности, и, если бы только было возможно, никто не стал бы «городить огород» ради, разве что, красивого названия.

О невозможности использования «максимальной» σ -алгебры мы еще скажем несколько слов, когда будем рассматривать геометрическую вероятность. Сейчас попробуем построить другую σ -алгебру, опираясь на более умеренные запросы. Итак, что бы мы хотели от σ -алгебры на прямой? Разумеет-

ся, основное требование к ней заключается в том, чтобы ей принадлежали всевозможные интервалы $[a, b]$, $[a, b)$, $(a, b]$, (a, b) . Минимальная σ -алгебра, удовлетворяющая этому требованию, носит название *борелевской σ -алгебры* и является тем объектом, на котором без всяких логических противоречий можно построить математически строгую теорию.

Все сказанное относительно прямой в полной мере относится и к пространствам элементарных исходов, представляющим собой плоскость, трехмерное пространство и пространства более высоких размерностей, а также их невырожденные части (отрезки, многоугольники, круги, шары и т. д.). В теории вероятностей такие пространства элементарных исходов называются непрерывными. \square

4. Вероятность

Приступим теперь к аксиоматическому определению последней составляющей вероятностного пространства $(\Omega, \mathfrak{B}, P)$ — вероятности или, как иногда говорят, вероятностной меры P .

Предположим сначала, что пространство элементарных исходов *конечно*. Пусть каждому *событию* A (т. е. подмножеству A пространства элементарных исходов Ω , принадлежащему σ -алгебре \mathfrak{B}) поставлено в соответствие число $P(A)$. Числовая функция $P(A)$ (заданная на σ -алгебре \mathfrak{B}) называется *вероятностью*, если она удовлетворяет следующим аксиомам:

P1. $P(A) \geq 0$ (*аксиома неотрицательности*);

P2. $P(\Omega) = 1$ (*аксиома нормированности*);

P3. $P(A + B) = P(A) + P(B)$ (*аксиома сложения*), если $A, B \in \mathfrak{B}$ и $A \cup B = \emptyset$.

Как говорилось во введении, аксиомы вероятности представляют собой не что иное, как математическое отражение основных свойств частоты.

Из аксиом P1–P3 можно вывести ряд очевидных свойств вероятности.

Поскольку $\Omega = A + \bar{A}$, то по аксиоме сложения $P(\Omega) = P(A) + P(\bar{A})$ или, с учетом аксиомы нормированности,

1. $P(\bar{A}) = 1 - P(A)$ (*вероятность дополнительного события*).

Далее, поскольку $A = A + \emptyset$, то из аксиомы сложения имеем

2. $P(\emptyset) = 0$ (*вероятность невозможного события*).

Пусть $A \subset B$. Тогда $B = A + (B \setminus A)$, по аксиоме сложения $P(B) = P(A) + P(B \setminus A)$, и из аксиомы неотрицательности получаем

3. $P(A) \leq P(B)$ (*«большему» событию соответствует большая вероятность*).

В частности, так как всегда $A \subset \Omega$, то, с учетом аксиомы неотрицательности,

4. $0 \leq P(A) \leq 1$ (*вероятность заключена между 0 и 1*).

Наконец, поскольку $A \cup B = A + (B \setminus A)$ и $B = (B \setminus A) + AB$, то из аксиомы сложения находим: $P(A \cup B) = P(A) + P(B \setminus A)$ и $P(B \setminus A) = P(B) - P(AB)$. Следовательно,

5. $P(A \cup B) = P(A) + P(B) - P(AB)$ (*вероятность объединения двух событий*).

Последнее свойство допускает очевидное, но весьма полезное обобщение на случай произвольного числа слагаемых

6. $P(A_1 \cup \dots \cup A_n) = P(A_1) + \dots + P(A_n) - P(A_1A_2) - P(A_1A_3) - \dots - P(A_{n-1}A_n) + P(A_1A_2A_3) + \dots + (-1)^{n+1}P(A_1A_2 \dots A_n).$

Свойство 6 доказывается индукцией по n . Так, для трех событий A , B и C

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B \cup C) - P(A(B \cup C)) = \\ &= P(A) + P(B) + P(C) - P(BC) - P(AB \cup BC) = \\ &= P(A) + P(B) + P(C) - P(BC) - P(AB) - P(AC) + P(ABC). \end{aligned}$$

Из свойств 6 и 2 имеем для любого числа n (попарно) непересекающихся событий A_1, \dots, A_n

7. $P(A_1 + \dots + A_n) = P(A_1) + \dots + P(A_n).$

В случае, когда Ω содержит конечное число n элементарных исходов, вероятность можно определить конструктивно.

Действительно, с одной стороны, пусть на пространстве элементарных исходов Ω задана некоторая вероятность $P(A)$. Обозначим через p_1, \dots, p_n вероятности элементарных исходов $\omega_1, \dots, \omega_n$. Тогда по аксиоме сложения Р3 вероятность любого события A определяется формулой $P(A) = \sum P(\omega_i) = \sum p_i$, где суммирование ведется по всем индексам i , соответствующим входящим в событие A элементарным исходам. В силу аксиом неотрицательности Р1 и нормированности Р2 числа p_1, \dots, p_n являются неотрицательными и удовлетворяют свойству $p_1 + \dots + p_n = 1$.

С другой стороны, пусть p_1, \dots, p_n — любой набор неотрицательных чисел, таких, что $p_1 + \dots + p_n = 1$. Поставим в соответствие каждому элементарному исходу ω_i ($i = 1, \dots, n$) число $P(\omega_i) = p_i$, а любому событию A — число $P(A) = \sum P(\omega_i) = \sum p_i$, где суммирование ведется по всем индексам i , соответствующим входящим в событие A элементарным исходам. Очевидно, достоверному событию Ω мы должны сопоставить число $P(\Omega) = p_1 + \dots + p_n = 1$. Нетрудно видеть, что определенная таким образом функция $P(A)$ удовлетворяет аксиомам Р1–Р3, т. е. является вероятностью.

Итак, существует взаимно однозначное соответствие между всеми вероятностями $\mathbf{P}(A)$ на Ω и наборами p_1, \dots, p_n неотрицательных чисел, удовлетворяющими условию $p_1 + \dots + p_n = 1$.

В частности, мы можем всем элементарным исходам ω_i приписать одну и ту же вероятность $p_i = 1/n$ ($i = 1, \dots, n$). В этом случае реализуется так называемый принцип классической вероятности, о котором мы подробно поговорим в следующей главе.

В случае произвольного (не обязательно конечного) пространства элементарных исходов Ω аксиому Р3 необходимо заменить более сильной *расширенной аксиомой сложения*

$$\text{Р3'}. \quad \mathbf{P}(A_1 + \dots + A_n + \dots) = \mathbf{P}(A_1) + \dots + \mathbf{P}(A_n) + \dots,$$

справедливой для счетного числа попарно несовместных событий.

Именно аксиомы Р1, Р2 и Р3' определяют аксиоматическое понятие вероятности.

Очевидно, что свойства вероятности 1–7 сохраняются и в этом случае.

Пример 21. Пусть Ω состоит из счетного числа элементарных исходов $\omega_1, \dots, \omega_n, \dots$. И в этом случае любую вероятностную меру $\mathbf{P}(A)$ можно получить, задав вероятности $p_1 = \mathbf{P}(\omega_1), \dots, p_n = \mathbf{P}(\omega_n), \dots$ элементарных исходов, причем последовательность p_1, \dots, p_n, \dots должна удовлетворять только условиям неотрицательности $p_i \geq 0$ ($i = 1, 2, \dots$) и нормированности $p_1 + \dots + p_n + \dots = 1$. По-прежнему вероятность любого события A определяется как сумма $\sum p_i$ вероятностей всех входящих в A элементарных исходов ω_i , однако если событие A содержит бесконечное число элементарных исходов, то и сумма будет бесконечной. \square

Пример 22. Пусть пространство элементарных исходов Ω представляет собой прямую $(-\infty, \infty)$ с борелевской σ -алгеброй на нем (см. пример 20). Теперь уже в наиболее интересных случаях мы не можем приписать каждому элементарному исходу ω иной вероятности, кроме $\mathbf{P}(\omega) = 0$, и, следовательно, определить вероятность любого события на основе вероятностей входящих в него элементарных исходов. Тем не менее и сейчас вероятность можно задать конструктивно.

Для того чтобы показать это, предположим сначала, что она каким-то образом уже задана для всех событий (элементов борелевской σ -алгебры), и рассмотрим функцию $F(x) = \mathbf{P}(A_x)$, равную вероятности события $A_x = (-\infty, x]$, состоящего из всех точек полупрямой $(-\infty, x]$. Как вероятность функция $F(x)$ обязана обладать определенными свойствами, которые мы сейчас опишем.

Во-первых, значения функции $F(x)$ как вероятности должны лежать между 0 и 1.

Во-вторых, так как для любых $x_1 < x_2$ событие $(-\infty, x_1]$ содержится в событии $(-\infty, x_2]$, то $F(x_1) \leq F(x_2)$. Иными словами, $F(x)$ – неубывающая функция аргумента x .

В-третьих, поскольку событие $(-\infty, -\infty)$ невозможно, а событие $(-\infty, \infty)$ достоверно, то $F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0$, $F(\infty) = \lim_{x \rightarrow \infty} F(x) = 1$.

Наконец, так как событие $(-\infty, x]$ представляет собой объединение счетного числа событий $(-\infty, x - 1/n)$, то из расширенной аксиомы сложения

и монотонности $F(x)$ можно вывести (см. параграф 2 гл. 5), что $F(x)$ — непрерывная слева функция.

Зная функцию $F(x)$, можно определить вероятности любых других событий. В частности, вероятность события $A = [x_1, x_2]$ ($x_1 < x_2$) определяется формулой $\mathbf{P}(A) = F(x_2) - F(x_1)$.

Таким образом, любая вероятность на прямой полностью определяется своей функцией $F(x)$, которая удовлетворяет перечисленным выше свойствам.

Справедливо и обратное. Любая неубывающая непрерывная слева функция $F(x)$, удовлетворяющая условиям $F(-\infty) = 0$ и $F(\infty) = 1$, задает некоторую вероятность на прямой $(-\infty, \infty)$. Действительно, достаточно сопоставить каждому событию $A_x = (-\infty, x)$ число $\mathbf{P}(A_x) = F(x)$, а событию $A = [x_1, x_2]$ — число $\mathbf{P}(A) = F(x_2) - F(x_1)$. Можно показать, что определенная таким образом для всех событий $A = [x_1, x_2]$ числовая функция $\mathbf{P}(A)$ будет удовлетворять трем аксиомам вероятности. Для любых других событий, составляющих борелевскую σ -алгебру на прямой, вероятность определяется единственным образом с помощью так называемой теоремы о продолжении меры. \square

Глава 2

КЛАССИЧЕСКАЯ И ГЕОМЕТРИЧЕСКАЯ ВЕРОЯТНОСТИ

В этой главе мы рассмотрим некоторые вероятностные пространства, объединенные интуитивным понятием симметрии или «равновероятности». В соответствии с тем, какое пространство элементарных исходов рассматривается — конечное или непрерывное, понятие «равновероятности» реализуется в двух схемах: *классической* и *геометрической вероятности*. Как всегда, право на жизнь вышеперечисленных схем определяется практикой. В различных учебниках приводятся результаты многочисленных статистических опытов, подтверждающих корректность понятия «равновероятность».

1. Классическая вероятность

Понятие классической вероятности мы рассмотрим сначала на примере нашей «палочки-выручалочки» — монеты. Предположим, что опыт состоит в однократном подбрасывании монеты. Как мы теперь знаем, пространство элементарных исходов Ω содержит два элементарных исхода: ω_1 — выпадение «герба» и ω_2 — выпадение «цифры», а σ -алгебра \mathfrak{B} насчитывает 4 события: \emptyset , $\{\omega_1\}$, $\{\omega_2\}$ и Ω . Ясно, что обычная монета обладает свойством симметрии, так как у нас нет оснований предпочесть «герб» «цифре», т. е. элементарный исход ω_1 элементарному исходу ω_2 . Поэтому естественно сопоставить обоим элементарным исходам одинаковую вероятность $P(\omega_1) = P(\omega_2)$. Так как согласно аксиоме сложения $P(\Omega) = P(\omega_1) + P(\omega_2)$, а в силу аксиомы нормированности $P(\Omega) = 1$, то получаем $P(\omega_1) = P(\omega_2) = 1/2$. Таким образом, каждому из четырех имеющихся в σ -алгебре \mathfrak{B} событий мы ставим в соответствие вероятности: $P(\emptyset) = 0$, $P(\omega_1) = P(\omega_2) = 1/2$, $P(\Omega) = 1$.

Перейдем теперь к общему случаю. Пусть пространство элементарных исходов Ω состоит из конечного числа n равнозначных исходов $\omega_1, \omega_2, \dots, \omega_n$ (σ -алгебра событий содержит 2^n событий). Тогда каждому элементарному исходу ω_i ($i = 1, \dots, n$) поставим в соответствие одну и ту же вероятность $P(\omega_i) = 1/n$. Ясно, что в силу аксиомы сложения для определения вероятности любого события A необходимо подсчитать число m элементарных исходов ω , содержащихся в A , и затем положить

$$P(A) = \frac{m}{n}.$$

Таким образом, в классической схеме вероятность любого события A определяется как отношение числа m благоприятных для события A элементарных исходов к общему числу элементарных исходов n .

Пример 1. Определим вероятность выпадения на игральной кости четного числа очков (событие A). В этом случае общее число элементарных исходов $n = 6$ (σ -алгебра \mathfrak{B} состоит из $2^6 = 64$ событий), а число благоприятных исходов $m = 3$ (выпадение «двойки», «четверки» и «шестерки»). Искомая вероятность $P(A) = 3/6 = 1/2$. \square

Пример 2. Производится трехкратное подбрасывание монеты. Определим вероятность события A , заключающегося в выпадении «герба» хотя бы один раз. Выпишем все элементарные исходы:

$$\begin{array}{ll} \omega_1 = \text{«герб»} - \text{«герб»} - \text{«герб»}, & \omega_2 = \text{«герб»} - \text{«цифра»} - \text{«герб»}, \\ \omega_3 = \text{«герб»} - \text{«герб»} - \text{«цифра»}, & \omega_4 = \text{«герб»} - \text{«цифра»} - \text{«цифра»}, \\ \omega_5 = \text{«цифра»} - \text{«герб»} - \text{«герб»}, & \omega_6 = \text{«цифра»} - \text{«цифра»} - \text{«герб»}, \\ \omega_7 = \text{«цифра»} - \text{«герб»} - \text{«цифра»}, & \omega_8 = \text{«цифра»} - \text{«цифра»} - \text{«цифра»}. \end{array}$$

Всего имеем $n = 8$ элементарных исходов (σ -алгебра \mathfrak{B} состоит из $2^8 = 256$ событий). Благоприятными из них для события A являются $m = 7$ исходов: $\omega_1, \omega_2, \dots, \omega_7$. Значит, $P(A) = 7/8$.

Вероятность $P(\bar{A})$ можно подсчитать и другим способом. Дополнительным к A будет событие \bar{A} , заключающееся в невыпадении ни одного «герба». Событие \bar{A} состоит только из одного элементарного исхода ω_8 , поэтому $P(\bar{A}) = 1/8$. Переходя снова к событию A , имеем $P(A) = 1 - P(\bar{A}) = 7/8$. Отметим, что привлечение дополнительного события позволяет иногда существенно упростить численный подсчет вероятности. \square

Пример 3. Найдем вероятность того, что при бросании двух игральных костей в сумме выпадает не менее четырех очков (событие A). Поскольку при бросании двух игральных костей может выпасть от 2 до 12 очков, а рассматриваемое событие A состоит в выпадении 4, 5, ..., 12 очков, то удобно перейти к дополнительному событию \bar{A} — выпадению двух или трех очков. Пространство элементарных исходов состоит из 36 исходов — пар (1,1), (1,2), (2,1), (1,3) и т. д. (заметим, что пары (1,2) и (2,1) представляют собой разные элементарные исходы, поскольку выпадение одного очка на первой кости и двух на второй — не то же самое, что двух очков на первой кости и одного очка на второй). Благоприятными для события \bar{A} будут элементарные исходы (1,1), (1,2) и (2,1). Значит, $m_{\bar{A}} = 3$, $P(\bar{A}) = 3/36 = 1/12$ и $P(A) = 1 - P(\bar{A}) = 11/12$. \square

Пример 4. Из колоды в 36 игральных карт наудачу выбирается одна. Определим вероятность того, что она окажется тузом (событие A). Из колоды мы можем выбрать любую из 36 карт ($n = 36$). Тузов в колоде 4 ($m = 4$). Таким образом, $P(A) = 4/36 = 1/9$. \square

2. Элементы комбинаторики в теории вероятностей

Примеры, рассмотренные в предыдущем параграфе, имели ту характерную особенность, что для них нетрудно было подсчитать как общее число элементарных исходов, так и число исходов, благоприятных для данного события. Однако именно этот подсчет и представляет наибольшую трудность при решении более сложных задач на классическую вероятность. Для того чтобы иметь некоторые стандартные приемы при

расчетах по схеме классической вероятности, приведем основную формулу комбинаторики и рассмотрим понятия перестановки, размещения и сочетания.

Пусть имеется k групп элементов, причем i -я группа состоит из n_i элементов. Выберем по одному элементу из каждой группы. Тогда общее число N способов, которыми можно произвести такой выбор, определяется соотношением

$$N = n_1 \cdot n_2 \cdots n_k,$$

называемым *основной формулой комбинаторики*.

Для доказательства этой формулы рассмотрим сначала случай $k = 2$ и перенумеруем все элементы первой группы числами от 1 до n_1 , а второй — от 1 до n_2 . Тогда каждый возможный способ выбора двух элементов отождествим с парой чисел (i, j) , где $i = 1, \dots, n_1$, $j = 1, \dots, n_2$. Очевидно, что таких пар $n_1 n_2$. Для окончания доказательства достаточно воспользоваться методом математической индукции. Так, для $k = 3$ всевозможные способы выбора трех элементов можно отождествить с тройками (i, j, l) . Поскольку первые два элемента можно выбрать $n_1 n_2$ способами, то все три элемента можно выбрать $N = (n_1 n_2) n_3 = n_1 n_2 n_3$ способами.

В том случае, когда все группы состоят из одинакового числа элементов, т. е. $n_1 = n_2 = \dots = n_k$, можно считать, что каждый раз выбор производится из одной и той же группы, причем элемент после выбора снова возвращается в группу. Тогда число всех способов выбора равно n^k . Такой способ выбора носит название *выборки с возвращением*.

Перестановкой из n элементов называется любой *упорядоченный набор* этих элементов. Так, всевозможными перестановками чисел 1, 2, 3 являются: (1, 2, 3), (1, 3, 2), (2, 1, 3), (2, 3, 1), (3, 1, 2) и (3, 2, 1). Для определения числа различных перестановок из n элементов, которое мы будем обозначать через P_n , заметим, что на первом месте перестановки может стоять любой из n элементов, на втором — любой из $n - 1$ оставшихся, на третьем — любой из остальных $n - 2$ и т. д. В силу основной формулы комбинаторики (в данном случае мы имеем n групп элементов размеров $n, n - 1, \dots, 1$) получаем

$$P_n = n(n - 1) \cdots 2 \cdot 1 = n!.$$

Размещением из n элементов по m называется любой *упорядоченный набор* из m различных элементов, выбранных из общей совокупности в n элементов. Выпишем для примера все размещения из четырех чисел 1, 2, 3, 4 по два: (1, 2), (1, 3), (1, 4), (2, 1), (2, 3), (2, 4), (3, 1), (3, 2), (3, 4), (4, 1), (4, 2), (4, 3). Число размещений $(n)_m$ (используется также запись A_n^m) подсчитывается точно так же, как и число перестановок: на первом месте может находиться любой из n элементов, на втором — любой из $n - 1$ оставшихся, ..., на m -м месте — любой из $n - m + 1$ элементов. Снова воспользовавшись основной формулой комбинатори-

ки (выбор осуществляется из групп размеров $n, n - 1, \dots, n - m + 1$), имеем

$$(n)_m = n(n - 1) \dots (n - m + 1) = \frac{n(n - 1) \dots 2 \cdot 1}{(n - m)(n - m - 1) \dots 2 \cdot 1} = \frac{n!}{(n - m)!}.$$

Заметим, что способ выбора, приводящий к перестановкам и размещениям, носит название *выборки без возвращения*.

Сочетанием из n элементов по m называется любой *неупорядоченный набор* из m различных элементов, выбранных из общей совокупности в n элементов. Сочетаниями из четырех чисел 1, 2, 3, 4 по два являются: (1,2), (1,3), (1,4), (2,3), (2,4) и (3,4). Для определения числа сочетаний $\binom{n}{m}$ (употребляется также запись C_n^m) заметим, что сочетание от размещения отличается только тем, что входящие в него элементы неупорядочены. Но, как мы знаем, m элементов можно упорядочить $m!$ способами. Значит, каждое сочетание соответствует $m!$ размещениям. Поэтому $(n)_m = m! \binom{n}{m}$ или

$$\binom{n}{m} = \frac{(n)_m}{m!} = \frac{n!}{(n - m)! m!}.$$

Вооружившись знанием формул для чисел перестановок, размещений и сочетаний, продолжим рассмотрение задач на классическую вероятность.

Пример 5. На четырех карточках написаны буквы *в*, *л*, *к* и *о*. Карточки перемешиваются и выкладываются в ряд. Найдем вероятность того, что образовавшееся слово будет «волк» (событие A). В соответствии с комбинаторными принципами для определения общего числа элементарных исходов нужно подсчитать число упорядоченных наборов из четырех букв. Мы имеем дело с числом перестановок, поэтому число элементарных исходов $n = 4! = 24$. Слово «волк» образует только одна перестановка, т. е. число благоприятных для события A элементарных исходов $m = 1$. Поэтому $P(A) = 1/24$. \square

Пример 6. Из колоды в 36 игральных карт вынимаются наудачу 3 карты. Найдем вероятность того, что все эти 3 карты будут одной масти (событие A). Для большей наглядности приведем два решения.

В первом решении будем предполагать, что выбор производится последовательно по одной карте и нужно учитывать его порядок. Тогда результат выбора отождествим с размещением из 36 карт по 3, и общее число элементарных исходов $n = (36)_3 = 36 \cdot 35 \cdot 34$. Для подсчета общего числа благоприятных исходов предположим сначала, что мы последовательно вынимаем карты пиковой масти. Поскольку «пик» в колоде 9, то число способов, которыми мы можем последовательно вынуть 3 карты пиковой масти, равно числу размещений из 9 карт по 3, т. е. $(9)_3 = 9 \cdot 8 \cdot 7$. Но благоприятными для нас будут также такие ситуации, при которых мы вынимаем 3 «трефы», 3 «бубны», 3 «червы». Поэтому для определения общего числа благоприятных исходов нужно число размещений из 9 по 3 умножить на 4: $m = 4 \cdot (9)_3 = 4 \cdot 9 \cdot 8 \cdot 7$. Значит, $P(A) = 4 \cdot (9)_3 / (36)_3 = 4/85$.

Во втором решении мы не будем учитывать порядок выбора карт. Тогда общее число элементарных исходов определяется уже как число сочетаний из

36 карт по 3, т. е. $n = \binom{36}{3} = \frac{36 \cdot 35 \cdot 34}{1 \cdot 2 \cdot 3}$. Однако и при подсчете числа благоприятных исходов мы должны помнить, что порядок выбора несуществен, т. е. $m = 4 \cdot \binom{9}{3} = \frac{4 \cdot 9 \cdot 8 \cdot 7}{1 \cdot 2 \cdot 3}$. Нетрудно видеть, что окончательное значение $P(A)$ будет тем же самым, что и в первом решении. Рекомендуем любознательному читателю еще раз разобрать этот пример и объяснить, почему в обоих случаях получился один и тот же ответ. \square

В заключение этого параграфа мы рассмотрим некоторые наиболее часто встречающиеся в физической практике задачи о распределении частиц по ячейкам.

Гипергеометрическое распределение. Предположим, что имеется $n = n_1 + \dots + n_k$ различных частиц, причем из них n_1 частиц первого типа, n_2 — второго типа, ..., n_k — k -го типа. Случайным образом из этих n частиц выбирается $m = m_1 + \dots + m_k$ частиц. Найдем вероятность того, что среди выбранных окажется ровно $m_1 \leq n_1$ частиц первого типа, $m_2 \leq n_2$ — второго типа, ..., $m_k \leq n_k$ — k -го типа (событие A). Поскольку порядок выбора несуществен, то при определении общего числа исходов и числа благоприятных исходов мы должны пользоваться числом сочетаний. Общее число элементарных исходов есть число сочетаний из n частиц по $m - \binom{n}{m}$. Далее, m_1 частиц первого типа мы можем выбрать $\binom{n_1}{m_1}$ способами, m_2 частиц второго типа $\binom{n_2}{m_2}$ способами, ..., m_k частиц k -го типа $\binom{n_k}{m_k}$ способами. При этом любой способ выбора частиц определенного типа комбинирует с любыми способами выбора частиц остальных типов и, значит, число благоприятных событию A исходов равно $\binom{n_1}{m_1} \binom{n_2}{m_2} \dots \binom{n_k}{m_k}$. Поэтому

$$P(A) = P(m_1, \dots, m_k) = \frac{\binom{n_1}{m_1} \binom{n_2}{m_2} \dots \binom{n_k}{m_k}}{\binom{n}{m}}.$$

Определенные последней формулой вероятности $P(m_1, \dots, m_k)$ носят название гипергеометрического распределения.

Пример 7. Найдем вероятность того, что в «Спортлото 6 из 49»¹⁾ будет угадано 3 номера (событие A_3), 4 номера (A_4), 5 номеров (A_5) и 6 номеров (A_6). Мы имеем дело с гипергеометрическим распределением, в котором $n = 49$,

¹⁾ В конце прошлого столетия в нашей стране были широко распространены разные варианты игры «Спортлото», наиболее популярным из которых являлся вариант «Спортлото 6 из 49». Купив карточку за 30 коп. (что примерно соответствует нынешним 30 руб.), нужно было вычеркнуть 6 номеров из 49 возможных. Аналогично, во время розыгрыша из 49 шаров выбирались 6. Угадавший от 3 до 6 номеров, мог получить выигрыш, который для всех 6 угаданных номеров составлял 10 000 тыс. руб.

$k = 2$, $n_1 = 6$, $n_2 = 43$ и $m = 6$. Для события A_3 $m_1 = 3$ и $m_2 = 3$, для события A_4 $m_1 = 4$ и $m_2 = 2$, для события A_5 $m_1 = 5$ и $m_2 = 1$ и, наконец, для события A_6 $m_1 = 6$ и $m_2 = 0$. Поэтому

$$\mathbf{P}(A_3) = \frac{\binom{6}{3} \binom{43}{3}}{\binom{49}{6}} \cong 0,0176, \quad \mathbf{P}(A_4) = \frac{\binom{6}{4} \binom{43}{2}}{\binom{49}{6}} \cong 0,00097,$$

$$\mathbf{P}(A_5) = \frac{\binom{6}{5} \binom{43}{1}}{\binom{49}{6}} \cong 0,000\,018, \quad \mathbf{P}(A_6) = \frac{\binom{6}{6} \binom{43}{0}}{\binom{49}{6}} \cong 7 \cdot 10^{-8}. \quad \square$$

Статистика Бозе–Эйнштейна. Предположим, что n неразличимых частиц распределяются по m ячейкам. Различными считаются распределения частиц по ячейкам, отличающиеся только числом попавших в каждую ячейку частиц. Такое распределение носит в физике название статистики Бозе–Эйнштейна. Найдем общее число различных размещений в статистике Бозе–Эйнштейна (число элементарных исходов). Для этого рассмотрим последовательность из $n + m - 1$ элементов (рис. 1) и выберем из них $m - 1$ «черный» элемент. Если

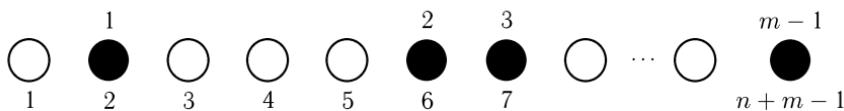


Рис. 1

считать «белый» элемент частицей, а «черный» — перегородкой, то, как нетрудно видеть, существует взаимно однозначное соответствие между способами выбора $m - 1$ «черного» элемента и размещениями частиц в статистике Бозе–Эйнштейна. Так, на приведенном рисунке в первую ячейку попала одна частица, во вторую — три, третья оказалась пустой, ..., последняя, m -я ячейка также оказалась пустой.

Поэтому общее число размещений равно $\binom{n+m-1}{m-1}$. Найдем теперь вероятность попадания в фиксированную ячейку ровно k частиц (событие A). Заметим, что если в этой фиксированной ячейке уже находится k частиц, то остальные $n - k$ частиц должны быть распределены по оставшимся $m - 1$ ячейкам, а это, как мы знаем, можно сделать $\binom{n-k+m-1-1}{m-1-1} = \binom{n+m-k-2}{m-2}$ способами. Следовательно, $\mathbf{P}(A) = \binom{n+m-k-2}{m-2} / \binom{n+m-1}{m-1}$. Отметим, что статистике Бозе–Эйнштейна подчиняются фотоны, атомные ядра и атомы, содержащие четное число частиц.

Статистика Ферми–Дирака. В статистике Ферми–Дирака так же, как и в статистике Бозе–Эйнштейна, n неразличимых частиц распре-

деляются по m ячейкам ($n \leq m$), однако в каждой ячейке не может находиться более одной частицы. Число различных размещений (элементарных исходов) совпадает с числом способов, которыми мы можем выбрать (без учета порядка выбора) n занятых ячеек из общего числа ячеек m , т. е. равно $\binom{m}{n}$. Пусть событие A — заняты фиксированные k ячеек ($k \leq n$). Тогда оставшиеся $m - k$ ячеек должны быть заполнены $n - k$ частицами, а это можно сделать $\binom{m - k}{n - k}$ способами. Поэтому $P(A) = \binom{m - k}{n - k} / \binom{m}{n}$. Статистике Ферми–Дирака подчиняются электроны, протоны и нейтроны.

Статистика Максвелла–Больцмана. Предполагая, что n различных частиц распределяются по m ячейкам без ограничений на число попавших в каждую ячейку частиц, получаем статистику Максвелла–Больцмана. Поскольку каждая из n частиц может попасть в любую из m ячеек, то общее число элементарных исходов равно m^n . Событие A заключается в том, что в первую ячейку попало n_1 частиц, во вторую — n_2 частиц, ..., в m -ю — n_m частиц ($n_1 + \dots + n_m = n$). Число благоприятных для события A исходов подсчитаем следующим образом. В первую ячейку могут попасть любые n_1 частиц из n имеющихся первоначально. Поскольку порядок выбора частиц несуществен, то это можно сделать $\binom{n}{n_1}$ способами. Как только первая ячейка заполнена, у нас остается $n - n_1$ частиц, и вторую ячейку мы можем заполнить $\binom{n - n_1}{n_2}$ различными способами. Продолжая эту процедуру и используя основную формулу комбинаторики, получаем, что число благоприятных событию A способов равно

$$\binom{n}{n_1} \binom{n - n_1}{n_2} \dots \binom{n - n_1 - \dots - n_{m-1}}{n_m} = \\ = \frac{n!}{n_1! (n - n_1)!} \frac{(n - n_1)!}{n_2! (n - n_1 - n_2)!} \dots \frac{(n - n_1 - \dots - n_{m-1})!}{n_m! 0!} = \frac{n!}{n_1! n_2! \dots n_m!}.$$

Таким образом,

$$P(A) = \frac{n!}{n_1! n_2! \dots n_m! \cdot m^n}.$$

Отметим, что статистика Максвелла–Больцмана представляет собой частный случай так называемой полиномиальной схемы, которую мы рассмотрим в параграфе 7 гл. 4. Статистике Максвелла–Больцмана подчиняется идеальный газ.

Пример 8. Поток из 4 частиц поступает в счетчик, состоящий из трех датчиков. Каждая частица с одинаковой вероятностью может попасть в один и только один из этих датчиков. Поток считается зарегистрированным, если он отмечен хотя бы двумя датчиками. Найдем вероятность события A ,

заключающегося в том, что поток будет зарегистрирован. Удобно перейти к дополнительному событию \bar{A} , которое происходит тогда и только тогда, когда все 4 частицы попадают либо в первый, либо во второй, либо в третий датчик. Но вероятность всем 4 частицам попасть в первый датчик определяется статистикой Максвелла–Больцмана, причем $n = 4$, $m = 3$, $n_1 = 4$, $n_2 = 0$, $n_3 = 0$. Учитывая, что вероятность попадания всех 4 частиц во второй и третий датчики точно такая же, как и в первый, получаем

$$\mathbf{P}(\bar{A}) = 3 \cdot \frac{1}{3^4} \cdot \frac{4!}{4! 0! 0!} = \frac{1}{27}$$

и $\mathbf{P}(A) = 1 - 1/27 = 26/27$. □

3. Геометрическая вероятность

Пусть теперь рассматривается непрерывная вероятностная схема, т. е. пространство элементарных исходов представляет собой некоторую ограниченную область (отрезок, многоугольник, круг, параллелепипед, шар и т. п.) k -мерного пространства (прямой, плоскости, трехмерного пространства и т. д.). Естественно желание обобщить принцип равновероятности элементарных исходов классической вероятности и на эту схему. Однако в непрерывном случае число элементарных исходов бесконечно и, воспользовавшись принципом равновероятности, мы не смогли бы приписать каждому элементарному исходу иной вероятности, кроме нуля. Поэтому подойдем к определению геометрической вероятности по-другому. Рассмотрим сначала отрезок $[0, 1]$ и предположим, что идеальная частица равномерно бросается на этот отрезок. Понятию равномерности придадим следующий смысл. Каждому отрезку $[a, b]$ ($0 \leq a \leq b \leq 1$), независимо от его расположения, поставим в соответствие одинаковую вероятность попадания частицы на этот отрезок, равную его длине: $\mathbf{P}[a, b] = b - a$, а затем эту вероятность попытаемся с помощью трех аксиом P1, P2 и P3' продолжить на любое подмножество точек отрезка $[0, 1]$. Очевидно, что вероятность попадания частицы в любую точку x равна нулю, вероятность попадания на любой интервал (a, b) или полуинтервал $(a, b]$, $[a, b)$ (на отрезке $[0, 1]$) равна $b - a$, вероятность попадание в любое множество точек на отрезке $[0, 1]$, состоящее из конечного и даже счетного объединения непересекающихся отрезков, интервалов и полуинтервалов равна сумме их длин, т. е. «длине», или лучше сказать, мере этого множества. В частности, вероятность попадания частицы в множество рациональных точек равна нулю.

Однако, как уже говорилось, имеется препятствие к такому продолжению, связанное с существованием подмножеств, которым разумным образом с сохранением трех аксиом вероятность мы никак не сможем приписать. Поэтому приходится ограничиваться только элементами борелевской σ -алгебры \mathfrak{B} , порожденной всевозможными интервалами (т. е. подмножествами, имеющими меру), что, впрочем, более чем достаточно для практических потребностей.

В общем случае геометрическая вероятность определяется совершенно аналогично. Пусть Ω — некоторая область, имеющая меру $\mu(\Omega)$ (длину, площадь, объем и т. д.), такую, что $0 < \mu(\Omega) < \infty$. Скажем, что точка *равномерным образом* попадает в Ω (*реализуется принцип геометрической вероятности*), если *вероятность* $P(A)$ попадания ее в каждую область A , являющуюся подобластью Ω , пропорциональна *мере* этой области $\mu(A)$ или в силу аксиомы нормированности

$$P(A) = \frac{\mu(A)}{\mu(\Omega)}.$$

Пример 9. В круг радиусом $R = 1$ равномерно бросается точка. Найдем вероятность события A , заключающегося в попадании этой точки в круг радиусом $r = 1/2$ с тем же центром (рис. 2).

Рассмотрим два способа решения этой задачи:

1) вероятность $P(A)$ определяется как отношение площади внутреннего круга к площади внешнего:

$$P(A) = \frac{\pi r^2}{\pi R^2} = \frac{1}{4};$$

2) заметим, что в силу принципа геометрической вероятности как угол φ , так и расстояние ρ от точки ω до центра O должно быть распределено равномерно. Поскольку точки, равноотстоящие от центра, все либо одновременно принадлежат меньшему кругу, либо нет, то вероятность попадания в этот круг равна отношению радиусов: $P(A) = r/R = 1/2$.

Итак, мы получили в одной и той же задаче два разных ответа. Причина кроется в том, что понятие геометрической вероятности *не инвариантно относительно преобразований* рассматриваемой области Ω . В частности, в нашем примере при втором способе решения мы считаем, что равновероятно попадание точки в области A_1 и A_2 , заштрихованные на рис. 3. Но с точки зрения обычного понятия площади, используемого при первом способе решения, это не так. Значит, вероятность существенно зависит от способа определения понятия «равновероятно» или, иными словами, от того, как мы задали меру $\mu(A)$. Именно на неинвариантности понятия геометрической вероятности относительно преобразований основаны многочисленные парадоксы, часто приводимые в различных учебниках.

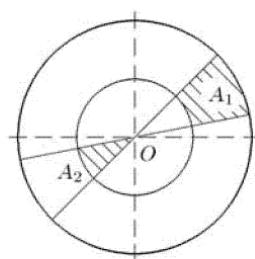


Рис. 3

Возвращаясь к рассматриваемому примеру, отметим, что в приведенной постановке задачи предпочтительным нужно считать первый способ. Однако не следует думать, что второе решение относится к числу математических фокусов. Представляемая этим решением модель сигнала с равномерно распределенными фазой (углом) и амплитудой (радиусом) находит широкое применение в статистической радиофизике. □

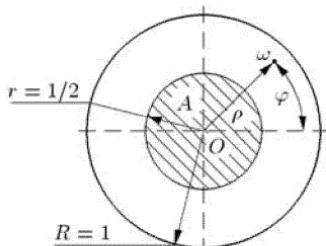


Рис. 2

Пример 10. На Землю параллельно плоскости экватора падает поток метеоритов. Найдем вероятность того, что упавший метеорит попадет между 15° и 45° северной широты (событие A). Естественно предполагать, что поток метеоритов равномерно распределен на плоскости, перпендикулярной плоскости экватора. Если мы теперь спроектируем земной шар на эту плоскость (рис. 4), то получим, что вероятность наступления события A пропорциональна площади $\mu(A)$ заштрихованной области A . Определим $\mu(A)$:

$$\mu(A) = 2R_3^2 \int_{\pi/12}^{\pi/4} \cos \varphi d\sin \varphi = 2R_3^2 \int_{\pi/12}^{\pi/4} \cos^2 \varphi d\varphi = \frac{R_3^2}{2} \int_{\pi/6}^{\pi/2} (1 + \cos \varphi) d\varphi = R_3^2 \left(\frac{\pi}{6} + \frac{1}{4} \right).$$

Окончательно получаем

$$\mathbb{P}(A) = R_3^2 \frac{\frac{\pi}{6} + \frac{1}{4}}{\pi R_3^2} = \frac{1}{4\pi} + \frac{1}{6} \cong 0,246.$$
□

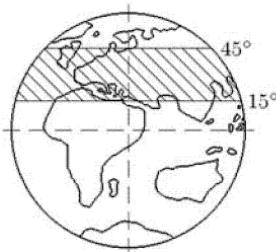


Рис. 4

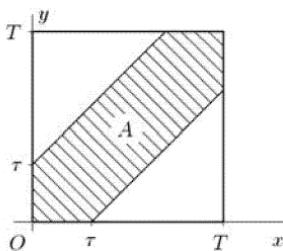


Рис. 5

Пример 11. Поступление каждого из двух сигналов в приемник равновозможно в любой момент промежутка времени T . Найдем вероятность того, что приемник будет «забит» (событие A), что происходит в том случае, когда промежуток времени между моментами поступления обоих сигналов меньше τ . Для этого обозначим моменты поступления сигналов через x и y . Ясно, что для наложения сигналов необходимо и достаточно, чтобы $|x - y| \leq \tau$. Изобразим x и y как точки внутри квадрата со сторонами T (рис. 5). Тогда исходы, благоприятные для наложения сигналов, представляются заштрихованной областью A . В силу принципа геометрической вероятности искомая вероятность равна отношению площади заштрихованной фигуры к площади всего квадрата:

$$\mathbb{P}(A) = \frac{T^2 - (T - \tau)^2}{T^2} = 1 - \left(1 - \frac{\tau}{T}\right)^2.$$
□

Пример 12 (задача Бюффона). Плоскость разграфлена параллельными прямыми, отстоящими друг от друга на расстояние $2a$. На плоскость наудачу бросается тонкая игла длиной $2l$ ($l < a$). Найдем вероятность того, что игла пересечет какую-нибудь прямую (рис. 6). Для этого прежде всего решим, что в данном случае соответствует понятию «наудачу». Ясно, что если игла бросается с достаточной высоты и ее начальное положение случайно, то под словом «наудачу» естественно подразумевать следующее: во-первых, центр иглы наудачу попадет на отрезок длиной $2a$, во-вторых, угол φ между игрой и прямой равномерно распределен на $[0, \pi]$ и, в-третьих, на величину угла

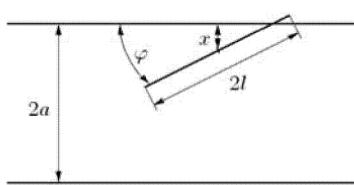


Рис. 6

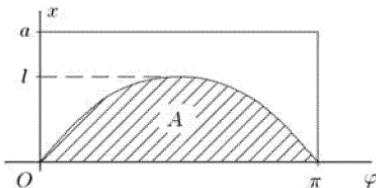


Рис. 7

не влияет расстояние от центра до прямой. Поэтому изобразим результат бросания точкой с координатами (φ, x) , лежащей внутри прямоугольника со сторонами π и a (рис. 7), где x — расстояние от центра иглы до ближайшей прямой. Из рис. 6 видно, что пересечение иглы с прямой происходит тогда и только тогда, когда $x < l \sin \varphi$. Искомая вероятность равна отношению площади заштрихованной области A к площади прямоугольника:

$$P(A) = \frac{1}{a\pi} \int_0^\pi l \sin \varphi \, d\varphi = \frac{2l}{a\pi}.$$

Задача Бюффона может быть использована для экспериментального определения числа π . Так, Вольф (Цюрих) бросал иглу 5000 раз и получил 2532 пересечения с прямыми; при этом $2a = 45$ мм, $2l = 36$ мм. Заменяя теперь $P(A)$ частотой пересечения прямых $f = 2532/5000 = 0,5064$, получаем эмпирическое значение числа π :

$$\pi = \frac{2l}{af} = 3,1596.$$

Более точное определение числа π таким путем вряд ли возможно, поскольку, с одной стороны, здесь влияют физические особенности опыта (толщина иглы, неточность при определении факта пересечения и т. д.), с другой стороны, как мы увидим дальше, необходимо производить очень большое число испытаний. \square

Г л а в а 3

УСЛОВНАЯ ВЕРОЯТНОСТЬ. НЕЗАВИСИМОСТЬ СОБЫТИЙ. ФОРМУЛЫ ПОЛНОЙ ВЕРОЯТНОСТИ И БАЙЕСА

Одним из основных понятий при аксиоматическом построении теории вероятностей является понятие условной вероятности. Именно условная вероятность оценивает то изменение в степени уверенности о наступлении некоторого события, которое происходит после получения дополнительной информации. С помощью условной вероятности определяется независимость событий, формализующая понятие не связанных между собой опытов. Использование условной вероятности в ряде случаев позволяет существенно упростить решение задачи (формулы умножения вероятностей, полной вероятности и Байеса).

1. Условная вероятность

Рассмотрим два события A и B . Пусть известно, что событие A наступило, но неизвестно, какой конкретно из элементарных исходов ω , составляющих событие A , произошел. Что можно сказать в этом случае о вероятности наступления события B ?

Пример 1. Событие A — выпадение четного числа очков на игральной кости, событие B — выпадение нечетного числа очков. Поскольку события A и B несовместны, то при наступлении события A событие B уже не может произойти и ему естественно приписать условную вероятность 0. \square

Пример 2. Событие A — выпадение 4 или 6 очков на игральной кости, событие B — выпадение четного числа очков. Так как событие A принадлежит событию B , то при наступлении события A событие B обязательно произойдет, т. е. событие B имеет условную вероятность 1. \square

Пример 3. Событие A — выпадение четного числа очков на игральной кости, событие B — выпадение не менее 5 очков. Если событие A наступило, то произошел один из трех элементарных исходов: выпало 2, 4 или 6 очков. Но из этих трех исходов только один исход (выпадение «шестерки») влечет за собой появление события B . В соответствии с принципом классической вероятности в данном случае естественно определить условную вероятность события B числом $1/3$. Заметим, что в этом примере условная вероятность появления события B совпадает с безусловной. \square

Из приведенных примеров видно, что условная вероятность может как совпадать с безусловной вероятностью, так и быть меньше или больше ее. Саму же условную вероятность $P(B|A)$ события B при условии события A в рамках классической схемы естественно определить как отношение числа исходов m_{AB} , благоприятных для

совместного осуществления событий A и B , к числу исходов m_A , благоприятных для события A , т. е.

$$\mathbf{P}(B | A) = \frac{m_{AB}}{m_A}.$$

Поделим теперь числитель и знаменатель полученного выражения на общее число n элементарных исходов:

$$\mathbf{P}(B | A) = \frac{\frac{m_{AB}}{n}}{\frac{m_A}{n}} = \frac{\mathbf{P}(AB)}{\mathbf{P}(A)}.$$

Последняя формула уже может служить общим определением условной вероятности при аксиоматическом подходе. Итак, *условной вероятностью события B при условии события A ($\mathbf{P}(A) \neq 0$) называется отношение вероятности пересечения событий A и B к вероятности события A :*

$$\mathbf{P}(B | A) = \frac{\mathbf{P}(AB)}{\mathbf{P}(A)}.$$

Нетрудно видеть, что условная вероятность обладает всеми свойствами безусловной вероятности. Так,

$$\mathbf{P}(\Omega | A) = 1, \quad \mathbf{P}(\emptyset | A) = 0, \quad \mathbf{P}(C + B | A) = \mathbf{P}(C | A) + \mathbf{P}(B | A).$$

Иногда бывает полезным равенство

$$\mathbf{P}(\bar{B} | A) = 1 - \mathbf{P}(B | A),$$

вытекающее из соотношения

$$1 = \mathbf{P}(\Omega | A) = \mathbf{P}(B + \bar{B} | A) = \mathbf{P}(B | A) + \mathbf{P}(\bar{B} | A).$$

Пример 4. При переписи населения Англии и Уэльса в 1891 г. оказалось, что темноволосые отцы и темноволосые сыновья составляют 5 % обследованных, темноволосые отцы и светловолосые сыновья — 7,9 %, светловолосые отцы и темноволосые сыновья — 8,9 %, светловолосые отцы и светловолосые сыновья — 78,2 %. Найдем условные вероятности рождения светловолосого сына у темноволосого и светловолосого отцов. Пусть событие A означает, что в случайно выбранной паре отец—сын светловолосым является отец, событие B — светловолосый сын. Тогда событие \bar{A} — темноволосый отец, событие \bar{B} — темноволосый сын и из результатов переписи следует, что $\mathbf{P}(\bar{A}\bar{B}) = 0,05$, $\mathbf{P}(\bar{A}B) = 0,079$, $\mathbf{P}(A\bar{B}) = 0,089$, $\mathbf{P}(AB) = 0,782$. Поэтому

$$\mathbf{P}(A) = \mathbf{P}(A\bar{B}) + \mathbf{P}(AB) = 0,871, \quad \mathbf{P}(\bar{A}) = 0,129,$$

и условная вероятность у темноволосого отца родиться светловолосому сыну

$$\mathbf{P}(B | \bar{A}) = \frac{\mathbf{P}(\bar{A}B)}{\mathbf{P}(\bar{A})} = \frac{0,079}{0,129} \cong 0,612,$$

а у светловолосого —

$$\mathbf{P}(B | A) = \frac{\mathbf{P}(AB)}{\mathbf{P}(A)} = \frac{0,782}{0,871} \cong 0,898. \quad \square$$

Пример 5. При трехкратном подбрасывании симметричной монеты выпало два «герба» (событие A). Определим условную вероятность того, что при втором подбрасывании выпал «герб» (событие B). В этом случае событие AB состоит в выпадении двух «гербов», один из которых выпал при втором подбрасывании, т. е. содержит $m_{AB} = 2$ элементарных исходов: «герб»—«герб»—«цифра» и «цифра»—«герб»—«герб». Поскольку всего у нас $n = 8$ элементарных исходов и мы находимся в рамках классической схемы, то $P(AB) = 1/4$. Аналогично, событию A благоприятствуют 3 исхода, и $P(A) = 3/8$. Итак, искомая вероятность

$$P(B|A) = \frac{P(AB)}{P(A)} = \frac{2}{3}.$$

□

Пример 6. Электрон может быть с вероятностью 0,6 обнаружен на одной из четырех орбит атома, причем с равной вероятностью — на любой из них, и с дополнительной вероятностью 0,4 не обнаружен вообще. На первых трех орбитах электрон не обнаружен (событие A). Найдем вероятность обнаружить его на четвертой орбите (событие B) при этом условии. Поскольку вероятность обнаружить электрон на каждой орбите равна $0,6/4 = 0,15$, то вероятность не обнаружить электрон на первых трех орбитах $P(A) = 1 - 3 \cdot 0,15 = 0,55$. Событие AB — электрон не обнаружен на первых трех орбитах и обнаружен на четвертой орбите — совпадает просто с событием B — электрон находится на четвертой орбите и имеет вероятность $P(AB) = P(B) = 0,15$. Искомая вероятность, равная условной вероятности события B при условии события A , в этом случае имеет вид $P(B|A) = P(B)/P(A) = 3/11$. □

2. Формула умножения вероятностей

На практике часто происходит так, что известны или достаточно просто определяются именно условные вероятности и с их помощью необходимо вычислить безусловную вероятность некоторого события. Простейшей формулой для решения задач такого типа является формула умножения вероятностей. Пусть имеются события A_1, A_2, \dots, A_n . Тогда, используя понятие условной вероятности, можно написать

$$P(A_n | A_1 A_2 \dots A_{n-1}) = \frac{P(A_1 A_2 \dots A_n)}{P(A_1 A_2 \dots A_{n-1})}$$

или, умножая на $P(A_1 A_2 \dots A_{n-1})$,

$$P(A_1 A_2 \dots A_n) = P(A_1 A_2 \dots A_{n-1}) \cdot P(A_n | A_1 A_2 \dots A_{n-1}).$$

Аналогичным образом

$$P(A_1 A_2 \dots A_{n-1}) = P(A_1 A_2 \dots A_{n-2}) \cdot P(A_{n-1} | A_1 A_2 \dots A_{n-2}).$$

Продолжая эту процедуру, получаем

$$P(A_1 A_2 \dots A_n) = P(A_1) P(A_2 | A_1) P(A_3 | A_1 A_2) \dots P(A_n | A_1 A_2 \dots A_{n-1}).$$

Последнее соотношение носит название *формулы умножения вероятностей*.

Пример 7. На 7 карточках написаны буквы $л, л, о, о, о, т, т$. Из них последовательно выбираются 4 и кладутся слева направо. Найдем вероятность того, что в результате образуется слово «лото» (событие A). Это — задача на классическую вероятность, и ее можно было бы решить стандартным образом, вычисляя общее число элементарных исходов и число благоприятных для события A исходов. Однако гораздо проще воспользоваться формулой умножения вероятностей. Введем события: A_1 — на первой вынутой карточке написана буква $л$, A_2 — на второй — буква $о$, A_3 — на третьей — буква $т$ и A_4 — на четвертой — буква $о$. Тогда событие A представляет собой пересечение событий A_1, A_2, A_3 и A_4 и по формуле умножения вероятностей

$$\mathbf{P}(A) = \mathbf{P}(A_1 A_2 A_3 A_4) = \mathbf{P}(A_1) \mathbf{P}(A_2 | A_1) \mathbf{P}(A_3 | A_1 A_2) \mathbf{P}(A_4 | A_1 A_2 A_3).$$

Заметим теперь, что в соответствии с принципом классической вероятности безусловная вероятность $\mathbf{P}(A_1)$ определяется как отношение числа карточек, на которых написана буква $л$, к общему числу карточек, т. е. $\mathbf{P}(A_1) = 2/7$. Далее, если событие A_1 произошло, то у нас осталось 6 карточек и на трех из них написана буква $о$. Поэтому $\mathbf{P}(A_2 | A_1) = 3/6$. Аналогично, если произошли события A_1 и A_2 , то из пяти оставшихся карточек на двух написана буква $т$, и, значит, $\mathbf{P}(A_3 | A_1 A_2) = 2/5$. Наконец, $\mathbf{P}(A_4 | A_1 A_2 A_3) = 1/2$, поскольку из четырех оставшихся карточек на двух написана буква $о$. Окончательно получаем

$$\mathbf{P}(A) = \mathbf{P}(A_1 A_2 A_3 A_4) = \frac{2}{7} \cdot \frac{1}{2} \cdot \frac{2}{5} \cdot \frac{1}{2} = \frac{1}{35}.$$

Любознательному читателю предлагаем решить эту задачу с помощью комбинаторных методов. \square

Пример 8. В некоторых сельских местностях России существовало когда-то следующее гадание. Девушка зажимает в руке шесть травинок так, чтобы концы травинок торчали сверху и снизу; подруга связывает эти травинки попарно между собой сверху и снизу в отдельности. Если при этом все шесть травинок оказывались связанными в одно кольцо, то это должно было означать, что девушка в текущем году выйдет замуж. Найдем вероятность того, что травинки при связывании наудачу образуют кольцо (событие A). Для этого предположим сначала, что связаны верхние концы травинок. Тогда у нас образуется 3 пары травинок, и нижние концы этих пар будем последовательно связывать. Пусть событие A_1 означает, что при первом связывании у нас не образуется кольца (рис. 1), т. е. не будут связаны между собой нижние концы одной и той же пары. Событие A_2 — после второго связывания не образуется кольцо. Тогда событие A представляет собой пересечение событий A_1 и A_2 , и по формуле умножения вероятностей $\mathbf{P}(A) = \mathbf{P}(A_1) \mathbf{P}(A_2 | A_1)$. Теперь заметим, что при первом связывании событие A_1 происходит только тогда, когда мы не связываем между собой концы одной пары, а это возможно в 4 случаях из 5. Поэтому $\mathbf{P}(A_1) = 4/5$. Если событие A_1 произошло, то для наступления события A_2 необходимо связать построенную цепь и оставшуюся пару, что в свою очередь можно сделать 2 способами из 3. Значит, $\mathbf{P}(A_2 | A_1) = 2/3$ и $\mathbf{P}(A) = 4/5 \cdot 2/3 = 8/15$.

Предоставляем читателю возможность решить эту задачу для $2n$ травинок. \square

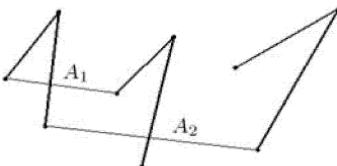


Рис. 1

3. Независимость событий

Следующим важным понятием теории вероятностей является независимость событий. События A и B называются независимыми, если условная вероятность события B при условии A совпадает с безусловной вероятностью события B , т. е. $\mathbf{P}(B|A) = \mathbf{P}(B)$.

Введенное таким образом понятие скорее можно было бы назвать независимостью события B от события A . Покажем, что на самом деле понятие независимости симметрично относительно перестановки событий A и B , т. е. если событие B не зависит от события A , то событие A также не зависит от события B . Для этого подставим вместо условной вероятности $\mathbf{P}(B|A)$ ее значение $\mathbf{P}(B|A) = \mathbf{P}(AB)/\mathbf{P}(A)$. Тогда независимость событий A и B эквивалентна выполнению равенства

$$\mathbf{P}(AB) = \mathbf{P}(A)\mathbf{P}(B),$$

откуда и следует симметричность понятия независимости. Отметим, что последнюю формулу также можно было бы взять в качестве определения независимости. Этим определением можно пользоваться даже в том случае, когда вероятность события A или B равна нулю.

Если события A и B независимы, то независимыми являются также пары событий \bar{A} и B , A и \bar{B} , \bar{A} и \bar{B} . В частности,

$$\mathbf{P}(\bar{A}|B) = 1 - \mathbf{P}(A|B) = 1 - \mathbf{P}(A) = \mathbf{P}(\bar{A}),$$

что означает независимость событий \bar{A} и B . Аналогично доказывается независимость остальных пар событий.

Пример 9. Проводится опыт, состоящий из двукратного подбрасывания симметричной монеты. Событие A — выпадение «герба» при первом подбрасывании, событие B — выпадение «герба» при втором. В этом случае, как мы знаем, пространство элементарных исходов Ω состоит из четырех исходов. Событию A благоприятствуют два исхода, событию B — два исхода и событию AB — один исход. Значит, $\mathbf{P}(A) = \mathbf{P}(B) = 1/2$, $\mathbf{P}(AB) = 1/4$. События A и B независимы, поскольку $\mathbf{P}(AB) = \mathbf{P}(A)\mathbf{P}(B)$. \square

Этот пример подтверждает с точки зрения классической вероятности интуитивный принцип «несвязности» результатов испытаний такого типа, как подбрасывание монеты, бросание игральной кости и т. д. Однако на практике поступают обычно обратным образом. Пусть имеются два различных испытания, причем в первом из них с вероятностью $\mathbf{P}(A)$ может произойти событие A , а во втором с вероятностью $\mathbf{P}(B)$ — событие B . Рассмотрим «обобщенный опыт», результатами которого являются пары $AB = (A, B)$, $\bar{A}B = (\bar{A}, B)$, $A\bar{B} = (A, \bar{B})$ и $\bar{A}\bar{B} = (\bar{A}, \bar{B})$. Если априори известно, что результаты испытаний не связаны между собой, то вполне естественно считать события A и B независимыми и положить $\mathbf{P}(AB) = \mathbf{P}(A)\mathbf{P}(B)$, $\mathbf{P}(\bar{A}B) = \mathbf{P}(\bar{A})\mathbf{P}(B)$ и т. д. В частности, в рассмотренном приме-

ре, исходя из двух однократных подбрасываний монеты и независимости результатов подбрасываний, нужно положить для события AB — выпадение «гербов» при обоих подбрасываниях вероятность $\mathbf{P}(AB) = \mathbf{P}(A)\mathbf{P}(B) = 1/2 \cdot 1/2 = 1/4$.

Пример 10. Известно, что вероятность двум близнецам быть одного пола примерно равна 0,64, причем вероятность рождения мальчика близка к 0,51. Найдем вероятность того, что второй из близнецов — мальчик (событие B), при условии, что первый из них — также мальчик (событие A). Так как пары мальчик–девочка и девочка–мальчик равновероятны, то вероятность события A — первый ребенок будет мальчиком, а второй — девочкой равна $\mathbf{P}(A\bar{B}) = (1 - 0,64)/2 = 0,36/2 = 0,18$. Поскольку из условия задачи $\mathbf{P}(A) = 0,51$, то $\mathbf{P}(\bar{B}|A) = 0,18/0,51 = 6/17$ и $\mathbf{P}(B|A) = 1 - \mathbf{P}(\bar{B}|A) = 11/17$, т. е. условная вероятность появления второго мальчика при условии появления первого мальчика существенно больше безусловной вероятности рождения мальчика. Поэтому гипотезу о независимости событий A и B следует признать ошибочной. С точки зрения генетики это связано с рождением так называемых одногенетических близнецов. \square

Определим теперь понятие независимости событий в совокупности. Назовем события A , B и C *независимыми (в совокупности)*, если $\mathbf{P}(AB) = \mathbf{P}(A)\mathbf{P}(B)$, $\mathbf{P}(AC) = \mathbf{P}(A)\mathbf{P}(C)$, $\mathbf{P}(BC) = \mathbf{P}(B)\mathbf{P}(C)$ и, кроме того, $\mathbf{P}(ABC) = \mathbf{P}(A)\mathbf{P}(B)\mathbf{P}(C)$. Аналогично определяется понятие независимости в совокупности и большего числа событий: вероятность пересечения любых двух различных событий равна произведению вероятностей этих событий; вероятность пересечения любых трех событий равна произведению их вероятностей; ...; вероятность пересечения всех событий равна произведению вероятностей. Из независимости событий в совокупности вытекает их попарная независимость. Однако из попарной независимости, вообще говоря, независимость в совокупности не следует, как показывает следующий простой пример (С. Н. Бернштейн).

Пример 11. Три грани правильного тетраэдра раскрашены в синий, красный и зеленый цвета, а в раскраске четвертой грани присутствуют все эти цвета. Событие A — тетраэдр упал на грань, в раскраске которой присутствует синий, событие B — красный и событие C — зеленый цвет. Поскольку каждый цвет присутствует в раскраске двух граней из четырех, то $\mathbf{P}(A) = \mathbf{P}(B) = \mathbf{P}(C) = 1/2$. Далее, два цвета присутствуют в раскраске одной грани. Поэтому $\mathbf{P}(AB) = \mathbf{P}(BC) = \mathbf{P}(AC) = 1/4$, и события A , B и C попарно независимы. Однако все три цвета присутствуют в раскраске также только одной грани и, значит, $\mathbf{P}(ABC) = 1/4 \neq \mathbf{P}(A)\mathbf{P}(B)\mathbf{P}(C) = 1/8$. Таким образом, хотя события A , B и C попарно независимы, они зависимы в совокупности. \square

Заметим, что когда говорят о независимости событий A_1, \dots, A_n , то подразумевают именно независимость событий в совокупности в отличие от попарной независимости событий A_1, \dots, A_n .

В заключение этого параграфа приведем формулу для вероятности объединения независимых событий. Пусть $A = A_1 \cup \dots \cup A_n$. Тогда в соответствии с формулой де Моргана $\bar{A} = \bar{A}_1 \dots \bar{A}_n$. Если события A_1, \dots, A_n независимы, то события $\bar{A}_1, \dots, \bar{A}_n$ также независимы и,

значит, $\mathbf{P}(\bar{A}) = \mathbf{P}(\bar{A}_1) \dots \mathbf{P}(\bar{A}_n)$. Отсюда окончательно получаем для вероятности объединения независимых событий формулу

$$\mathbf{P}(A) = 1 - [1 - \mathbf{P}(A_1)] \dots [1 - \mathbf{P}(A_n)].$$

Пример 12. Пусть некоторое техническое устройство состоит из n элементов. Назовем соединение этих элементов *последовательным* (рис. 2), если отказ любого элемента влечет за собой отказ всего устройства. Пусть A_i



Рис. 2

($i = 1, \dots, n$) — событие, заключающееся в отказе i -го элемента. Тогда событие A — отказ всего устройства может быть записано в виде $A = A_1 \cup \dots \cup A_n$. Предполагая, что отказы отдельных элементов независимы, получаем для вероятности отказа всего устройства выражение

$$\mathbf{P}(A) = 1 - [1 - \mathbf{P}(A_1)] \dots [1 - \mathbf{P}(A_n)]. \quad \square$$

Наряду с последовательным соединением элементов применяется также *параллельное соединение* (рис. 3), при котором отказ устройства происходит только при отказе всех элементов. Очевидно, что при параллельном соединении событие A — отказ всего устройства представляет собой пересечение событий A_i ($i = 1, \dots, n$) — отказов отдельных элементов, и в случае независимости отказов

$$\mathbf{P}(A) = \mathbf{P}(A_1) \dots \mathbf{P}(A_n).$$

Разумеется, надежность реальных технических устройств определяется более сложными структурными схемами, чем последовательное или параллельное соединение элементов. Однако довольно часто встречаются устройства, структурную схему надежности которых можно представить в виде отдельных компонентов, содержащих только последовательные или только параллельные соединения элементов. В таком случае расчет надежности устройства можно произвести поэтапно, объединяя на каждом этапе либо последовательно, либо параллельно соединенные элементы и заменяя их одним «обобщенным» элементом.

Например, пусть нужно определить вероятность отказа состоящего из трех элементов устройства, структурная схема надежности которого приведена на рис. 4. Известно, что элементы отказывают независимо друг от друга и вероятности отказов элементов 1, 2, 3 за заданное время работы равны соответственно $\mathbf{P}(A_1) = \mathbf{P}(A_2) = 0,1$, $\mathbf{P}(A_3) = 0,01$. Для нахождения вероятности $\mathbf{P}(A)$ отказа всего устройства заменим сначала параллельно соединенные элементы 1 и 2 «обобщенным» элементом 4, вероятность отказа которого в соответствии с приведенным ранее правилом равна

$$\mathbf{P}(A_4) = \mathbf{P}(A_1) \mathbf{P}(A_2) = 0,1^2 = 0,01.$$

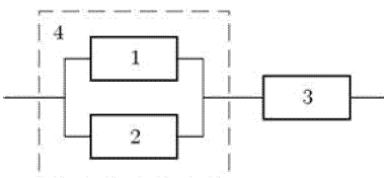


Рис. 4

Теперь осталось заметить, что структурная схема надежности всего устройства представляет собой последовательное соединение элемента 3 и «обобщенного» элемента 4. Значит,

$$\mathbf{P}(A) = 1 - [1 - \mathbf{P}(A_3)][1 - \mathbf{P}(A_4)] = 1 - (1 - 0,01)^2 \cong 0,02.$$

Отметим, что при вычислении вероятности отказа устройства мы фактически пользовались следующим представлением события A :

$$A = (A_1 A_2) \cup A_3.$$

Отсюда в силу формул де Моргана для события \bar{A} — безотказной работы устройства на заданном промежутке времени — получаем выражение

$$\bar{A} = (\bar{A}_1 \cup \bar{A}_2) \bar{A}_3,$$

где событие \bar{A}_i ($i = 1, 2, 3$) — безотказная работа i -го элемента.

4. Формула полной вероятности

Предположим, что в результате опыта может произойти одно из n событий, во-первых, *несовместных* и, во-вторых, составляющих *полную группу событий* H_i , т. е. $H_i H_j = \emptyset$ при $i \neq j$ и $H_1 + \dots + H_n = \Omega$. События, удовлетворяющие этим двум требованиям, будем называть

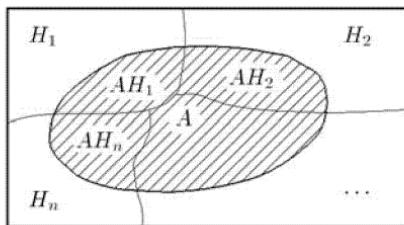


Рис. 5

гипотезами. Пусть также имеется некоторое событие A , и нам известны $\mathbf{P}(H_1), \dots, \mathbf{P}(H_n), \mathbf{P}(A|H_1), \dots, \mathbf{P}(A|H_n)$. Задача состоит в определении безусловной вероятности $\mathbf{P}(A)$. Для решения представим событие A в следующем виде:

$$A = A\Omega = A(H_1 + \dots + H_n) = AH_1 + \dots + AH_n$$

(см. рис. 5). Тогда

$$\mathbf{P}(A) = \mathbf{P}(AH_1) + \dots + \mathbf{P}(AH_n).$$

По формуле умножения вероятностей

$$\mathbf{P}(AH_1) = \mathbf{P}(H_1)\mathbf{P}(A|H_1), \dots, \mathbf{P}(AH_n) = \mathbf{P}(H_n)\mathbf{P}(A|H_n).$$

Поэтому

$$\mathbf{P}(A) = \mathbf{P}(H_1)\mathbf{P}(A|H_1) + \dots + \mathbf{P}(H_n)\mathbf{P}(A|H_n).$$

Выведенная формула носит название *формулы полной вероятности*. При всей своей простоте она играет весьма существенную роль в теории вероятностей.

Пример 13. Частица пролетает мимо трех счетчиков, причем она может попасть в каждый из них с вероятностями 0,3, 0,2 и 0,4. В свою очередь, если частица попадает в первый счетчик, то она регистрируется с вероятностью 0,6, во второй — с вероятностью 0,5 и в третий — с вероятностью 0,55. Найдем вероятность того, что частица будет зарегистрирована (событие A). В нашем случае гипотеза H_1 — частица попадает в первый счетчик ($P(H_1) = 0,3$), H_2 — во второй ($P(H_2) = 0,2$) и H_3 — в третий ($P(H_3) = 0,4$). События H_1 , H_2 и H_3 не пересекаются, однако они не составляют полной группы событий. Для того чтобы получить полную группу событий, нужно добавить событие H_4 , заключающееся в том, что частица не попадет ни в один счетчик. Ясно, что $P(H_4) = 1 - P(H_1) - P(H_2) - P(H_3) = 1 - 0,3 - 0,2 - 0,4 = 0,1$. Условные вероятности события A при условии каждой гипотезы равны: $P(A|H_1) = 0,6$, $P(A|H_2) = 0,5$, $P(A|H_3) = 0,55$ и $P(A|H_4) = 0$. По формуле полной вероятности имеем

$$P(A) = 0,3 \cdot 0,6 + 0,2 \cdot 0,5 + 0,4 \cdot 0,55 + 0,1 \cdot 0 = 0,5. \quad \square$$

5. Формула Байеса

Во многих приложениях теории вероятностей встречается следующая задача. Пусть до опыта имеются гипотезы H_1, \dots, H_n . После опыта становится известной информация о его результатах, но не полная. А именно, результаты наблюдений показывают, не какой конкретно элементарный исход ω из пространства элементарных исходов Ω произошел, а что наступило некоторое событие A . Считая, что до опыта были известны (априорные) вероятности гипотез $P(H_1), \dots, P(H_n)$ и условные вероятности $P(A|H_1), \dots, P(A|H_n)$, необходимо определить (апостериорные) вероятности гипотез $P(H_1|A), \dots, P(H_n|A)$.

Для решения поставленной задачи вспомним, что по определению условной вероятности

$$P(H_i | A) = \frac{P(AH_i)}{P(A)}$$

и по формуле умножения вероятностей

$$P(AH_i) = P(H_i)P(A|H_i).$$

Поэтому

$$P(H_i | A) = \frac{P(H_i)P(A|H_i)}{P(A)}.$$

Полученное выражение носит название *формулы Байеса*.

Подставляя вместо вероятности $\mathbf{P}(A)$ ее значение, вычисленное по формуле полной вероятности, формулу Байеса можно переписать также в виде

$$\mathbf{P}(H_i | A) = \frac{\mathbf{P}(H_i)\mathbf{P}(A | H_i)}{\mathbf{P}(H_1)\mathbf{P}(A | H_1) + \dots + \mathbf{P}(H_n)\mathbf{P}(A | H_n)}.$$

Формула Байеса находит широкое применение в математической статистике.

Пример 14. Три завода выпускают одинаковые изделия, причем первый завод производит 50%, второй — 20% и третий — 30% всей продукции. Первый завод выпускает 1% брака, второй завод — 8% и третий — 3%. Наудачу выбранное изделие оказалось бракованым (событие A). Найдем вероятность того, что оно изготовлено на втором заводе. У нас имеется три гипотезы: H_1 — изделие изготовлено на первом заводе, H_2 — на втором заводе и H_3 — на третьем. По условию задачи $\mathbf{P}(H_1) = 0,5$, $\mathbf{P}(H_2) = 0,2$, $\mathbf{P}(H_3) = 0,3$, $\mathbf{P}(A | H_1) = 0,01$, $\mathbf{P}(A | H_2) = 0,08$, $\mathbf{P}(A | H_3) = 0,03$. Условная вероятность того, что бракованное изделие изготовлено на втором заводе, определяется формулой Байеса:

$$\begin{aligned} \mathbf{P}(H_2 | A) &= \frac{\mathbf{P}(H_2)\mathbf{P}(A | H_2)}{\mathbf{P}(H_1)\mathbf{P}(A | H_1) + \mathbf{P}(H_2)\mathbf{P}(A | H_2) + \mathbf{P}(H_3)\mathbf{P}(A | H_3)} = \\ &= \frac{0,2 \cdot 0,08}{0,5 \cdot 0,01 + 0,2 \cdot 0,08 + 0,3 \cdot 0,03} = \frac{8}{15}. \end{aligned}$$

Итак, несмотря на то что продукция второго завода составляет $1/5$, его доля в браке больше половины. \square

Пример 15. Два охотника одновременно стреляют одинаковыми пулями в медведя. В результате медведь был убит одной пулей (событие A). Как охотники должны поделить шкуру убитого медведя, если известно, что вероятность попадания у первого охотника 0,3, а у второго — 0,6? Снова воспользуемся формулой Байеса. Однако предварительно определим гипотезы. Гипотеза H_1 : попал первый охотник, второй промахнулся; гипотеза H_2 : попал второй, первый промахнулся. События H_1 и H_2 несовместны, однако не составляют полной группы событий. Поэтому введем еще две гипотезы: H_3 — попали оба охотника и H_4 — оба промахнулись. Заметим, что событие A может происходить тогда и только тогда, когда произошла либо гипотеза H_1 , либо гипотеза H_2 , т. е. $\mathbf{P}(A | H_1) = \mathbf{P}(A | H_2) = 1$, $\mathbf{P}(A | H_3) = \mathbf{P}(A | H_4) = 0$. Кроме того, делая естественное предположение, что попадания охотников в медведя не зависят друг от друга, получаем: $\mathbf{P}(H_3) = 0,3 \cdot 0,6 = 0,18$, $\mathbf{P}(H_4) = (1 - 0,3) \times (1 - 0,6) = 0,28$, $\mathbf{P}(H_1) = 0,3 \cdot (1 - 0,6) = 0,12$, $\mathbf{P}(H_2) = (1 - 0,3) \times 0,6 = 0,42$. Теперь мы в состоянии применить формулу Байеса, согласно которой

$$\mathbf{P}(H_1 | A) = \frac{0,12 \cdot 1}{0,12 \cdot 1 + 0,42 \cdot 1 + 0,18 \cdot 0 + 0,28 \cdot 0} = \frac{2}{9},$$

$$\mathbf{P}(H_2 | A) = \frac{0,42 \cdot 1}{0,12 \cdot 1 + 0,42 \cdot 1 + 0,18 \cdot 0 + 0,28 \cdot 0} = \frac{7}{9}.$$

Таким образом, при справедливом дележе первый охотник должен получить $2/9$ шкуры, т. е. меньше $1/4$, в то время как, на первый взгляд, казалось, что ему причитается $1/3$ шкуры. \square

Пример 16. В группе 15 студентов. Из них: 5 «отличников», 7 «четверочников» и 3 «троечника». Известно, что «отличник» с вероятностью 0,9 получает на каждом экзамене «отлично» и с вероятностью 0,1 — «хорошо». Аналогично, «четверочник» с вероятностью 0,1 получает «отлично», с вероятностью 0,7 — «хорошо» и с вероятностью 0,2 — «удовлетворительно». Наконец, «троечник» получает с вероятностью 0,1 «отлично», с вероятностью 0,2 — «хорошо» и с вероятностью 0,7 — «удовлетворительно». Один из студентов из этой группы получил на первом экзамене «хорошо». Найдем вероятность того, что на следующем экзамене он получит «отлично». Итак, опыт состоит в последовательной сдаче двух экзаменов. Событие A — студент сдал первый экзамен на «хорошо», событие B — второй экзамен на «отлично». Ясно, что в качестве гипотез надо взять: H_1 — студент «отличник», H_2 — «четверочник» и H_3 — «троечник». Если бы нам необходимо было найти просто безусловную вероятность события B , то нужно было бы воспользоваться формулой полной вероятности, в которой $P(H_1) = 5/15 = 1/3$, $P(H_2) = 7/15$, $P(H_3) = 3/15 = 1/5$, $P(B|H_1) = 0,9$, $P(B|H_2) = 0,1$ и $P(B|H_3) = 0,1$. Однако нас интересует условная вероятность события B при условии A , поэтому мы сначала с помощью формулы Байеса найдем условные вероятности гипотез H_1 , H_2 и H_3 при условии A . Поскольку $P(A|H_1) = 0,1$, $P(A|H_2) = 0,7$ и $P(A|H_3) = 0,2$, то имеем:

$$P(H_1|A) = \frac{\frac{1}{3} \cdot 0,1}{\frac{1}{3} \cdot 0,1 + \frac{7}{15} \cdot 0,7 + \frac{1}{5} \cdot 0,2} = \frac{1}{12},$$

$$P(H_2|A) = \frac{\frac{7}{15} \cdot 0,7}{\frac{1}{3} \cdot 0,1 + \frac{7}{15} \cdot 0,7 + \frac{1}{5} \cdot 0,2} = \frac{49}{60},$$

$$P(H_3|A) = 1 - P(H_1|A) - P(H_2|A) = \frac{1}{10}.$$

Таким образом, по полученной на первом экзамене оценке мы обязаны присвоить нашему студенту новые вероятности $1/12$, $49/60$ и $1/10$ того, что он «отличник», «четверочник» и «троечник». Теперь для вычисления условной вероятности $P(B|A)$ воспользуемся формулой полной вероятности:

$$P(B|A) = P(H_1|A)P(B|H_1A) + P(H_2|A)P(B|H_2A) + P(H_3|A)P(B|H_3A),$$

в которой вместо вероятностей $P(B)$ и $P(H_i)$ ($i = 1, 2, 3$) взяты условные вероятности $P(B|A)$ и $P(H_i|A)$, а вместо условных вероятностей $P(B|H_i)$ — условные вероятности $P(B|H_iA)$ (справедливость этой формулы мы предлагаем читателю проверить самостоятельно). Тогда, предполагая, что для студента одной успеваемости результат следующего экзамена не зависит от результата предыдущего, т. е. $P(B|H_iA) = P(B|H_i)$, получаем окончательно

$$\begin{aligned} P(B|A) &= P(H_1|A)P(B|H_1) + P(H_2|A)P(B|H_2) + P(H_3|A)P(B|H_3) = \\ &= \frac{1}{12} \cdot 0,9 + \frac{49}{60} \cdot 0,1 + \frac{1}{10} \cdot 0,1 = \frac{1}{6}. \end{aligned}$$

Для лучшего усвоения этого примера рекомендуем читателю построить пространство элементарных исходов Ω (в данном случае оно будет состоять из 27 «троек» — «успеваемость—оценка на первом экзамене—оценка на втором экзамене») и определить на нем вероятность. При определении вероятности нужно учитывать, что оценки на каждом экзамене для студентов одной успеваемости независимы.

Вообще говоря, можно отказаться и от условия независимости оценок на каждом экзамене. Тогда в формуле полной вероятности нужно вместо условных вероятностей $P(B|H_i)$ использовать условные вероятности $P(B|AH_i)$, т. е. учитывать не только успеваемость студента, но и его предыдущую оценку. Разумеется, модель экзаменов должна быть тоже более тонкой, т. е. необходимо задать для студента каждой успеваемости вероятность получения им любых оценок на двух последовательных экзаменах. \square

Г л а в а 4

СХЕМА БЕРНУЛЛИ

Во введении мы говорили, что теория вероятностей имеет дело с такими явлениями, при которых испытания можно повторять, по крайней мере теоретически, бесконечное число раз, и при этом появление или непоявление некоторого наблюдаемого события в каждом испытании не будет зависеть от исходов предыдущих испытаний. Сейчас мы рассмотрим более подробно эту схему, носящую в теории вероятностей название *последовательности независимых одинаковых испытаний*, или *схемы Бернулли*, основываясь на аксиоматическом определении вероятности и уже введенном нами понятии независимости событий.

Итак, опыт состоит в n -кратном повторении одинаковых испытаний, в каждом из которых может с вероятностью p наступить некоторое событие (будем говорить в этом случае, что произошел «успех») или с вероятностью $q = 1 - p$ не наступить (произошла «неудача»). Результат каждого опыта можно записать в виде последовательности УНН...У, состоящей из n букв У и Н, причем буква У (или Н) на i -м месте означает, что в i -м испытании произошел успех (или неудача). Пространство элементарных исходов Ω состоит из 2^n исходов, каждый из которых отождествляется с определенной последовательностью УНН...У (σ -алгебра событий \mathfrak{B} включает 2^n событий!). Заметим теперь, что в силу независимости испытаний мы обязаны сопоставить каждому элементарному исходу $\omega = \dots$ вероятность $P(\omega) = P(\dots) = pqq\dots p$, причем буква p (или q) в произведении повторяется столько раз, сколько раз произошел успех (или неудача). Типичным представителем схемы Бернулли является n -кратное подбрасывание несимметричной монеты, причем, например, «успех» означает выпадение «герба», а «неудача» — «цифры».

1. Формула Бернулли

Вычислим вероятность $P_n(m)$ получить в n испытаниях ровно m успехов. Событие A_m — в n испытаниях произошло ровно m успехов — состоит из тех элементарных исходов, в которых буква У появляется ровно m раз. Для того чтобы подсчитать число таких исходов, заметим, что оно совпадает с числом способов, которыми можно расставить m букв У на n местах. Поскольку порядок, в котором мы расставляем буквы, нас не интересует, то мы имеем дело с сочетаниями и число способов равно $\binom{n}{m}$. С другой стороны, каждый элементарный исход, в котором интересующая нас буква У встречается ровно m раз, как мы знаем, имеет вероятность $p^m q^{n-m}$. Окончательно получаем для вероятности $P(A_m) = P_n(m)$ наступления m успехов в n независимых испытаниях *формулу Бернулли*:

$$P_n(m) = \binom{n}{m} p^m q^{n-m} \quad (m = 0, 1, 2, \dots, n).$$

Данное выражение носит также название *биномиального закона*, поскольку $P_n(m)$ можно получить как коэффициент при z^m разложения по степеням z бинома $(pz + q)^n$. Заметим, что последнее выражение представляет собой производящую функцию (z -преобразование) для биномиального закона (см. параграф 3 гл. 8); аппарат производящих функций широко используется в теории вероятностей.

Пример 1. В шаре радиусом R находится n молекул идеального газа. Вычислим вероятность того, что ровно m из них будут находиться на расстоянии, меньшем $\rho = \lambda R$ ($0 \leq \lambda \leq 1$) от центра этого шара. Поскольку каждая из n молекул может находиться в любой точке шара независимо от остальных, искомая вероятность определяется с помощью формулы Бернулли. При этом для вычисления вероятности p попадания одной молекулы в шар радиусом ρ (успех) можно воспользоваться схемой геометрической вероятности, т. е.

$$p = \frac{\frac{4}{3}\pi\lambda^3 R^3}{\frac{4}{3}\pi R^3} = \lambda^3.$$

Тогда по формуле Бернулли окончательно получаем

$$P_n(m) = \binom{n}{m} \lambda^{3m} (1 - \lambda^3)^{n-m}. \quad \square$$

Пример 2. Частица пролетает последовательно мимо 6 счетчиков. Каждый счетчик независимо от остальных отмечает ее пролет с вероятностью $p = 0,8$. Частица считается зарегистрированной (событие A), если она отмечена не менее чем двумя счетчиками. Найдем вероятность зарегистрировать частицу. В соответствии с аксиомой сложения искомую вероятность $P(A)$ можно представить в виде

$$P(A) = P(A_2) + P(A_3) + P(A_4) + P(A_5) + P(A_6),$$

где A_i — событие, заключающееся в том, что частица отмечена ровно i счетчиками. Теперь для определения $P(A_i)$ можно было бы воспользоваться формулой Бернулли, однако мы предварительно перейдем к дополнительному событию \bar{A} — частица либо не отмечена ни одним счетчиком, либо отмечена только одним. Тогда

$$P(\bar{A}) = P(A_0) + P(A_1) = P_6(0) + P_6(1) =$$

$$= \binom{6}{0} p^0 q^6 + \binom{6}{1} p^1 q^5 = (0,2)^6 + 6 \cdot 0,8 \cdot (0,2)^5 = 0,0016$$

и

$$P(A) = 1 - P(\bar{A}) = 0,9984. \quad \square$$

2. Формула Пуассона

Предположим, что мы хотим определить вероятность выпадения ровно 5100 «гербов» при 10000 бросаний монеты. Ясно, что при таком большом числе испытаний использование формулы Бернулли весьма затруднительно с точки зрения вычислений. Поэтому возникает естественное желание иметь простые, но достаточно точные приближенные формулы для вычисления $P_n(m)$ при больших n . Такие формулы

существуют и определяются теоремой Пуассона, а также локальной и интегральной теоремами Муавра–Лапласа.

Начнем с наиболее простой из них — формулы Пуассона. Формула Пуассона применяется тогда, когда наряду с большим значением числа испытаний n мала вероятность успеха p . Рекомендации по применению предельных формул (в том числе и формулы Пуассона) мы дадим ниже. Сейчас же сформулируем теорему Пуассона. Строго математически теорема Пуассона опирается на довольно сложное понятие схемы серий, поэтому ниже мы приведем «инженерную» интерпретацию этой теоремы.

Теорема Пуассона. Пусть число испытаний n в схеме Бернулли велико, а вероятность успеха p в одном испытании мала, причем мало также произведение $\lambda = np$. Тогда $P_n(m)$ определяется по приближенной формуле (формула Пуассона)

$$P_n(m) \approx \frac{\lambda^m}{m!} e^{-\lambda} \quad (m = 0, 1, 2, \dots, n).$$

Доказательство. Запишем формулу Бернулли

$$P_n(m) = \binom{n}{m} p^m (1-p)^{n-m} = \frac{n(n-1)\dots(n-m+1)}{m!} p^m (1-p)^{n-m}$$

или, с учетом обозначения $\lambda = np$,

$$P_n(m) = \frac{\lambda^m}{m!} \left(1 - \frac{\lambda}{n}\right)^n \frac{n-1}{n} \cdot \frac{n-2}{n} \cdots \frac{n-m+1}{n} \left(1 - \frac{\lambda}{n}\right)^{-m}.$$

Как известно, $(1 - \lambda/n)^n \approx e^{-\lambda}$ при больших n . Кроме того, если n велико, то $(n-1)/n \approx 1$, $(n-2)/n \approx 1, \dots$, $(n-m+1)/n \approx 1$ и $(1 - \lambda/n)^{-m} \approx 1$. Поэтому

$$P_n(m) \approx \frac{\lambda^m}{m!} e^{-\lambda},$$

что и требовалось доказать. \square

Совокупность вероятностей $P(m; \lambda) = \lambda^m e^{-\lambda} / m!$ ($m = 0, 1, \dots$) называется *распределением Пуассона*.

Значения функции $P(m; \lambda)$ для некоторых λ приведены в табл. 1 приложения.

Отметим, что формула Пуассона справедлива также по отношению к числу неудач, но только в том случае, когда мало $\lambda' = nq$.

3. Формулы Муавра–Лапласа

Если в схеме Бернулли наряду с числом испытаний n велики также значения np и nq , то следует применять формулы Муавра–Лапласа — локальную или интегральную. При этом локальную формулу Муавра–Лапласа, как следует из самого названия, необходимо применять в том случае, когда нас интересует вероятность получить ровно m успехов в n испытаниях, а интегральную — если определяется вероятность получить число успехов, заключенное в пределах от m_1 до m_2 .

Так же, как и теорема Пуассона, локальная и интегральная теоремы Муавра–Лапласа приводятся в «инженерной» трактовке.

Локальная теорема Муавра–Лапласа. Если в схеме Бернулли число испытаний n велико, то для всех m справедлива приближенная формула (локальная формула Муавра–Лапласа)

$$\sqrt{npq} P_n(m) \approx \varphi(x),$$

где $x = (m - np)/\sqrt{npq}$, а

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Одно из первых доказательств теоремы было основано на известной из курса математического анализа формуле Стирлинга

$$n! e^n n^{-n} (2\pi n)^{-\frac{1}{2}} \underset{n \rightarrow \infty}{\longrightarrow} 1.$$

Не вдаваясь в математические подробности доказательства, изложим основную его суть. Считая, что n и $n - m$ достаточно велики, и подставляя в формулу Бернулли вместо $n!$, $m!$ и $(n - m)!$ их приближенные значения, вычисленные по формуле Стирлинга, получаем

$$\sqrt{npq} P_n(m) \approx \frac{A}{\sqrt{2\pi}} \sqrt{\frac{n^2 pq}{m(n-m)}},$$

где

$$A = \frac{p^m q^{n-m} n^n}{m^m (n-m)^{n-m}}.$$

Положим $x = (m - np)/\sqrt{npq}$. Тогда $m = np + x\sqrt{npq}$, $n - m = nq - x\sqrt{npq}$. Логарифмируя теперь A , имеем

$$\begin{aligned} \ln A = - & \left\{ (np + x\sqrt{npq}) \ln \left(1 + x\sqrt{\frac{q}{np}} \right) + \right. \\ & \left. + (nq - x\sqrt{npq}) \ln \left(1 - x\sqrt{\frac{p}{nq}} \right) \right\}. \end{aligned}$$

Поскольку $\sqrt{q/(np)}$ и $\sqrt{p/(nq)}$ малы при больших n , разложим логарифмы в ряд Маклорена по степеням x до второго члена. Тогда

$$\begin{aligned} \ln A \approx - & \left\{ (np + x\sqrt{npq}) \left(x\sqrt{\frac{q}{np}} - \frac{x^2}{2} \cdot \frac{q}{np} \right) + \right. \\ & \left. + (nq - x\sqrt{npq}) \left(-x\sqrt{\frac{p}{nq}} - \frac{x^2}{2} \cdot \frac{p}{nq} \right) \right\} \approx -\frac{x^2}{2} (p+q) = -\frac{x^2}{2}. \end{aligned}$$

Следовательно,

$$A \approx e^{-x^2/2}.$$

Наконец, учитывая, что $m/(np) \approx 1$ и $(n-m)/(nq) \approx 1$ при фиксированном x и больших n , получаем

$$\sqrt{\frac{n^2 pq}{m(n-m)}} \approx 1,$$

откуда и вытекает утверждение теоремы.

Значения функции $\varphi(x)$, которую называют *плотностью стандартного нормального, или гауссова, распределения*, для некоторых x приведены в табл. 2 приложения. Поскольку функция $\varphi(x)$ является четной, то при определении $\varphi(x)$ для отрицательных x нужно воспользоваться равенством $\varphi(x) = \varphi(-x)$.

Интегральная теорема Муавра–Лапласа. Если в схеме Бернулли число испытаний n велико, то для вероятности $P\{m_1 \leq \mu \leq m_2\}$ того, что число успехов μ заключено в пределах от m_1 до m_2 , справедливо приближенное соотношение (интегральная формула Муавра–Лапласа)

$$P\{m_1 \leq \mu \leq m_2\} \approx \Phi(x_2) - \Phi(x_1),$$

где $x_1 = (m_1 - np)/\sqrt{npq}$, $x_2 = (m_2 - np)/\sqrt{npq}$,

$$\Phi(x) = \int_{-\infty}^x \varphi(y) dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy.$$

Доказательство теоремы проведем, опираясь на локальную теорему Муавра–Лапласа (здесь мы также опускаем отдельные технические детали доказательства). Если $a_n \rightarrow \infty$ при $n \rightarrow \infty$, то в силу теоремы о среднем и равномерной непрерывности функции $\varphi(x)$

$$\varphi(x) - a_n \int_{x-1/(2a_n)}^{x+1/(2a_n)} \varphi(y) dy \xrightarrow[n \rightarrow \infty]{} 0$$

равномерно по x . Поэтому, полагая $a_n = \sqrt{npq}$, из локальной теоремы Муавра–Лапласа находим

$$P_n(m) = \frac{1}{\sqrt{npq}} \varphi(x) + o\left(\frac{1}{\sqrt{n}}\right) = \int_{x-1/(2\sqrt{npq})}^{x+1/(2\sqrt{npq})} \varphi(y) dy + o\left(\frac{1}{\sqrt{n}}\right),$$

где $x = (m - np)/\sqrt{npq}$. Суммируя по всем m от m_1 до m_2 , окончательно получаем

$$P\{m_1 \leq \mu \leq m_2\} = \sum_{m=m_1}^{m_2} P_n(m) \approx \int_{x_1-1/(2\sqrt{npq})}^{x_2+1/(2\sqrt{npq})} \varphi(y) dy \approx \int_{x_1}^{x_2} \varphi(y) dy.$$

Тем самым утверждение теоремы доказано. \square

Отметим, что интегральная теорема Муавра–Лапласа является следствием более общей центральной предельной теоремы, которую мы докажем другим, более простым способом в гл. 8.

Функция $\Phi(x)$, фигурирующая в интегральной формуле Муавра–Лапласа, носит название *функции стандартного нормального, или гауссова, распределения*. В силу четности $\varphi(x)$ функция стандартного нормального распределения обладает свойством $\Phi(-x) = 1 - \Phi(x)$. Поэтому в табл. 3 приложения приведены значения не $\Phi(x)$, а *интеграла Лапласа* $\Phi_0(x) = \int_0^x \varphi(y) dy$ и только для положительных x . Ясно, что $\Phi_0(x)$ является нечетной функцией, т. е. $\Phi_0(-x) = -\Phi_0(x)$ и, кроме того, $\Phi(x) = \Phi_0(x) + 1/2$. В терминах интеграла Лапласа интегральная формула Муавра–Лапласа имеет вид

$$\mathbf{P}\{m_1 \leqslant \mu \leqslant m_2\} \approx \Phi_0(x_2) - \Phi_0(x_1).$$

Именно этой формулой мы будем пользоваться при расчетах в примерах следующего параграфа.

Замечание. Распределение Пуассона, плотность и функция стандартного нормального распределения играют в приложениях столь существенную роль, что таблицы их значений содержатся практически в любом справочнике, учебнике или задачнике по теории вероятностей или математической статистике (см. также приложение). Однако следует еще раз обратить внимание на то, что довольно часто в таблицах приводятся не значения функции стандартного нормального распределения, а значения интеграла Лапласа $\Phi_0(x)$ или даже функции $\Phi^*(x) = \int_x^\infty \varphi(y) dy = 1 - \Phi(x)$. Поэтому, прежде чем пользоваться таблицей, необходимо внимательно посмотреть, значения какой функции даны в этой таблице.

4. Применение приближенных формул Пуассона и Муавра–Лапласа

В этом параграфе мы дадим некоторые рекомендации (носящие, вообще говоря, условный характер) по применению приближенных формул и рассмотрим примеры расчетов с их помощью.

Если число испытаний $n = 10–20$, то приближенные формулы используются для грубых прикидочных расчетов. При этом формула Пуассона применяется в том случае, когда $\lambda = np$ или $\lambda' = nq$ изменяются в пределах от 0–2 (при $n = 10$) до 0–3 (при $n = 20$); в противном случае необходимо пользоваться формулами Муавра–Лапласа.

Если $n = 20–100$, то приближенные формулы уже можно использовать для прикладных инженерных расчетов. Формулу Пуассона рекомендуется применять, когда λ или λ' заключены в следующих пределах: 0–3 ($n = 20$) – 0–5 ($n = 100$).

Если $n = 100–1000$, то практически при любых инженерных расчетах можно обойтись приближенными формулами. Формула Пуассона

используется, когда λ или λ' изменяются в следующих пределах: 0—5 ($n = 100$) — 0—10 ($n = 1000$).

Наконец, при $n > 1000$ даже специальные таблицы рассчитываются с помощью приближенных формул (правда, для увеличения точности используют специальные поправки). В этом случае для применения формулы Пуассона необходимо, чтобы λ или λ' лежали в пределах 0—10 и более.

Прежде чем переходить к рассмотрению примеров, скажем несколько слов о погрешностях, возникающих при использовании приближенных формул. Для этого отметим, что знак \approx приближенного равенства использовался нами в том смысле, что при увеличении n разность между величинами, связанными этим знаком, стремится к нулю. Иными словами, приближенные формулы гарантируют только *малую абсолютную погрешность*, но *относительная погрешность*, т. е. отношение величин, связанных знаком \approx , может быть сколь угодно *большой*. Так, при использовании формулы Пуассона для вычисления биномиальных вероятностей $P_n(m)$ относительная погрешность имеет тенденцию к увеличению с ростом m . Аналогично, в формулах Муавра—Лапласа относительная погрешность увеличивается с ростом абсолютного значения $|x| = |(m - np)/\sqrt{npq}|$, причем в интегральной формуле Муавра—Лапласа такое увеличение происходит только в том случае, когда $x_1 = (m_1 - np)/\sqrt{npq}$ и $x_2 = (m_2 - np)/\sqrt{npq}$ имеют одинаковый знак. Читатель сможет убедиться в этом из рассмотренных ниже примеров.

Следует отметить, что к настоящему времени доказаны теоремы, позволяющие не только получать более точные приближения биномиальных вероятностей $P_n(m)$, но и оценивать возникающие при этом погрешности. Однако эти результаты довольно сложны и весьма редко используются в инженерной практике.

Пример 3. Счетчик регистрирует попадающие в него частицы с вероятностью $p = 0,9$. Найдем вероятность $P_{10}(m)$ того, что он зарегистрировал m ($m = 0, 1, \dots, 10$) частиц при условии, что в него попало 10 частиц.

Воспользуемся сначала точной формулой Бернулли, в соответствии с которой $P_{10}(m) = \binom{10}{m} 0,9^m (1 - 0,9)^{10-m}$. Результаты расчетов приведены в табл. 1 в графе «Точное значение $P_{10}(m)$ ».

Постараемся применить теперь приближенные формулы. В нашем случае $\lambda = np = 9$ велико, однако $\lambda' = nq = 1$ мало и, значит, рекомендации советуют воспользоваться формулой Пуассона, но по отношению к незарегистрированным частицам. В соответствии с этой формулой

$$P_{10}(m) \approx \frac{1^{10-m} e^{-1}}{10 - m} = P(10 - m; 1).$$

По табл. 1 приложения находим

$$P_{10}(10) \approx P(0; 1) = 0,36788, \quad P_{10}(9) \approx P(1; 1) = 0,36788$$

и т. д. Приближенные значения приведены в табл. 1 в графе «Значение, вычисленное по формуле Пуассона». Для сравнения полученных результатов

в эту же таблицу включена графа «Погрешность». Анализируя приведенные в этой графе числа, видим, что максимальная абсолютная погрешность 0,01954 невелика, чего нельзя сказать о максимальной относительной погрешности. В частности, приближенное значение 0,00051 вероятности $P_{10}(4)$, вычисленное по формуле Пуассона, почти в 4 раза больше истинного значения 0,00014 этой вероятности.

Наконец, покажем, как воспользоваться локальной формулой Муавра–Лапласа, хотя здесь это и не рекомендуется делать. Тогда $P_{10}(m)$ мы должны заменить числом $\varphi(x_m)/\sqrt{10 \cdot 0,9 \cdot 0,1}$, где $x_m = (m - 10 \cdot 0,9)/\sqrt{10 \cdot 0,9 \cdot 0,1}$, а $\varphi(x)$ — плотность стандартного нормального распределения. Результаты вычислений с использованием табл. 2 приложения приведены в табл. 2. Как и следовало ожидать, локальная формула Муавра–Лапласа дает существенно большие погрешности. \square

Таблица 1

m	Точное значение $P_{10}(m)$	Значение, вычисленное по формуле Пуассона	Погрешность
0			
1			
2		0,00001	0,00001
3	0,00001	0,00007	0,00006
4	0,00014	0,00051	0,00037
5	0,00149	0,00307	0,00158
6	0,01116	0,01533	0,00417
7	0,05740	0,06131	0,00391
8	0,19371	0,18394	-0,00977
9	0,38742	0,36788	-0,01954
10	0,34868	0,36788	0,01920

Таблица 2

m	Точное значение $P_{10}(m)$	x_m	$\varphi(x_m)$	Значение, вычисленное по локальной формуле Муавра–Лапласа	Погрешность
0		-9,49			
1		-8,43			
2		-7,38			
3	0,00001	-6,32			-0,00001
4	0,00014	-5,27			-0,00014
5	0,00149	-4,22	0,00006	0,00006	-0,00143
6	0,01116	-3,16	0,00271	0,00286	-0,00830
7	0,05740	-2,11	0,04307	0,04540	-0,01200
8	0,19371	-1,05	0,22988	0,24231	0,04860
9	0,38742	0,00	0,39894	0,42052	0,03310
10	0,34868	1,05	0,22988	0,24231	-0,10637

Пример 4. Производится 10 подбрасываний симметричной монеты. Найдем вероятность того, что выпадет ровно m ($m = 0, 1, \dots, 10$) «гербов».

Так же, как и в предыдущем примере, сначала по формуле Бернулли $P_{10}(m) = \binom{10}{m} 0,5^{10}$ вычислим точные значения этих вероятностей (см. табл. 3, графа «Точное значение $P_{10}(m)$ »).

Воспользуемся теперь приближенными формулами. В данном примере $\lambda = 5$ и $\lambda' = 5$, поэтому применим формулу Муавра–Лапласа. Так как нас интересует вероятность выпадения ровно m «гербов», то необходимо применить локальную формулу Муавра–Лапласа, в которой

$$\sqrt{npq} = \sqrt{10 \cdot 0,5 \cdot 0,5} \cong 1,5811,$$

и для m «гербов»

$$x_m = \frac{m - 5}{\sqrt{10 \cdot 0,5 \cdot 0,5}} \cong \frac{m - 5}{1,5811}.$$

Результаты вычислений также приведены в табл. 3. Из табл. 3 видно, что и в этом случае максимальная абсолютная погрешность 0,00622 по отношению к максимальному значению $P_{10}(5) = 0,24609$ вероятности $P_{10}(m)$ достаточно мала.

Таблица 3

m	Точное значение $P_{10}(m)$	x_m	$\varphi(x_m)$	Значение, вычисленное по локальной формуле Муавра–Лапласа	Погрешность
0	0,00098	-3,16	0,00271	0,00171	0,00073
1	0,00977	-2,53	0,01625	0,01028	0,00051
2	0,04395	-1,90	0,06562	0,04150	-0,00245
3	0,11719	-1,26	0,18037	0,11408	-0,00311
4	0,20508	-0,63	0,32713	0,20690	0,00182
5	0,24609	0,00	0,39894	0,25231	0,00622
6	0,20508	0,63	0,32713	0,20690	0,00182
7	0,11719	1,26	0,18037	0,11408	-0,00311
8	0,04395	1,90	0,06562	0,04150	-0,00245
9	0,00977	2,53	0,01625	0,01028	0,00051
10	0,00098	3,16	0,00271	0,00171	0,00073

Читателю советуем самостоятельно применить формулу Пуассона и убедиться в том, что она дает существенно большие погрешности. \square

Пример 5. В тираже «Спортлото 6 из 49» участвуют 10 000 000 карточек. Найдем вероятность события A — хотя бы в одной из этих карточек зачеркнуты 6 выигрышных номеров (максимальный выигрыш). Естественно сразу же перейти к дополнительному событию \bar{A} — ни на одну карточку не выпадет максимальный выигрыш. Считая, что в каждой из карточек номера зачеркиваются случайным образом и независимо от остальных, видим, что число карточек, на которые выпал максимальный выигрыш, подчиняется биномиальному закону с параметрами $n = 10 000 000$ и $p = 7 \cdot 10^{-8}$ (см. пример 7 в гл. 2). Поскольку $\lambda = np = 0,7$, то для определения вероятности $P(\bar{A})$ воспользуемся формулой Пуассона. Тогда $P(\bar{A}) = P_{10 000 000}(0) \approx P(0; 0,7)$. Из табл. 1 приложения имеем

$P(0; 0,7) = 0,49659$, и, значит, $P(A) \approx 0,50341$. Таким образом, вероятность того, что хотя бы одна из 10 000 000 карточек окажется выигрышной, чуть больше 1/2. \square

Пример 6. Для определения числа π экспериментальным путем (см. пример 12 в гл. 2) производится $n = 10\,000$ бросаний иглы длиной $l = a/2$. Значение π определяется при этом формулой $\pi = n/\mu$, где μ — число пересечений иглой одной из прямых. Найдем вероятность того, что определенное таким образом приближенное значение π будет заключено в пределах от 3,14 до 3,15 (событие A). Событие A происходит тогда и только тогда, когда число пересечений будет лежать в пределах от $m_1 = 3175$ до $m_2 = 3184$. Вероятность успеха (пересечение иглой одной из прямых при $l = a/2$) $p = 1/\pi \cong 0,3183$, неудачи — $q = 1 - p \cong 0,6817$. Поскольку n , p и np велики, а по условию задачи нас интересует вероятность попадания на интервал, воспользуемся интегральной формулой Муавра–Лапласа, в которой $x_1 \cong (3175 - 3183)/\sqrt{2169} \cong -0,17$ и $x_2 \cong (3184 - 3183)/\sqrt{2169} \cong 0,02$. По табл. 3 приложения получаем, что $\Phi_0(x_1) = -0,06749$, $\Phi_0(x_2) = 0,00798$, и, значит, $P\{3175 \leq \mu \leq 3184\} \approx \Phi_0(x_2) - \Phi_0(x_1) = 0,07547$. Отсюда можно сделать вывод: 10 000 бросаний иглы явно мало для того, чтобы определенное экспериментально число π с достаточно большой вероятностью лежало в пределах от 3,14 до 3,15. \square

Пример 7. В сосуде содержится $n = 5,4 \cdot 10^{22}$ молекул газа. В некоторый момент времени сосуд делят непроницаемой перегородкой на две части одинаковых объемов. Считая, что каждая молекула с одинаковой вероятностью ($p = q = 1/2$) может находиться в любой из двух частей, определим вероятность события A — в одной из частей сосуда будет содержаться молекул по крайней мере на $10^{-8}\%$ (от общего числа молекул в сосуде) больше, чем в другой. Переходим к дополнительному событию \bar{A} , которое, как нетрудно видеть, состоит в том, что число молекул в первой части сосуда заключено в пределах от $m_1 = 2,7(10^{22} - 10^{12})$ до $m_2 = 2,7(10^{22} + 10^{12})$. Поскольку n весьма велико, а m_1 и m_2 различны, воспользуемся интегральной формулой Муавра–Лапласа, в которой $x_1 = -23,3$ и $x_2 = 23,3$. Следовательно, $P(\bar{A}) \approx \Phi_0(23,3) - \Phi_0(-23,3) = 2\Phi_0(23,3)$. В табл. 3 приложения, как и в большинстве других таблиц, значения функции $\Phi_0(x)$ приводятся для x , не превосходящих 5. Учитывая монотонное возрастание функции $\Phi_0(x)$, можно сделать следующую оценку для вероятности $P(\bar{A})$: $P(\bar{A}) > 2\Phi_0(5) = 2 \cdot 0,499\,999\,7 = 0,999\,999\,4$. Значит, $P(A) < 1 - 0,999\,999\,4 = 6 \cdot 10^{-7}$. Более точные расчеты показывают, что $2\Phi_0(23,3) < 10^{-100}$. Как мы видим, вероятность числу молекул в одной части сосуда отличаться от числа молекул в другой даже на $10^{-8}\%$ ничтожно мала. \square

Пример 8. Определим, какое число n подбрасываний симметричной монеты надо произвести, чтобы наблюденная частота $f = \mu/n$ выпадения «герба» отличалась от вероятности $p = 1/2$ выпадения «герба» не более чем на 0,01 с вероятностью 0,99. Предположим, что мы произвели n испытаний. Тогда число μ выпадений «герба» должно быть заключено в пределах от $m_1 = n/2 - 0,01n$ до $m_2 = n/2 + 0,01n$. Воспользуемся интегральной формулой Муавра–Лапласа, в которой $p = q = 1/2$, $x_1 = -0,01n/(0,5\sqrt{n}) = -0,02\sqrt{n}$, $x_2 = 0,01n/(0,5\sqrt{n}) = 0,02\sqrt{n}$. Тогда $P\{m_1 \leq \mu \leq m_2\} \approx \Phi_0(0,02\sqrt{n}) - \Phi_0(-0,02\sqrt{n}) = 2\Phi_0(0,02\sqrt{n})$. Поскольку из условия задачи известно, что $P\{m_1 \leq \mu \leq m_2\} = 0,99$, то получаем

$\Phi(0,02\sqrt{n}) = 0,495$. Из табл. 3 приложения находим $0,02\sqrt{n} \cong 2,58$. Отсюда следует, что $\sqrt{n} = 129$ и $n = 16641$. Таким образом, необходимо произвести около 17 000 подбрасываний.

Определим теперь, сколько нужно произвести подбрасываний симметричной монеты, чтобы частота выпадения «герба» отличалась от вероятности не более чем на 0,005 (а не на 0,01) с той же вероятностью 0,99. Иными словами, считая, что наблюденная частота f выпадения «герба» является оценкой вероятности $p = 1/2$ выпадения «герба», мы хотим узнать, во сколько раз нужно увеличить число испытаний, чтобы повысить точность оценки в два раза. Производя аналогичные вычисления, получаем $\sqrt{n} = 258$ и $n = 66564$. Значит, число подбрасываний нужно увеличить в 4 раза. Если бы мы захотели повысить точность оценки в 3 раза, то нам пришлось бы увеличить число подбрасываний уже в 9 раз. Нетрудно видеть, что выведенный закон присущ схеме Бернулли с любой вероятностью успеха p : оценивая вероятность p с помощью наблюденной частоты f , для улучшения оценки в α раз мы должны увеличить число наблюдений в α^2 раз. \square

5. Теорема Бернулли

Предположим, что мы произвели большое число n испытаний Бернулли с вероятностью успеха p . По числу полученных успехов μ определим наблюденную частоту $f = \mu/n$. Спрашивается, как сильно может отличаться наблюденная частота успеха f от вероятности успеха p ? Ясно, что, вообще говоря, частота f может принимать любые значения от 0 до 1. Так, мы вполне можем получить в n испытаниях одни неудачи. Но, как мы знаем, вероятность такого события равна q^n , и при больших n она будет весьма мала. Поэтому естественно ожидать, что при больших n частота f с большой вероятностью группируется вокруг вероятности p , что мы сейчас и установим, исходя из интегральной теоремы Муавра–Лапласа.

Пусть $\varepsilon > 0$ — любое сколь угодно малое число, а событие A заключается в том, что наблюдаемая в n испытаниях частота f отличается от вероятности по модулю не больше чем на ε . Иными словами, событие A происходит тогда и только тогда, когда число успехов в n испытаниях заключено в пределах от $m_1 = np - n\varepsilon$ до $m_2 = np + n\varepsilon$. Теперь, если n велико, то мы можем воспользоваться интегральной теоремой Муавра–Лапласа, в которой $x_1 = -\varepsilon\sqrt{n/(pq)}$, $x_2 = \varepsilon\sqrt{n/(pq)}$, и, значит,

$$\mathbf{P}(A) \approx \Phi(x_2) - \Phi(x_1).$$

Но с ростом n $x_1 \rightarrow -\infty$, а $x_2 \rightarrow \infty$. В свою очередь, $\Phi(-\infty) = 0$, $\Phi(\infty) = 1$. Поэтому, каково бы ни было $\varepsilon > 0$, с ростом n вероятность того, что частота f отличается от вероятности p не более чем на ε , стремится к 1.

Установленный нами факт предельного постоянства частоты впервые был обнаружен Я. Бернулли, он носит название (слабого) закона больших чисел или теоремы Бернулли. Закон больших чисел и его многочисленные обобщения являются звеном, позволяющим связать

аксиоматическое построение теории вероятностей с эмпирическим законом постоянства частоты, с которого мы начали путешествие в теорию вероятностей. Именно он позволяет обосновать то широкое применение методов теории вероятностей на практике, которое мы имеем в настоящее время.

Однако если произвести более строгий логический анализ, то окажется, что слабый закон больших чисел также не вполне удовлетворяет нашим исходным предпосылкам, поскольку когда мы говорим о стабильности частоты, то имеем в виду процесс, протекающий во времени. Слабый закон больших чисел утверждает только, что при большом, но фиксированном числе испытаний частота мало отличается от вероятности. Слабый закон больших чисел еще не исключает значительных, но редких отклонений частоты от вероятности при последовательном проведении испытаний. Здесь мы пока только отметим, что имеет место усиленный закон больших чисел, который в определенной степени устраняет это логическое несовершенство слабого закона больших чисел.

Более подробно закон больших чисел, как и интегральная теорема Муавра–Лапласа, будут обсуждены нами в гл. 8, посвященной предельным теоремам теории вероятностей.

6. Вычисление определенных интегралов методом Монте-Карло

В этом параграфе мы рассмотрим применение предельных теорем для вычисления многократных интегралов.

Начнем с повторения материала, излагаемого в любом курсе математического анализа.

Пусть непрерывная элементарная (т. е. выраженная в элементарных функциях) функция $y = f(x)$, заданная на отрезке $[a, b]$, ограничена снизу и сверху числами c и C соответственно, и необходимо вычислить определенный интеграл $\int_a^b f(x) dx$. Поскольку линейные замены $u = (x - a)/(b - a)$ и $v = (y - c)/(C - c)$ приводят к интегрированию функции, заданной на отрезке $[0, 1]$ и ограниченной снизу нулем и сверху единицей, то в дальнейшем будем считать, что $a = 0$, $b = 1$ и $0 \leq f(x) \leq 1$.

Для вычисления определенного интеграла $\int_0^1 f(x) dx$, как должно быть известно читателю, необходимо сначала найти какую-либо первообразную $F(x)$ функции $f(x)$ и затем воспользоваться формулой Ньютона–Лейбница, согласно которой $\int_0^1 f(x) dx = F(1) - F(0)$.

Однако хорошо известно и то, что интеграл далеко не от всякой элементарной функции $f(x)$ может быть выражен также в элемен-

тарных функциях. Тем не менее к настоящему времени разработаны методы приближенного численного интегрирования (простейшими из них являются формулы трапеций и Симпсона), для применения которых нужно вычислить значения функции $f(x)$ в определенных точках x_1, \dots, x_m (узлах интегрирования) и затем сложить полученные результаты с некоторыми весовыми коэффициентами. Обычно для достижения требуемой точности достаточно взять весьма небольшое число m узлов интегрирования. Пусть, например, $m = 10$. Тогда, даже если $f(x)$ задается сложным аналитическим выражением, современная ЭВМ легко справится с поставленной задачей и почти мгновенно подсчитает значение определенного интеграла.

Усложним задачу и предположим, что нужно вычислить k -кратный интеграл $\int \dots \int_{0 \leq x_1, \dots, x_k \leq 1} f(x_1, \dots, x_k) dx_1 \dots dx_k$, где $f(x_1, \dots, x_k)$ также заключена между 0 и 1. Теперь уже для применения численных методов необходимо взять m узлов интегрирования x_{i1}, \dots, x_{im} для каждой переменной интегрирования x_i ($i = 1, \dots, k$) и найти значения функции $f(x_1, \dots, x_k)$ для всех комбинаций $x_{1j_1}, \dots, x_{kj_k}$ (общих узлов интегрирования). Поскольку x_{ij_i} для каждого i может принимать m различных значений, то всего общих узлов интегрирования будет m^k .

В частности, если $m = 10$, то для вычисления 10-кратного интеграла необходимо подсчитать значения $f(x_1, \dots, x_{10})$ в 10^{10} точках, что представляет собой уже весьма трудоемкую задачу даже для самых современных ЭВМ.

Попробуем теперь решить те же самые задачи, привлекая вероятностные соображения. Для этого снова вернемся к однократному интегралу

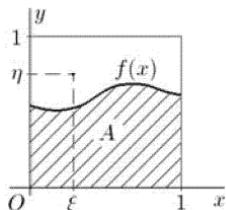


Рис. 1

$\int_0^1 f(x) dx$ и вспомним, что геометрически он представляет собой площадь области A , ограниченной графиком функции $f(x)$ (рис. 1).

Проведем опыт, заключающийся в бросании случайным образом (т. е. в соответствии с принципом геометрической вероятности) двух точек на отрезок $[0, 1]$. Обозначим координату одной из них через ξ , а другой — через η и отложим ξ и η по осям абсцисс и ординат соответственно (см. рис. 1). Проверим выполнение неравенства $\eta < f(\xi)$. Справедливость этого неравенства означает, что точка (ξ, η) попала в область A . Но в соответствии с принципом геометрической вероятности вероятность $P(A)$ попадания точки (ξ, η) в область A есть отношение площади A к площади единичного квадрата, т. е. $P(A) = \int_0^1 f(x) dx$.

Повторим описанный выше опыт n раз и по результатам наблюдений определим частоту $f = n_A/n$ появления события A , т. е. попадания точки (ξ, η) в область A . Поскольку по теореме Бернулли частота f

с ростом n стремится к вероятности $\mathbf{P}(A)$, то, подставляя вместо вероятности $\mathbf{P}(A)$ ее значение, получаем приближенное равенство

$$\int_0^1 f(x) dx \approx f = \frac{n_A}{n},$$

которое и служит для оценки интеграла по результатам случайных испытаний.

Описанный метод приближенного вычисления определенного интеграла носит название *метода статистических испытаний* или *метода Монте-Карло* (город Монте-Карло — место сосредоточения всемирно известных игорных домов). Название «метод Монте-Карло» связано с тем, что проводимые испытания очень напоминают подбрасывание монеты, бросание игральной кости или игру в рулетку.

Имеется существенное качественное различие между погрешностями, возникающими при применении методов численного интегрирования и метода Монте-Карло. В первом случае при выполнении соответствующих условий можно дать гарантированную оценку точности, т. е. указать достоверные границы, в которых обязательно будет заключено истинное значение вычисляемого интеграла. Во втором случае гарантированную оценку нельзя дать в принципе, а можно сказать только, что отклонение значения интеграла, вычисленного методом Монте-Карло, от истинного значения этого же интеграла не превосходит некоторой величины с определенной вероятностью.

Для определения количественного значения погрешности при применении метода Монте-Карло обычно пользуются интегральной формулой Муавра–Лапласа. Пусть истинное значение интеграла

$$\int_0^1 f(x) dx = \mathbf{P}(A) = p.$$

Тогда для заданного ε событие $\{$ приближенное значение $f = n_A/n$ интеграла удовлетворяет неравенству $|f - p| < \varepsilon\}$ совпадает с событием $\{\text{число успехов } n \text{ в схеме Бернулли с вероятностью успеха } p \text{ удовлетворяет неравенству } |n_A - np| < \varepsilon n\}$ или, что то же самое, с событием $\{\text{число успехов } n_A \text{ заключено в пределах от } np - \varepsilon n \text{ до } np + \varepsilon n\}$. Воспользовавшись теперь интегральной формулой Муавра–Лапласа, получаем, что вероятность p_ε того, что значение интеграла $\int_0^1 f(x) dx$, вычисленного методом Монте-Карло, отличается от истинного значения p этого интеграла не более чем на ε , задается выражением

$$p_\varepsilon \approx \Phi\left(\varepsilon \sqrt{\frac{n}{pq}}\right) - \Phi\left(-\varepsilon \sqrt{\frac{n}{pq}}\right) = 2\Phi_0\left(\varepsilon \sqrt{\frac{n}{pq}}\right).$$

Из последней формулы видно, что если мы хотим уменьшить погрешность ε в α раз с сохранением той же вероятности p_ε , то мы должны произвести в α^2 раз больше опытов (см. пример 8).

При вычислении вероятности p_ε часто возникает затруднение, связанное с тем, что значение p нам неизвестно. В этом случае обычно идут двумя путями. Первый путь заключается в том, что величина pq заменяется ее верхней оценкой $1/4$. Второй путь состоит в замене в формуле для p_ε вероятности p ее приближенным значением f .

Более точную оценку погрешности при применении метода Монте-Карло можно получить на основе результатов математической статистики (раздел «Доверительные интервалы»).

Метод Монте-Карло очевидным образом переносится и на тот случай, когда нужно вычислить k -кратный интеграл

$$\int_{0 \leqslant x_1, \dots, x_k \leqslant 1} \dots \int f(x_1, \dots, x_k) dx_1 \dots dx_k.$$

Единственное отличие заключается в том, что на отрезок $[0, 1]$ необходимо бросать уже не две, а $k + 1$ точек $\xi_1, \dots, \xi_k, \eta$ и проверять выполнение неравенства $\eta < f(\xi_1, \dots, \xi_k)$. Все остальные результаты, в том числе окончательное выражение

$$\int_{0 \leqslant x_1, \dots, x_k \leqslant 1} \dots \int f(x_1, \dots, x_k) dx_1 \dots dx_k \approx f = \frac{n_A}{n}$$

для приближенного значения интеграла и формула для оценки вероятности p_ε погрешности, полностью сохраняются и в этом случае.

Естественно, возникает вопрос: *в каких случаях следует применять методы численного интегрирования, а в каких — метод Монте-Карло?* Для этого вернемся к началу параграфа и снова предположим, что для достижения заданной точности вычисления однократного интеграла необходимо взять $m = 10$ узлов интегрирования x_i , т. е. фактически произвести 10 «обобщенных» операций, состоящих в определении $f(x_i)$. Опыт показывает (см. также пример 8), что для достижения аналогичной точности методом Монте-Карло требуются десятки, а то и сотни тысяч испытаний. Пусть для определенности необходимое число испытаний $n = 10^5$. Поскольку, как правило, основное время при вычислениях занимает нахождение значения $f(\xi)$, то трудоемкость метода Монте-Карло оценивается 10^5 «обобщенными» операциями, т. е. метод Монте-Карло существенно уступает численным методам. Перейдем к вычислению интеграла кратности k . Тогда, как мы уже говорили, при применении численных методов трудоемкость вычислений составит уже 10^k операций. Однако использование метода Монте-Карло ведет лишь к незначительному увеличению трудоемкости. В частности, при вычислении 5-кратного интеграла трудоемкость обоих методов практически совпадает, а при вычислении интегралов большей кратности

численные методы уже начинают проигрывать методу Монте-Карло, причем тем больше, чем больше кратность интеграла.

Резюмируя изложенное выше, можно сказать, что применение метода Монте-Карло оправдано только при вычислении кратных интегралов, причем в случае большой кратности метод Монте-Карло просто не имеет конкурентов со стороны методов численного интегрирования.

К достоинствам метода Монте-Карло можно отнести и то, что погрешность при вычислении интеграла с его помощью вообще не зависит от свойств гладкости функции $f(x)$, в то время как для получения заданной точности при численном интегрировании необходимо выполнение ограничений на производные функции $f(x)$. В частности, метод Монте-Карло с одинаковым успехом применим для интегрирования как непрерывных функций, так и функций, терпящих разрывы.

Конечно, при вычислении интегралов методом Монте-Карло никто не производит физического бросания точки на отрезок. Для этой цели служат специальные *программы* или *датчики случайных* или, точнее говоря, «*псевдослучайных*» чисел.

Следует обратить внимание на то, что метод Монте-Карло не позволяет беспрепятственно уменьшать погрешность вычислений. И дело здесь даже не в том, что необходимо проводить очень большое число испытаний. «Псевдослучайные» числа, как вытекает из их названия, не удовлетворяют полностью свойству случайности. Поэтому, прежде чем использовать какой-либо датчик «псевдослучайных» чисел, обычно производят многочисленные проверки (на отсутствие периодичности, на равномерность, на независимость и т.д.) с помощью различных критериев. Однако даже выдержавшие самые строгие проверки датчики генерируют числа, которые не могут быть отнесены к разряду «случайных» в полном смысле этого слова.

7. Полиномиальная схема

В заключение скажем несколько слов о так называемой полиномиальной схеме. Если схема Бернулли интерпретируется как подбрасывание несимметричной монеты, то полиномиальную схему можно трактовать как обобщение статистики Максвелла–Больцмана на тот случай, когда вероятности попадания каждой частицы в различные ячейки неодинаковы. Итак, предположим, что опыт состоит из n независимых одинаковых испытаний, в каждом из которых может произойти одно и только одно из m несовместных событий A_1, \dots, A_m , причем событие A_i наступает с вероятностью p_i . Тогда вероятность $P(n_1, \dots, n_m)$ того, что в n испытаниях событие A_1 произойдет ровно n_1 раз, ..., событие A_m произойдет ровно n_m раз ($n_1 + \dots + n_m = n$), определяется выражением

$$P(n_1, \dots, n_m) = \frac{n!}{n_1! \dots n_m!} p_1^{n_1} \dots p_m^{n_m}.$$

Последняя формула носит название *полиномиального распределения*. Ее вывод аналогичен выводу формулы Бернулли с учетом комбинаторных соотношений, используемых при рассмотрении статистики Максвелла–Больцмана, и любознательный читатель вполне может привести его самостоятельно. Полиномиальную вероятность $P(n_1, \dots, n_m)$ можно получить также как коэффициент при $x_1^{n_1} \dots x_m^{n_m}$ в разложении полинома $(p_1x_1 + \dots + p_mx_m)^n$ в ряд по степеням x_1, \dots, x_m .

Полиномиальная схема обладает теми же предельными свойствами, что и схема Бернулли. Так, если устремить n к бесконечности, то вероятность $P(n_1, \dots, n_m)$ приближенно вычисляется с помощью так называемой многомерной нормальной плотности. Вводя в рассмотрение вектор (f_1, \dots, f_m) наблюденных частот появлений событий A_1, \dots, A_m , можно доказать, что при большом числе испытаний n этот вектор мало отличается от вектора вероятностей (p_1, \dots, p_m) .

Пример 9. В магазине висит 1 костюм второго роста, 2 костюма третьего роста, 3 костюма четвертого роста. Костюм второго роста спрашивается с вероятностью 0,2, костюм третьего роста — с вероятностью 0,3, костюм четвертого роста — с вероятностью 0,5. В магазин обратились 3 покупателя. Найдем вероятность того, что хотя бы один из них ушел без покупки (событие A). Представим событие A в виде суммы несовместных событий:

$$A = A_{300} + A_{210} + A_{201} + A_{030},$$

где A_{ijk} означает, что i покупателей спросили костюм второго роста, j — третьего роста и k — четвертого. Воспользовавшись теперь полиномиальным распределением, получаем

$$\begin{aligned} P(A) &= P(3, 0, 0) + P(2, 1, 0) + P(2, 0, 1) + P(0, 3, 0) = \\ &= \frac{3!}{3! 0! 0!} (0,2)^3 + \frac{3!}{2! 1! 0!} (0,2)^2 \cdot 0,3 + \frac{3!}{2! 0! 1!} (0,2)^2 \cdot 0,5 + \frac{3!}{0! 3! 0!} (0,3)^3 = 0,131. \end{aligned}$$

Таким образом, искомая вероятность $P(A) = 0,131$. □

Глава 5

СЛУЧАЙНЫЕ ВЕЛИЧИНЫ И ИХ РАСПРЕДЕЛЕНИЯ

В предыдущих главах мы рассмотрели аксиоматическое построение вероятностного пространства (Ω, \mathcal{B}, P) , а также разобрали некоторые простейшие вероятностные схемы. Однако теория вероятностей не достигла бы такого расцвета и не была бы столь широко используема на практике, если бы занималась только лишь случайными событиями. Возвращаясь к истокам возникновения теории вероятностей, вспомним, что уже в азартных играх интерес играющих вызывает не наступление случайного исхода, а связанный с ним выигрыш или проигрыш, т. е. определенная *числовая величина*, поставленная в соответствие этому исходу. Вполне естественно такую числовую величину назвать *случайной величиной*. Изучением последнего понятия мы сейчас и займемся.

1. Случайная величина

Рассмотрим вероятностное пространство (Ω, \mathcal{B}, P) , т. е. пространство элементарных исходов Ω , σ -алгебру событий \mathcal{B} из Ω и определенную на ней вероятность P .

Случайной величиной ξ называется функция, ставящая в соответствие каждому элементарному исходу ω число $\xi = \xi(\omega)$. Для того чтобы такое определение было математически корректным, необходимо добавить следующее требование: для любого числа x множество $\{\omega: \xi(\omega) < x\}$ элементарных исходов ω , для которых $\xi(\omega) < x$, является событием, или, иными словами, принадлежит σ -алгебре \mathcal{B} (это свойство носит название измеримости функции $\xi = \xi(\omega)$ относительно σ -алгебры \mathcal{B}). Таким образом, с точки зрения функционального анализа случайная величина представляет собой не что иное, как обычную числовую функцию, заданную на пространстве элементарных исходов Ω (и измеримую относительно σ -алгебры \mathcal{B}). Специфика теории вероятностей проявляется в том, что на Ω задана также вероятность P . Случайные величины будем обозначать греческими буквами, снабжая их при необходимости индексами: ξ , η_1 , μ_2 и т. д.

Для краткости условимся в дальнейшем вместо записи $\{\omega: \xi(\omega) < x\}$ использовать запись $\{\xi(\omega) < x\}$, если необходимо подчеркнуть связь случайной величины с пространством элементарных исходов Ω , или даже запись $\{\xi < x\}$, если не акцентируется внимание на этой связи.

Пример 1. Два игрока играют в «орлянку» на следующих условиях: если при подбрасывании монеты выпадает «герб», то первый игрок платит второму 1 руб., если «цифра», то второй игрок платит первому 2 руб. Опишем случайную величину ξ , равную выигрышу первого игрока в этой игре (при одном подбрасывании монеты). Как мы знаем, пространство элементарных исходов Ω состоит из двух исходов: ω_1 — выпадение «герба» и ω_2 — «цифры». σ -алгебра событий \mathfrak{B} насчитывает 4 события: \emptyset , $\{\omega_1\}$, $\{\omega_2\}$ и Ω . Предполагая, что монета симметричная, найдем вероятности всех событий из \mathfrak{B} : $P(\emptyset) = 0$, $P(\omega_1) = P(\omega_2) = 1/2$, $P(\Omega) = 1$. Итак, вероятностное пространство $(\Omega, \mathfrak{B}, P)$ нами определено. Осталось заметить, что случайная величина ξ принимает значение -1 , если выпал «герб» ($\xi(\omega_1) = -1$), и 2 , если выпала «цифра» ($\xi(\omega_2) = 2$). Измеримость функции $\xi(\omega)$ очевидна, поскольку при $x \leq -1$ множество $\{\xi(\omega) < x\}$ является невозможным событием \emptyset , при $-1 < x \leq 2$ множество $\{\xi(\omega) < x\}$ состоит из элементарного исхода ω_1 и, наконец, при $x > 2$ множество $\{\xi(\omega) < x\}$ состоит из двух исходов ω_1 и ω_2 , т. е. представляет собой достоверное событие Ω . \square

Пример 2. Рассмотрим схему Бернуlli с вероятностью успеха p . Сопоставим каждому элементарному исходу $\omega = \text{УНН...У}$ функцию $\mu(\omega)$, равную числу успехов при этом исходе, т. е. числу букв У, содержащихся в последовательности УНН...У. Измеримость функции $\mu(\omega)$ проверяется точно так же, как в предыдущем примере. Таким образом, μ — случайная величина. Отметим, что значения $\mu(\omega)$ совпадают для тех элементарных исходов ω , при которых произошло одинаковое число успехов. \square

Пример 3. Случайными величинами будут число очков, выпавших при бросании игральной кости, а также суммарные числа очков, выпавших при бросании двух, трех и более костей. Советуем читателю для этого примера самостоятельно построить пространство элементарных исходов Ω , σ -алгебру \mathfrak{B} , вероятность P определить случайную величину ξ . \square

Пример 4. На отрезок $[0, 1]$ в соответствии с принципом геометрической вероятности падает идеальная точка. Пусть $\xi(\omega) = \omega$ — координата ее падения. Пространство элементарных исходов Ω в данном случае совпадает с отрезком $[0, 1]$, σ -алгебра \mathfrak{B} является борелевской (порожденной всевозможными интервалами) σ -алгеброй на этом отрезке, а вероятность попадания на каждый интервал внутри отрезка $[0, 1]$ равна его длине. Измеримость функции $\xi(\omega)$ вытекает из того, что множество $\{\xi(\omega) < x\}$ при $x \leq 0$ пусто, при $0 < x \leq 1$ совпадает с интервалом $[0, x]$ (а все интервалы, как мы знаем, принадлежат борелевской σ -алгебре) и, наконец, при $x > 1$ совпадает со всем отрезком $[0, 1]$, т. е. является достоверным событием. Таким образом, ξ — случайная величина.

Отметим, что в этом примере мы имеем дело с интересным явлением: значение случайной величины $\xi = \xi(\omega)$ для каждого элементарного исхода ω совпадает с самим этим исходом (в данном случае, наверное, лучше было бы сказать с «номером» этого исхода). Оказывается, такая ситуация встречается очень часто и связано это с тем, что, как правило, исследователь наблюдает именно случайную величину ξ , значит, для него понятие «элементарный исход» отождествлено с понятием «значение случайной величины». \square

Пример 5. На плоский экран падает частица. Будем считать, что нам известна вероятность попадания частицы в каждое (измеримое) множество на экране. Случайными величинами в данном случае будут, например, расстояние

от центра экрана до точки падения, квадрат этого расстояния, угол в полярной системе координат и т. д. \square

Прежде чем перейти к дальнейшему изучению случайных величин, отметим, что все известные из курса математического анализа функции, как и вообще все функции, встречающиеся в реальной жизни, являются измеримыми. Поэтому в дальнейшем понятие измеримости мы нигде больше использовать не будем.

2. Функция распределения случайной величины

Функцией распределения (вероятностей) случайной величины ξ называется функция $F(x)$, значение которой в точке x равно вероятности события $\{\xi < x\}$, т. е. события, состоящего из тех и только тех элементарных исходов ω , для которых $\xi(\omega) < x$:

$$F(x) = \mathbf{P}\{\xi < x\}.$$

Обычно говорят, что значение функции распределения в точке x равно вероятности случайной величине ξ принять значение, меньшее x . Правда, в таком определении имеется маленькая стилистическая неточность, связанная с тем, что слово «вероятность» мы обязаны употреблять только вместе со словом «событие».

Выведем некоторые очевидные свойства функции распределения. Так как, по определению, функция распределения является вероятностью, то

1. $0 \leq F(x) \leq 1$.

Далее, если $x_1 < x_2$, то событие $\{\xi < x_1\}$ принадлежит событию $\{\xi < x_2\}$ и, значит,

2. $F(x_1) \leq F(x_2)$ при $x_1 < x_2$ ($F(x)$ — неубывающая функция).

Положим

$$F(-\infty) = \lim_{x \rightarrow -\infty} F(x) \quad \text{и} \quad F(+\infty) = \lim_{x \rightarrow \infty} F(x).$$

Поскольку событие $\{\xi < -\infty\}$ является невозможным, а $\{\xi < \infty\}$ — достоверным, то имеем

3. $F(-\infty) = 0, \quad F(\infty) = 1$.

Событие $\{\xi < x_2\}$ при $x_1 < x_2$ представляет собой объединение двух непересекающихся событий: $\{\xi < x_1\}$ — случайная величина ξ приняла значение, меньшее x_1 , и $\{x_1 \leq \xi < x_2\}$ — случайная величина ξ приняла значение, лежащее в интервале $[x_1, x_2]$. Поэтому из аксиомы сложения получаем

4. $\mathbf{P}\{x_1 \leq \xi < x_2\} = F(x_2) - F(x_1)$.

Наконец, пусть x_1, \dots, x_n, \dots — возрастающая последовательность чисел, сходящаяся к x . Тогда событие $A = \{\xi < x\}$ является счетным объединением несовместных событий

$$A_1 = \{\xi < x_1\}, \quad A_n = \{x_{n-1} \leq \xi < x_n\} \quad (n = 2, 3, \dots),$$

т. е.

$$A = A_1 + \dots + A_n + \dots$$

В силу расширенной аксиомы сложения

$$\mathbf{P}(A) = \mathbf{P}(A_1) + \mathbf{P}(A_2) + \dots + \mathbf{P}(A_n) + \dots =$$

$$= F(x_1) + [F(x_2) - F(x_1)] + \dots + [F(x_n) - F(x_{n-1})] + \dots =$$

$$= \lim_{n \rightarrow \infty} F(x_n).$$

Следовательно,

5. $F(x) = F(x - 0)$ ($F(x)$ — непрерывная слева функция).

Типичный вид функции распределения приведен на рис. 1.

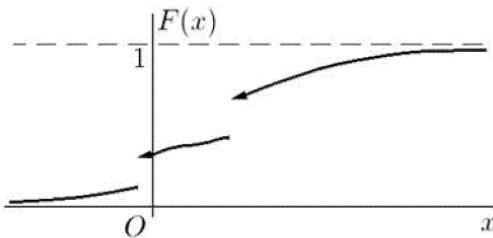


Рис. 1

Заметим, что, зная функцию распределения $F(x)$, можно однозначно определить вероятность попадания случайной величины ξ не только на интервал $[x_1, x_2]$, но и в любое измеримое (борелевское) множество на прямой.

Итак, с любой случайной величиной связана ее функция распределения. Отметим, что справедливо и обратное. Любая неубывающая непрерывная слева функция $F(x)$, удовлетворяющая условиям $F(-\infty) = 0$ и $F(\infty) = 1$, является функцией распределения некоторой случайной величины ξ . Действительно, можно рассмотреть падение идеальной точки на прямую $(-\infty, \infty)$, которая в этом случае принимается в качестве пространства элементарных исходов (см. также пример 22 в гл. 1). Поставим в соответствие каждому событию, заключающемуся в том, что точка попала на интервал $(-\infty, x)$ на прямой, число $F(x)$, а событию, заключающемуся в попадании точки на интервал $[x_1, x_2]$, — число $F(x_2) - F(x_1)$. Определенная таким образом для всех событий, связанных с попаданием точки на интервал $[x_1, x_2]$, числовая функция будет удовлетворять трем аксиомам вероятности. Для любых

других событий, составляющих σ -алгебру boreлевских множеств на прямой, вероятность определяется единственным образом с помощью теоремы о продолжении меры. Если теперь взять в качестве случайной величины ξ координату падения, т. е. положить $\xi(\omega) = \omega$, то $F(x)$ будет являться функцией распределения ξ .

В дальнейшем иногда для того, чтобы подчеркнуть, какой именно случайной величине принадлежит функция распределения $F(x)$, будем к функции распределения приписывать нижний индекс, обозначающий эту случайную величину: $F_\xi(x) = P\{\xi < x\}$.

В некоторых учебниках функцией распределения называют вероятность события $\{\xi \leq x\}$. Такое определение ничего не меняет в наших рассуждениях. Единственное изменение касается свойства 5: функция $F(x)$ будет непрерывна справа.

Обычно (и мы будем также придерживаться этой традиции) для того, чтобы избежать сложного для неподготовленного читателя понятия интеграла Стильеса, при знакомстве с теорией вероятностей ограничиваются изучением так называемых дискретных и непрерывных случайных величин. Дискретную случайную величину вкратце можно охарактеризовать как случайную величину, все возможные значения которой можно пересчитать. В свою очередь, для непрерывной случайной величины вероятность попадания на «малый» интервал $[x, x + \Delta]$ приближенно пропорциональна длине этого интервала Δ с коэффициентом пропорциональности $p(x)$, зависящим от x и носящим название плотности распределения.

Однако существуют случайные величины, не относящиеся ни к одному из этих двух типов. Простейшим примером такой случайной величины является время работы электрической лампочки. Купленная лампочка может с ненулевой вероятностью оказаться бракованной, т. е. время ее работы будет равно нулю, и в этом смысле его необходимо отнести к дискретным случайным величинам. Если же лампочка окажется исправной, то мы не сможем пересчитать все моменты времени, в которые она может отказать, и тогда время ее безотказной работы естественно считать непрерывной случайной величиной. Для данного примера можно сказать, что мы имеем дело со «смесью» дискретной и непрерывной случайных величин.

Существуют и более сложные примеры, в которых случайные величины уже не являются «смесью» дискретной и непрерывной компонент, но с точки зрения практики эти примеры представляют собой математическую абстракцию.

Часто поведение случайной величины удобно характеризовать не с помощью функции распределения, а как-то иначе. Если при этом возможно однозначно восстановить функцию распределения, то такая характеристика называется *законом распределения* (или просто *распределением*) случайной величины. Примерами законов распределения являются ряд распределения и плотность распределения, которые мы рассмотрим в следующих параграфах.

3. Дискретные случайные величины

Как уже говорилось, дискретной называется случайная величина, которая каждому элементарному исходу ω ставит в соответствие одно из конечного (или в общем случае счетного) набора чисел X_1, X_2, \dots, X_n ($X_1, X_2, \dots, X_n, \dots$). Дискретную случайную величину удобно характеризовать рядом распределения.

Рядом распределения (вероятностей) дискретной случайной величины называется таблица (табл. 1), состоящая из двух строк: в верхней строке перечислены все возможные значения случайной величины, а в нижней — вероятности $p_i = P\{\xi = X_i\}$ того, что случайная величина примет эти значения.

Таблица 1

ξ	X_1	X_2	\dots	X_i	\dots	X_n
P	p_1	p_2	\dots	p_i	\dots	p_n

Иногда для того, чтобы подчеркнуть, что ряд распределения относится именно к случайной величине ξ , будем наряду с записью p_i употреблять запись $p_{\xi i}$.

Рассмотрим некоторые наиболее часто встречающиеся на практике распределения дискретных случайных величин.

Биномиальное распределение. Дискретная случайная величина μ распределена по биномиальному закону, если она принимает значения $0, 1, 2, \dots, n$ в соответствии с рядом распределения, представленным в табл. 2, где $0 < p, q < 1$ и $p + q = 1$.

Таблица 2

μ	0	1	\dots	i	\dots	n
P	q^n	$\binom{n}{1}pq^{n-1}$	\dots	$\binom{n}{i}p^iq^{n-i}$	\dots	p^n

Биномиальное распределение является не чем иным, как распределением числа успехов μ в n испытаниях Бернулли с вероятностью успеха p и неудачи $q = 1 - p$ (см. пример 2).

Пуассоновское распределение. Дискретная случайная величина ξ распределена по закону Пуассона, если она принимает целые неотрицательные значения с вероятностями, представленными рядом распределения в табл. 3, где $\lambda > 0$ — параметр пуассоновского распределения.

С распределением Пуассона мы тоже уже встречались в предельной теореме Пуассона. Распределение Пуассона носит также название *закона редких событий*, поскольку оно всегда появляется там, где производится большое число испытаний, в каждом из которых с ма-

Таблица 3

ξ	0	1	2	...	n	...
P	$e^{-\lambda}$	$\lambda e^{-\lambda}$	$\frac{\lambda^2}{2!} e^{-\lambda}$...	$\frac{\lambda^n}{n!} e^{-\lambda}$...

лой вероятностью происходит «редкое» событие. По закону Пуассона распределены, например, число вызовов, поступивших на телефонную станцию; число метеоритов, упавших в определенном районе; число распавшихся нестабильных частиц и т. д.

Геометрическое распределение. Снова рассмотрим схему Бернулли. Пусть ξ — число испытаний, которое необходимо провести, прежде чем появится первый успех. Тогда ξ — дискретная случайная величина, принимающая значения $0, 1, 2, \dots, n, \dots$. Определим вероятность события $\{\xi = n\}$. Очевидно, что $\xi = 0$, если в первом же испытании произойдет успех. Поэтому $P\{\xi = 0\} = p$. Далее, $\xi = 1$ в том случае, когда в первом испытании произошла неудача, а во втором — успех. Но вероятность такого события, как мы знаем, равна qp , т. е. $P\{\xi = 1\} = qp$. Аналогично, $\xi = 2$, если в первых двух испытаниях произошли неудачи, а в третьем — успех, и, значит, $P\{\xi = 2\} = qqp$. Продолжая эту процедуру, получаем ряд распределения, представленный в табл. 4.

Таблица 4

ξ	0	1	2	...	n	...
P	p	qp	q^2p	...	$q^n p$...

Случайная величина с таким рядом распределения называется распределенной по геометрическому закону.

Покажем теперь, как по ряду распределения дискретной случайной величины построить ее функцию распределения $F(x)$. Пусть ξ — дискретная случайная величина, заданная своим рядом распределения, причем значения X_1, X_2, \dots, X_n расположены в порядке возрастания. Тогда для всех $x \leq X_1$ событие $\{\xi < x\}$ является невозможным и поэтому в соответствии с определением $F(x) = 0$ (рис. 2). Если

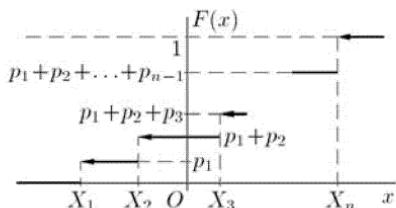


Рис. 2

$X_1 < x \leq X_2$, то событие $\{\xi < x\}$ состоит из тех и только тех элементарных исходов ω , для которых $\xi(\omega) = X_1$, и, следовательно, $F(x) = p_1$. Аналогично, при $X_2 < x \leq X_3$ событие $\{\xi < x\}$ состоит из тех элементарных исходов ω , для которых либо $\xi(\omega) = X_1$, либо $\xi(\omega) = X_2$, т. е. $\{\xi < x\} = \{\xi = X_1\} + \{\xi = X_2\}$, а, значит, $F(x) = p_1 + p_2$ и т. д. Наконец, при $x > X_n$ событие $\{\xi < x\}$ достоверно и $F(x) = 1$. Таким образом, функция распределения дискретной случайной величины является кусочно-постоянной функцией, принимающей на интервале $(-\infty, X_1]$ значение 0, на интервалах $(X_i, X_{i+1}]$ ($1 \leq i < n$) — значение $p_1 + \dots + p_i$ и на интервале (X_n, ∞) — значение 1.

Пример 6. На зачете студент получил $n = 4$ задачи. Вероятность решить правильно каждую задачу $p = 0,8$. Определим ряд распределения и построим функцию распределения случайной величины μ — числа правильно решенных задач. В данном случае мы имеем дело с биномиальным законом. Подставляя в ряд распределения, заданный табл. 2, $n = 4$, $p = 0,8$ и $q = 0,2$, получаем ряд распределения и функцию распределения, представленные в табл. 5 и на рис. 3. \square

Таблица 5

μ	0	1	2	3	4
P	0,0016	0,0256	0,1536	0,4096	0,4096

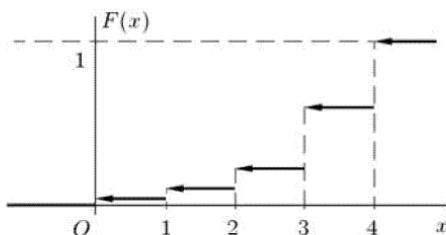


Рис. 3

Пример 7. Вероятность получить заданный эффект в физическом опыте $p = 0,4$. Определим ряд распределения и построим функцию распределения случайной величины ξ , равной числу «пустых» опытов, которые должен произвести экспериментатор, прежде чем он получит необходимый эффект. Случайная величина ξ распределена по геометрическому закону. Поэтому, воспользовавшись табл. 4, имеем ряд распределения и функцию распределения, представленные в табл. 6 и на рис. 4. \square

Таблица 6

μ	0	1	2	3	...
P	0,4	0,24	0,144	0,0864	...

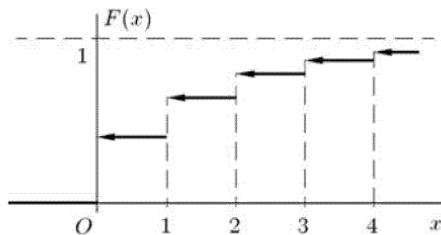


Рис. 4

Пример 8. Выпишем ряд распределения случайной величины ξ — числа угаданных номеров в «Спортлото 6 из 49». Как мы знаем (см. пример 7 в гл. 2), число угаданных номеров распределено по гипергеометрическому закону. В этом же примере были найдены $p_3 = P\{\xi = 3\} = P(A_3)$, $p_4 = P\{\xi = 4\} = P(A_4)$, $p_5 = P\{\xi = 5\} = P(A_5)$ и $p_6 = P\{\xi = 6\} = P(A_6)$. Аналогично определяются и вероятности $p_0 = P\{\xi = 0\}$ — не угадать ни одного номера, $p_1 = P\{\xi = 1\}$ — угадать ровно один номер и $p_2 = P\{\xi = 2\}$ — угадать ровно два номера:

$$p_0 = \frac{\binom{6}{0} \binom{43}{6}}{\binom{49}{6}} \cong 0,4360, \quad p_1 = \frac{\binom{6}{1} \binom{43}{5}}{\binom{49}{6}} \cong 0,4130, \quad p_2 = \frac{\binom{6}{2} \binom{43}{4}}{\binom{49}{6}} \cong 0,1324.$$

Таблица 7

ξ	0	1	2	3	4	5	6
P	0,4360	0,4130	0,1324	0,0176	0,00097	$2 \cdot 10^{-5}$	$7 \cdot 10^{-8}$

Окончательно получаем ряд распределения, представленный в табл. 7. \square

4. Непрерывные случайные величины

Непрерывной называется случайная величина ξ , функцию распределения которой $F(x)$ можно представить в виде

$$F(x) = \int_{-\infty}^x p(y) dy.$$

Функция $p(x)$ называется *плотностью распределения (вероятностей) случайной величины ξ* . Так же, как и прежде, иногда для того, чтобы подчеркнуть принадлежность плотности распределения случайной величине ξ , будем наряду с записью $p(x)$ употреблять запись $p_\xi(x)$. Отметим, что все реально встречающиеся плотности распределения являются непрерывными (за исключением, быть может, конечного числа

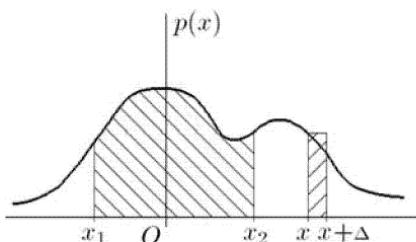


Рис. 5

точек) функциями и, следовательно, для них плотность распределения $p(x)$ представляет собой производную функции распределения $F(x)$, т. е. $p(x) = F'(x)$. Для простоты изложения в дальнейшем будем рассматривать только такие плотности распределения. Типичный вид плотности распределения изображен на рис. 5.

Выведем простейшие свойства плотности распределения.

Поскольку плотность распределения является производной от функции распределения, а функция распределения — неубывающая дифференцируемая функция, то

1. $p(x) \geq 0$.

Далее, $\mathbf{P}\{x_1 \leq \xi < x_2\} = F(x_2) - F(x_1)$. Значит, в силу определения непрерывной случайной величины

$$\mathbf{2.} \quad \mathbf{P}\{x_1 \leq \xi < x_2\} = \int_{x_1}^{x_2} p(y) dy.$$

Таким образом, вероятность попадания случайной величины на интервал $[x_1, x_2]$ численно равна площади криволинейной трапеции, заштрихованной на рис. 5.

В частности, если $x_1 = -\infty$, $x_2 = \infty$, то событие $\{-\infty < \xi < \infty\}$ является достоверным, и поэтому

$$\mathbf{3.} \quad \int_{-\infty}^{\infty} p(x) dx = 1.$$

Иными словами, площадь, целиком заключенная под всей кривой, изображающей плотность распределения, равна единице.

Наконец, часто бывает полезна следующая трактовка плотности распределения. Как видно из рис. 5, если Δ мало, то вероятность попадания на интервал $[x, x + \Delta]$ приближенно совпадает с площадью прямоугольника со сторонами Δ и $p(x)$. Значит, с точностью до $o(\Delta)$

4. $\mathbf{P}\{\xi \leq \xi < x + \Delta\} \approx p(x)\Delta$.

Отсюда, в частности, следует, что

5. $\mathbf{P}\{\xi = x\} = 0$,

т. е. вероятность попадания в любую (заданную до опыта) точку для непрерывной случайной величины равна нулю. Здесь снова возникает кажущееся логическое противоречие: до опыта каждому возможному значению непрерывной случайной величины мы можем присвоить только вероятность, равную нулю, однако после опыта случайная величина все же принимает некоторое значение. Решение этого противоречия

опять связано с непрерывной структурой числовой прямой. Для того чтобы гарантировать ненулевую вероятность попадания в некоторое подмножество прямой, необходимо, чтобы это подмножество содержало более чем конечное и даже более чем счетное число точек. Из свойства 5 вытекает также, что в свойствах 2 и 4 знак \leqslant можно заменить на знак строгого неравенства $<$. Например, свойство 2 можно переписать в виде

$$\mathbf{2'}. \quad \mathbb{P}\{x_1 < \xi < x_2\} = \int_{x_1}^{x_2} p(y) dy.$$

Рассмотрим некоторые наиболее важные распределения непрерывных случайных величин.

Равномерное распределение. Равномерно распределенная на отрезке $[a, b]$ случайная величина имеет плотность распределения

$$p(x) = \begin{cases} \frac{1}{b-a}, & \text{если } a \leqslant x \leqslant b, \\ 0, & \text{если } x < a \text{ или } x > b. \end{cases}$$

Легко видеть, что функция распределения в этом случае определяется выражением

$$F(x) = \begin{cases} 0, & \text{если } x < a, \\ \frac{x-a}{b-a}, & \text{если } a \leqslant x \leqslant b, \\ 1, & \text{если } x > b. \end{cases}$$

Графики плотности распределения $p(x)$ и функции распределения $F(x)$ приведены на рис. 6 и рис. 7.

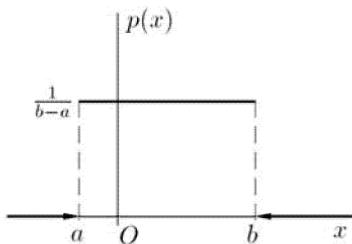


Рис. 6

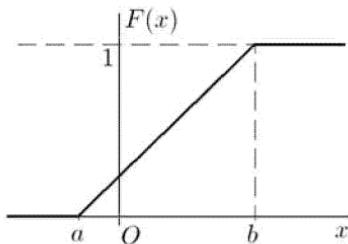


Рис. 7

Вероятность попадания равномерно распределенной случайной величины на интервал (x_1, x_2) , лежащий внутри отрезка $[a, b]$, равна $F(x_2) - F(x_1) = (x_2 - x_1)/(b - a)$, т. е. пропорциональна длине этого интервала. Таким образом, равномерное распределение реализует принцип геометрической вероятности при бросании точки на отрезок $[a, b]$.

Заметим, что в примере 4 случайная величина ξ равномерно распределена на отрезке $[0, 1]$.

Экспоненциальное распределение. Случайная величина подчиняется экспоненциальному (показательному) закону, если она имеет плотность распределения

$$p(x) = \begin{cases} 0, & \text{если } x < 0, \\ \lambda e^{-\lambda x}, & \text{если } x \geq 0, \end{cases}$$

где $\lambda > 0$ — параметр экспоненциального распределения. Для функции распределения в данном случае нетрудно получить следующее выражение:

$$F(x) = \begin{cases} 0, & \text{если } x < 0, \\ 1 - e^{-\lambda x}, & \text{если } x \geq 0. \end{cases}$$

Графики плотности распределения и функции распределения экспоненциальной случайной величины приведены на рис. 8 и рис. 9.

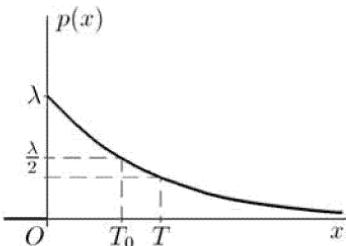


Рис. 8

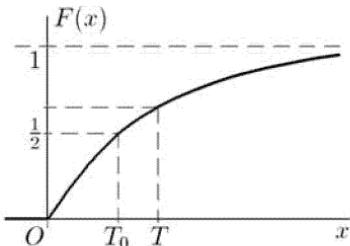


Рис. 9

Экспоненциально распределенная случайная величина может принимать только положительные значения. Экспоненциальному распределению подчинено время распада атомов различных элементов. При этом число $T = 1/\lambda$ носит название среднего времени распада. Кроме того, употребляют также число $T_0 = \ln 2/\lambda$, называемое периодом полураспада. Название «период полураспада» основано на следующем физическом соображении. Пусть у нас первоначально имелось n атомов вещества. Тогда через время T_0 каждый атом распадается с вероятностью $p = F(T_0) = 1 - e^{-\lambda \ln 2/\lambda} = 1 - 1/2 = 1/2$. Поэтому в силу независимости отдельных распадов число распавшихся за время T_0 атомов имеет биномиальное распределение с $p = q = 1/2$. Но как мы знаем из теоремы Бернулли (см. параграф 5, гл. 4), при больших n это число будет примерно равно $n/2$, т. е. период полураспада T_0 представляет собой не что иное, как время, в течение которого распадается половина имеющегося вещества.

Экспоненциально распределенная случайная величина ξ обладает весьма важным свойством, которое естественно назвать *отсут-*

ствием последействия. Трактуя ξ как время распада атома, рассмотрим событие $A = \{x_1 < \xi < x_1 + x_2\}$ и найдем условную вероятность этого события при условии выполнения события $B = \{\xi > x_1\}$. По определению условной вероятности $P(A|B) = P(AB)/P(B)$. Но событие AB , как нетрудно видеть, совпадает с событием A . Поэтому $P(A|B) = P(A)/P(B)$. Далее,

$$\begin{aligned} P(A) &= P\{x_1 < \xi < x_1 + x_2\} = (1 - e^{-\lambda(x_1+x_2)}) - (1 - e^{-\lambda x_1}) = \\ &= e^{-\lambda x_1}(1 - e^{-\lambda x_2}), \end{aligned}$$

$$P(B) = P\{\xi > x_1\} = 1 - P\{\xi < x_1\} = e^{-\lambda x_1}.$$

Значит,

$$P(A|B) = \frac{e^{-\lambda x_1}(1 - e^{-\lambda x_2})}{e^{-\lambda x_1}} = 1 - e^{-\lambda x_2}.$$

Мы получили, что вероятность распада атома за время x_2 при условии, что перед этим он уже прожил время x_1 , совпадает с безусловной вероятностью распада того же самого атома за время x_2 . Именно это свойство и представляет собой отсутствие последействия. Допуская некоторую вольность речи, отсутствие последействия можно трактовать как независимость остаточного времени жизни атома от того, сколько он уже прожил. Можно показать и обратное: если случайная величина ξ обладает свойством отсутствия последействия, то она обязана иметь экспоненциальное распределение. Таким образом, *отсутствие последействия является характеристическим свойством экспоненциально распределенных случайных величин*.

Практика показывает, что экспоненциальное распределение имеют и другие физические величины, например времена между падениями метеоритов в определенный район, времена между соседними поступлениями вызовов на телефонную станцию и т. д. Экспоненциальное распределение тесно связано с распределением Пуассона, а именно: если времена между последовательными наступлениями некоторого события представляют собой независимые (определение независимости случайных величин будет дано в гл. 6) экспоненциально распределенные (с одним и тем же параметром λ) случайные величины, то число наступлений этого события за время t распределено по закону Пуассона с параметром λt . Отметим также, что дискретным аналогом экспоненциального распределения является геометрическое распределение.

Нормальное распределение. Случайная величина распределена по нормальному, или гауссову, закону, если она имеет плотность распределения

$$\varphi_{m,\sigma}(x) = \Phi'_{m,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} \quad (-\infty < m < \infty, \sigma > 0).$$

Нормальное распределение зависит от двух параметров: m — математического ожидания, или среднего значения, нормального закона, и σ — среднего квадратичного отклонения. Графики плотности $\varphi_{m,\sigma}(x)$ и функции $\Phi_{m,\sigma}(x)$ нормального распределения в зависимости от m и σ приведены на рис. 10 и рис. 11.

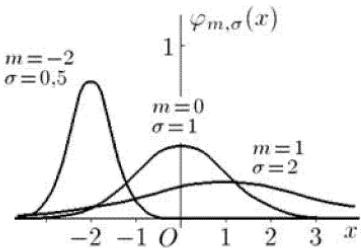


Рис. 10

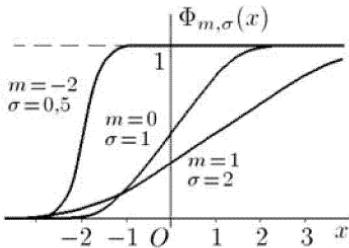


Рис. 11

Как видно из этих рисунков, параметр m определяет положение центра плотности нормального распределения, а σ — разброс относительно центра. Если $m = 0$, $\sigma = 1$, то такой нормальный закон называется стандартным и его функция распределения обозначается через $\Phi(x)$. С плотностью и функцией стандартного нормального распределения мы уже встречались в локальной и интегральной теоремах Муавра–Лапласа. Нормальный закон также очень часто встречается на практике. Как мы увидим в гл. 8, посвященной предельным теоремам, он обычно возникает в явлениях, подверженных действию большого числа малых случайных воздействий.

Распределение Вейбулла. Случайная величина распределена по закону Вейбулла, если она имеет плотность распределения

$$p(x) = \begin{cases} 0, & \text{если } x < 0, \\ \alpha\beta x^{\beta-1} e^{-\alpha x^\beta}, & \text{если } x \geq 0 \quad (\alpha > 0, \beta > 0). \end{cases}$$

Нетрудно проверить, что функция распределения в этом случае определяется следующим выражением:

$$F(x) = \begin{cases} 0, & \text{если } x < 0, \\ 1 - e^{-\alpha x^\beta}, & \text{если } x \geq 0. \end{cases}$$

Семейство распределений Вейбулла является двухпараметрическим и описывает положительные случайные величины. Графики плотности и функции распределения Вейбулла представлены на рис. 12 и рис. 13.

Считается, что распределению Вейбулла подчиняются времена безотказной работы многих технических устройств. Если $\beta = 1$, то распределение Вейбулла превращается в экспоненциальное распределение, а если $\beta = 2$ — в так называемое распределение Рэлея.

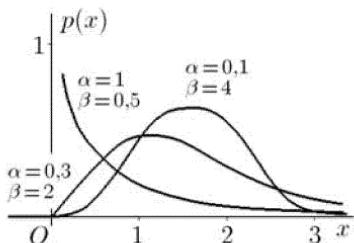


Рис. 12

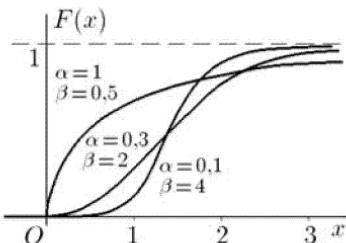


Рис. 13

Гамма-распределение. Другим распределением, также достаточно хорошо описывающим времена безотказной работы различных технических устройств, является гамма-распределение с плотностью распределения

$$p(x) = \begin{cases} 0, & \text{если } x < 0, \\ \frac{\lambda^\gamma x^{\gamma-1}}{\Gamma(\gamma)} e^{-\lambda x}, & \text{если } x \geq 0 \quad (\lambda > 0, \gamma > 0), \end{cases}$$

где $\Gamma(\gamma) = \int_0^\infty x^{\gamma-1} e^{-x} dx$ — гамма-функция Эйлера (следующие свойства гамма-функции являются весьма полезными при изучении гамма-распределения: $\Gamma(\gamma + 1) = \gamma\Gamma(\gamma)$ и $\Gamma(n) = (n - 1)!$ для целых n). Графики плотности и функции гамма-распределения изображены на рис. 14 и рис. 15.

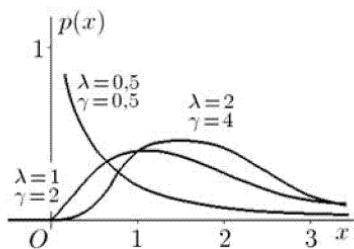


Рис. 14

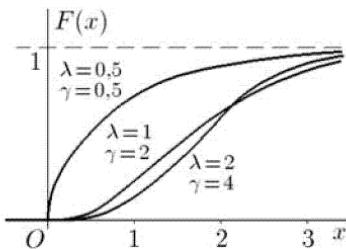


Рис. 15

Как видно из рисунков 12–15, распределение Вейбулла и гамма-распределение весьма близки между собой. Основным преимуществом распределения Вейбулла перед гамма-распределением является то, что его функция распределения выражается в явном виде. Поэтому раньше, когда ЭВМ еще не были достаточно распространены, распределение Вейбулла использовалось гораздо чаще, чем гамма-распределение. Хотя в общем случае гамма-распределение и не выражается в явном виде, оно обладает некоторыми весьма важными свойствами. Так, если $\gamma = k$, т. е. γ принимает целые значения, то мы получаем распределение

Эрланга, находящее важные применения в теории массового обслуживания. Если же $\gamma = k/2$ — полуцелое, а $\lambda = 1/2$, то гамма-распределение превращается в так называемое распределение χ^2 (хи-квадрат), роль которого в математической статистике невозможно переоценить; параметр k называется в этом случае числом степеней свободы распределения χ^2 . Наконец, при $\gamma = 1$ мы имеем дело все с тем же экспоненциальным распределением. Гамма-распределение имеет и другие интересные особенности, которых мы здесь не будем касаться.

5. Функции от случайной величины

Пусть на вероятностном пространстве $(\Omega, \mathfrak{B}, \mathbf{P})$ задана случайная величина $\xi = \xi(\omega)$. Возьмем обычную (измеримую) числовую функцию $g(x)$ числового аргумента x . Сопоставляя каждому элементарному исходу ω число $\eta(\omega)$ по формуле $\eta(\omega) = g(\xi(\omega))$, мы получим новую случайную величину η , которую естественно назвать *функцией $g(\xi)$ от случайной величины ξ* .

Функция $\eta = g(\xi)$ от дискретной случайной величины также является дискретной случайной величиной, поскольку она не может принимать больше значений, чем случайная величина ξ . Очевидно, что, если случайная величина ξ имеет ряд распределения, представленный в табл. 1, то ряд распределения случайной величины $\eta = g(\xi)$ определяется табл. 8.

Таблица 8

η	$g(X_1)$	$g(X_2)$	\dots	$g(X_n)$
\mathbf{P}	p_1	p_2	\dots	p_n

При этом, если в верхней строке табл. 8 появляются одинаковые значения $g(X_i)$, то соответствующие столбцы надо объединить в один, приспав им суммарную вероятность.

Пример 9. Снова рассмотрим игру «Спортлото 6 из 49». Поставив на некоторые фиксированные номера, мы в результате розыгрыша получим случайную величину ξ — число угаданных нами номеров (напомним, что пространство элементарных исходов состоит из всевозможных сочетаний по 6 номеров из 49), причем каждому элементарному исходу ω в соответствии с принципом классической вероятности сопоставлена вероятность $\mathbf{P}(\omega) = 1/\binom{49}{6}$ (ряд распределения случайной величины ξ представлен в табл. 7). Однако сама случайная величина ξ нас не интересует, для нас представляет интерес выигрыш, связанный с числом угаданных номеров ξ . Рассмотрим идеализированный вариант игры, при котором, не угадав ни одного или угадав один или два номера, мы проигрываем (с учетом платы за билет) 0,3 руб., угадав 3 номера, получаем выигрыш 2,7 руб., угадав 4 номера — 54,7 руб., 5 номеров — 699,7 руб. и 6 номеров — 9999,7 руб. Выигрыш η зависит

только лишь от числа угаданных номеров, т. е. представляет собой функцию от случайной величины ξ : $\eta = g(\xi)$, причем числовая функция $g(x)$ определена формулами: $g(0) = g(1) = g(2) = -0,3$, $g(3) = 2,7$, $g(4) = 54,7$, $g(5) = 699,7$ и $g(6) = 9999,7$. Ряд распределения случайной величины η получается из ряда распределения ξ (табл. 7) заменой в верхней строке чисел $i = 0, 1, \dots, 6$ на соответствующие значения $g(i)$ (табл. 9).

Таблица 9

η	-0,3	-0,3	-0,3	2,7	54,7	699,7	9999,7
P	0,4360	0,4130	0,1324	0,0176	0,00097	$2 \cdot 10^{-5}$	$7 \cdot 10^{-8}$

Осталось заметить, что в табл. 9 три первых столбца имеют одинаковые значения η , равные $-0,3$. Поэтому их надо объединить в один. Окончательный ряд распределения представлен в табл. 10.

Таблица 10

η	-0,3	2,7	54,7	699,7	9999,7
P	0,9814	0,0176	0,00097	$2 \cdot 10^{-5}$	$7 \cdot 10^{-8}$

Реально при игре в «Спортлото» выигрыш η зависит от числа играющих, поставивших на ту или иную комбинацию, и в этом случае его нельзя считать функцией от числа угаданных номеров ξ , а необходимо рассматривать более сложную модель, учитывающую вероятности (частоты) использования различных комбинаций номеров. В частности, мы не можем (без обращения к «поговоренным» силам) изменить вероятность угадывания определенного числа номеров, но мы можем увеличить выигрыш, ставя на «непопулярные» комбинации, которые хотя и появляются с той же частотой, что и остальные, но приносят больший выигрыш. Поиск «непопулярных» комбинаций относится к сфере психологии, а не теории вероятностей. \square

Функция $\eta = g(\xi)$ от непрерывной случайной величины ξ может быть как непрерывной, так и дискретной (дискретной она будет, например, если множество значений функции $g(x)$ не более чем счетно). Найдем функцию распределения $F_\eta(x)$ по заданной плотности распределения $p_\xi(x)$. По самому определению, $F_\eta(x)$ представляет собой вероятность события $\{\eta < x\}$, состоящего из тех элементарных исходов ω , для которых $g(\xi(\omega)) < x$. В свою очередь, вероятность события $\{g(\xi(\omega)) < x\}$ можно определить, используя аксиому сложения вероятностей, «просуммировав» вероятности всех возможных значений y случайной величины ξ , для которых $g(y) < x$. Поскольку, как мы знаем, вероятность случайной величине ξ принять значение в промежутке от y до $y + dy$ приближенно равна $p_\xi(y) dy$, то, заменяя сумму на интеграл, получаем

$$F_\eta(x) = \int_{g(y) < x} p_\xi(y) dy.$$

Последняя запись означает, что интегрирование производится по всем тем значениям y , для которых $g(y) < x$.

Пример 10. Случайная величина ξ распределена по стандартному нормальному закону. Найдем распределение случайной величины $\eta = \xi^2$. В данном случае $g(y) = y^2$, поэтому

$$F_\eta(x) = \int_{y^2 < x} p_\xi(y) dy = \frac{1}{\sqrt{2\pi}} \int_{y^2 < x} e^{-y^2/2} dy.$$

Поскольку при $x \leq 0$ нет ни одного y , для которого $y^2 < x$, или, что то же самое, не существует ни одного ω , для которого $g(\xi(\omega)) < x$, то $F_\eta(x) = 0$ при $x \leq 0$. Если же $x > 0$, то область $\{y^2 < x\}$ совпадает с областью $\{-\sqrt{x} < y < \sqrt{x}\}$ и, значит,

$$F_\eta(x) = \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{x}}^{\sqrt{x}} e^{-y^2/2} dy$$

или в силу четности $p_\xi(y)$

$$F_\eta(x) = \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{x}} e^{-y^2/2} dy.$$

Делая теперь замену $z = y^2$, окончательно получаем при $x > 0$

$$F_\eta(x) = \frac{1}{\sqrt{2\pi}} \int_0^x \frac{1}{\sqrt{z}} e^{-z/2} dz.$$

Нетрудно видеть, что случайная величина η имеет плотность распределения

$$p_\eta(x) = \begin{cases} 0, & \text{если } x \leq 0, \\ \frac{1}{\sqrt{2\pi x}} e^{-x/2}, & \text{если } x > 0, \end{cases}$$

являющуюся плотностью гамма-распределения с параметрами $\lambda = 1/2$, $\gamma = 1/2$, или, иными словами, η распределена по закону χ^2 с одной степенью свободы. Именно как распределение квадрата стандартной случайной величины и появляется распределение χ^2 в математической статистике. \square

Особенно просто находится функция распределения случайной величины $\eta = g(\xi)$, если $g(x)$ — монотонно возрастающая функция. В этом случае событие $\{g(\xi(\omega)) < x\}$ эквивалентно событию $\{\xi(\omega) < g^{-1}(x)\}$, где $g^{-1}(x)$ — функция, обратная $g(x)$, и, значит,

$$F_\eta(x) = \mathbf{P}\{g(\xi) < x\} = \mathbf{P}\{\xi < g^{-1}(x)\} = F_\xi(g^{-1}(x)).$$

Пример 11. Пусть плотность $p(x)$ распределения случайной величины ξ положительна для всех x . Рассмотрим случайную величину $\eta = F(\xi)$, где $F(x)$ — функция распределения ξ . В силу сделанного предположения $F(x)$ — монотонно возрастающая функция, принимающая значения от 0 до 1, и, след-

довательно,

$$F_\eta(x) = \begin{cases} 0, & \text{если } x < 0, \\ F(F^{-1}(x)) = x, & \text{если } 0 \leq x \leq 1, \\ 1, & \text{если } x > 1. \end{cases}$$

Таким образом, случайная величина η распределена равномерно на отрезке $[0, 1]$.

Полученный результат находит широкое применение при моделировании случайных величин с заданной функцией распределения $F(x)$. Дело в том, что практически все используемые в настоящее время датчики случайных (более правильно, «псевдослучайных») чисел инициируют величины η , распределенные равномерно на отрезке $[0, 1]$. Тогда, если функция $F(x)$ имеет достаточно просто вычисляемую обратную функцию $F^{-1}(x)$, то, полагая $\xi = F^{-1}(\eta)$, получаем случайную величину ξ с заданной функцией распределения $F(x)$. Отметим, что такой способ может быть применен даже для моделирования некоторых дискретных случайных величин. \square

Если же, кроме того, ξ — непрерывная случайная величина, а $g^{-1}(x)$ имеет производную $(g^{-1}(x))'$, то η является также непрерывной и ее плотность распределения определяется с помощью формулы дифференцирования сложной функции:

$$p_\eta(x) = F'_\eta(x) = F'_\xi(g^{-1}(x)) (g^{-1}(x))' = p_\xi(g^{-1}(x)) (g^{-1}(x))'.$$

Случай монотонно убывающей функции $g(x)$ предоставляется разобрать читателю.

Пример 12. Случайная величина ξ распределена по нормальному закону с параметрами m и σ . Найдем распределение случайной величины $\eta = e^\xi$. В данном примере $g(x) = e^x$, $g^{-1}(x) = \ln x$. В силу свойства экспоненциальной функции случайная величина η может принимать только положительные значения. Далее, при $x > 0$ функция $g^{-1}(x)$ дифференцируема, причем $(g^{-1}(x))' = (\ln x)' = 1/x$. Таким образом,

$$p_\eta(x) = \varphi_{m,\sigma}(\ln x) \frac{1}{x} = \begin{cases} 0, & \text{если } x \leq 0, \\ \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - m)^2}{2\sigma^2}}, & \text{если } x > 0. \end{cases}$$

Распределение с плотностью $p_\eta(x)$ носит название *логнормального* (поскольку логарифм случайной величины η распределен по нормальному закону). Двухпараметрическое логнормальное семейство наряду с распределением Вейбулла и гамма-распределением также довольно часто используется при описании времени безотказной работы различных технических устройств. \square

Пример 13. Положительная случайная величина ξ распределена по экспонциальному закону с параметром $\lambda = 1/2$ (распределение χ^2 с двумя степенями свободы). Покажите самостоятельно, что случайная величина $\eta = \sqrt{\xi}$ имеет распределение Вейбулла с параметрами $\alpha = 1/2$, $\beta = 2$ (распределение Рэлея). Воспользуйтесь монотонностью функции $g(x) = \sqrt{x}$ при $x > 0$. \square

Часто в дальнейшем нам будут встречаться линейные преобразования случайных величин: $\eta = a\xi + b$ ($a \neq 0$). Тогда

$$g(x) = ax + b, \quad g^{-1}(x) = \frac{x - b}{a}, \quad (g^{-1}(x))' = \frac{1}{a}$$

и, значит, при $a > 0$

$$F_\eta(x) = F_\xi(g^{-1}(x)) = F_\xi\left(\frac{x - b}{a}\right), \quad p_\eta(x) = \frac{1}{a} p_\xi\left(\frac{x - b}{a}\right).$$

Пример 14. Случайная величина ξ распределена по стандартному нормальному закону. Найдем распределение случайной величины $\eta = \sigma\xi + m$ ($\sigma > 0$). Тогда в формулах для линейного преобразования $a = \sigma$, $b = m$ и

$$p_\eta(x) = \frac{1}{\sigma} \varphi\left(\frac{x - m}{\sigma}\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} = \varphi_{m,\sigma}(x).$$

Итак, из случайной величины ξ , распределенной по стандартному нормальному закону, с помощью линейного преобразования $\eta = \sigma\xi + m$ получается нормально распределенная случайная величина с произвольными параметрами m и σ .

Читателю предоставляем показать обратное: если η — нормально распределенная случайная величина с параметрами m и σ , то случайная величина $\xi = (\eta - m)/\sigma$ распределена по стандартномуциальному закону. Из последнего свойства, в частности, следует

$$\Phi_{m,\sigma}(x) = P\{\eta < x\} = P\left\{\frac{\eta - m}{\sigma} < \frac{x - m}{\sigma}\right\} = P\left\{\xi < \frac{x - m}{\sigma}\right\} = \Phi\left(\frac{x - m}{\sigma}\right).$$

Эта формула позволяет вычислять значение функции нормального распределения $\Phi_{m,\sigma}(x)$ при любых значениях параметров m и σ через значение функции стандартного нормального распределения $\Phi(x)$ и оправдывает тот факт, что во всех справочниках приведены только таблицы значений функции стандартного нормального распределения. \square

Глава 6

МНОГОМЕРНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ И ИХ СВОЙСТВА

Разумеется, на каждом вероятностном пространстве (Ω, \mathcal{B}, P) , содержащем хотя бы один элементарный исход ω , можно определить не одну случайную величину. На практике необходимость учета непредсказуемых воздействий также весьма редко приводит к рассмотрению только одной случайной величины. Обычно выделяют несколько случайных компонент и на их основе строят модель исследуемого явления. Классическим примером такого рода, описанным многими великими художниками слова, снова являются азартные игры. Игрок, даже имея самое твердое намерение сыграть только один раз, впадал в азарт (откуда и пошло само название «азартные игры») и, как правило, не уходил из-за игорного стола, пока совсем не разорялся. Здесь суммарный выигрыш игрока состоит из отдельных случайных величин — выигравшей его в каждой партии.

Итак, в этой главе мы обобщим результаты предыдущей главы на случай нескольких случайных величин. Естественно, для того чтобы над случайными величинами можно было производить различные действия (сложение, вычитание, умножение и т. п.), необходимо их задание на одном вероятностном пространстве.

1. Многомерная случайная величина

Пусть на одном и том же вероятностном пространстве (Ω, \mathcal{B}, P) задано n случайных величин $\xi_1 = \xi_1(\omega), \dots, \xi_n = \xi_n(\omega)$. Совокупность случайных величин (ξ_1, \dots, ξ_n) назовем *многомерной* (*n-мерной*) *случайной величиной* или *случайным вектором*.

В дальнейшем для сокращения записи будем пользоваться также обозначением $\xi = (\xi_1, \dots, \xi_n)$.

Пример 1. На скачках, в которых участвуют 2 лошади, n человек заключают пари (между собой, группами и т. д.). Считая, что результат скачек случаен, причем с вероятностью $P(\omega_i)$ происходит элементарный исход ω_i ($i = 1, 2$) — к финишу первой пришла лошадь с номером i , получаем, что каждому элементарному исходу ω_i соответствует вектор $(\xi_1(\omega_i), \dots, \xi_n(\omega_i))$, компонентами которого являются выигрыши 1-го, ..., n -го участников pari. \square

Пример 2. Рассмотрим схему Бернулли, состоящую из n испытаний. Каждому элементарному исходу $\omega = \dots$ поставим в соответствие число μ_i , равное нулю, если в последовательности УНН...У на i -м месте стоит буква Н (в i -м испытании произошла неудача), и единице, если стоит буква У (произошел успех). Тогда случайная величина μ_i есть не что иное, как число успехов в i -м испытании. Совокупность случайных величин (μ_1, \dots, μ_n) представляет собой *n*-мерный случайный вектор. Нетрудно видеть, что в данном

случае существует взаимно однозначное соответствие между элементарными исходами ω и значениями случайного вектора (μ_1, \dots, μ_n) . Действительно, элементарному исходу $\omega = \dots$ соответствует значение $(1, 0, 0, \dots, 1)$ случайного вектора (μ_1, \dots, μ_n) и, наоборот, случайный вектор (μ_1, \dots, μ_n) принимает значение $(1, 0, 0, \dots, 1)$ только для элементарного исхода $\omega = \dots$. Ясно также, что суммарное число успехов μ в n испытаниях Бернулли (см. пример 2 в гл. 5) представляет собой сумму чисел успехов в каждом испытании: $\mu = \mu_1 + \dots + \mu_n$. \square

Пример 3. Двумерной случайной величиной является вектор (ξ_1, ξ_2) , где ξ_1 — сумма очков, выпавших при бросании двух игральных костей, а ξ_2 — сумма их квадратов. В этом случае различным элементарным исходам могут соответствовать одинаковые значения случайного вектора (ξ_1, ξ_2) . Например, двум элементарным исходам $\omega_1 = (2, 3)$ и $\omega_2 = (3, 2)$ соответствует одно и то же значение вектора $(\xi_1(\omega_1), \xi_2(\omega_1)) = (\xi_1(\omega_2), \xi_2(\omega_2)) = (5, 13)$. \square

Пример 4. Двумерными случайными величинами являются декартовы (ξ_1, ξ_2) или полярные (η_1, η_2) координаты точки падения частицы на плоский экран (см. пример 5 в гл. 5). Случайные величины (ξ_1, ξ_2) и (η_1, η_2) связаны соотношениями: $\xi_1 = \eta_1 \cos \eta_2$, $\xi_2 = \eta_1 \sin \eta_2$. \square

Пример 5. Широта ξ_1 и долгота ξ_2 места падения метеорита на Землю представляют собой двумерный случайный вектор (ξ_1, ξ_2) . В эту модель можно ввести также третью координату ξ_3 — время от начала наблюдений до момента падения первого метеорита на Землю ((ξ_1, ξ_2) — координаты падения этого метеорита), если рассматривать процесс, протекающий во времени. \square

2. Совместная функция распределения

Функцией распределения (n -мерного) случайного вектора $\xi = (\xi_1, \dots, \xi_n)$ называется функция $F(x_1, \dots, x_n) = F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n)$, значение которой в точке (x_1, \dots, x_n) равно вероятности совместного осуществления (пересечения) событий $\{\xi_1 < x_1\}, \dots, \{\xi_n < x_n\}$, т. е.

$$F(x_1, \dots, x_n) = F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = P\{\xi_1 < x_1, \dots, \xi_n < x_n\}.$$

Функцию $F(x_1, \dots, x_n)$ называют также *совместной функцией распределения случайных величин* ξ_1, \dots, ξ_n .

Для сокращения записи в дальнейшем для функции распределения $F(x_1, \dots, x_n) = F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n)$, будем использовать также обозначения $F(\mathbf{x})$ и $F_{\xi}(\mathbf{x})$.

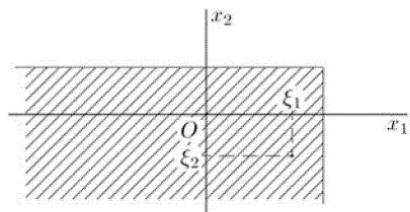


Рис. 1

Далее мы будем в основном рассматривать двумерный случай. В большинстве приводимых далее результатов переход от двумерного к общему n -мерному случаю не должен вызвать каких-либо трудностей, а там, где могут возникнуть затруднения, приводятся дополнительные пояснения.

Совместная функция распределения двумерной случайной величины (ξ_1, ξ_2) , по определению, представляет собой не что иное, как вероятность попадания точки с координатами (ξ_1, ξ_2) в область, заштрихованную на рис. 1.

Выведем свойства совместной функции распределения, аналогичные свойствам функции распределения одномерной случайной величины.

Как и прежде, поскольку $F(x_1, x_2)$ — вероятность, то

1. $0 \leq F(x_1, x_2) \leq 1$.

Следующее свойство является очевидным обобщением свойства 2 функции распределения одномерной случайной величины:

2. $F(x_1, x_2)$ — неубывающая функция по каждому из аргументов x_1 и x_2 .

Поскольку события $\{\xi_1 < -\infty\}$ и $\{\xi_2 < -\infty\}$ невозможны, а в результате пересечения невозможного события с любым событием, как мы знаем, также получается невозможное событие, то

3. $F(-\infty, x_2) = F(x_1, -\infty) = 0$.

События $\{\xi_1 < \infty\}$ и $\{\xi_2 < \infty\}$ так же, как и их пересечение, достоверны. Значит,

4. $F(\infty, \infty) = 1$.

Найдем вероятность попадания двумерной случайной величины (ξ_1, ξ_2) в прямоугольник $\{a_1 \leq x_1 < b_1, a_2 \leq x_2 < b_2\}$ (рис. 2). Для этого сначала определим вероятность попадания в полуполосу $\{x_1 < b_1, a_2 \leq x_2 < b_2\}$.

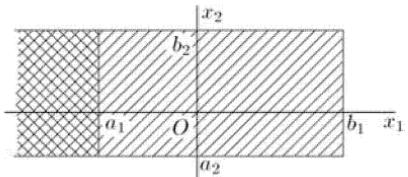


Рис. 2

Но эта вероятность представляет собой вероятность попадания в квадрант $\{x_1 < b_1, x_2 < b_2\}$ за вычетом вероятности попадания в квадрант $\{x_1 < b_1, x_2 < a_2\}$, т. е.

$$\mathbf{P}\{\xi_1 < b_1, a_2 \leq \xi_2 < b_2\} = F(b_1, b_2) - F(b_1, a_2).$$

Теперь осталось заметить, что вероятность попадания в прямоугольник $\{a_1 \leq x_1 < b_1, a_2 \leq x_2 < b_2\}$ есть вероятность попадания в полуполосу $\{x_1 < b_1, a_2 \leq x_2 < b_2\}$, из которой вычтена вероятность попадания в полуполосу $\{x_1 < a_1, a_2 \leq x_2 < b_2\}$, равная $F(a_1, b_2) - F(a_1, a_2)$. Окончательно получаем

5. $\mathbf{P}\{a_1 \leq \xi_1 < b_1, a_2 \leq \xi_2 < b_2\} = F(b_1, b_2) - F(b_1, a_2) - F(a_1, b_2) + F(a_1, a_2)$.

Подобно одномерному случаю доказывается и следующее свойство:

6. $F(x_1, x_2)$ — непрерывная слева по каждому из аргументов x_1 и x_2 функция.

Наконец, последнее свойство устанавливает естественную связь между распределением случайного вектора (ξ_1, ξ_2) и распределениями

случайных величин ξ_1 и ξ_2 . Событие $\{\xi_2 < \infty\}$ достоверно, поэтому $\{\xi_1 < x_1\} \cap \{\xi_2 < \infty\} = \{\xi_1 < x_1\}$. Аналогично $\{\xi_1 < \infty\} \cap \{\xi_2 < x_2\} = \{\xi_2 < x_2\}$. Значит,

$$7. F_{\xi_1, \xi_2}(x, \infty) = F_{\xi_1}(x), \quad F_{\xi_1, \xi_2}(\infty, x) = F_{\xi_2}(x).$$

Отметим, что, в отличие от одномерного случая, не любая функция $F(x_1, x_2)$, удовлетворяющая условиям 2–4 и 6, может являться совместной функцией распределения некоторой двумерной случайной величины. Для того чтобы это было возможно, необходимо потребовать также выполнение для любых $x_1, y_1, x_2, y_2, \quad x_1 \leq y_1, \quad x_2 \leq y_2$, следующего дополнительного условия:

$$F(y_1, y_2) - F(y_1, x_2) - F(x_1, y_2) + F(x_1, x_2) \leq 0.$$

3. Дискретные двумерные случайные величины

Двумерная случайная величина (ξ, η) называется дискретной, если каждая из случайных величин ξ и η является дискретной. Ясно, что если случайная величина ξ может принимать только значения X_1, \dots, X_n (для простоты мы ограничимся только конечным множеством значений), а случайная величина η — значения Y_1, \dots, Y_m , то двумерный случайный вектор (ξ, η) может принимать только пары значений (X_i, Y_j) ($i = 1, \dots, n, \quad j = 1, \dots, m$).

Так же, как и в одномерном случае, распределение двумерной дискретной случайной величины естественно описать с помощью перечисления всевозможных пар (X_i, Y_j) значений случайного вектора (ξ, η) и соответствующих вероятностей, с которыми эти пары значений принимают случайные величины ξ и η . Такое перечисление удобно реализовать таблицей (табл. 1) с двумя входами. В верхней строке этой таблицы перечислены все возможные значения Y_1, \dots, Y_m случайной величины η , а в левом столбце — значения X_1, \dots, X_n случайной величины ξ . На пересечении столбца « Y_j » со строкой « X_i » приведена вероятность $p_{ij} = P\{\xi = X_i, \eta = Y_j\}$ совместного осуществления событий $\{\xi = X_i\}$ и $\{\eta = Y_j\}$. К этой таблице обычно добавляют еще одну строку « P_η » и столбец « P_ξ ». На пересечении столбца « P_ξ » со строкой « X_i » записывают число $p_{\xi i} = p_{i1} + \dots + p_{im}$. Но $p_{\xi i}$ представляет собой не что иное, как вероятность случайной величине ξ принять значение X_i , т. е. $p_{\xi i} = P\{\xi = X_i\}$. Таким образом, первый и последний столбцы таблицы дают нам ряд распределения случайной величины ξ . Аналогично в строке « P_η » $p_{\eta j} = p_{1j} + \dots + p_{nj}$, а первая и последняя строки в совокупности задают ряд распределения случайной величины η . Для контроля правильности составления таблицы рекомендуется просуммировать элементы последней строки и последнего столбца. Если хотя бы одна из этих сумм не будет равна единице, то это означает, что при составлении таблицы была допущена ошибка.

Таблица 1

ξ	η				
	Y_1	Y_2	\dots	Y_m	P_ξ
X_1	p_{11}	p_{12}	\dots	p_{1m}	$p_{\xi 1}$
X_2	p_{21}	p_{22}	\dots	p_{2m}	$p_{\xi 2}$
\dots	\dots	\dots	\dots	\dots	\dots
X_n	p_{n1}	p_{n2}	\dots	p_{nm}	$p_{\xi n}$
P_η	$p_{\eta 1}$	$p_{\eta 2}$	\dots	$p_{\eta m}$	1

По табл. 1 нетрудно определить функцию распределения $F_{\xi\eta}(x, y)$. Ясно, что для этого необходимо просуммировать p_{ij} по всем тем значениям i и j , для которых $X_i < x$, $Y_j < y$. Иными словами,

$$F(x, y) = \sum_{\substack{i: X_i < x, \\ j: Y_j < y}} p_{ij}.$$

Пример 6. В схеме Бернулли производится 2 испытания с вероятностью успеха p и вероятностью неудачи $q = 1 - p$. Выпишем распределение двумерного случайного вектора (μ_1, μ_2) , где μ_i ($i = 1, 2$) — число успехов в i -м испытании (см. пример 2). Каждая из случайных величин μ_1 и μ_2 может принимать 2 значения: 0 или 1. Числа успехов в обоих испытаниях равны нулю тогда, когда произойдут две неудачи, а это в силу независимости испытаний происходит с вероятностью $q \cdot q = q^2$. Поэтому $P\{\mu_1 = 0, \mu_2 = 0\} = q^2$ и на пересечении столбца «0» со строкой «0» нужно написать q^2 (табл. 2). Далее, $\mu_1 = 1$ и $\mu_2 = 0$, если в первом испытании произошел успех, а во втором — неудача, и, значит, $P\{\mu_1 = 1, \mu_2 = 0\} = pq$. Аналогично заполняется второй столбец: $P\{\mu_1 = 0, \mu_2 = 1\} = qp$, $P\{\mu_1 = 1, \mu_2 = 1\} = p^2$. Наконец, на пересечении столбца « P_{μ_1} » и строки «0» должно стоять $P\{\mu_1 = 0\} = q^2 + qp = q(q + p) = q$, а на пересечении столбца « P_{μ_1} » и строки «1» — $P\{\mu_1 = 1\} = pq + p^2 = p(q + p) = p$. Таким же образом выписываем последнюю строку: $P\{\mu_2 = 0\} = q^2 + pq = q$, $P\{\mu_2 = 1\} = qp + p^2 = p$. Проверяем правильность составления таблицы: сумма элементов последнего столбца $p + q = 1$, последней строки $p + q = 1$. Значит, есть надежда, что таблица составлена правильно.

Построим теперь совместную функцию распределения $F(x_1, x_2) = P\{\mu_1 < x_1, \mu_2 < x_2\}$ случайных величин μ_1 и μ_2 . Поскольку при $x_1 \leq 0$ или $x_2 \leq 0$ нет ни одного элементарного исхода ω , для которого $\mu_1(\omega) < x_1$ или $\mu_2(\omega) < x_2$, то событие $\{\mu_1 < x_1, \mu_2 < x_2\}$ невозможно и, значит, $F(x_1, x_2) = 0$ при $x_1 \leq 0$ или $x_2 \leq 0$. Далее, если $0 < x_1 \leq 1$ и $0 < x_2 \leq 1$, то событие $\{\mu_1 < x_1, \mu_2 < x_2\}$ эквивалентно событию $\{\mu_1 = 0, \mu_2 = 0\}$, которое, как

Таблица 2

μ_1	μ_2		
	0	1	P_{μ_1}
0	q^2	qp	q
1	pq	p^2	p
P_{μ_2}	q	p	1

также равный 1/6 (на верхней грани выпало 2 очка, на нижней 5) и т. д. Столбец « P_ξ » и строку « P_η » находим, суммируя соответствующие строки и столбцы; как и должно было быть, мы получаем в них одинаковые значения 1/6, соответствующие принципу классической вероятности.

Предоставляем интересующемуся читателю самостоятельно построить совместную функцию распределения случайных величин ξ и η . \square

4. Непрерывные двумерные случайные величины

Непрерывной двумерной случайной величиной (ξ, η) называется такая двумерная случайная величина (ξ, η) , функция распределения которой $F(x_1, x_2) = P\{\xi < x_1, \eta < x_2\}$ может быть представлена в виде

$$F(x_1, x_2) = \iint_{\substack{-\infty < y_1 < x_1, \\ -\infty < y_2 < x_2}} p(y_1, y_2) dy_1 dy_2.$$

Здесь имеется в виду двукратный интеграл по области $\{y_1 < x_1, y_2 < x_2\}$; имеет место теорема Фубини, которая говорит, что этот двукратный интеграл можно представить в виде повторного, причем в любом порядке:

$$F(x_1, x_2) = \int_{-\infty}^{x_1} dy_1 \int_{-\infty}^{x_2} p(y_1, y_2) dy_2 = \int_{-\infty}^{x_2} dy_2 \int_{-\infty}^{x_1} p(y_1, y_2) dy_1.$$

Функция $p(x_1, x_2) = p_{\xi, \eta}(x_1, x_2)$ называется *совместной плотностью распределения случайных величин ξ и η .*

Так же, как и в одномерном случае, будем предполагать, что $p(x_1, x_2)$ непрерывная (или «почти» непрерывная) функция по обоим аргументам. Тогда совместная плотность распределения представляет собой смешанную производную совместной функции распределения:

$$p(x_1, x_2) = \frac{\partial^2}{\partial x_1 \partial x_2} F(x_1, x_2) = \frac{\partial^2}{\partial x_2 \partial x_1} F(x_1, x_2).$$

Нетрудно вывести следующие свойства совместной плотности распределения:

1. $p(x_1, x_2) \geq 0.$

2. $P\{a_1 < \xi < b_1, a_2 < \eta < b_2\} = \int_{a_1}^{b_1} dx_1 \int_{a_2}^{b_2} p(x_1, x_2) dx_2.$

3. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x_1, x_2) dx_1 dx_2 = 1.$

4. $P\{x_1 < \xi < x_1 + \Delta_1, x_2 < \eta < x_2 + \Delta_2\} \approx p(x_1, x_2) \Delta_1 \Delta_2.$

5. $P\{\xi = x_1, \eta = x_2\} = 0.$

Кроме этих свойств, повторяющих свойства плотности распределения одномерной случайной величины, укажем дополнительные свойства совместной плотности распределения.

Пусть D — некоторая область на плоскости (рис. 4).

Тогда, как следует из свойства 4, вероятность попадания двумерной случайной величины (ξ, η) в малый прямоугольник $\{a < \xi_1 < a + \Delta_1, b < \xi_2 < b + \Delta_2\}$ приближенно равна $p(a, b)\Delta_1\Delta_2$.

Поскольку попадания в непересекающиеся прямоугольники являются несовместными событиями, то, для того чтобы найти полную вероятность попадания двумерной случайной величины (ξ, η) в область D , нужно просуммировать вероятности попадания во все «малые» прямоугольники, входящие в область D . Переходя к пределу, получаем для $\mathbf{P}\{(\xi, \eta) \in D\}$ — вероятности попадания (ξ, η) в область D — формулу

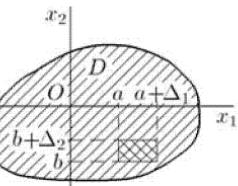


Рис. 4

получаем для $\mathbf{P}\{(\xi, \eta) \in D\}$ — вероятности попадания (ξ, η) в область D — формулу

$$6. \mathbf{P}\{(\xi, \eta) \in D\} = \iint_D p(x_1, x_2) dx_1 dx_2.$$

Далее, из свойства 7 совместной функции распределения и определения совместной плотности распределения имеем

$$F_\xi(x) = F_{\xi, \eta}(x, \infty) = \int_{-\infty}^x dy_1 \int_{-\infty}^{\infty} p_{\xi, \eta}(y_1, y_2) dy_2,$$

$$F_\eta(x) = F_{\xi, \eta}(\infty, x) = \int_{-\infty}^x dy_2 \int_{-\infty}^{\infty} p_{\xi, \eta}(y_1, y_2) dy_1,$$

откуда, дифференцируя по x , получаем выражения для одномерных плотностей распределения случайных величин ξ и η

$$7. p_\xi(x) = \int_{-\infty}^{\infty} p_{\xi, \eta}(x, y) dy, \quad p_\eta(y) = \int_{-\infty}^{\infty} p_{\xi, \eta}(x, y) dx.$$

Пример 8. Предположим, что в соответствии с принципом геометрической вероятности мы бросаем точку случайным образом в круг радиусом R с центром в начале координат (см. пример 10 в гл. 2). Пусть случайная величина ξ — абсцисса точки падения, а η — ордината. Естественно, поскольку точка не может попасть за пределы круга, то $p(x_1, x_2) = 0$ при $x_1^2 + x_2^2 > R^2$. Для каждой области внутри круга (в частности, прямоугольника $\{a < \xi < a + \Delta_1, b < \eta < b + \Delta_2\}$) вероятность попадания пропорциональна площади этой области (равна $A\Delta_1\Delta_2$, где A — коэффициент пропорциональности). Поэтому из свойства 4 совместной плотности распределения имеем: $p(x_1, x_2) = A$ при $x_1^2 + x_2^2 \leq R^2$, т. е. плотность распределения постоянна внутри

круга. Для определения постоянной A воспользуемся свойством 3 совместной плотности распределения. Поскольку $p(x_1, x_2) = 0$ при $x_1^2 + x_2^2 > R^2$, то

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x_1, x_2) dx_1 dx_2 = \iint_{x_1^2 + x_2^2 \leq R^2} A dx_1 dx_2 = \pi A R^2 = 1$$

и, значит, $A = 1/(\pi R^2)$. Итак,

$$p(x_1, x_2) = \begin{cases} 0, & \text{если } x_1^2 + x_2^2 > R^2, \\ \frac{1}{\pi R^2}, & \text{если } x_1^2 + x_2^2 \leq R^2. \end{cases}$$

Нетрудно найти плотность распределения случайной величины ξ :

$$p_{\xi}(x) = \int_{-\infty}^{\infty} p(x, y) dy = \int_{-\sqrt{R^2 - x^2}}^{\sqrt{R^2 - x^2}} \frac{1}{\pi R^2} dy = \begin{cases} 0, & \text{если } |x| > R, \\ \frac{2\sqrt{R^2 - x^2}}{\pi R^2}, & \text{если } |x| \leq R. \end{cases}$$

Аналогичное выражение получается и для $p_{\eta}(y)$.

Для нахождения совместной функции распределения $F(x_1, x_2)$ заметим, что в силу определения совместной плотности распределения

$$F(x_1, x_2) = \frac{1}{\pi R^2} \iint_D dy_1 dy_2,$$

где область D представляет собой пересечение квадранта $\{y_1 < x_1, y_2 < x_2\}$ и круга $x_1^2 + x_2^2 \leq R^2$, т. е. $F(x_1, x_2)$ с точностью до множителя $1/(\pi R^2)$ совпадает с площадью области D , имеющей двойную штриховку на рис. 5.

Мы думаем, читатель достаточно хорошо знаком с основами интегрального исчисления и может определить площадь области D для различных x_1 и x_2 самостоятельно.

Предложенный выше вариант определения совместной плотности распределения $p(x_1, x_2)$, которую естественно назвать равномерной плотностью распределения внутри круга $x_1^2 + x_2^2 \leq R^2$, реализует первое решение примера 10 в гл. 2. Во втором решении вероятность попадания одна и та же для областей, имеющих одинаковые прращения полярных координат: радиуса $\Delta\rho$ и угла $\Delta\varphi$. Переходя к декартовым координатам, получаем, что совместная плотность распределения ξ и η в этом случае должна иметь вид

$$p_{\xi, \eta}(x_1, x_2) = \begin{cases} 0, & \text{если } x_1^2 + x_2^2 > R^2, \\ \frac{B}{\sqrt{x_1^2 + x_2^2}}, & \text{если } x_1^2 + x_2^2 \leq R^2. \end{cases}$$

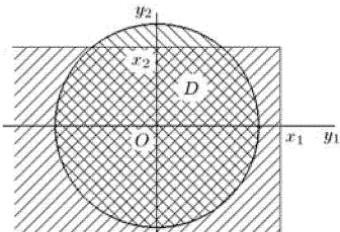


Рис. 5

Предоставляем читателю определить нормировочную константу B , а также плотности распределения случайных величин ξ и η и совместную функцию распределения $F_{\xi, \eta}(x_1, x_2)$. \square

В заключение этого параграфа рассмотрим наиболее часто встречающееся на практике распределение непрерывной n -мерной случайной величины.

Многомерное нормальное (гауссово) распределение. Пусть $A = (a_{ij})$ — положительно определенная симметрическая квадратная матрица порядка n (т. е. $a_{ij} = a_{ji}$, а все собственные значения матрицы A положительны). Обозначим через A^{-1} матрицу, обратную матрице A , а через $A^{-1}(\mathbf{x}, \mathbf{x})$ — квадратичную форму, порожденную матрицей A^{-1} , т. е. $A^{-1}(\mathbf{x}, \mathbf{x}) = \sum_{i,j=1}^n a_{ij}^{(-1)} x_i x_j$, где $a_{ij}^{(-1)}$ — элементы матрицы A^{-1} , а x_i — i -я координата вектора $\mathbf{x} = (x_1, \dots, x_n)$. Пусть также задан n -мерный вектор $\mathbf{m} = (m_1, \dots, m_n)$.

Скажем, что n -мерный непрерывный случайный вектор $\xi = (\xi_1, \dots, \xi_n)$ распределен по (невырожденному) нормальному закону, если его совместная плотность распределения $p(x_1, \dots, x_n)$ (определяется точно так же, как и совместная плотность распределения одномерной случайной величины):

$$p(x_1, \dots, x_n) = \frac{\partial^n}{\partial x_1 \dots \partial x_n} F(x_1, \dots, x_n)$$

задается формулой

$$p(x_1, \dots, x_n) = p(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} |A|^{1/2}} e^{-\frac{1}{2} A^{-1}(\mathbf{x} - \mathbf{m}, \mathbf{x} - \mathbf{m})}.$$

(Здесь $|A|$ — определитель матрицы A .) Матрица A носит название ковариационной матрицы (матрицы ковариаций), а вектор \mathbf{m} — вектора средних. Если матрица A (а значит, и матрица A^{-1}) совпадает с единичной матрицей I , а вектор $\mathbf{m} = (0, \dots, 0)$, то

$$p(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2} (x_1^2 + \dots + x_n^2)}.$$

Такая плотность распределения по аналогии с одномерным случаем называется плотностью стандартного n -мерного нормального распределения (соответственно $F(x_1, \dots, x_n)$ — функция стандартного n -мерного нормального распределения).

Пусть (ξ_1, \dots, ξ_n) — n -мерный случайный вектор, распределенный по (произвольному) нормальному закону. Тогда $n - 1$ -мерный случайный вектор $(\xi_1, \dots, \xi_{n-1})$ распределен также поциальному закону с вектором средних (m_1, \dots, m_{n-1}) и ковариационной матрицей A' , получаемой из матрицы A вычеркиванием последних строки и столбца (это можно показать непосредственно с помощью многомерного аналога свойства 7 совместной плотности распределения двумерной

случайной величины:

$$p_{\xi_1, \dots, \xi_{n-1}}(x_1, \dots, x_{n-1}) = \int_{-\infty}^{\infty} p_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) dx_n;$$

однако существенно проще для этой цели воспользоваться аппаратом характеристических функций, который частично будет рассмотрен нами в гл. 8). В частности, каждая из случайных величин ξ_i распределена по нормальному закону со средним m_i и средним квадратичным отклонением $\sigma_i = \sqrt{a_{ii}}$.

Рассмотрим теперь уравнение $p(x_1, \dots, x_n) = c$, которое для плотности распределения n -мерного нормального распределения эквивалентно уравнению

$$A^{-1}(\mathbf{x} - \mathbf{m}, \mathbf{x} - \mathbf{m}) = c_1 = -2 \ln(c(2\pi)^{n/2} |A|^{1/2}).$$

Для всех $c_1 > 0$ это уравнение в силу положительной определенности матрицы A^{-1} представляет собой уравнение n -мерного эллипсоида, называемого эллипсоидом рассеивания; его главные оси называются осями рассеивания (рис. 6).

Будем трактовать n -мерный случайный вектор (ξ_1, \dots, ξ_n) как координаты случайной точки ξ в n -мерном пространстве. Тогда, если мы выберем в n -мерном пространстве новую ортонормированную систему координат x'_1, \dots, x'_n , связанную с главными осями, то

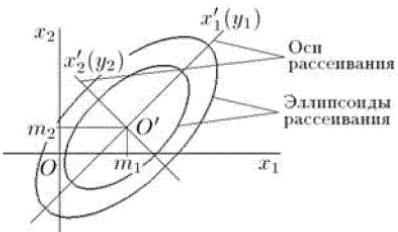


Рис. 6

в этой системе новые координаты (ξ'_1, \dots, ξ'_n) случайной точки ξ снова будут описываться n -мерным нормальным законом, имеющим нулевой вектор средних \mathbf{m}' и диагональную матрицу ковариаций A' , причем диагональные элементы σ_i^2 матрицы A' будут пропорциональны квадратам коэффициентов растяжения k_i эллипса распределения по соответствующим осям рассеивания. Еще раз вводя новые координаты $y_i = \sigma_i x'_i$, получаем, что в этой последней системе y_1, \dots, y_n координаты случайной точки ξ будут распределены по стандартному нормальному закону. Таким образом, делая обратные преобразования, можно трактовать (невырожденный) нормально распределенный вектор ξ с произвольными вектором средних \mathbf{m} и матрицей ковариаций A как координаты случайной точки ξ , имеющей стандартное нормальное распределение, в некоторой (вообще говоря, не ортонормированной и даже не ортогональной) системе координат.

Разумеется, встречаются многомерные случайные величины, которые нельзя отнести ни к дискретному, ни к непрерывному типу. Так, у двумерной случайной величины (ξ, η) одна координата (допустим η) может быть дискретной, а другая ξ — непрерывной. Такую двумер-

ную случайную величину удобно характеризовать набором функций $p_i(x) = \partial \mathbf{P}\{\xi < x, \eta = Y_i\} / \partial x$.

Пример 9. Придя в кассу «Аэрофлота», клиент застает очередь из η человек. Ясно, что η — дискретная случайная величина, принимающая значения 0, 1, 2, ... Наряду с длиной очереди η естественно рассмотреть и непрерывную случайную величину ξ — общее время, проведенное клиентом в кассе. Если интервалы времени между приходами клиентов независимы (см. параграф 6) и имеют одно и то же экспоненциальное распределение с параметром λ , а длительность обслуживания каждого клиента кассиром также распределена экспоненциально с параметром μ , то, как показано в теории массового обслуживания, совместное распределение случайных величин ξ и η в установившемся режиме работы задается функциями

$$p_i(x) = \frac{\partial}{\partial x} \mathbf{P}\{\xi < x, \eta = i\} = (1 - \rho) \rho^i \frac{\mu^{i+1} x^i}{i!} e^{-\mu x}$$

$$x > 0, \quad i = 0, 1, 2, \dots; \quad \rho = \frac{\lambda}{\mu} < 1, \quad \lambda, \mu > 0.$$

Ряд распределения случайной величины η находим, проинтегрировав $p_i(x)$ по x :

$$p_{\eta i} = \mathbf{P}\{\eta = i\} = \int_{-\infty}^{\infty} p_i(x) dx = (1 - \rho) \rho^i \int_0^{\infty} \frac{\mu^{i+1} x^i}{i!} e^{-\mu x} dx = (1 - \rho) \rho^i$$

$$(i = 0, 1, 2, \dots).$$

Итак, длина очереди η распределена по геометрическому закону. Аналогично, плотность распределения случайной величины ξ получаем, суммируя $p_i(x)$ по i :

$$p_\xi(x) = F'_\xi(x) = \sum_{i=0}^{\infty} (1 - \rho) \rho^i \frac{\mu^{i+1} x^i}{i!} e^{-\mu x} = (1 - \rho) \mu e^{-\mu(1-\rho)x} \quad (x > 0),$$

т. е. время ξ пребывания клиента в кассе имеет экспоненциальное распределение с параметром $\mu(1 - \rho)$. \square

Еще более интересные явления возникают, если рассматривать случайные величины, имеющие явную функциональную зависимость.

Пример 10. Пусть ξ — непрерывная случайная величина с функцией распределения $F(x)$ и плотностью распределения $p(x)$. Рассмотрим двумерный случайный вектор (ξ, ξ) с одинаковыми координатами ξ и совместной функцией распределения $F_{\xi, \xi}(x_1, x_2) = \mathbf{P}\{\xi < x_1, \xi < x_2\}$. Ясно, что вектор (ξ, ξ) не является дискретным. Покажем, что (ξ, ξ) не может быть и непрерывной двумерной случайной величиной. Для этого заметим, что при $x_1 < x_2$ событие $\{\xi < x_1, \xi < x_2\}$ совпадает с событием $\{\xi < x_1\}$ и, значит, $F_{\xi, \xi}(x_1, x_2) = F(x_1)$. Но тогда

$$p_{\xi, \xi}(x_1, x_2) = \frac{\partial^2}{\partial x_1 \partial x_2} F_{\xi, \xi}(x_1, x_2) = 0$$

при $x_1 < x_2$. Аналогично, $F_{\xi, \xi}(x_1, x_2) = F(x_2)$ и

$$p_{\xi, \xi}(x_1, x_2) = \frac{\partial^2}{\partial x_1 \partial x_2} F_{\xi, \xi}(x_1, x_2) = 0$$

при $x_2 < x_1$.

Таким образом, если бы у вектора (ξ, ξ) существовала плотность распределения, то она равнялась бы нулю всюду, кроме биссектрисы $x_1 = x_2$, и, значит,

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_{\xi, \xi}(x_1, x_2) dx_1 dx_2 = 0,$$

что противоречит свойству 3 совместной плотности распределения. Ясно, что причина этого явления кроется в том, что значения двумерного случайного вектора (ξ, ξ) полностью сосредоточены на биссектрисе $x_1 = x_2$. \square

5. Условные распределения¹⁾

Рассмотрим двумерную случайную величину (ξ, η) с совместной функцией распределения $F_{\xi, \eta}(x, y)$. Пусть известно, что случайная величина η приняла значение y . Естественно задать вопрос: а что можно сказать при этом условии о распределении случайной величины ξ ? Ответ на этот вопрос дает условная функция распределения случайной величины ξ при условии $\eta = y$. Как видно из самой постановки задачи, понятие условного распределения весьма схоже с понятием условной вероятности, разобранным в параграфе 1 гл. 3. Именно исходя из понятия условной вероятности, мы введем понятие условного распределения.

Начнем с наиболее простого случая. Пусть случайная величина η является дискретной. Назовем *условной функцией распределения* $F_{\xi}(x | \eta = Y_i)$ случайной величины ξ при условии $\eta = Y_i$ условную вероятность события $\{\xi < x\}$ при условии события $\{\eta = Y_i\}$, т. е. в соответствии с определением условной вероятности,

$$F_{\xi}(x | \eta = Y_i) = \mathbf{P}\{\xi < x | \eta = Y_i\} = \frac{\mathbf{P}\{\xi < x, \eta = Y_i\}}{\mathbf{P}\{\eta = Y_i\}}.$$

Условная функция распределения обладает всеми теми свойствами, которые присущи обычной (безусловной) функции распределения.

Пример 11. Найдем условное распределение времени пребывания ξ клиента в кассе «Аэрофлота» (пример 9) при условии, что в момент прихода он застает очередь η из i человек. В этом случае, как мы знаем,

$$\mathbf{P}\{\xi < x, \eta = i\} = (1 - \rho) \rho^i \int_0^x \frac{\mu^{i+1} y^i}{i!} e^{-\mu y} dy, \quad \mathbf{P}\{\eta = i\} = (1 - \rho) \rho^i.$$

Таким образом,

$$F_{\xi}(x | \eta = i) = \int_0^x \frac{\mu^{i+1} y^i}{i!} e^{-\mu y} dy$$

¹⁾ Результаты настоящего параграфа в дальнейшем нигде, кроме параграфа 5 гл. 7, использоваться не будут, и при первом прочтении его можно пропустить.

и, значит, условная функция распределения $F_\xi(x \mid \eta = i)$ имеет *условную плотность распределения*

$$p_\xi(x \mid \eta = i) = F'_\xi(x \mid \eta = i) = \frac{\mu^{i+1} x^i}{i!} e^{-\mu x},$$

представляющую собой плотность гамма-распределения с параметрами μ , $i + 1$ или, что то же самое, плотность распределения Эрланга порядка $i + 1$. \square

Если ξ — также дискретная случайная величина, причем $P\{\xi = X_i, \eta = Y_j\} = p_{ij}$, то удобно рассматривать *условную вероятность* π_{ij} того, что случайная величина ξ примет значение X_i при условии $\eta = Y_j$, определяемую как условную вероятность события $\{\xi = X_i\}$ при условии события $\{\eta = Y_j\}$, т. е.

$$\pi_{ij} = P\{\xi = X_i \mid \eta = Y_j\} = \frac{P\{\xi = X_i, \eta = Y_j\}}{P\{\eta = Y_j\}} = \frac{p_{ij}}{p_{\eta j}}.$$

Обычно условное распределение дискретной случайной величины ξ при условии дискретной случайной величины η описывает табл. 4. Ясно, что элементы π_{ij} табл. 4 получаются из элементов табл. 1 по формуле $\pi_{ij} = p_{ij}/p_{\eta j}$ и, наоборот, $p_{ij} = \pi_{ij}p_{\eta j}$.

Таблица 4

ξ	η				
	Y_1	Y_2	\dots	Y_m	P_ξ
X_1	π_{11}	π_{12}	\dots	π_{1m}	$p_{\xi 1}$
X_2	π_{21}	π_{22}	\dots	π_{2m}	$p_{\xi 2}$
\dots	\dots	\dots	\dots	\dots	\dots
X_n	π_{n1}	π_{n2}	\dots	π_{nm}	$p_{\xi n}$
P_η	$p_{\eta 1}$	$p_{\eta 2}$	\dots	$p_{\eta m}$	$1 \diagdown 1$

Таблица 5

μ_1	μ_2		
	0	1	P_{μ_1}
0	q	q	q
1	p	p	p
P_{μ_2}	q	p	$1 \diagdown 1$

Пример 12. Условное распределение случайной величины μ_1 (числа успехов в первом испытании) при условии $\mu_2 = j$ ($j = 0, 1$) (числа успехов во втором испытании) из примера 6 задается табл. 5. \square

Пример 13. Условное распределение случайной величины ξ (числа очков, выпавших на верхней грани игральной кости) при условии $\eta = j$ ($j = 1, \dots, 6$) (числа очков, выпавших на нижней грани игральной кости) из примера 7 представлено в табл. 6. \square

В общем случае условную функцию распределения случайной величины ξ при условии $\eta = y$ также естественно было бы определить формулой

$$F_\xi(x \mid \eta = y) = \frac{P\{\xi < x, \eta = y\}}{P\{\eta = y\}}.$$

Однако это не всегда возможно (например, событие $\{\eta = y\}$ для непрерывной случайной величины имеет нулевую вероятность: $P\{\eta = y\} = 0$). Поэтому попытаемся воспользоваться предельным переходом,

Таблица 6

ξ	η						
	1	2	3	4	5	6	P_ξ
1	0	0	0	0	0	1	$1/6$
2	0	0	0	0	1	0	$1/6$
3	0	0	0	1	0	0	$1/6$
4	0	0	1	0	0	0	$1/6$
5	0	1	0	0	0	0	$1/6$
6	1	0	0	0	0	0	$1/6$
P_η	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	1

рассматривая вместо события $\{\eta = y\}$ событие $\{y \leq \eta < y + \Delta\}$ и устремляя Δ к нулю. Итак, определим сначала условную вероятность

$$P\{\xi < x \mid y \leq \eta < y + \Delta\} =$$

$$= \frac{P\{\xi < x, y \leq \eta < y + \Delta\}}{P\{y \leq \eta < y + \Delta\}} = \frac{F_{\xi,\eta}(x, y + \Delta) - F_{\xi,\eta}(x, y)}{F_\eta(y + \Delta) - F_\eta(y)}$$

и назовем *условной функцией распределения* $F_\xi(x \mid \eta = y)$ предел

$$F_\xi(x \mid \eta = y) = \lim_{\Delta \rightarrow 0} P\{\xi < x \mid y \leq \eta < y + \Delta\}.$$

Оказывается, такой предел всегда существует (правда, в определенном смысле: это производная Радона–Никодима одной меры по другой). Производная Радона–Никодима определяется не однозначно, а с точностью до множества точек на прямой, вероятность попадания в которое случайной величины η равна нулю. Впрочем, с неоднозначностью определения условной функции распределения мы фактически уже встречались в случае дискретной случайной величины η ; действительно, условную функцию распределения $F_\xi(x \mid \eta = y)$ мы определяли только для $y = Y_i$, для остальных значений y мы могли бы задать $F_\xi(x \mid \eta = y)$ совершенно произвольным образом, поскольку все равно случайная величина η такие значения не принимает).

Если же случайная величина η непрерывна, то условную функцию распределения можно определить следующим выражением:

$$F_\xi(x \mid \eta = y) = \frac{\frac{\partial}{\partial y} F_{\xi,\eta}(x, y)}{p_\eta(y)},$$

формально получаемым, если в выражении для $P\{\xi < x \mid y \leq \eta < y + \Delta\}$ поделить числитель и знаменатель на Δ и устремить Δ к нулю.

Пример 14. Найдем условное распределение длины очереди η в кассу «Аэрофлота» в момент прихода клиента при условии, что общее время ξ , прове-

денное им в кассе, составило x (пример 9). Ответ дадим в терминах *условного ряда распределения* $\mathbf{P}\{\eta = i \mid \xi = x\}$. Тогда в обозначениях примера 9

$$\mathbf{P}\{\eta = i \mid \xi = x\} = \frac{\frac{\partial}{\partial x} \mathbf{P}\{\eta = i, \xi < x\}}{p_\xi(x)} = \frac{p_i(x)}{p_\xi(x)} = \frac{(\mu x)^i}{i!} e^{-\mu x} = \frac{(\lambda x)^i}{i!} e^{-\lambda x}.$$

Таким образом, условное распределение очереди η при условии $\xi = x$ представляет собой распределение Пуассона с параметром λx . \square

В наиболее важных для приложений случаях вектор (ξ, η) представляет собой двумерную непрерывную случайную величину с совместной плотностью распределения $p_{\xi, \eta}(x, y)$. Тогда

$$F_{\xi, \eta}(x, y) = \int_{-\infty}^y dv \int_{-\infty}^x p_{\xi, \eta}(u, v) du, \quad \frac{\partial}{\partial y} F_{\xi, \eta}(x, y) = \int_{-\infty}^x p_{\xi, \eta}(u, y) du$$

и, значит,

$$F_\xi(x \mid \eta = y) = \frac{\int_x^\infty p_{\xi, \eta}(u, y) du}{p_\eta(y)}.$$

Нетрудно видеть, что условная функция распределения $F_\xi(x \mid \eta = y)$ имеет производную по x , т. е. существует *условная плотность распределения случайной величины ξ при условии $\eta = y$* :

$$p_\xi(x \mid \eta = y) = \frac{\partial}{\partial x} F_\xi(x \mid \eta = y) = \frac{p_{\xi, \eta}(x, y)}{p_\eta(y)}.$$

Пример 15. Найдем условную плотность распределения случайной величины ξ — абсциссы точки падения (из примера 8) при условии, что ордината η приняла значение y . Тогда, как мы знаем,

$$p_{\xi, \eta}(x, y) = \begin{cases} 0, & \text{если } |x| > \sqrt{R^2 - y^2}, \\ \frac{1}{\pi R^2}, & \text{если } |x| \leqslant \sqrt{R^2 - y^2}, \end{cases}$$

$$p_\eta(y) = \begin{cases} 0, & \text{если } |y| > R, \\ \frac{2\sqrt{R^2 - y^2}}{\pi R^2}, & \text{если } |y| \leqslant R, \end{cases}$$

и при $|y| \leqslant R$

$$p_\xi(x \mid \eta = y) = \begin{cases} 0, & \text{если } |x| > \sqrt{R^2 - y^2}, \\ \frac{1}{2\sqrt{R^2 - y^2}}, & \text{если } |x| \leqslant \sqrt{R^2 - y^2}. \end{cases}$$

Таким образом, случайная величина ξ при условии $\eta = y$ равномерно распределена на отрезке $[-\sqrt{R^2 - y^2}, \sqrt{R^2 - y^2}]$. Если $|y| > R$, то условная плотность распределения $p_\xi(x \mid \eta = y)$ не определена; но это нас не должно волновать, поскольку случайная величина η не может принимать значение, по модулю большее R .

Рекомендуем читателю самостоятельно решить эту задачу для второго варианта определения плотности распределения из примера 8. \square

Пример 16. Пусть (ξ, η) — двумерный нормальный вектор с матрицей ковариаций

$$A = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \quad (\sigma_1, \sigma_2 > 0, -1 < \rho < 1)$$

и вектором средних (m_1, m_2) . Найдем условную плотность распределения случайной величины ξ при условии $\eta = y$. Для этого сначала определим $|A|$ и A^{-1} :

$$|A| = \sigma_1^2\sigma_2^2(1 - \rho^2), \quad A^{-1} = \begin{pmatrix} \frac{1}{\sigma_1^2(1 - \rho^2)} & \frac{-\rho}{\sigma_1\sigma_2(1 - \rho^2)} \\ \frac{-\rho}{\sigma_1\sigma_2(1 - \rho^2)} & \frac{1}{\sigma_2^2(1 - \rho^2)} \end{pmatrix}.$$

Теперь мы можем выписать совместную плотность распределения случайных величин ξ и η :

$$p_{\xi, \eta}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1 - \rho^2}} e^{-\frac{1}{2(1 - \rho^2)} \left[\frac{(x - m_1)^2}{\sigma_1^2} - \frac{2\rho(x - m_1)(y - m_2)}{\sigma_1\sigma_2} + \frac{(y - m_2)^2}{\sigma_2^2} \right]}.$$

Далее, как нам известно,

$$p_\eta(y) = \frac{1}{\sigma_2\sqrt{2\pi}} e^{-\frac{(y - m_2)^2}{2\sigma_2^2}}$$

и, значит,

$$p_\xi(x | \eta = y) = \frac{p_{\xi, \eta}(x, y)}{p_\eta(y)} = \frac{1}{\sigma_1\sqrt{2\pi(1 - \rho^2)}} e^{-\frac{1}{2\sigma_1^2(1 - \rho^2)} \left[x - \left(m_1 + \frac{\rho\sigma_1(y - m_2)}{\sigma_2} \right) \right]^2}.$$

Таким образом, условное распределение ξ при условии $\eta = y$ снова является нормальным со средним значением $m_1 + \rho\sigma_1(y - m_2)/\sigma_2$ и средним квадратичным отклонением $\sigma_1\sqrt{1 - \rho^2}$. \square

6. Независимые случайные величины

Назовем случайные величины ξ и η независимыми, если совместная функция распределения $F_{\xi, \eta}(x, y)$ представляется в виде произведения одномерных функций распределения $F_\xi(x)$ и $F_\eta(y)$:

$$F_{\xi, \eta}(x, y) = F_\xi(x)F_\eta(y).$$

Понятие независимости случайных величин представляет собой перенос понятия независимости событий на случайные величины и опять-таки отражает отсутствие связи между случайными величинами ξ и η (хотя, повторяем еще раз, и ξ , и η заданы на одном и том же вероятностном пространстве $(\Omega, \mathcal{B}, \mathbf{P})$ и в совокупности определяют двумерный случайный вектор (ξ, η)). Иными словами, независимость случайных величин ξ и η можно охарактеризовать следующим образом: зная значение, которое приняла случайная величина η , мы никакой новой информации о распределении случайной величины ξ не получим. Отметим, что обычно независимость случайных величин вводится

несколько иначе; в этом случае приведенное здесь определение выступает в роли необходимого и достаточного условия независимости.

Совершенно аналогично определяется независимость произвольного числа случайных величин. *Случайные величины ξ_1, \dots, ξ_n называются независимыми (в совокупности), если*

$$F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = F_{\xi_1}(x_1) \dots F_{\xi_n}(x_n).$$

Разумеется, так же, как и для событий, из попарной независимости не следует независимость случайных величин в совокупности.

Пример 17. Свяжем с бросанием правильного раскрашенного тетраэдра (пример 11 в гл. 3) три случайные величины: ξ_1 , ξ_2 и ξ_3 , каждая из которых может принимать значения 0 или 1, причем $\xi_1 = 1$, если тетраэдр упал на грань, в раскраске которой присутствует синий цвет, и $\xi_1 = 0$ в противном случае. Аналогично, ξ_2 характеризует наличие красного цвета, а ξ_3 — зеленого. Нетрудно видеть, что случайные величины ξ_1 , ξ_2 и ξ_3 будут попарно независимы, но зависимы в совокупности. \square

Пример 18. Рассмотрим широту ξ_1 , долготу ξ_2 места падения метеорита на Землю и время ξ_3 от начала наблюдений до момента его падения (см. пример 5). Предполагая, что каждый падающий метеорит имеет случайное направление, мы должны считать случайные величины ξ_1 , ξ_2 и ξ_3 независимыми. Однако если Земля проходит через метеоритный поток определенного направления, уже нельзя пользоваться моделью с независимыми ξ_1 , ξ_2 и ξ_3 , поскольку долгота места падения связана с тем, какой стороной к потоку метеоритов обращена Земля, т. е. в конечном счете со временем. Тем не менее и в этом случае широта ξ_1 не будет зависеть от времени ξ_3 , т. е. имеет место независимость случайных величин ξ_1 и ξ_3 , хотя случайные величины ξ_1 , ξ_2 и ξ_3 зависимы в совокупности. Предоставляем читателю самостоятельно поразмышлять над вопросом: какое направление должен иметь поток метеоритов, чтобы были независимыми между собой случайная величина ξ_1 и случайный вектор (ξ_2, ξ_3) (широта не зависит от долготы и времени падения). \square

Пусть ξ и η — независимые случайные величины. Рассмотрим событие $\{x_1 \leq \xi < x_2\}$, связанное со случайной величиной ξ , и событие $\{y_1 \leq \eta < y_2\}$, связанное со случайной величиной η . В силу независимости ξ и η

$$\begin{aligned} \mathbf{P}\{x_1 \leq \xi < x_2, y_1 \leq \eta < y_2\} &= \\ &= F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1) = \\ &= F_\xi(x_2)F_\eta(y_2) - F_\xi(x_1)F_\eta(y_2) - F_\xi(x_2)F_\eta(y_1) + F_\xi(x_1)F_\eta(y_1) = \\ &= [F_\xi(x_2) - F_\xi(x_1)][F_\eta(y_2) - F_\eta(y_1)] = \\ &= \mathbf{P}\{x_1 \leq \xi < x_2\}\mathbf{P}\{y_1 \leq \eta < y_2\}. \end{aligned}$$

Таким образом, для независимых случайных величин независимы между собой не только события, связанные с попаданием случайных величин ξ и η на интервалы $(-\infty, x)$ и $(-\infty, y)$, но и на любые интервалы $[x_1, x_2]$ и $[y_1, y_2]$ и, более того, в любые (измеримые) одномерные множества A и B .

Для проверки независимости компонент многомерных дискретных и непрерывных случайных векторов обычно бывают удобными другие эквивалентные определения независимости (доказательство эквивалентности приводимых ниже определений предоставляем читателю). Так, *дискретные случайные величины ξ и η независимы, если для всех возможных значений X_i и Y_j*

$$p_{ij} = \mathbf{P}\{\xi = X_i, \eta = Y_j\} = \mathbf{P}\{\xi = X_i\}\mathbf{P}\{\eta = Y_j\} = p_{\xi i}p_{\eta j}.$$

Пример 19. В схеме Бернулли с двумя испытаниями (см. пример 6)

$$\begin{aligned}\mathbf{P}\{\mu_1 = 0, \mu_2 = 0\} &= q^2 = \mathbf{P}\{\mu_1 = 0\}\mathbf{P}\{\mu_2 = 0\}, \\ \mathbf{P}\{\mu_1 = 0, \mu_2 = 1\} &= qp = \mathbf{P}\{\mu_1 = 0\}\mathbf{P}\{\mu_2 = 1\}, \\ \mathbf{P}\{\mu_1 = 1, \mu_2 = 0\} &= pq = \mathbf{P}\{\mu_1 = 1\}\mathbf{P}\{\mu_2 = 0\}, \\ \mathbf{P}\{\mu_1 = 1, \mu_2 = 1\} &= p^2 = \mathbf{P}\{\mu_1 = 1\}\mathbf{P}\{\mu_2 = 1\}.\end{aligned}$$

Таким образом, числа успехов μ_1 и μ_2 в первом и втором испытаниях представляют собой независимые случайные величины. Впрочем, иного и нельзя было ожидать из самого определения схемы Бернулли. Читатель может самостоятельно убедиться в том, что независимы в совокупности случайные величины μ_1, \dots, μ_n — числа успехов в первом, ..., n -м испытаниях Бернулли. \square

Непрерывные случайные величины ξ и η независимы, если для всех x и y

$$p_{\xi, \eta}(x, y) = p_{\xi}(x)p_{\eta}(y).$$

Отметим здесь же, что если независимые случайные величины ξ и η непрерывны, то и двумерная случайная величина (ξ, η) обязана быть непрерывной (ср. с примером 10).

Пример 20. Координаты двумерного нормального случайного вектора (см. пример 16) независимы тогда и только тогда, когда $\rho = 0$, т. е. матрицы A^{-1} и A — диагональные. В этом можно без труда убедиться, сравнивая выражения для $p_{\xi, \eta}(x, y)$ с произведением $p_{\xi}(x)p_{\eta}(y)$. Аналогично, случайные величины ξ_1, \dots, ξ_n , имеющие совместное n -мерное нормальное распределение, независимы (в совокупности) тогда и только тогда, когда матрица A^{-1} , а значит, и ковариационная матрица A диагональны. Интересно отметить, что попарная независимость всех компонент ξ_i и ξ_j нормального вектора (ξ_1, \dots, ξ_n) влечет за собой независимость случайных величин ξ_1, \dots, ξ_n в совокупности (ср. с примером 17). \square

Для читателя, ознакомившегося с понятием условного распределения (параграф 5), приведем еще один критерий независимости случайных величин ξ и η : *случайные величины ξ и η независимы тогда и только тогда, когда условное распределение (функция распределения $F_{\xi}(x | \eta = y)$, плотность распределения $p_{\xi}(x | \eta = y)$) случайной величины ξ при условии $\eta = y$ совпадает с безусловным распределением (функцией распределения $F_{\xi}(x)$, плотностью распределения $p_{\xi}(x)$) случайной величины ξ при всех значениях x и y .* В частности, дискретные величины ξ и η независимы тогда и только тогда, когда все условные вероятности $\pi_{ij} = \mathbf{P}\{\xi = X_i | \eta = Y_j\}$ совпадают с безусловными вероятностями $p_{\xi i} = \mathbf{P}\{\xi = X_i\}$, т. е. все столбцы табл. 4 совпадают с последним.

Пример 21. В схеме Бернулли с двумя испытаниями (см. пример 6) числа успехов μ_1 и μ_2 в первом и втором испытаниях независимы, поскольку в табл. 5 все три столбца совпадают. Этот факт нами уже был установлен другим способом в примере 19. \square

Пример 22. Число очков ξ , выпавших на верхней грани игральной кости, и число очков η на нижней грани (см. пример 7) — зависимые случайные величины, поскольку вообще ни один из первых шести столбцов табл. 6 не совпадает с последним. \square

Пример 23. В примере 11 показано, что условное распределение времени пребывания ξ клиента в кассе «Аэрофлота» (см. пример 9) при условии, что в момент прихода он застает очередь η из i человек, является распределением Эрланга порядка $i+1$, в то время как случайная величина ξ распределена по экспоненциальному закону. Таким образом, ξ и η — зависимые случайные величины. \square

Пример 24. Условная плотность распределения случайной величины ξ (абсциссы точки падения при равномерном бросании в круг, см. пример 8) при условии $\eta = y$ (ординаты точки падения), как следует из примера 15, равномерна, в то время как безусловная плотность распределения случайной величины ξ таковой не является. И в этом примере случайные величины ξ и η зависимы между собой. \square

7. Функции от многомерных случайных величин

Функция от многомерной случайной величины определяется точно так же, как и функция от одномерной. Рассмотрим это понятие на примере двумерной случайной величины. Пусть на вероятностном пространстве $(\Omega, \mathfrak{B}, P)$ задана двумерная случайная величина (ξ_1, ξ_2) . Предположим, что у нас имеется (измеримая) числовая функция $g(x_1, x_2)$ числовых аргументов x_1 и x_2 . Случайную величину $\eta = g(\xi_1, \xi_2) = g(\xi_1(\omega), \xi_2(\omega))$ назовем *функцией от двумерной случайной величины* (ξ_1, ξ_2) .

Функция $g(\xi_1, \xi_2)$ от двумерной дискретной случайной величины (ξ_1, ξ_2) снова является дискретной случайной величиной, принимающей значения $g(X_i, Y_j)$ с вероятностью $p_{ij} = P\{\xi_1 = X_i, \xi_2 = Y_j\}$, где X_i — значения случайной величины ξ_1 , а Y_j — случайной величины ξ_2 . Разумеется, для того чтобы построить ряд распределения случайной величины $\eta = g(\xi_1, \xi_2)$, необходимо, во-первых, исключить все те значения $g(X_i, Y_j)$, которые принимаются с вероятностью, равной нулю, а, во-вторых, объединить в один столбец все одинаковые значения $g(X_i, Y_j)$, приписав этому столбцу суммарную вероятность.

Пример 25. Рассмотрим случайную величину μ — суммарное число успехов в двух испытаниях Бернулли. Тогда $\mu = \mu_1 + \mu_2$ (см. пример 2) и $g(x_1, x_2) = x_1 + x_2$. Поскольку случайные величины μ_1 и μ_2 принимают только два значения 0 или 1, то случайная величина μ может принимать 4 значения: $g(0,0) = 0 + 0 = 0$, $g(1,0) = 1 + 0 = 1$, $g(0,1) = 0 + 1 = 1$ и $g(1,1) = 1 + 1 = 2$ с вероятностями (см. пример 6) q^2 , pq , qp и p^2 соответственно (табл. 7). Осталось заметить, что двум средним столбцам соответствуют

Таблица 7

μ	$g(0,0) = 0$	$g(1,0) = 1$	$g(0,1) = 1$	$g(1,1) = 2$
P_μ	q^2	pq	qp	p^2

ет одно и то же значение 1 случайной величины μ и их необходимо объединить. Окончательно получаем ряд распределения случайной величины μ , представленный в табл. 8.

Естественно, мы получили, что суммарное число успехов μ в двух испытаниях имеет распределение Бернулли. \square

Таблица 8

μ	0	1	2
P_μ	q^2	$2pq$	p^2

Таблица 9

ζ	7
P_ζ	1

Пример 26. Пусть ξ — число очков, выпавших на верхней грани игральной кости, а η — на нижней (пример 7). Рассмотрим случайную величину $\zeta = \xi + \eta$ — суммарное число очков, выпавших на верхней и нижней гранях ($g(x, y) = x + y$). Тогда ζ может принять любое целочисленное значение от 2 до 12. Так, например, значению $\zeta = 4$ соответствует выпадение 1–3, 2–2 и 3–1 очков на верхней и нижней гранях. Однако нетрудно видеть из табл. 3, что все значения, кроме значения, равного семи, случайная величина ζ принимает с вероятностью, равной нулю, и их необходимо изъять из ряда распределения, а значение 7 случайная величина ζ принимает в 6 случаях (1–6, 2–5, 3–4, 4–3, 5–2 и 6–1), причем каждый из этих случаев реализуется с вероятностью $1/6$. Поэтому случайная величина ζ может принимать всего одно значение — 7 с вероятностью, равной единице, т. е. она имеет ряд распределения, представленный в табл. 9.

Собственно говоря, это было очевидно с самого начала, поскольку игральная кость размечена таким образом, что сумма очков на противоположных гранях равна семи. \square

В том случае, когда (ξ_1, ξ_2) — двумерная непрерывная случайная величина с плотностью распределения $p_{\xi_1, \xi_2}(x_1, x_2)$, функция распределения случайной величины $\eta = g(\xi_1, \xi_2)$ определяется формулой

$$F_\eta(x) = \iint_{g(x_1, x_2) < x} p_{\xi_1, \xi_2}(x_1, x_2) dx_1 dx_2.$$

Область интегрирования в последней формуле состоит из всех x_1 и x_2 , для которых $g(x_1, x_2) < x$.

Пример 27. Пусть (ξ_1, ξ_2) — двумерный случайный вектор, распределенный по стандартному нормальному закону. Найдем распределение случайной величины $\eta = \sqrt{\xi_1^2 + \xi_2^2}$. В этом случае $g(x_1, x_2) = \sqrt{x_1^2 + x_2^2}$. Очевидно, что

$F_\eta(x) = 0$ при $x \leq 0$, а при $x > 0$

$$F_\eta(x) = \iint_{\sqrt{x_1^2+x_2^2} < x} \frac{1}{2\pi} e^{-\frac{1}{2}(x_1^2+x_2^2)} dx_1 dx_2.$$

Последний интеграл удобно вычислить, переходя к полярным координатам ρ и φ . Тогда $x_1^2 + x_2^2 = \rho^2$, $dx_1 dx_2 = \rho d\varphi d\rho$, а область интегрирования превращается в круг $\rho < x$:

$$F_\eta(x) = \int_0^x d\rho \int_0^{2\pi} \frac{1}{2\pi} e^{-\frac{1}{2}\rho^2} \rho d\varphi = \int_0^x \rho e^{-\frac{1}{2}\rho^2} d\rho = 1 - e^{-x^2/2} \quad (x > 0).$$

Это уже известное нам распределение Рэлея.

Полученный результат допускает многочисленные физические трактовки. Приведем одну из них. Если движущаяся в плоскости частица имеет случайные составляющие скорости, распределенные по двумерному стандартному нормальному закону, то абсолютная величина скорости распределена по закону Рэлея. Трехмерным аналогом распределения Рэлея (абсолютная величина скорости частицы, движущейся в трехмерном пространстве, причем составляющие скорости распределены по трехмерному стандартному нормальному закону) является распределение Максвелла, представляющее собой распределение случайной величины $\eta = \sqrt{\xi}$, где ξ — случайная величина, распределенная по закону χ^2 с тремя степенями свободы. \square

Особо важным для теории вероятностей представляется случай, когда ξ_1 и ξ_2 — независимые случайные величины, а $\eta = \xi_1 + \xi_2$ — их сумма. Тогда $g(x_1, x_2) = x_1 + x_2$ и в силу независимости ξ_1 и ξ_2

$$\begin{aligned} F_\eta(x) &= \iint_{x_1+x_2 < x} p_{\xi_1, \xi_2}(x_1, x_2) dx_1 dx_2 = \iint_{x_1+x_2 < x} p_{\xi_1}(x_1) p_{\xi_2}(x_2) dx_1 dx_2 = \\ &= \int_{-\infty}^{\infty} p_{\xi_1}(x_1) dx_1 \int_{-\infty}^{x-x_1} p_{\xi_2}(x_2) dx_2 = \int_{-\infty}^{\infty} F_{\xi_2}(x-y) p_{\xi_1}(y) dy. \end{aligned}$$

Дифференцируя последнюю формулу под знаком интеграла, получаем выражение для плотности $p_\eta(x)$ распределения суммы ξ_1 и ξ_2

$$p_\eta(x) = \int_{-\infty}^{\infty} p_{\xi_2}(x-y) p_{\xi_1}(y) dy.$$

Последнее выражение носит название *формулы свертки* и должно быть хорошо известно читателю, знакомому с элементами теории преобразований Фурье.

Пример 28. Случайные величины ξ_1 и ξ_2 независимы и имеют гамма-распределения с параметрами λ, γ_1 и λ, γ_2 соответственно. Найдем плотность распределения суммы $\eta = \xi_1 + \xi_2$. Ясно, что поскольку ξ_1 и ξ_2 — положительные случайные величины, то случайная величина η также положительна

и $p_\eta(x) = 0$ при $x < 0$. При $x > 0$, учитывая, что $p_{\xi_1}(y) = 0$ при $y < 0$ и $p_{\xi_2}(x-y) = 0$ при $y > x$, имеем по формуле свертки

$$\begin{aligned} p_\eta(x) &= \int_{-\infty}^{\infty} p_{\xi_2}(x-y)p_{\xi_1}(y) dy = \int_0^x \frac{\lambda^{\gamma_2}(x-y)^{\gamma_2-1}}{\Gamma(\gamma_2)} e^{-\lambda(x-y)} \frac{\lambda^{\gamma_1}y^{\gamma_1-1}}{\Gamma(\gamma_1)} e^{-\lambda y} dy = \\ &= \lambda^{\gamma_1+\gamma_2} e^{-\lambda x} \int_0^x \frac{(x-y)^{\gamma_2-1}y^{\gamma_1-1}}{\Gamma(\gamma_2)\Gamma(\gamma_1)} dy, \end{aligned}$$

$\Gamma(x)$ введена в гл. 5, параграф 4. Делая замену $y = xz$, получаем

$$p_\eta(x) = \lambda^{\gamma_1+\gamma_2} x^{\gamma_1+\gamma_2-1} e^{-\lambda x} \int_0^1 \frac{(1-z)^{\gamma_2-1} z^{\gamma_1-1}}{\Gamma(\gamma_2)\Gamma(\gamma_1)} dz.$$

Интеграл, стоящий в последнем выражении, представляет собой так называемый β -интеграл, хорошо известный в теории специальных функций. Однако мы для его вычисления воспользуемся просто условием нормировки

$$\int_0^\infty p_\eta(x) dx = 1,$$

откуда

$$\int_0^1 \frac{(1-z)^{\gamma_2-1} z^{\gamma_1-1}}{\Gamma(\gamma_2)\Gamma(\gamma_1)} dz = \frac{1}{\Gamma(\gamma_1 + \gamma_2)},$$

и получаем окончательно, что случайная величина η также имеет гамма-распределение с параметрами $\lambda, \gamma_1 + \gamma_2$. Отсюда, в частности, следует: если независимые случайные величины имеют распределения Эрланга порядков k_1 и k_2 соответственно (с одинаковым параметром λ), то их сумма также распределена по закону Эрланга порядка $k_1 + k_2$ (также с параметром λ); если независимые случайные величины имеют распределения χ^2 с k_1 и k_2 степенями свободы, то их сумма также имеет распределение χ^2 с $k_1 + k_2$ степенями свободы. \square

Пусть теперь с двумерным случайным вектором (ξ_1, ξ_2) связана не одна, а две (можно рассматривать и большее количество) случайные величины $\eta_1 = g_1(\xi_1, \xi_2)$ и $\eta_2 = g_2(\xi_1, \xi_2)$. Тогда мы можем определить совместную функцию распределения случайных величин η_1 и η_2 . Так, для непрерывного двумерного случайного вектора (ξ_1, ξ_2)

$$F_{\eta_1, \eta_2}(x_1, x_2) = \iint_{\substack{1(y_1, y_2) < x_1, \\ g_2(y_1, y_2) < x_2}} p_{\xi_1, \xi_2}(y_1, y_2) dy_1 dy_2.$$

Если $g_1(x_1, x_2)$ и $g_2(x_1, x_2)$ задают взаимно однозначное преобразование плоскости саму в себя (или в некоторую область G), причем обратные преобразования $h_1(y_1, y_2)$ и $h_2(y_1, y_2)$ имеют непрерывные частные производные по y_1 и y_2 , то плотности распределения случай-

ных векторов (ξ_1, ξ_2) и (η_1, η_2) связаны между собой соотношениями

$$p_{\eta_1, \eta_2}(y_1, y_2) = p_{\xi_1, \xi_2}(h_1(y_1, y_2), h_2(y_1, y_2)) |J|,$$

где

$$J = \begin{vmatrix} \frac{\partial}{\partial y_1} h_1(y_1, y_2) & \frac{\partial}{\partial y_2} h_1(y_1, y_2) \\ \frac{\partial}{\partial y_1} h_2(y_1, y_2) & \frac{\partial}{\partial y_2} h_2(y_1, y_2) \end{vmatrix}$$

— якобиан преобразования $[h_1(y_1, y_2), h_2(y_1, y_2)]$. Это свойство вытекает из того факта, что $P\{y_1 < \eta_1 < y_1 + \Delta_1, y_2 < \eta_2 < y_2 + \Delta_2\}$, с одной стороны, приближено равна $p_{\eta_1, \eta_2}(y_1, y_2)\Delta_1\Delta_2$, а с другой — приближено равна $p_{\xi_1, \xi_2}(h_1(y_1, y_2), h_2(y_1, y_2))S$, где S — площадь прообраза прямоугольника со сторонами $(y_1, y_1 + \Delta_1)$ и $(y_2, y_2 + \Delta_2)$, которая, как известно из курса математического анализа, в свою очередь приближенно равна $|J|\Delta_1\Delta_2$.

В частности, пусть $g_i(x_1, x_2) = b_{i1}x_1 + b_{i2}x_2 + c_i$, $i = 1, 2$, т. е. преобразование линейное, причем матрица $B = (b_{ij})$ невырожденная. Тогда

$$p_{\eta}(y) = p_{\xi}(B^{-1}(y - \mathbf{c})) \frac{1}{\|B\|},$$

где B^{-1} — обратная к B матрица, а $\|B\|$ — модуль определителя матрицы B .

Пример 29. Пусть $\xi = (\xi_1, \dots, \xi_n)$ — (невырожденный) n -мерный случайный вектор, распределенный по нормальному закону с матрицей ковариаций A_{ξ} и вектором средних $\mathbf{m}_{\xi} = (0, \dots, 0)$ (этого всегда можно добиться, вводя случайный вектор $\xi^* = \xi - \mathbf{m}$). Введем новый случайный вектор $\eta = (\eta_1, \dots, \eta_n) = B\xi$, где B — некоторая невырожденная квадратная матрица порядка n (в этом случае случайные величины $\eta_i = b_{i1}\xi_1 + \dots + b_{in}\xi_n$ задаются линейными функциями $g_i(x_1, \dots, x_n) = b_{i1}x_1 + \dots + b_{in}x_n$). Тогда случайный вектор η имеет плотность распределения

$$\begin{aligned} p_{\eta}(y) &= p_{\xi}(B^{-1}y) \frac{1}{\|B\|} = \frac{1}{(2\pi)^{n/2}|A_{\xi}|^{1/2}\|B\|} e^{-\frac{1}{2}A_{\xi}^{-1}(B^{-1}y, B^{-1}y)} = \\ &= \frac{1}{(2\pi)^{n/2}|BA_{\xi}B|^{1/2}} e^{-\frac{1}{2}(B)^{-1}A_{\xi}^{-1}B^{-1}(y, y)}, \end{aligned}$$

т. е. также распределен поциальному закону с нулевым вектором средних \mathbf{m}_{η} и матрицей ковариаций $A_{\eta} = BA_{\xi}B$, где B — матрица линейного преобразования (b_{ij}) , а B — транспонированная к B матрица. Из курса линейной алгебры известно, что всегда можно подобрать невырожденное преобразование B таким образом, чтобы матрица A_{η} была единичной, т. е. вектор η имел бы стандартное нормальное распределение. Таким образом, мы нашли второй способ (ср. с результатом параграфа 4) задания n -мерного нормального случайного вектора ξ , имеющего произвольный вектор средних \mathbf{m} и матрицу ковариаций A , с помощью n -мерного стандартного нормального вектора η : $\xi = B^{-1}\eta + \mathbf{m}$. Отметим также, что если матрица B^{-1} вырождена, то мы получаем так называемый вырожденный нормальный закон. \square

В дальнейшем нам понадобится следующее почти очевидное свойство функций от случайных величин. Пусть случайные величины ξ_1 и ξ_2 независимы, а функции $g_1(x_1, x_2)$ и $g_2(x_1, x_2)$ таковы, что $g_1(x_1, x_2) = g_1(x_1)$ и $g_2(x_1, x_2) = g_2(x_2)$. Тогда случайные величины $\eta_1 = g_1(\xi_1, \xi_2) = g_1(\xi_1)$ и $\eta_2 = g_2(\xi_1, \xi_2) = g_2(\xi_2)$ также независимы. Действительно (предполагая, например, что ξ_1 и ξ_2 — непрерывные случайные величины), имеем

$$\begin{aligned} F_{\eta_1, \eta_2}(x_1, x_2) &= \iint_{\substack{(y_1, y_2) < x_1, \\ g_1(y_1, y_2) < x_1, \\ g_2(y_1, y_2) < x_2}} p_{\xi_1, \xi_2}(y_1, y_2) dy_1 dy_2 = \\ &= \iint_{\substack{(y_1, y_2) < x_1, \\ g_1(y_1, y_2) < x_1, \\ g_2(y_1, y_2) < x_2}} p_{\xi_1}(y_1) p_{\xi_2}(y_2) dy_1 dy_2 = \\ &= \left(\int_{g_1(y_1) < x_1} p_{\xi_1}(y_1) dy_1 \right) \left(\int_{g_2(y_2) < x_2} p_{\xi_2}(y_2) dy_2 \right) = F_{\eta_1}(x_1) F_{\eta_2}(x_2). \end{aligned}$$

Как и все остальные свойства, рассмотренные в настоящем параграфе, это свойство без всяких комментариев переносится на случай произвольной размерности n случайного вектора ξ .

Пример 30. Пусть ξ_1, \dots, ξ_n — независимые случайные величины, распределенные по стандартному нормальному закону. Тогда случайные величины ξ_1^2, \dots, ξ_n^2 также независимы и распределены по закону χ^2 с одной степенью свободы (см. пример 12 в гл. 5) и, как следует из примера 28, случайная величина $\eta = \xi_1^2 + \dots + \xi_n^2$ имеет распределение χ^2 с n степенями свободы. Извлекая квадратный корень из η , получаем при $n = 2$ и $n = 3$ распределения Рэлея и Максвелла (см. пример 27). \square

Г л а в а 7

ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ СЛУЧАЙНЫХ ВЕЛИЧИН

Мы теперь знаем, что каждая случайная величина характеризуется своей функцией распределения. С точки зрения наблюдателя, две случайные величины, имеющие одинаковые функции распределения, неразличимы, несмотря на то, что они могут быть заданы на различных вероятностных пространствах и описывать разные явления. Так, при игре в «орлянку» все равно, какая (симметричная) монета бросается, и, если кто из играющих и пытается сменить монету, то это скорее дань предрассудку, нежели возможность поправить свои дела. Однако в том случае, когда случайные величины имеют различные функции распределения и их необходимо сравнить, возникает определенная трудность. Иногда эта трудность легко преодолима. Например, если в схеме Бернулли нас интересует число успехов, то из двух схем Бернулли естественно выбрать ту, в которой больше вероятность успеха. В общем же случае непонятно, как сравнивать две функции распределения, а поэтому хотелось бы характеризовать каждую случайную величину некоторым (неслучайным) числом (возможно, несколькими числами), которое и позволило бы произвести упорядочение случайных величин в определенном смысле.

Такие числовые характеристики будут рассмотрены нами в этой главе. Отметим, что основную роль на практике играют математическое ожидание, характеризующее «центральное» значение случайной величины, и дисперсия, характеризующая «разброс» вокруг математического ожидания; их роль более подробно будет выяснена в следующей главе. Среди остальных характеристик можно выделить те, которые применяются в специальных вероятностных дисциплинах (например, квантили широко используются в математической статистике), и те, которые носят ярко выраженный теоретический характер (моменты высших порядков).

1. Математическое ожидание случайной величины

Математическим ожиданием (средним значением) $M\xi$ дискретной случайной величины ξ называется сумма произведений значений X_i случайной величины на вероятности $p_i = P\{\xi = X_i\}$, с которыми эти значения принимаются:

$$M\xi = \sum_i X_i p_i.$$

При этом, если случайная величина ξ принимает счетное число значений, то необходимо, чтобы

$$\sum_{i=1}^{\infty} |X_i| p_i < \infty;$$

в противном случае говорят, что математическое ожидание случайной величины ξ не существует.

Математическое ожидание дискретной случайной величины имеет аналог в теоретической механике. Пусть на прямой расположена система материальных точек с массами p_i и пусть X_i — координата i -й точки. Тогда центр тяжести системы будет иметь координату

$$\overline{X} = \frac{\sum_i X_i p_i}{\sum_i p_i} = \frac{\sum_i X_i p_i}{1} = \sum_i X_i p_i,$$

совпадающую с математическим ожиданием $M\xi$ случайной величины ξ .

Пример 1. Пусть ξ — число угаданных номеров в «Спортлото 6 из 49» (см. пример 8 в гл. 5). В соответствии с рядом распределения в табл. 7 гл. 5 имеем

$$\begin{aligned} M\xi &= 0 \cdot p_0 + 1 \cdot p_1 + 2 \cdot p_2 + 3 \cdot p_3 + 4 \cdot p_4 + 5 \cdot p_5 + 6 \cdot p_6 \cong \\ &\cong 0 \cdot 0,436 + 1 \cdot 0,413 + 2 \cdot 0,1324 + 3 \cdot 0,0176 + 4 \cdot 0,00097 + \\ &\quad + 5 \cdot 2 \cdot 10^{-5} + 6 \cdot 7 \cdot 10^{-8} \cong 0,735. \end{aligned}$$

Таким образом, среднее число угаданных номеров равно 0,735. \square

Пример 2. Найдем математическое ожидание случайной величины μ , распределенной по биномиальному закону (число успехов в n испытаниях Бернулли с вероятностью успеха p):

$$\begin{aligned} M\mu &= \sum_{i=0}^n i P_n(i) = \sum_{i=0}^n i \binom{n}{i} p^i q^{n-i} = \sum_{i=0}^n i \frac{n!}{i! (n-i)!} p^i q^{n-i} = \\ &= \sum_{i=1}^n np \frac{(n-1)!}{(i-1)! (n-i)!} p^{i-1} q^{n-i} = np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j q^{n-1-j} = \\ &= np \sum_{j=0}^{n-1} P_{n-1}(j) = np. \end{aligned} \quad \square$$

Пример 3. Пусть ξ имеет распределение Пуассона. Тогда

$$M\xi = \sum_{i=0}^{\infty} i \frac{\lambda^i}{i!} e^{-\lambda} = \lambda \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} e^{-\lambda} = \lambda \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} e^{-\lambda} = \lambda e^{\lambda} e^{-\lambda} = \lambda.$$

Таким образом, параметр λ пуассоновского распределения совпадает с математическим ожиданием. \square

Пример 4. Математическое ожидание геометрически распределенной случайной величины ξ имеет вид

$$M\xi = \sum_{i=0}^{\infty} ipq^i = pq \sum_{i=0}^{\infty} iq^{i-1} = pq \left(\sum_{i=0}^{\infty} q^i \right)'_q = pq \left(\frac{1}{1-q} \right)'_q = \frac{pq}{(1-q)^2} = \frac{pq}{p^2} = \frac{q}{p}. \quad \square$$

Пример 5. Положительная целочисленная случайная величина ξ имеет закон распределения $p_i = \mathbf{P}\{\xi = i\} = 1/[i(i+1)]$ ($i = 1, 2, \dots$). Тогда

$$\sum_{i=1}^{\infty} ip_i = \sum_{i=1}^{\infty} \frac{i}{i(i+1)} = \sum_{i=1}^{\infty} \frac{1}{i+1} = \infty$$

и, значит, математическое ожидание случайной величины ξ не существует. \square

Для определения математического ожидания непрерывной случайной величины ξ , имеющей плотность распределения $p(x)$, заметим, допуская некоторую вольность изложения, что случайная величина ξ принимает значение x с вероятностью $p(x) dx$. Заменяя сумму на интеграл, получаем: *математическим ожиданием (средним значением) $\mathbf{M}\xi$ непрерывной случайной величины ξ называется интеграл*

$$\mathbf{M}\xi = \int_{-\infty}^{\infty} x p(x) dx.$$

В этом случае условием существования математического ожидания является

$$\int_{-\infty}^{\infty} |x| p(x) dx < \infty.$$

Так же, как и в дискретном случае, математическое ожидание непрерывной случайной величины можно интерпретировать как центр тяжести стержня, плотность массы которого в точке x равна $p(x)$.

Пример 6. Найдем математическое ожидание равномерно распределенной на отрезке $[a, b]$ случайной величины ξ . Поскольку в этом случае $p(x) = 0$ при $x < a$ и $x > b$, то

$$\mathbf{M}\xi = \int_{-\infty}^{\infty} x p(x) dx = \int_a^b \frac{x}{b-a} dx = \frac{1}{b-a} \cdot \frac{1}{2} (b^2 - a^2) = \frac{b+a}{2}.$$

Как и следовало ожидать, $\mathbf{M}\xi$ совпадает с центром отрезка $[a, b]$. \square

Пример 7. Математическое ожидание случайной величины ξ , распределенной по нормальному закону со средним m и средним квадратичным отклонением σ , определяется формулой

$$\mathbf{M}\xi = \int_{-\infty}^{\infty} x \varphi_{m,\sigma}(x) dx = \int_{-\infty}^{\infty} \frac{x}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} dx.$$

Делая замену $y = (x - m)/\sigma$, получаем

$$\begin{aligned} \mathbf{M}\xi &= \int_{-\infty}^{\infty} \frac{\sigma y + m}{\sqrt{2\pi}} e^{-y^2/2} dy = \int_{-\infty}^{\infty} \frac{\sigma y}{\sqrt{2\pi}} e^{-y^2/2} dy + m \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy = \\ &= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y e^{-y^2/2} dy + m \int_{-\infty}^{\infty} \varphi(y) dy. \end{aligned}$$

Нетрудно видеть, что первый интеграл равен нулю, а второй равен единице как интеграл от плотности стандартного нормального распределения. Таким образом, $M\xi = m$, откуда и пошло название параметра m — среднее значение (второе название математического ожидания). \square

Пример 8. Пусть ξ — случайная величина, распределенная по закону Вейбулла. Тогда, поскольку $p(x) = 0$ при $x < 0$, то

$$M\xi = \int_{-\infty}^{\infty} x p(x) dx = \int_0^{\infty} \alpha\beta x^{\beta} e^{-\alpha x^{\beta}} dx.$$

Делая замену $y = \alpha x^{\beta}$, получаем

$$M\xi = \int_0^{\infty} y e^{-y} \alpha^{-\frac{1}{\beta}} y^{\frac{1}{\beta}-1} dy = \alpha^{-\frac{1}{\beta}} \int_0^{\infty} y^{1/\beta} e^{-y} dy = \alpha^{-\frac{1}{\beta}} \Gamma\left(\frac{1}{\beta} + 1\right). \quad \square$$

Пример 9. Математическое ожидание случайной величины ξ , имеющей гамма-распределение, задается выражением

$$M\xi = \int_0^{\infty} \frac{\lambda^{\gamma} x^{\gamma}}{\Gamma(\gamma)} e^{-\lambda x} dx.$$

Делая замену $y = \lambda x$, получаем

$$M\xi = \frac{1}{\lambda \Gamma(\gamma)} \int_0^{\infty} y^{\gamma} e^{-y} dy = \frac{\Gamma(\gamma+1)}{\lambda \Gamma(\gamma)} = \frac{\gamma}{\lambda}. \quad \square$$

Пример 10. Случайная величина ξ имеет плотность распределения Коши $p(x) = 1/[\pi(1+x)^2]$. Тогда

$$\int_{-\infty}^{\infty} \frac{|x| dx}{\pi(1+x^2)} = \infty$$

и математическое ожидание ξ не существует. \square

В общем случае математическое ожидание случайной величины задается выражением

$$M\xi = \int_{-\infty}^{\infty} x dF(x),$$

где интеграл понимается в так называемом смысле Стильеса. Поскольку мы рассматриваем только дискретные и непрерывные случайные величины, последнее выражение можно трактовать как обобщенную запись формул для математических ожиданий дискретной и непрерывной случайных величин.

2. Математическое ожидание функции от случайной величины. Свойства математического ожидания

Прежде чем переходить к описанию свойств математического ожидания случайной величины, позволяющих, как будет видно из примеров, в ряде случаев существенно упростить его вычисление, определим математическое ожидание функции от случайной величины (случайно-

го вектора). Итак, пусть $\eta = g(\xi)$ — функция от случайной величины. Для определения $M\eta = Mg(\xi)$ можно было бы сначала по формулам параграфа 5 гл. 5 найти распределение случайной величины η и затем уже, воспользовавшись определениями предыдущего параграфа, вычислить $M\eta$. Однако мы применим другой, более удобный подход. Рассмотрим сначала дискретную случайную величину ξ , принимающую значения X_1, \dots, X_n . Тогда случайная величина $\eta = g(\xi)$, как мы знаем, принимает значения $g(X_1), \dots, g(X_n)$ с теми же вероятностями $p_i = P\{\xi = X_i\}$, и ее математическое ожидание определяется формулой

$$M\eta = Mg(\xi) = \sum_{i=1}^n g(X_i) p_i.$$

Если случайная величина ξ принимает счетное число значений, то математическое ожидание ξ определяется формулой

$$M\eta = Mg(\xi) = \sum_{i=1}^{\infty} g(X_i) p_i,$$

но при этом для существования математического ожидания необходимо выполнение условия

$$\sum_{i=1}^{\infty} |g(X_i)| p_i < \infty.$$

Пример 11. Определим математическое ожидание выигрыша η в «Спортлото 6 из 49» (см. пример 9 в гл. 5). Поскольку η является функцией от случайной величины ξ — числа угаданных номеров, то, воспользовавшись формулой для математического ожидания функции от случайной величины и рядом распределения в табл. 7 из гл. 5, имеем

$$\begin{aligned} M\eta = Mg(\xi) &= g(0)P\{\xi = 0\} + g(1)P\{\xi = 1\} + g(2)P\{\xi = 2\} + \\ &+ g(3)P\{\xi = 3\} + g(4)P\{\xi = 4\} + g(5)P\{\xi = 5\} + g(6)P\{\xi = 6\} \cong \\ &\cong -0,3 \cdot (0,436 + 0,413 + 0,1324) + 2,7 \cdot 0,0176 + 54,7 \cdot 0,00097 + \\ &+ 699,7 \cdot 2 \cdot 10^{-5} + 9999,7 \cdot 7 \cdot 10^{-8} \cong -0,179. \end{aligned}$$

Таким образом, математическое ожидание выигрыша отрицательно и равно примерно 18 коп., а это значит, что играющий в среднем проигрывает больше половины стоимости билета (30 коп.). Естественно, мы получим то же значение $M\eta$, если воспользуемся рядом распределения случайной величины η , представленным в табл. 10 из гл. 5. \square

Аналогично, для непрерывной случайной величины ξ , имеющей плотность распределения $p(x)$, математическое ожидание случайной величины $\eta = g(\xi)$ определяется выражением

$$M\eta = Mg(\xi) = \int_{-\infty}^{\infty} g(x) p(x) dx,$$

причем и здесь должно быть выполнено условие

$$\int_{-\infty}^{\infty} |g(x)| p(x) dx < \infty.$$

В дальнейшем, чтобы каждый раз не оговаривать условие существования математического ожидания, будем предполагать, что соответствующие сумма или интеграл сходятся абсолютно.

Математическое ожидание функции от многомерной случайной величины определяется точно так же. Так, математическое ожидание $\mathbf{M}\eta$ функции $\eta = g(\xi_1, \xi_2)$ от дискретной и непрерывной двумерных случайных величин (ξ_1, ξ_2) задается выражениями

$$\begin{aligned}\mathbf{M}\eta &= \mathbf{M}g(\xi_1, \xi_2) = \sum_{i,j} g(X_i, Y_j) p_{ij}, \\ \mathbf{M}\eta &= \mathbf{M}g(\xi_1, \xi_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) p_{\xi_1, \xi_2}(x, y) dx dy.\end{aligned}$$

Теперь мы в состоянии вывести свойства математического ожидания.

Если случайная величина ξ принимает всего одно значение c с вероятностью, равной единице (т. е. по сути дела является неслучайной величиной), то

$$\mathbf{1. M}c = c \cdot 1 = c.$$

Далее, найдем математическое ожидание случайной величины $\eta = a\xi + b$ ($g(x) = ax + b$). Рассматривая, например, непрерывный случай, имеем

$$\mathbf{M}\eta = \mathbf{M}(a\xi + b) = \int_{-\infty}^{\infty} (ax + b) p_{\xi}(x) dx = a \int_{-\infty}^{\infty} x p_{\xi}(x) dx + b \int_{-\infty}^{\infty} p_{\xi}(x) dx,$$

т. е.

$$\mathbf{2. M}(a\xi + b) = a\mathbf{M}\xi + b.$$

Аналогично свойство 2 доказывается для дискретной случайной величины ξ .

Пусть теперь $\eta = \xi_1 + \xi_2$ ($g(x_1, x_2) = x_1 + x_2$). Тогда (теперь уже на примере дискретной случайной величины)

$$\begin{aligned}\mathbf{M}\eta &= \mathbf{M}(\xi_1 + \xi_2) = \sum_{i,j} (X_i + Y_j) p_{ij} = \sum_{i,j} X_i p_{ij} + \sum_{i,j} Y_j p_{ij} = \\ &= \sum_i X_i \sum_j p_{ij} + \sum_j Y_j \sum_i p_{ij} = \sum_i X_i p_{\xi_1 i} + \sum_j Y_j p_{\xi_2 j}\end{aligned}$$

и, значит,

$$\mathbf{3. M}(\xi_1 + \xi_2) = \mathbf{M}\xi_1 + \mathbf{M}\xi_2.$$

К свойству 3 можно сделать следующее замечание: математическое ожидание суммы случайных величин может существовать даже тогда,

когда математические ожидания обоих слагаемых не существуют. Так, $M(\xi_1 - \xi_2) = M0 = 0$ несмотря на то, что $M\xi$ может и не существовать. Очевидно также, что свойство 3 обобщается на случай произвольного числа n слагаемых.

Наконец, если ξ_1 и ξ_2 независимы, то для математического ожидания их произведения $\eta = \xi_1 \xi_2$ ($g(x_1, x_2) = x_1 x_2$) имеет место формула (снова обращаясь к непрерывному случаю)

$$\begin{aligned} M\eta = M(\xi_1 \xi_2) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 p_{\xi_1, \xi_2}(x_1, x_2) dx_1 dx_2 = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 p_{\xi_1}(x_1) p_{\xi_2}(x_2) dx_1 dx_2 = \\ &= \left(\int_{-\infty}^{\infty} x_1 p_{\xi_1}(x_1) dx_1 \right) \left(\int_{-\infty}^{\infty} x_2 p_{\xi_2}(x_2) dx_2 \right). \end{aligned}$$

Поэтому

4. $M(\xi_1 \xi_2) = M\xi_1 M\xi_2$ для независимых случайных величин ξ_1 и ξ_2 .

И это свойство допускает обобщение на произведение любого числа независимых (в совокупности) сомножителей.

Пример 12. Математическое ожидание случайной величины ξ , распределенной по стандартному нормальному закону, имеет вид

$$M\xi = \int_{-\infty}^{\infty} x \varphi(x) dx = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 0.$$

В примере 13 из гл. 5 показано, что случайная величина $\eta = \sigma\xi + m$ распределена по нормальному закону с параметрами m и σ . Таким образом, по свойству 2 математического ожидания $M\eta = \sigma M\xi + m = m$ и соответственно параметр m нормального закона является математическим ожиданием. Этот же результат был нами получен прямыми вычислениями в примере 7. \square

Пример 13. Представим число успехов μ в схеме Бернулли из n испытаний в виде $\mu = \mu_1 + \dots + \mu_n$, где μ_i — число успехов в i -м испытании. Нетрудно видеть, что

$$M\mu_i = 0 \cdot q + 1 \cdot p = p.$$

Значит, по свойству 3, $M\mu = M\mu_1 + \dots + M\mu_n = np$, что совпадает с результатами примера 2, но получено с минимальными вычислениями. \square

3. Дисперсия. Моменты высших порядков

Математическое ожидание не всегда является достаточно удовлетворительной характеристикой случайной величины. Например, играть в орлянку можно по копейке, по рублю и даже по 1000 руб. При любой (одинаковой) ставке игроков и симметричной монете математическое ожидание выигрыша каждого игрока равно нулю, однако далеко не каждый не только рискует, но и просто сможет вверить «глупой»

монете 1000 руб. Поэтому наряду со средним значением нужно иметь и число, характеризующее «разброс» случайной величины вокруг своего среднего. Такой характеристикой обычно служит дисперсия. И дело тут не только в том, что дисперсия является единственной мерой степени разброса (существует бесконечно много таких характеристик, в частности, центральные моменты любого четного порядка, которые также будут определены в этом параграфе; кроме того, сама дисперсия не всегда является хорошим показателем степени разброса). Использование дисперсии и других характеристик второго порядка (ковариаций) позволяет применить в теории вероятностей сильно развитый аппарат гильбертовых пространств. Особо важную роль этот аппарат играет в теории так называемых стационарных в широком смысле случайных процессов, которые в свою очередь являются основной математической моделью в ряде практических приложений.

Вторым (начальным) моментом m_2 случайной величины ξ называется математическое ожидание квадрата ξ ($g(x) = x^2$):

$$m_2 = \mathbf{M}\xi^2 = \sum_i X_i^2 p_i$$

для дискретной величины ξ и

$$m_2 = \mathbf{M}\xi^2 = \int_{-\infty}^{\infty} x^2 p(x) dx$$

— для непрерывной.

Дисперсия $D\xi$ дискретной и непрерывной случайных величин ξ определяется соответственно формулами ($g(x) = (x - \mathbf{M}\xi)^2$):

$$D\xi = \mathbf{M}(\xi - \mathbf{M}\xi)^2 = \sum_i (X_i - \mathbf{M}\xi)^2 p_i$$

и

$$D\xi = \mathbf{M}(\xi - \mathbf{M}\xi)^2 = \int_{-\infty}^{\infty} (x - \mathbf{M}\xi)^2 p(x) dx.$$

Дисперсия $D\xi$ представляет собой второй момент случайной величины ξ , из которой вычтено ее математическое ожидание $\mathbf{M}\xi$, т. е. центрированной (имеющей нулевое математическое ожидание) случайной величины $\hat{\xi} = \xi - \mathbf{M}\xi$. Поэтому иногда дисперсию называют вторым центральным моментом.

Дисперсия также имеет аналог в теоретической механике — центральный (относительно центра тяжести) момент инерции, характеризующий разброс массы относительно центра тяжести.

Выведем некоторые свойства дисперсии.

Если случайная величина ξ с вероятностью, равной единице, принимает всего одно значение c , то из свойства 1 математического ожидания ($\mathbf{M}\xi = c$) получаем

$$1. D\xi = M(\xi - c)^2 = (c - c)^2 \cdot 1 = 0.$$

Можно показать, что справедливо и обратное: дисперсия случайной величины ξ равна нулю тогда и только тогда, когда ξ с вероятностью, равной единице, принимает всего одно значение.

Определим дисперсию случайной величины $\eta = a\xi + b$. Используя свойство 2 математического ожидания, имеем

$$\begin{aligned} D\eta &= M(\eta - M\eta)^2 = M[a\xi + b - M(a\xi + b)]^2 = M(a\xi + b - aM\xi - b)^2 = \\ &= M[a(\xi - M\xi)]^2 = M[a^2(\xi - M\xi)^2] = a^2 M(\xi - M\xi)^2. \end{aligned}$$

Поэтому

$$2. D(a\xi + b) = a^2 D\xi.$$

Далее из свойств 2 и 3 математического ожидания получаем

$$D\xi = M(\xi - M\xi)^2 = M[\xi^2 - 2\xi M\xi + (M\xi)^2] = M\xi^2 - 2(M\xi)^2 + (M\xi)^2,$$

т. е.

$$3. D\xi = m_2 - (M\xi)^2.$$

Свойство 3 дает весьма удобную формулу для расчета дисперсии дискретной случайной величины с помощью ЭВМ или микрокалькулятора. Действительно, если бы мы производили вычисления дисперсии по первоначальной формуле, нам пришлось бы два раза суммировать по i : первый раз при подсчете математического ожидания и второй — дисперсии. Свойство 3 позволяет обходиться одним циклом: мы можем одновременно суммировать с весами p_i и сами значения случайной величины, и их квадраты.

Наконец, пусть ξ и η — независимые случайные величины. Тогда, используя независимость случайных величин $\xi = \xi - M\xi$ и $\eta = \eta - M\eta$, а также свойства 2–4 математического ожидания, получаем

$$\begin{aligned} D(\xi + \eta) &= M[\xi + \eta - M(\xi + \eta)]^2 = M[(\xi - M\xi) + (\eta - M\eta)]^2 = \\ &= M(\xi - M\xi)^2 + 2M[(\xi - M\xi)(\eta - M\eta)] + M(\eta - M\eta)^2 = \\ &= D\xi + 2(M\xi \cdot M\eta) + D\eta. \end{aligned}$$

Значит,

$$4. D(\xi + \eta) = D\xi + D\eta \text{ для независимых случайных величин } \xi \text{ и } \eta.$$

Очевидно, что свойство 4 справедливо для суммы не только двух, но и любого числа n попарно независимых слагаемых.

Заметим, что дисперсия $D\xi$ имеет размерность квадрата размерности случайной величины ξ . Для практических же целей удобно иметь меру разброса, размерность которой совпадает с размерностью ξ . В качестве такой меры естественно использовать $\sigma = \sqrt{D\xi}$, которую называют *средним квадратичным отклонением* случайной величины ξ (иногда также *стандартом* или *стандартным отклонением*).

Пример 14. Найдем дисперсию случайной величины ξ , распределенной по закону Пуассона. Для этого воспользуемся свойством 3 дисперсии. Математическое ожидание $M\xi = \lambda$ было найдено в примере 3. Определим второй момент:

$$\begin{aligned} M\xi^2 &= \sum_{i=0}^{\infty} i^2 \frac{\lambda^i}{i!} e^{-\lambda} = \lambda \sum_{i=1}^{\infty} i \frac{\lambda^{i-1}}{(i-1)!} e^{-\lambda} = \lambda \sum_{j=0}^{\infty} (j+1) \frac{\lambda^j}{j!} e^{-\lambda} = \\ &= \lambda \left(\sum_{j=0}^{\infty} j \frac{\lambda^j}{j!} e^{-\lambda} + \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} e^{-\lambda} \right) = \lambda (M\xi + 1) = \lambda^2 + \lambda. \end{aligned}$$

Таким образом,

$$D\xi = \lambda^2 + \lambda - \lambda^2 = \lambda$$

и, значит, дисперсия ξ , так же, как и математическое ожидание, совпадает с параметром λ . \square

Пример 15. Пусть μ — число успехов в n испытаниях Бернулли. Дисперсию μ можно подсчитать так же, как в примере 2 было подсчитано математическое ожидание: воспользовавшись непосредственным определением дисперсии. Однако мы поступим другим образом. Для этого снова (см. пример 13) представим μ в виде суммы $\mu = \mu_1 + \dots + \mu_n$. Дисперсия каждого слагаемого равна

$$\begin{aligned} D\mu_i &= (0 - M\mu_i)^2 q + (1 - M\mu_i)^2 p = (-p)^2 q + (1-p)^2 p = \\ &= p^2 q + q^2 p = pq(p+q) = pq. \end{aligned}$$

Учитывая, что μ_i независимы, и воспользовавшись свойством 4 дисперсии, получаем

$$D\mu = D\mu_1 + \dots + D\mu_n = npq. \quad \square$$

Пример 16. Дисперсия равномерно распределенной на отрезке $[a, b]$ случайной величины ξ определяется формулой

$$\begin{aligned} D\xi &= \int_a^b \left(x - \frac{b+a}{2} \right)^2 \frac{1}{b-a} dx = \\ &= \frac{1}{3(b-a)} \left[\left(b - \frac{b+a}{2} \right)^3 - \left(a - \frac{b+a}{2} \right)^3 \right] = \frac{(b-a)^3}{12(b-a)} = \frac{(b-a)^2}{12}. \quad \square \end{aligned}$$

Пример 17. Дисперсия случайной величины ξ , распределенной по нормальному закону с параметрами m и σ , имеет вид

$$D\xi = \int_{-\infty}^{\infty} (x-m)^2 \varphi_{m,\sigma}(x) dx = \int_{-\infty}^{\infty} \frac{(x-m)^2}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} dx.$$

Делая замену $y = (x-m)/\sigma$, получаем

$$D\xi = \sigma^2 \int_{-\infty}^{\infty} \frac{y^2}{\sqrt{2\pi}} e^{-y^2/2} dy.$$

Полагая $v = y/\sqrt{2\pi}$, $du = y e^{-y^2/2} dy$ и интегрируя по частям, находим

$$\mathbf{D}\xi = \sigma^2 \int_{-\infty}^{\infty} \varphi(y) dy = \sigma^2.$$

Таким образом, дисперсия нормально распределенной случайной величины совпадает с квадратом второго параметра. Этого и следовало ожидать, поскольку σ носит название среднего квадратичного отклонения. \square

Пример 18. Для определения дисперсии случайной величины ξ , имеющей гамма-распределение, воспользуемся свойством 3 дисперсии. Тогда

$$\mathbf{M}\xi^2 = \int_0^{\infty} \frac{\lambda^{\gamma} x^{\gamma+1}}{\Gamma(\gamma)} e^{-\lambda x} dx$$

или после замены $y = \lambda x$

$$\mathbf{M}\xi^2 = \frac{1}{\lambda^2 \Gamma(\gamma)} \int_0^{\infty} y^{\gamma+1} e^{-y} dy = \frac{\Gamma(\gamma+2)}{\lambda^2 \Gamma(\gamma)} = \frac{\gamma(\gamma+1)}{\lambda^2}.$$

Вспоминая (см. пример 9), что $\mathbf{M}\xi = \gamma/\lambda$, окончательно получаем

$$\mathbf{D}\xi = \mathbf{M}\xi^2 - (\mathbf{M}\xi)^2 = \frac{\gamma(\gamma+1)}{\lambda^2} - \frac{\gamma^2}{\lambda^2} = \frac{\gamma}{\lambda^2}. \quad \square$$

Пример 19. Пусть ξ — случайная величина, имеющая дисперсию $\mathbf{D}\xi$. Введем новую случайную величину $\eta = \xi - a$. Найдем число a , доставляющее минимум $\mathbf{M}\eta^2$. Воспользовавшись свойствами 2 и 3 дисперсии, имеем

$$\mathbf{M}\eta^2 = \mathbf{D}\eta + (\mathbf{M}\eta)^2 = \mathbf{D}\xi + (\mathbf{M}\xi - a)^2.$$

Первое слагаемое от a не зависит, а второе принимает минимальное значение, равное 0, при $a = \mathbf{M}\xi$. Таким образом, в качестве a нужно взять $\mathbf{M}\xi$, а само минимальное значение $\mathbf{M}\eta^2$ совпадает с дисперсией $\mathbf{D}\xi$. \square

В некоторых теоретических исследованиях встречаются моменты высших порядков.

Моментом m_k порядка k (k -м моментом) называется математическое ожидание k -й степени случайной величины ξ ($g(x) = x^k$):

$$m_k = \mathbf{M}\xi^k = \sum_i X_i^k p_i,$$

если ξ — дискретная случайная величина, и

$$m_k = \mathbf{M}\xi^k = \int_{-\infty}^{\infty} x^k p(x) dx,$$

если ξ непрерывна. Иногда k -й момент называют также начальным моментом k -го порядка.

Центральным моментом \bar{m}_k порядка k (k -м центральным моментом) называется математическое ожидание k -й степени случайной величины $\xi = \xi - \mathbf{M}\xi$ ($g(x) = (x - \mathbf{M}\xi)^k$):

$$\bar{m}_k = \mathbf{M}(\xi - \mathbf{M}\xi)^k = \sum_i (X_i - \mathbf{M}\xi)^k p_i$$

и

$$\hat{m}_k = \mathbf{M}(\xi - \mathbf{M}\xi)^k = \int_{-\infty}^{\infty} (x - \mathbf{M}\xi)^k p(x) dx$$

соответственно для дискретной и непрерывной случайных величин ξ .

Момент первого порядка совпадает с математическим ожиданием, центральный момент первого порядка равен нулю, центральный момент второго порядка является дисперсией. Отметим также, что в теоретических изысканиях рассматриваются моменты не обязательно целого порядка k .

4. Ковариация и корреляция случайных величин

Пусть (ξ_1, ξ_2) — двумерный случайный вектор. Будем называть *ковариацией* $\text{cov}(\xi_1, \xi_2)$ случайных величин ξ_1 и ξ_2 *математическое ожидание произведения центрированных случайных величин* $\ddot{\xi}_1 = \xi_1 - \mathbf{M}\xi_1$ и $\ddot{\xi}_2 = \xi_2 - \mathbf{M}\xi_2$:

$$\text{cov}(\xi_1, \xi_2) = \mathbf{M}\ddot{\xi}_1\ddot{\xi}_2 = \mathbf{M}[(\xi_1 - \mathbf{M}\xi_1)(\xi_2 - \mathbf{M}\xi_2)].$$

Как обычно, выпишем последнее выражение для дискретного и непрерывного случайных векторов (ξ_1, ξ_2) :

$$\text{cov}(\xi_1, \xi_2) = \sum_{i,j} (X_i - \mathbf{M}\xi_1)(Y_j - \mathbf{M}\xi_2) p_{ij},$$

$$\text{cov}(\xi_1, \xi_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 - \mathbf{M}\xi_1)(x_2 - \mathbf{M}\xi_2) p_{\xi_1, \xi_2}(x_1, x_2) dx_1 dx_2.$$

Ковариация обладает следующими свойствами:

1. $\text{cov}(\xi, \xi) = \mathbf{M}(\xi - \mathbf{M}\xi)^2 = \mathbf{D}\xi$.

Если ξ_1 и ξ_2 независимы (и имеют математические ожидания), то

$$\text{cov}(\xi_1, \xi_2) = \mathbf{M}[(\xi_1 - \mathbf{M}\xi_1)(\xi_2 - \mathbf{M}\xi_2)] = [\mathbf{M}(\xi_1 - \mathbf{M}\xi_1)][\mathbf{M}(\xi_2 - \mathbf{M}\xi_2)]$$

и, значит,

2. $\text{cov}(\xi_1, \xi_2) = 0$ для независимых случайных величин ξ_1 и ξ_2 .

Как видно из свойства 2, ковариация независимых случайных величин равна нулю. Однако обратное, вообще говоря, неверно. Существуют зависимые случайные величины, ковариация которых также равна нулю.

Далее, пусть $\eta_1 = a_1\xi_1 + b_1$, $\eta_2 = a_2\xi_2 + b_2$. Тогда

$$\begin{aligned} \text{cov}(\eta_1, \eta_2) &= \mathbf{M}[(\eta_1 - \mathbf{M}\eta_1)(\eta_2 - \mathbf{M}\eta_2)] = \\ &= \mathbf{M}(a_1\xi_1 + b_1 - a_1\mathbf{M}\xi_1 - b_1)(a_2\xi_2 + b_2 - a_2\mathbf{M}\xi_2 - b_2) = \\ &= \mathbf{M}[a_1a_2(\xi_1 - \mathbf{M}\xi_1)(\xi_2 - \mathbf{M}\xi_2)]. \end{aligned}$$

Поэтому

$$\mathbf{3. cov}(\eta_1, \eta_2) = a_1 a_2 \mathbf{cov}(\xi_1, \xi_2).$$

Наконец, рассмотрим дисперсию случайной величины $\eta_x = x\xi_1 - \xi_2$, где x — произвольное число. По свойствам дисперсии

$$\mathbf{D}\eta_x = x^2 \mathbf{D}\xi_1 - 2x \mathbf{cov}(\xi_1, \xi_2) + \mathbf{D}\xi_2.$$

Как функция от x дисперсия $\mathbf{D}\eta_x$ представляет собой квадратный трехчлен. Но дисперсия любой случайной величины не может быть меньше нуля, а это означает, что дискриминант $[2 \mathbf{cov}(\xi_1, \xi_2)]^2 - 4\mathbf{D}\xi_1 \mathbf{D}\xi_2$ квадратного трехчлена $\mathbf{D}\eta_x$ неположителен, т. е.

$$\mathbf{4. } -\sqrt{\mathbf{D}\xi_1 \mathbf{D}\xi_2} \leq \mathbf{cov}(\xi_1, \xi_2) \leq \sqrt{\mathbf{D}\xi_1 \mathbf{D}\xi_2}.$$

Более того, если дискриминант равен нулю, то уравнение

$$x^2 \mathbf{D}\xi_1 - 2x \mathbf{cov}(\xi_1, \xi_2) + \mathbf{D}\xi_2 = 0$$

имеет решение x_0 . Тогда $\mathbf{D}\eta_{x_0} = 0$ и, значит, $\eta_{x_0} \equiv c$, $\xi_2 \equiv x_0 \xi_1 - c$. В этом случае нетрудно видеть из свойства 3, что если коэффициент пропорциональности x_0 положителен, то $\mathbf{cov}(\xi_1, \xi_2) = \sqrt{\mathbf{D}\xi_1 \mathbf{D}\xi_2}$, а если отрицателен, то $\mathbf{cov}(\xi_1, \xi_2) = -\sqrt{\mathbf{D}\xi_1 \mathbf{D}\xi_2}$. Таким образом,

$$\mathbf{5. cov}(\xi_1, \xi_2) = \pm \sqrt{\mathbf{D}\xi_1 \mathbf{D}\xi_2} \text{ тогда и только тогда, когда случайные величины } \xi_1 \text{ и } \xi_2 \text{ линейно зависимы.}$$

Наконец, раскрывая скобки в формуле, определяющей ковариацию, и используя свойства математического ожидания, получаем

$$\mathbf{6. cov}(\xi_1, \xi_2) = \mathbf{M}\xi_1 \xi_2 - \mathbf{M}\xi_1 \mathbf{M}\xi_2.$$

Последнее свойство часто бывает полезным при численном подсчете ковариации.

Заметим теперь, что введение ковариации позволяет выписать выражение для дисперсии произвольных (а не только независимых) случайных величин и к четырем уже известным **свойствам дисперсии** (см. параграф 3) можно добавить еще одно:

$$\mathbf{5. D}(\xi + \eta) = \mathbf{D}\xi + \mathbf{D}\eta + 2 \mathbf{cov}(\xi, \eta),$$

справедливое для произвольных (а не только независимых) случайных величин ξ и η .

В общем случае n -мерного случайного вектора (ξ_1, \dots, ξ_n) **матрицей ковариаций (ковариационной матрицей)** называется матрица $A = (\mathbf{cov}(\xi_i, \xi_j))$, состоящая из ковариаций случайных величин ξ_i и ξ_j .

Пример 20. Рассмотрим двумерную случайную величину (ξ, η) , распределенную по нормальному закону (см. пример 16 в гл. 6). Тогда

$$\mathbf{cov}(\xi, \eta) =$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{(x - m_1)(y - m_2)}{2\pi\sigma_1\sigma_2\sqrt{1 - \rho^2}} e^{-\frac{1}{2(1 - \rho^2)} \left[\frac{(x - m_1)^2}{\sigma_1^2} - \frac{2\rho(x - m_1)(y - m_2)}{\sigma_1\sigma_2} + \frac{(y - m_2)^2}{\sigma_2^2} \right]} dx dy.$$

Делая замену $u = (x - m_1)/\sigma_1$, $v = (y - m_2)/\sigma_2$, получаем

$$\begin{aligned} \text{cov}(\xi, \eta) &= \sigma_1 \sigma_2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{uv}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(u^2-2\rho uv+v^2)} du dv = \\ &= \sigma_1 \sigma_2 \int_{-\infty}^{\infty} \frac{u}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \int_{-\infty}^{\infty} \frac{v}{\sqrt{2\pi(1-\rho^2)}} e^{-\frac{(v-\rho u)^2}{2(1-\rho^2)}} dv du. \end{aligned}$$

Внутренний интеграл равен ρu . Поэтому

$$\text{cov}(\xi, \eta) = \rho \sigma_1 \sigma_2 \int_{-\infty}^{\infty} \frac{u^2}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du = \rho \sigma_1 \sigma_2.$$

Поскольку $D\xi = \sigma_1^2$, $D\eta = \sigma_2^2$, то матрица A представляет собой ковариационную матрицу (в данном случае название «ковариационная матрица» мы ввели раньше, нежели выяснили смысл этого понятия).

Аналогично, в общем случае n -мерного нормального случайного вектора (ξ_1, \dots, ξ_n) элементы a_{ij} ковариационной матрицы A являются ковариациями случайных величин ξ_i и ξ_j . \square

Пример 21. Пусть ξ_1 — число очков, выпавших на одной игральной кости, а ξ_2 — на другой. Рассмотрим случайные величины $\eta_1 = \xi_1 + \xi_2$ (сумма очков на обеих kostях) и $\eta_2 = \xi_1 - \xi_2$ (разность очков). Тогда

$$\begin{aligned} \text{cov}(\eta_1, \eta_2) &= M[(\xi_1 + \xi_2 - M\xi_1 - M\xi_2)(\xi_1 - \xi_2 - M\xi_1 + M\xi_2)] = \\ &= M[(\xi_1 - M\xi_1)^2 - (\xi_2 - M\xi_2)^2] = D\xi_1 - D\xi_2. \end{aligned}$$

Случайные величины ξ_1 и ξ_2 одинаково распределены (и даже независимы), и, значит, $D\xi_1 = D\xi_2$, $\text{cov}(\eta_1, \eta_2) = 0$. Однако несмотря на это, η_1 и η_2 зависимы, поскольку, например, из равенства $\eta_1 = 2$ (такое может произойти только при выпадении по одному очку на каждой kostи) обязательно следует равенство $\eta_2 = 0$. \square

Итак, ковариацию можно считать мерой независимости случайных величин (хотя и не очень хорошей, так как можно ввести другие, заведомо лучшие показатели независимости; оправданием повсеместного применения ковариации служит то, что она также относится к числу характеристик второго порядка). Существенным недостатком ковариации является то, что ее размерность совпадает с произведением размерностей случайных величин. Естественно, нужно иметь безразмерную характеристику независимости. Ее очень просто получить, для этого достаточно поделить ковариацию на произведение средних квадратичных отклонений.

Коэффициентом корреляции случайных величин ξ и η называется число $\rho = \rho(\xi, \eta)$, определяемое выражением

$$\rho = \frac{\text{cov}(\xi, \eta)}{\sqrt{D\xi \cdot D\eta}}.$$

Коэффициент корреляции является уже безразмерной величиной.

Выпишем его свойства, аналогичные свойствам ковариации:

$$1. \rho(\xi, \xi) = \frac{\text{D}\xi}{\sqrt{\text{D}\xi \cdot \text{D}\xi}} = 1.$$

Если ξ и η независимы (и существуют $\text{D}\xi > 0$ и $\text{D}\eta > 0$), то

$$2. \rho(\xi, \eta) = 0.$$

Далее, пусть $\eta_1 = a_1\xi_1 + b_1$, $\eta_2 = a_2\xi_2 + b_2$ ($a_1 \neq 0$, $a_2 \neq 0$). Тогда

$$3. \rho(\eta_1, \eta_2) = a_1 a_2 \text{cov}(\xi_1, \xi_2) / \sqrt{a_1^2 \text{D}\xi_1 \cdot a_2^2 \text{D}\xi_2} = \pm \rho(\xi_1, \xi_2).$$

При этом знак плюс надо брать тогда, когда a_1 и a_2 имеют одинаковые знаки, и минус — в противном случае.

Наконец,

$$4. -1 \leq \rho(\xi, \eta) \leq 1.$$

5. $\rho(\xi, \eta) = \pm 1$ тогда и только тогда, когда случайные величины ξ и η линейно зависимы.

Можно сказать, что коэффициент корреляции ρ отражает степень линейной зависимости случайных величин ξ и η . С возрастанием ξ случайная величина η имеет тенденцию к увеличению при $\rho > 0$ и к уменьшению при $\rho < 0$. Поэтому при $\rho > 0$ говорят о *положительной* корреляционной зависимости ξ и η , при $\rho < 0$ — об *отрицательной*. Например, рост и вес человека связаны положительной корреляционной зависимостью, а температура и время сохранности продукта — отрицательной. Если $\rho = 0$, то случайные величины ξ и η называются *некоррелированными*.

Пример 22. Найдем коэффициент корреляции случайных величин ξ — числа очков, выпавших на верхней грани игральной кости, и η — на нижней (пример 7 из гл. 6). Для этого сначала вычислим $\text{M}\xi$, $\text{M}\eta$, $\text{D}\xi$, $\text{D}\eta$ и $\text{cov}(\xi, \eta)$. Воспользовавшись табл. 3 из гл. 6, имеем:

$$\text{M}\xi = \text{M}\eta = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3,5,$$

$$\begin{aligned} \text{D}\xi = \text{D}\eta = (1 - 3,5)^2 \cdot \frac{1}{6} + (2 - 3,5)^2 \cdot \frac{1}{6} + (3 - 3,5)^2 \cdot \frac{1}{6} + \\ + (4 - 3,5)^2 \cdot \frac{1}{6} + (5 - 3,5)^2 \cdot \frac{1}{6} + (6 - 3,5)^2 \cdot \frac{1}{6} = \frac{35}{12}, \end{aligned}$$

$$\begin{aligned} \text{cov}(\xi, \eta) = (1 - 3,5)(1 - 3,5) \cdot 0 + (2 - 3,5)(1 - 3,5) \cdot 0 + \dots + \\ + (6 - 3,5)(1 - 3,5) \cdot \frac{1}{6} + (1 - 3,5)(2 - 3,5) \cdot 0 + \dots + \\ + (5 - 3,5)(2 - 3,5) \cdot \frac{1}{6} + (6 - 3,5)(2 - 3,5) \cdot 0 + \dots + \\ + (4 - 3,5)(3 - 3,5) \cdot \frac{1}{6} + \dots + (3 - 3,5)(4 - 3,5) \cdot \frac{1}{6} + \dots + \\ + (2 - 3,5)(5 - 3,5) \cdot \frac{1}{6} + \dots + (1 - 3,5)(6 - 3,5) \cdot \frac{1}{6} = -\frac{35}{12}. \end{aligned}$$

Таким образом, $\rho = (-35/12)/(35/12) = -1$. Впрочем, это мы могли бы установить и без всяких вычислений по свойству 5 коэффициента корреляции, если

бы вспомнили, что сумма чисел очков на противоположных гранях равна семи и, значит, $\xi = 7 - \eta$ (ξ и η связаны линейной зависимостью с отрицательным коэффициентом пропорциональности). \square

Пример 23. (Задача о наилучшем линейном прогнозе). Пусть (ξ_1, ξ_2) — двумерная случайная величина. Рассмотрим новую случайную величину $\eta = \xi_2 - (x\xi_1 + a)$.

Попытаемся подобрать числа x и a , доставляющие $M\eta^2$ минимальное значение. Из свойств дисперсии и математического ожидания имеем

$$M\eta^2 = D\eta + (M\eta)^2 = [D\xi_1 - 2x \operatorname{cov}(\xi_1, \xi_2) + x^2 D\xi_2] + [M\xi_2 - (xM\xi_2 + a)]^2.$$

Дифференцируя $D\eta$ по x и приравнивая производную нулю, получаем, что минимальное значение $D\eta$ достигается при $x_0 = \operatorname{cov}(\xi_1, \xi_2)/D\xi_2$ и равно

$$D\xi_1 - \frac{[\operatorname{cov}(\xi_1, \xi_2)]^2}{D\xi_2} = [1 - \rho^2(\xi_1, \xi_2)]D\xi_1.$$

Полагая теперь $a_0 = M\xi_2 - x_0 M\xi_2$, получаем минимальное значение второго слагаемого, равное нулю.

Окончательно имеем: минимальное значение $M\eta^2$, равное

$$[1 - \rho^2(\xi_1, \xi_2)]D\xi_1,$$

достигается при

$$x = x_0 = \frac{\operatorname{cov}(\xi_1, \xi_2)}{D\xi_2} \quad \text{и} \quad a = a_0 = M\xi_2 - (x_0 M\xi_2).$$

Таким образом, коэффициент корреляции тесно связан с задачей наилучшего линейного приближения одной случайной величины другой, о чем говорилось выше. \square

Замечание. Рассмотренная задача имеет очень простую трактовку в терминах наилучшего линейного прогноза. Действительно, пусть нам известно значение случайной величины ξ_1 и мы хотим построить по этому значению наилучший в смысле среднего квадратичного отклонения линейный прогноз ξ_2 случайной величины ξ_2 . Тогда этот прогноз определяется формулой $\tilde{\eta} = x_0 \xi_1 + a_0$, где x_0 и a_0 определены выше и выражаются только через моменты случайных величин ξ_1 и ξ_2 первого и второго порядка. В частности, при $\rho(\xi_1, \xi_2) = 1$ прогноз будет точным ($M\eta^2 = 0$), а при $\rho(\xi_1, \xi_2) = 0$ — состоит только в указании среднего значения $a_0 = M\xi_2$ ($x_0 = 0$). Это еще раз подтверждает тезис о том, что коэффициент корреляции является мерой линейной зависимости случайных величин ξ_1 и ξ_2 .

5. Условное математическое ожидание. Регрессия

Пусть (ξ, η) — двумерная случайная величина. В соответствии с результатами параграфа 5 гл. 6 (так как мы рекомендовали при первом прочтении пропустить этот параграф, то необходимо вернуться к нему и изучить изложенный там материал) можно определить условную функцию распределения случайной величины ξ при условии, что случайная величина η приняла определенное значение y . Поскольку условная функция распределения обладает всеми свойствами обычной (безусловной) функции распределения, то по ней можно определить

математическое ожидание, которое естественно назвать *условным математическим ожиданием*. Для простоты изложения ограничимся здесь только случаями дискретной и непрерывной двумерных случайных величин (ξ, η).

Начнем со случая дискретной случайной величины (ξ, η) . Пусть случайная величина ξ может принимать только значения X_1, \dots, X_n , а случайная величина η — только значения Y_1, \dots, Y_m . При каждом j рассмотрим условные вероятности $\pi_{ij} = P\{\xi = X_i \mid \eta = Y_j\} = P\{\xi = X_i, \eta = Y_j\}/P\{\eta = Y_j\} = p_{ij}/p_{\eta j}$ случайной величине ξ принять значение X_i при условии $\eta = Y_j$ (см. параграф 5 гл. 6). Назовем значением *условного математического ожидания* случайной величины ξ при условии $\eta = Y_j$ число

$$\mathbf{M}(\xi \mid \eta = Y_j) = \sum_{i=1}^n X_i \pi_{ij}.$$

По аналогии с (безусловным) математическим ожиданием значение условного математического ожидания при условии $\eta = Y_j$ описывает «среднее» значение случайной величины ξ , но только при условии, что случайная величина η приняла значение Y_j .

Из приведенного определения видно, что значение $\mathbf{M}(\xi | \eta = Y_j)$ условного математического ожидания зависит только от значения Y_j случайной величины η . Поэтому само *условное математическое ожидание случайной величины ξ относительно случайной величины η* естественно определить как функцию $\mathbf{M}(\xi | \eta) = g(\eta)$ от случайной величины, т.е. тоже как случайную величину. Область определения функции $g(y)$ совпадает со значениями Y_1, \dots, Y_m случайной величины η , а каждому значению Y_j аргумента y ставится в соответствие число $g(Y_j) = \mathbf{M}(\xi | \eta = Y_j)$.

Пример 24. Пусть μ_1 и μ_2 — числа успехов в первом и втором испытаниях Бернулли с вероятностью успеха p . Найдем $M(\mu_1 | \mu_2)$. Воспользовавшись табл. 5 из гл. 6, имеем

$$\mathbf{M}(\mu_1 \mid \mu_2 = 0) = 0 \cdot q + 1 \cdot p = p,$$

Пример 25. Найдем условное математическое ожидание $M(\xi | \eta)$ случайной величины ξ — числа очков, выпавших на верхней грани игральной кости, относительно случайной величины η — числа очков, выпавших на нижней грани (см. пример 13 из гл. 6). В соответствии с приведенной там же табл. 6, получаем:

$$\begin{aligned} \mathbf{M}(\xi \mid \eta = 1) &= 1 \cdot 0 + 2 \cdot 0 + 3 \cdot 0 + 4 \cdot 0 + 5 \cdot 0 + 6 \cdot 1 = 6, \\ \mathbf{M}(\xi \mid \eta = 2) &= 1 \cdot 0 + 2 \cdot 0 + 3 \cdot 0 + 4 \cdot 0 + 5 \cdot 1 + 6 \cdot 0 = 5, \\ &\vdots \\ \mathbf{M}(\xi \mid \eta = 6) &= 1 \cdot 1 + 2 \cdot 0 + 3 \cdot 0 + 4 \cdot 0 + 5 \cdot 0 + 6 \cdot 0 = 1. \end{aligned}$$

Условное математическое ожидание $M(\xi | \eta)$ обладает следующими свойствами:

1. $\mathbf{M}(c | \eta) \equiv c.$
2. $\mathbf{M}(a\xi + b | \eta) = a\mathbf{M}(\xi | \eta) + b.$
3. $\mathbf{M}(\xi_1 + \xi_2 | \eta) = \mathbf{M}(\xi_1 | \eta) + \mathbf{M}(\xi_2 | \eta).$

Они аналогичны свойствам безусловного математического ожидания (разумеется, арифметические действия понимаются теперь уже не как действия над числами, а как действия над функциями, определенными для всех значений случайной величины η).

Свойство

4. $\mathbf{M}(\xi_1 \cdot \xi_2 | \eta) = \mathbf{M}(\xi_1 | \eta)\mathbf{M}(\xi_2 | \eta)$

также имеет место, но при этом требование независимости случайных величин ξ_1 и ξ_2 нужно заменить требованием, которое называется *условной независимостью случайных величин ξ_1 и ξ_2 при условии случайной величины η* .

Кроме того, справедливы дополнительные свойства:

5. $\mathbf{M}\xi = \mathbf{M}[\mathbf{M}(\xi | \eta)].$
6. $\mathbf{M}[f(\xi) \cdot h(\eta) | \eta] = h(\eta)\mathbf{M}[f(\xi) | \eta]$, где $f(\xi)$ и $h(\eta)$ — (произвольные) функции от случайных величин ξ и η .
7. $\mathbf{M}(\xi | \eta) \equiv \mathbf{M}\xi$, если ξ и η независимы.

Докажем последние три свойства.

Действительно, из определений математического ожидания и условного математического ожидания имеем

$$\begin{aligned} \mathbf{M}[\mathbf{M}(\xi | \eta)] &= \sum_{j=1}^m \mathbf{M}(\xi | \eta = Y_j) p_{\eta j} = \sum_{j=1}^m p_{\eta j} \sum_{i=1}^n X_i \pi_{ij} = \\ &= \sum_{j=1}^m p_{\eta j} \sum_{i=1}^n X_i \frac{p_{ij}}{p_{\eta j}} = \sum_{i=1}^n \sum_{j=1}^m X_i p_{ij} = \mathbf{M}\xi, \end{aligned}$$

что доказывает свойство 5.

Далее, случайная величина $f(\xi) \cdot h(\eta)$ принимает значение $f(X_i) \times h(Y_j)$, когда ξ принимает значение X_i и η — значение Y_j , и, следовательно, для каждого j

$$\begin{aligned} \mathbf{M}[f(\xi) \cdot h(\eta) | \eta = Y_j] &= \sum_{i=1}^n f(X_i) \cdot h(Y_j) \pi_{ij} = \\ &= h(Y_j) \sum_{i=1}^n f(X_i) \pi_{ij} = h(Y_j) \mathbf{M}[f(\xi) | \eta = Y_j], \end{aligned}$$

откуда вытекает свойство 6.

Наконец, используя условие независимости случайных величин ξ и η , выраженное в терминах условного распределения (см. параграф 6 гл. 6), находим

$$\mathbf{M}(\xi \mid \eta = Y_j) = \sum_{i=1}^n X_i \pi_{ij} = \sum_{i=1}^n X_i p_{\xi i} = \mathbf{M}\xi,$$

откуда следует справедливость свойства 7.

Пример 26. Еще раз вычислим $\mathbf{M}(\mu_1 \mid \mu_2)$ (см. пример 24), но теперь уже воспользуемся свойством 7 условного математического ожидания. Тогда, поскольку μ_1 и μ_2 независимы, то $\mathbf{M}(\mu_1 \mid \mu_2) \equiv \mathbf{M}\mu_1 = p$. \square

Пример 27. Снова обратимся к примеру 25. Так как сумма очков на противоположных гранях игральной кости равна 7, то $\xi = 7 - \eta$. Представим ξ в виде $\xi = 1 \cdot (7 - \eta)$. Воспользовавшись теперь свойством 6, в котором положено $f(x) = 1$, $h(y) = 7 - y$, получаем

$$\mathbf{M}(\xi \mid \eta) = \mathbf{M}[1 \cdot (7 - \eta) \mid \eta] = (7 - \eta)\mathbf{M}(1 \mid \eta) = 7 - \eta,$$

т. е. мы пришли к тому же результату, что и ранее, но практически без вычислений. \square

В случае непрерывной двумерной случайной величины (ξ, η) значение условного математического ожидания случайной величины ξ при условии $\eta = y$ определяется формулой

$$\mathbf{M}(\xi \mid \eta = y) = \int_{-\infty}^{\infty} x p_{\xi}(x \mid \eta = y) dx,$$

где $p_{\xi}(x \mid \eta = y) = p(x, y)/p_{\eta}(y)$ — условная плотность распределения случайной величины ξ при условии $\eta = y$. И в этом случае условное математическое ожидание $\mathbf{M}(\xi \mid \eta)$ случайной величины ξ относительно случайной величины η определяется как функция $g(\eta) = \mathbf{M}(\xi \mid \eta)$ от случайной величины η , принимающая значение $g(y) = \mathbf{M}(\xi \mid \eta = y)$ при $\eta = y$. Читателю советуем самостоятельно проверить, что свойства условного математического ожидания, выведенные для дискретного случая, остаются справедливыми и в непрерывном.

Пример 28. Пусть (ξ, η) — двумерная нормальная случайная величина (пример 16 из гл. 6). Найдем условное математическое ожидание $\mathbf{M}(\xi \mid \eta)$. Тогда, как было показано в том же примере, условное распределение ξ при условии $\eta = y$ является нормальным со средним значением $m_1 + \rho\sigma_1(y - m_2)/\sigma_2$ и, согласно определению, $\mathbf{M}(\xi \mid \eta) = m_1 + \rho\sigma_1(\eta - m_2)/\sigma_2$. \square

Резюмируя вышеизложенное, можно сказать, что зависимость поведения «в среднем» случайной величины ξ от значения случайной величины η характеризуется функцией $g(y) = \mathbf{M}(\xi \mid \eta = y)$. Функция $g(y)$ называется также *функцией регрессии* или просто *регрессией случайной величины ξ на случайную величину η* , а ее график — *линией регрессии (случайной величины) ξ на (случайную величину) η* . Линия регрессии дает наглядное изображение зависимости «в среднем» случайной величины ξ от значения случайной величины η .

Пример 29. Регрессия $g(y) = \mathbf{M}(\xi \mid \eta = y)$ случайной величины ξ на случайную величину η для двумерной нормальной случайной величины (ξ, η) (см. пример 28) является линейной функцией $a + by$, где $a = m_1 - m_2 \rho \sigma_1 / \sigma_2$, а $b = \rho \sigma_1 / \sigma_2$. Очевидно, линия регрессии ξ на η представляет собой прямую. \square

6. Другие числовые характеристики случайных величин

В этом параграфе мы дадим краткое описание некоторых других применяемых на практике числовых характеристик случайных величин. Отметим, что эти характеристики, как и все остальные, рассматриваемые в настоящей главе, по сути дела являются *характеристиками распределений* случайных величин.

Асимметрией γ_1 случайной величины ξ называется отношение третьего центрального момента \hat{m}_3 к кубу среднего квадратичного отклонения σ : $\gamma_1 = \hat{m}_3 / \sigma^3$. Нетрудно видеть, что (при условии существования третьего момента) для симметрично распределенной относительно математического ожидания ($\mathbf{P}\{\xi < \mathbf{M}\xi - x\} = \mathbf{P}\{\xi > \mathbf{M}\xi + x\}$ для любого x) случайной величины ξ асимметрия равна нулю.

Эксцессом γ_2 случайной величины ξ называется отношение четвертого центрального момента \hat{m}_4 к квадрату дисперсии за вычетом числа 3: $\gamma_2 = \hat{m}_4 / \sigma^4 - 3$.

Ясно, что асимметрия и эксцесс являются безразмерными величинами.

Пример 30. Вычислим асимметрию и эксцесс нормального закона. По определению,

$$\begin{aligned}\hat{m}_3 &= \int_{-\infty}^{\infty} (x - m)^3 \varphi_{m,\sigma}(x) dx = \int_{-\infty}^{\infty} \frac{(x - m)^3}{\sigma \sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} dx, \\ \hat{m}_4 &= \int_{-\infty}^{\infty} (x - m)^4 \varphi_{m,\sigma}(x) dx = \int_{-\infty}^{\infty} \frac{(x - m)^4}{\sigma \sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} dx.\end{aligned}$$

Делая замену $y = x - m$, имеем

$$\hat{m}_3 = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{\infty} y^3 e^{-\frac{y^2}{2\sigma^2}} dy,$$

откуда в силу нечетности подынтегральной функции следует, что $\hat{m}_3 = 0$ и асимметрия $\gamma_1 = 0$.

Для того чтобы найти \hat{m}_4 , применим формулу интегрирования по частям. Полагая $u = (x - m)^3 / \sqrt{2\pi}$ и $dv = (x - m)e^{-(x-m)^2/(2\sigma^2)} dx / \sigma$, имеем

$$\hat{m}_4 = 3\sigma^2 \int_{-\infty}^{\infty} \frac{(x - m)^2}{\sigma \sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} dx.$$

Воспользовавшись теперь результатом примера 17, окончательно получаем, что $\hat{m}_4 = 3\sigma^4$ и, следовательно, эксцесс $\gamma_2 = 0$.

Таким образом, для нормального закона асимметрия и эксцесс равны нулю. В математической статистике асимметрия и эксцесс обычно служат для первой проверки распределения случайной величины на нормальность. \square

α -квантилью Q_α ($0 < \alpha < 1$) случайной величины ξ называется число, удовлетворяющее неравенствам $P\{\xi < Q_\alpha\} \leq \alpha$ и $P\{\xi > Q_\alpha\} \leq 1 - \alpha$. Квантили находят самое широкое применение в математической статистике при построении доверительных интервалов и проверке статистических гипотез. $1/2$ -квантиль называется также *медианой* M случайной величины ξ .

Пример 31. Найдем α -квантиль экспоненциального распределения. В этом случае Q_α представляет собой решение уравнения $F(Q_\alpha) = \alpha$, т. е. уравнения $1 - e^{-\lambda Q_\alpha} = \alpha$ (рис. 1). Поэтому $Q_\alpha = -\ln(1 - \alpha)/\lambda$. Ясно, что медиана экспоненциального распределения $M = \ln 2/\lambda$. Если трактовать экспоненциальное распределение как распределение времени распада атома (см. параграф 4 гл. 5), то медиана представляет собой период полураспада. \square

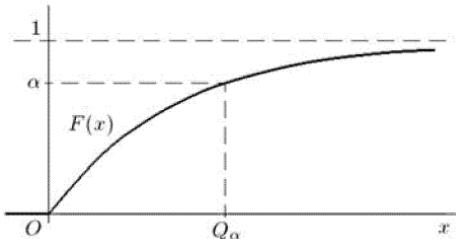


Рис. 1

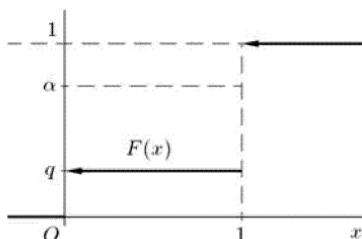


Рис. 2

Пример 32. Пусть случайная величина μ представляет собой число успехов в одном испытании Бернулли с вероятностью успеха p . Тогда (рис. 2) $Q_\alpha = 0$ при $0 < \alpha < q$, $Q_\alpha = 1$ при $q < \alpha < 1$, а q -квантилью является любое число от 0 до 1.

Этот пример показывает, что, во-первых, квантили могут совпадать для разных α , во-вторых, для некоторых α соответствующие квантили могут определяться неоднозначно. \square

Модой непрерывной случайной величины называется точка (локального) максимума плотности распределения $p(x)$. Различают *унимодальные* (имеющие одну моду), *бимодальные* (имеющие две моды) и *мультимодальные* (имеющие несколько мод) распределения.

Для определения моды дискретной случайной величины предположим сначала, что ее значения X_1, \dots, X_n расположены в порядке возрастания. Тогда *модой дискретной случайной величины* называется такое значение X_i , что $p_{i-1} < p_i$ и $p_{i+1} < p_i$. И в дискретном случае распределения могут быть *унимодальными*, *бимодальными* и *мультимодальными*.

Наивероятнейшим значением называется мода, доставляющая глобальный максимум вероятности (дискретной случайной величины) или

плотности распределения (непрерывной случайной величины). Если распределение унимодально, то мода также будет наивероятнейшим значением.

Мода и наивероятнейшее значение введены скорее для наглядности, чем для каких-то практических целей.

Пример 33. Плотность нормального распределения имеет единственный максимум в точке t . Поэтому мода нормального закона совпадает с математическим ожиданием. Она же является наивероятнейшим значением и медианой. \square

Пример 34. Найдем моду биномиального распределения. Для этого заметим, что

$$\frac{P_n(i)}{P_n(i+1)} = \frac{\binom{n}{i} p^i q^{n-i}}{\binom{n}{i+1} p^{i+1} q^{n-(i+1)}} = \frac{(i+1)q}{(n-i)p}.$$

Отсюда нетрудно вывести, что отношение $P_n(i)/P_n(i+1)$ меньше 1 при $i < np - q$ и больше 1 при $i > np - q$. Таким образом, если $np - q$ не является целым, то максимальное i , для которого $i > np - q$, является модой и наивероятнейшим значением. Если же $np - q$ — целое, то биномиальный закон имеет две моды и два наивероятнейших значения: $np - q$ и $np - q + 1$. \square

Энтропия $H = H(\xi)$ дискретной случайной величины ξ определяется формулой

$$H = H(\xi) = - \sum_{i=1}^n p_i \ln p_i.$$

Отметим, что энтропия не зависит от значений X_i случайной величины ξ , а зависит только от вероятностей p_i , с которыми эти значения принимаются. Энтропия является мерой априорной неопределенности случайной величины. Максимальное значение $H_{\max} = \ln n$ энтропия дискретной случайной величины достигает тогда, когда все n возможных значений случайная величина принимает с одной и той же вероятностью $p_i = 1/n$, минимальное $H_{\min} = 0$ — когда случайная величина принимает единственное значение с вероятностью, равной единице.

Энтропия $H(\xi, \eta)$ двумерной дискретной случайной величины (ξ, η) определяется формулой

$$H(\xi, \eta) = - \sum_{i,j} p_{ij} \ln p_{ij}.$$

Поскольку для независимых случайных величин $\ln p_{ij} = \ln p_{\xi j} + \ln p_{\eta j}$, то, как нетрудно видеть, энтропия случайной величины (ξ, η) с независимыми компонентами ξ и η представляет собой сумму энтропий: $H(\xi, \eta) = H(\xi) + H(\eta)$; в случае зависимости ξ и η энтропия $H(\xi, \eta)$ всегда меньше суммы $H(\xi) + H(\eta)$.

Энтропия играет важную роль в теории информации, она в некотором смысле представляет собой минимальный объем памяти, необходимый для записи информации, содержащейся в случайной величине.

Поскольку информация записывается обычно в двоичной системе, то основанием логарифма берется число 2.

Энтропия $H = H(\xi)$ непрерывной случайной величины ξ и энтропия $H(\xi, \eta)$ двумерной непрерывной случайной величины (ξ, η) задаются выражениями

$$H = H(\xi) = - \int_{-\infty}^{\infty} p_{\xi}(x) \ln p_{\xi}(x) dx,$$

$$H(\xi, \eta) = - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_{\xi, \eta}(x, y) \ln p_{\xi, \eta}(x, y) dx dy.$$

И в непрерывном случае энтропия $H(\xi, \eta)$ двумерной случайной величины совпадает с суммой $H(\xi) + H(\eta)$ энтропий компонент тогда и только тогда, когда ξ и η независимы; иначе $H(\xi, \eta) < H(\xi) + H(\eta)$. Однако в отличие, например, от математического ожидания энтропию непрерывной случайной величины нельзя получить предельным переходом от дискретного случая. Отметим также, что при заданной дисперсии σ^2 максимальную энтропию $\ln \sqrt{2\pi e \sigma^2}$ имеет нормально распределенная случайная величина.

Глава 8

ПРЕДЕЛЬНЫЕ ТЕОРЕМЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ

На практике довольно часто встречается ситуация, когда многократно наблюдается, как сказал бы не знакомый с теорией вероятностей человек, «одна и та же» случайная величина, а конечный результат представляет собой сумму наблюденных в каждом испытании значений этих величин. Наряду с уже упоминавшимися азартными играми сюда можно отнести: повторные замеры одного и того же параметра с последующим осреднением результатов с целью повышения точности измерений; многократное воздействие однородных причин на некоторый протекающий во времени физический процесс и т. д. Пройденный нами путь по теории вероятностей и приобретенный при этом опыт применения основных ее понятий позволяют подойти к описанию всех этих явлений с единых вероятностных позиций на основе следующей схемы: имеется последовательность независимых одинаково распределенных случайных величин и из нее образуется среднее арифметическое первых n членов. Спрашивается, как будет вести себя это среднее арифметическое, если n велико? Оказывается, при большом n оно теряет свойство случайности и приближается к математическому ожиданию каждого слагаемого (отсюда второе название математического ожидания — среднее значение). Этот факт носит название закона больших чисел. Собственно говоря, с частным случаем закона больших чисел мы уже встречались, когда рассматривали схему Бернулли (теорема Бернулли). Следуя исторической традиции, доказательство закона больших чисел мы сначала проведем, опираясь на неравенство Чебышева, которое является родоначальником многих других неравенств, широко применяемых в современной теории вероятностей.

Дальнейшее уточнение закона больших чисел происходило в двух направлениях. Первое связано с динамикой поведения средних арифметических. К основным результатам этого направления следует отнести усиленный закон больших чисел и закон повторного логарифма, полученные А. Н. Колмогоровым. Исходным пунктом второго направления, называемого иногда центральной предельной проблемой, являются теоремы Муавра—Лапласа. Решение центральной предельной проблемы позволило описать класс всех распределений, которые могут выступать в качестве предельных для функций распределения сумм независимых случайных величин в том случае, когда вкладом каждого слагаемого можно пренебречь, найти необходимые и достаточные условия сходимости к каждому распределению этого класса, оценить скорость сходимости. Простейшим примером результатов такого типа является центральная предельная теорема для одинаково распределенных слагаемых, имеющих дисперсию, которую мы докажем в параграфе 4. Основным математическим аппаратом при решении центральной предельной проблемы служит преобразование Фурье, носящее в теории вероятностей название характеристической функции, хотя в последнее время все чаще применяются другие методы.

1. Неравенство Чебышева. Закон больших чисел

Рассмотрим случайную величину ξ , имеющую дисперсию $D\xi = \sigma^2$. Мы уже говорили, что дисперсия является показателем разброса ξ вокруг математического ожидания $M\xi$. Однако с точки зрения исследователя разброс естественнее характеризовать вероятностью $P\{|\xi - M\xi| \geq \varepsilon\}$ отклонения случайной величины ξ от $M\xi$ на величину, большую некоторого заданного ε . Следующее неравенство позволяет оценить эту вероятность через дисперсию σ^2 .

Неравенство Чебышева. Для каждой случайной величины ξ , имеющей дисперсию $D\xi = \sigma^2$, при любом $\varepsilon > 0$ справедливо неравенство

$$P\{|\xi - M\xi| \geq \varepsilon\} \leq \frac{\sigma^2}{\varepsilon^2}.$$

Доказательство проведем для непрерывной случайной величины ξ с плотностью распределения $p(x)$. По определению,

$$D\xi = \int_{-\infty}^{\infty} (x - M\xi)^2 p(x) dx.$$

Поскольку подынтегральное выражение неотрицательно, то при уменьшении области интегрирования интеграл может только уменьшиться, поэтому

$$D\xi \geq \int_{-\infty}^{M\xi - \varepsilon} (x - M\xi)^2 p(x) dx + \int_{M\xi + \varepsilon}^{\infty} (x - M\xi)^2 p(x) dx - \int_{|x - M\xi| \geq \varepsilon} (x - M\xi)^2 p(x) dx.$$

Учитывая теперь, что $(x - M\xi)^2 \geq \varepsilon^2$, если $|x - M\xi| \geq \varepsilon$, получаем

$$\int_{|x - M\xi| \geq \varepsilon} (x - M\xi)^2 p(x) dx \geq \varepsilon^2 \int_{|x - M\xi| \geq \varepsilon} p(x) dx.$$

Остается заметить, что последний интеграл представляет собой вероятность события $|\xi - M\xi| \geq \varepsilon$ и, значит,

$$D\xi \geq \varepsilon^2 P\{|\xi - M\xi| \geq \varepsilon\},$$

откуда и вытекает неравенство Чебышева.

Аналогично неравенство Чебышева доказывается и для дискретного случая, при этом нужно только заменить интеграл на сумму. \square

Ясно, что применять неравенство Чебышева имеет смысл только тогда, когда $\varepsilon > \sigma$; в противном случае оно дает тривиальную оценку.

Пример 1. Пусть случайная величина ξ имеет плотность распределения $p(x) = e^{-|x|}/2$. Тогда $M\xi = 0$, $\sigma^2 = D\xi = 2$. Воспользовавшись неравенством Чебышева, оценим $p_\varepsilon = P\{|\xi| \geq \varepsilon\}$ для $\varepsilon = 2; 5; 10$. В результате получим

$$p_2 \leq 0,5, \quad p_5 \leq 0,08, \quad p_{10} \leq 0,02.$$

С другой стороны, поскольку $p_\varepsilon = 1 - F(\varepsilon) + F(-\varepsilon) = e^{-\varepsilon}$, имеем

$$p_2 = \frac{1}{e^2} \cong 0,1353, \quad p_5 = \frac{1}{e^5} \cong 0,0067, \quad p_{10} = \frac{1}{e^{10}} \cong 0,000\,045.$$

Таким образом, в этом примере неравенство Чебышева дает очень грубую оценку вероятности p_ε . \square

Пример 2. Пусть случайная величина ξ принимает только два значения 1 и -1 с одинаковыми вероятностями $1/2$. Тогда $M\xi = 0$, $D\xi = 1$. Применяя неравенство Чебышева, получаем $p_1 = P\{|\xi| \geq 1\} \leq 1$. С другой стороны, поскольку оба возможных значения ξ равны по модулю единице, то $p_1 = 1$.

Этот пример показывает, что если не делать никаких дополнительных предположений относительно случайной величины ξ , то неравенство Чебышева дает неулучшаемую оценку p_ε . \square

Рассмотрим теперь последовательность $\xi_1, \xi_2, \dots, \xi_n, \dots$ независимых одинаково распределенных случайных величин (так как случайные величины ξ_i одинаково распределены, то все их числовые характеристики, в частности математические ожидания и дисперсии, равны между собой). Скажем, что эта *последовательность удовлетворяет (слабому) закону больших чисел, если для некоторого a и любого $\varepsilon > 0$*

$$P\left\{\left|\frac{1}{n} \sum_{i=1}^n \xi_i - a\right| \geq \varepsilon\right\} \xrightarrow{n \rightarrow \infty} 0.$$

Иными словами, выполнение закона больших чисел отражает предельную устойчивость средних арифметических случайных величин: при большом числе испытаний они практически перестают быть случайными и с большой степенью достоверности могут быть предсказаны.

Иногда вместо выражения «последовательность $\xi_1, \xi_2, \dots, \xi_n, \dots$ удовлетворяет закону больших чисел» говорят «*среднее арифметическое случайных величин $\xi_1, \xi_2, \dots, \xi_n, \dots$ сходится по вероятности к некоторой предельной постоянной a* ».

Теорема (закон больших чисел). *Если последовательность $\xi_1, \xi_2, \dots, \xi_n, \dots$ независимых одинаково распределенных случайных величин такова, что существуют $M\xi_n = m$ и $D\xi_n = \sigma^2$, то для любого $\varepsilon > 0$*

$$P\left\{\left|\frac{1}{n} \sum_{i=1}^n \xi_i - m\right| > \varepsilon\right\} \xrightarrow{n \rightarrow \infty} 0.$$

Доказательство является элементарным следствием неравенства Чебышева. Действительно, по свойствам математического ожидания и дисперсии

$$M\left(\frac{1}{n} \sum_{i=1}^n \xi_i\right) = m, \quad D\left(\frac{1}{n} \sum_{i=1}^n \xi_i\right) = \frac{\sigma^2}{n}.$$

Воспользовавшись теперь неравенством Чебышева, получаем, что для любого $\varepsilon > 0$

$$P\left\{\left|\frac{1}{n} \sum_{i=1}^n \xi_i - m\right| \geq \varepsilon\right\} \leq \frac{\sigma^2}{n\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0. \quad \square$$

Таким образом, мы показали, что для последовательности $\xi_1, \xi_2, \dots, \xi_n, \dots$ выполняется закон больших чисел, причем постоянная a совпадает с математическим ожиданием $M\xi_n = m$.

Пример 3. Пусть μ_i — число успехов в i -м испытании Бернулли (см. пример 2 в гл. 6). Тогда $f_n = \sum_{i=1}^n \mu_i/n$ представляет собой частоту успехов в n испытаниях. Как мы знаем, $M\mu_i = p$ и $D\mu_i = pq$. Таким образом, в силу доказанной теоремы

$$P\{|f_n - p| \geq \varepsilon\} \xrightarrow[n \rightarrow \infty]{} 0.$$

Но последнее соотношение есть не что иное, как теорема Бернулли (см. параграф 5 гл. 4). \square

Замечание. Вообще говоря, существование дисперсии является достаточным, но не необходимым условием для выполнения закона больших чисел. В дальнейшем (пример 15) будет показано, что достаточным условием является просто существование математического ожидания, которое в этом случае выступает в качестве предельной постоянной a . Более того, существуют последовательности независимых одинаково распределенных случайных величин, даже не имеющих математического ожидания, но тем не менее удовлетворяющих закону больших чисел.

2. Усиленный закон больших чисел. Закон повторного логарифма

Пусть по-прежнему $\xi_1, \xi_2, \dots, \xi_n, \dots$ — последовательность независимых одинаково распределенных случайных величин, к которой применим закон больших чисел, т. е. при больших n среднее арифметическое $\eta_n = \sum_{i=1}^n \xi_i/n$ «почти совпадает» с некоторой постоянной a . Однако если мы будем последовательно наблюдать случайные величины $\eta_1, \eta_2, \dots, \eta_n, \dots$, то закон больших чисел *еще не гарантирует*, что η_n будет стремиться к a для любого элементарного исхода ω .

Пример 4. Рассмотрим схему Бернулли с равными вероятностями успеха и неудачи $p = q = 1/2$, в которой производится бесконечное число испытаний. Тогда последовательность $\mu_1, \mu_2, \dots, \mu_n, \dots$, где μ_i — число успехов в i -м испытании, будет представлять собой простейший вариант последовательности независимых одинаково распределенных случайных величин, для которой в соответствии с теоремой Бернулли выполнен закон больших чисел и при этом $a = M\mu_i = p = 1/2$.

Пространство элементарных исходов Ω состоит из всевозможных (бесконечных) последовательностей УНН...УН... В отличие от случая конечного числа испытаний Ω уже не будет дискретным (более того, Ω «почти» эквивалентно отрезку $[0, 1]$ с равномерной вероятностью на нем; для доказательства этого достаточно отождествить последовательность $\mu_1, \mu_2, \dots, \mu_n, \dots$ с двоичным представлением некоторого числа, заключенного между нулем и единицей), и каждый элементарный исход ω имеет вероятность $P(\omega) = 1/2 \cdots 1/2 \times \dots = 0$. Возьмем элементарный исход $\omega_0 = \text{УУ...У}$. Ясно, что для него $\eta_1(\omega_0) = \eta_2(\omega_0) = \dots = 1$, т. е. средние арифметические равны единице и не могут стремиться к $a = 1/2$. Читатель без труда может привести примеры и дру-

гих элементарных исходов, для которых последовательность $\eta_1, \eta_2, \dots, \eta_n, \dots$ либо будет сходиться к отличному от $1/2$ числу, либо вообще не будет сходиться. \square

Из приведенного выше примера видно, что могут существовать элементарные исходы, для которых $\eta_n(\omega)$ не сходится к a . Выделим эти элементарные исходы в отдельное событие A . Хотелось бы, чтобы при выполнении закона больших чисел вероятность события A равнялась нулю; в этом случае говорят, что для последовательности $\xi_1, \xi_2, \dots, \xi_n, \dots$ выполнен *усиленный закон больших чисел*. Но это не всегда так. Если заранее не предполагать существование математического ожидания $M\xi_i$, то можно привести примеры последовательностей независимых одинаково распределенных случайных величин, для которых выполнен закон больших чисел, но не выполнен усиленный закон больших чисел (более подробно об этом см. ниже, пример 15). Условие выполнения усиленного закона больших чисел содержится в следующей теореме, доказанной А. Н. Колмогоровым.

Теорема (усиленный закон больших чисел). *Существование математического ожидания является необходимым и достаточным условием выполнения усиленного закона больших чисел для последовательности независимых одинаково распределенных случайных величин $\xi_1, \xi_2, \dots, \xi_n, \dots$. Постоянная a в этом случае совпадает с математическим ожиданием $M\xi_i$.*

Доказательство теоремы Колмогорова мы здесь не приводим.

Отметим, что выполнение усиленного закона больших чисел естественно влечет за собой выполнение закона больших чисел.

Если теперь перейти от средних арифметических к накопленным суммам $S_n = \xi_1 + \dots + \xi_n$, то усиленный закон больших чисел гласит, что последовательность $S_n - nt$ для почти всех элементарных исходов при любом $\varepsilon > 0$ может находиться выше уровня $n\varepsilon$ или ниже уровня $-n\varepsilon$ не более конечного числа раз. Зададим теперь вопрос: нельзя ли указать более точные границы изменения последовательности $S_n - nt$? Естественно, если мы хотим больше получить, то мы должны также больше потребовать от случайных величин ξ_i . А именно, будем предполагать, что ξ_i имеют математическое ожидание $M\xi_i = t$ и дисперсию $D\xi_i = \sigma^2$. Для простоты изложения предположим, что $\sigma^2 = 1$ и $t = 0$. Более того, будем считать, что абсолютные значения случайных величин ξ_i ограничены, т. е. существует такое число X , что $|\xi_i(\omega)| < X$ для всех элементарных исходов ω ¹⁾. Из последнего предположения, в частности следует, что случайные величины ξ_i имеют не только математическое ожидание и дисперсию, но и моменты всех порядков.

¹⁾ Можно потребовать выполнение менее ограничительных условий, однако они более сложно формулируются.

Для произвольного ε обозначим через B'_ε , B''_ε и B_ε события, состоящие соответственно из тех элементарных исходов ω , для которых $S_n(\omega) > (1 + \varepsilon)\sqrt{2n \ln \ln n}$ бесконечное число раз, $S_n(\omega) < -(1 + \varepsilon)\sqrt{2n \ln \ln n}$ бесконечное число раз и $|S_n(\omega)| > (1 + \varepsilon)\sqrt{2n \ln \ln n}$ бесконечное число раз.

Теорема (закон повторного логарифма). $P(B_\varepsilon) = P(B'_\varepsilon) = P(B''_\varepsilon) = 0$ для любых $\varepsilon > 0$ и $P(B_\varepsilon) = P(B'_\varepsilon) = P(B''_\varepsilon) = 1$ для любых $\varepsilon < 0$.

Доказательство теоремы опирается на очень тонкие (так называемые показательные) вероятностные неравенства и также здесь не приводятся.

Смысл закона повторного логарифма заключается в следующем. Для любого сколь угодно малого $\varepsilon > 0$ накопленные суммы S_n , начиная с некоторого n , будут лежать в пределах от $-(1 + \varepsilon)\sqrt{2n \ln \ln n}$ до $(1 + \varepsilon)\sqrt{2n \ln \ln n}$. В то же время суммы S_n будут бесконечное число раз выходить за границы области, находящейся между кривыми $-(1 - \varepsilon)\sqrt{2n \ln \ln n}$ и $(1 - \varepsilon)\sqrt{2n \ln \ln n}$, причем как вверх, так и вниз (рис. 1).

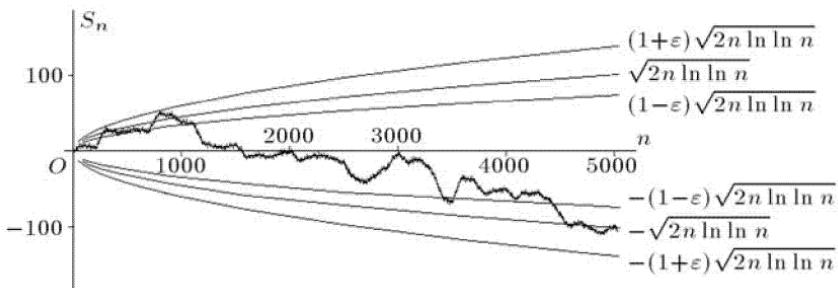


Рис. 1

Для читателя, знакомого с понятием верхнего и нижнего пределов (\limsup и \liminf) последовательности, можно предложить следующую формулировку закона повторного логарифма: для почти всех элементарных исходов ω

$$\limsup \frac{S_n(\omega)}{\sqrt{2n \ln \ln n}} = 1$$

и

$$\liminf \frac{S_n(\omega)}{\sqrt{2n \ln \ln n}} = -1.$$

Отметим, что само название «закон повторного логарифма» происходит от выражения $\ln \ln n$.

3. Характеристическая функция

Для дальнейших исследований нам понадобится понятие характеристической функции. *Характеристической функцией* $f(t) = f_\xi(t)$ случайной величины ξ называется математическое ожидание случайной величины $e^{it\xi}$, где $i = \sqrt{-1}$ — мнимая единица, а t — произвольное (действительное) число. Здесь мы имеем дело с комплексной случайной величиной, которая определяется так же, как и действительная, с той лишь разницей, что каждому элементарному исходу ставится в соответствие комплексное число, а не действительное. Используя общее правило вычисления математического ожидания и формулу Эйлера, получаем

$$f(t) = \mathbf{M}e^{it\xi} = \sum_j e^{itX_j} p_j = \sum_j p_j \cos(tX_j) + i \sum_j p_j \sin(tX_j)$$

для дискретной величины ξ и

$$f(t) = \mathbf{M}e^{it\xi} = \int_{-\infty}^{\infty} e^{itx} p(x) dx = \int_{-\infty}^{\infty} \cos(tx) p(x) dx + i \int_{-\infty}^{\infty} \sin(tx) p(x) dx$$

для непрерывной. Поскольку $|e^{itx}| = 1$, то характеристическая функция существует при всех (действительных) t для каждой случайной величины. Отметим, что характеристическая функция определяется не собственно случайной величиной, а ее функцией распределения, т. е. по существу характеризует именно распределение случайной величины. Читатель, знакомый с преобразованием Фурье, сразу же заметит, что характеристическая функция непрерывной случайной величины отличается от преобразования Фурье плотности распределения этой случайной величины только лишь отсутствием множителя $1/\sqrt{2\pi}$, что, как будет видно из дальнейшего, представляет определенное удобство при действиях над случайными величинами.

Пример 5. Найдем характеристическую функцию случайной величины μ , распределенной по биномциальному закону. Поскольку μ — дискретная случайная величина, принимающая значения $0, 1, \dots, n$, то

$$f(t) = \sum_{j=0}^n e^{itj} \binom{n}{j} p^j q^{n-j} = \sum_{j=0}^n \binom{n}{j} (pe^{it})^j q^{n-j} = (q + pe^{it})^n. \quad \square$$

Пример 6. Характеристическая функция случайной величины ξ , распределенной по экспоненциальному закону, имеет вид

$$f(t) = \int_0^{\infty} e^{itx} \lambda e^{-\lambda x} dx = \lambda \int_0^{\infty} e^{(it-\lambda)x} dx = \frac{\lambda}{\lambda - it}. \quad \square$$

Пример 7. Пусть случайная величина ξ распределена по стандартному нормальному закону. Тогда

$$f(t) = \int_{-\infty}^{\infty} e^{itx} \varphi(x) dx = \int_{-\infty}^{\infty} e^{itx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

Делая замену $y = x - it$, получаем

$$f(t) = \int_{-\infty-it}^{+\infty-it} \frac{1}{\sqrt{2\pi}} e^{-(y^2+t^2)/2} dy = \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \int_{-\infty-it}^{+\infty-it} e^{-y^2/2} dy.$$

Из теории функций комплексного переменного известно, что

$$\int_{-\infty-it}^{+\infty-it} e^{-y^2/2} dy = \sqrt{2\pi}.$$

Поэтому окончательно получаем

$$f(t) = e^{-t^2/2}. \quad \square$$

Выведем некоторые почти очевидные свойства характеристических функций.

1. $f(0) = 1$, $f(t)$ — непрерывная функция.

В самом деле (рассмотрим, например, дискретный случай)

$$f(0) = \sum_j e^{it0} p_j = \sum_j p_j = 1.$$

Непрерывность $f(t)$ следует из непрерывности функции e^{itX_j} и абсолютной сходимости ряда $\sum_j p_j$.

2. Если $\eta = a\xi + b$, то $f_\eta(t) = f_\xi(at)e^{ibt}$.

Действительно, из определения характеристической функции и свойств математического ожидания вытекает, что

$$f_\eta(t) = \mathbf{M}e^{it\eta} = \mathbf{M}e^{it(a\xi+b)} = e^{ibt}\mathbf{M}e^{iat\xi} = e^{ibt}f_\xi(at).$$

Пусть ξ_1 и ξ_2 — независимые случайные величины и $\eta = \xi_1 + \xi_2$. Тогда

3. $f_\eta(t) = f_{\xi_1}(t)f_{\xi_2}(t)$.

В самом деле, поскольку ξ_1 и ξ_2 независимы, то независимы случайные величины $e^{it\xi_1}$ и $e^{it\xi_2}$. Отсюда следует, что

$$\mathbf{M}e^{it\eta} = \mathbf{M}e^{it(\xi_1+\xi_2)} = \mathbf{M}(e^{it\xi_1}e^{it\xi_2}) = \mathbf{M}e^{it\xi_1}\mathbf{M}e^{it\xi_2}.$$

Именно свойство 3 является тем основным свойством, благодаря которому характеристические функции нашли такое широкое применение в теории вероятностей. При суммировании независимых случайных величин их плотности распределения преобразуются по формуле свертки.

Но формула свертки весьма неудобна для исследования, гораздо проще заменить ее простым перемножением характеристических функций.

Если случайная величина ξ имеет момент n -го порядка m_n , то характеристическая функция ξ дифференцируема n раз, причем для $k \leq n$

$$\mathbf{4.} f^{(k)}(0) = i^k m_k.$$

В самом деле, формальное дифференцирование характеристической функции k раз ($k \leq n$) дает (например, в непрерывном случае)

$$f^{(k)}(t) = i^k \int_{-\infty}^{\infty} x^k e^{itx} p(x) dx.$$

Законность дифференцирования определяется тем фактом, что

$$\left| \int_{-\infty}^{\infty} x^k e^{itx} p(x) dx \right| \leq \int_{-\infty}^{\infty} |x^k| p(x) dx,$$

и существованием момента n -го порядка. Заметим, что при четном n справедливо и обратное: если характеристическая функция имеет производную $f^{(n)}(0)$, то существуют моменты m_k всех порядков k до n -го включительно и $m_k = i^{-k} f^{(k)}(0)$.

Пример 8. Как мы знаем, если случайная величина ξ распределена по стандартному нормальному закону, то случайная величина $\eta = \sigma\xi + t$ распределена по нормальному закону с параметрами t и σ . Тогда характеристические функции $f_{m,\sigma}(t)$ и $f(t)$ случайных величин η и ξ связаны по свойству 2 соотношением

$$f_{m,\sigma}(t) = e^{itm} f(\sigma t)$$

или, если учесть результат примера 7,

$$f_{m,\sigma}(t) = e^{itm - \sigma^2 t^2 / 2}. \quad \square$$

Пример 9. Вычислим момент n -го порядка случайной величины ξ , распределенной по экспоненциальному закону. Воспользовавшись свойством 4 и результатом примера 6, получаем

$$m_n = i^{-n} f^{(n)}(0) = i^{-n} \left(\frac{\lambda}{\lambda - it} \right)^{(n)} \Big|_{t=0} = i^{-n} i^n \frac{\lambda n!}{\lambda^{n+1}} = \frac{n!}{\lambda^n}. \quad \square$$

Пример 10. Найдем еще раз характеристическую функцию числа успехов μ в n испытаниях, однако в отличие от примера 5 воспользуемся тем, что $\mu = \mu_1 + \dots + \mu_n$, где μ_j — число успехов в j -м испытании. Тогда

$$f_{\mu_j}(t) = q e^{it0} + p e^{it1} = q + p e^{it}$$

и, значит, по свойству 3

$$f_{\mu}(t) = f_{\mu_1}(t) \dots f_{\mu_n}(t) = (q + p e^{it})^n. \quad \square$$

Важнейшей особенностью характеристической функции $f(t)$ является тот факт, что она однозначно определяет функцию распределения $F(x)$. А именно, справедлива следующая формула.

Формула обращения. Для любых точек непрерывности x_1 и x_2 функции распределения $F(x)$ приращение

$$F(x_2) - F(x_1) = \frac{1}{2\pi} \lim_{T \rightarrow \infty} \int_{-T}^T \frac{e^{-itx_1} - e^{-itx_2}}{it} f(t) dt.$$

Отметим, что формула обращения справедлива и в точках разрыва $F(x)$, если считать, что в этих точках $F(x) = [F(x+0) - F(x-0)]/2$.

Не давая строгого математического доказательства, покажем, что по своей сути формула обращения представляет собой разновидность обратного преобразования Фурье. Действительно, формально применяя обратное преобразование Фурье к характеристической функции $f(t)$, получаем следующее выражение для плотности распределения $p(x)$:

$$p(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} f(t) dt.$$

Далее, вспоминая соотношение между плотностью распределения $p(x)$ и функцией распределения $F(x)$, имеем

$$F(x_2) - F(x_1) = \int_{x_1}^{x_2} p(x) dx = \frac{1}{2\pi} \int_{x_1}^{x_2} dx \int_{-\infty}^{\infty} e^{-itx} f(t) dt.$$

Еще раз производя формальную операцию — перестановку интегралов, получаем

$$F(x_2) - F(x_1) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t) dt \int_{x_1}^{x_2} e^{-itx} dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-itx_1} - e^{-itx_2}}{it} f(t) dt.$$

Хотя каждая из произведенных формальных операций, вообще говоря, математически необоснована (в частности, дискретные случайные величины вообще не имеют плотности распределения), как это часто бывает, конечный результат верен, если только понимать последний интеграл в том смысле, как написано в формуле обращения (в смысле главного значения).

Определение характеристической функции вместе с формулой обращения устанавливают *взаимно однозначное соответствие между функцией распределения и характеристической функцией*.

Пример 11. Пусть ξ_1 и ξ_2 — независимые случайные величины, распределенные по нормальному закону с параметрами m_1, σ_1 и m_2, σ_2 соответственно. Рассмотрим случайную величину $\eta = \xi_1 + \xi_2$. Тогда, как было показано в примере 8, $f_{\xi_1}(t) = e^{im_1 t - \sigma_1^2 t^2/2}$, $f_{\xi_2}(t) = e^{im_2 t - \sigma_2^2 t^2/2}$ и по свойству 3 характеристических функций получаем, что $f_{\eta}(t) = e^{i(m_1+m_2)t - (\sigma_1^2 + \sigma_2^2)t^2/2}$. Но ха-

рактеристическую функцию $f_\eta(t)$ имеет случайная величина, распределенная по нормальному закону с параметрами $m_1 + m_2$, $\sqrt{\sigma_1^2 + \sigma_2^2}$. Поэтому в силу взаимно однозначного соответствия между функцией распределения и характеристической функцией случайная величина η также распределена нормально (с параметрами $m_1 + m_2$, $\sqrt{\sigma_1^2 + \sigma_2^2}$). \square

Пример 12. Рассмотрим независимые случайные величины ξ_1 и ξ_2 , распределенные по закону Пуассона с параметрами λ_1 и λ_2 . Их характеристические функции задаются формулами

$$f_{\xi_1}(t) = \sum_{n=0}^{\infty} e^{itn} \frac{\lambda_1^n}{n!} e^{-\lambda_1} = e^{-\lambda_1} \sum_{n=0}^{\infty} \frac{(\lambda_1 e^{it})^n}{n!} = e^{\lambda_1(e^{it}-1)},$$

$$f_{\xi_2}(t) = e^{\lambda_2(e^{it}-1)}.$$

Пусть $\eta = \xi_1 + \xi_2$. Тогда $f_\eta(t) = e^{(\lambda_1+\lambda_2)(e^{it}-1)}$ и опять-таки в силу взаимно однозначного соответствия между функцией распределения и характеристической функцией случайная величина η распределена по закону Пуассона с параметром $\lambda_1 + \lambda_2$. \square

Основную роль при решении центральной предельной проблемы играет теорема о связи сходимости последовательности функций распределения со сходимостью последовательности соответствующих характеристических функций, или теорема непрерывности. Однако прежде чем перейти к формулировке этой теоремы, скажем несколько слов вообще о сходимости функций распределения. Будем говорить, что *последовательность функций распределения $F_1(x), \dots, F_n(x), \dots$ сходится к предельной функции распределения $F(x)$, если $F_n(x) \xrightarrow{n \rightarrow \infty} F(x)$ для любых x , являющихся точками непрерывности $F(x)$.* Такая сходимость называется *слабой сходимостью функций распределения* (обозначается $F_n(x) \xrightarrow{n \rightarrow \infty} F(x)$). Слабая сходимость является наиболее естественной для функций распределения, и в дальнейшем мы будем рассматривать только ее.

К определению слабой сходимости можно сделать несколько замечаний.

Во-первых, из слабой сходимости последовательности функций распределения еще нельзя сделать вывод о какой-либо сходимости последовательности самих случайных величин, так как даже одинаково распределенные случайные величины могут быть заданы на совершенно разных вероятностных пространствах.

Во-вторых, требование сходимости в любой точке непрерывности $F(x)$ нельзя заменить более сильным требованием сходимости во всех точках x . Это подтверждает следующий пример.

Пример 13. Пусть на одном и том же вероятностном пространстве (Ω, \mathcal{B}, P) задана последовательность случайных величин $\xi_1, \xi_2, \dots, \xi_n, \dots$, причем каждая случайная величина ξ_n принимает всего одно значение $-1/n$. Тогда последовательность $\xi_1, \xi_2, \dots, \xi_n, \dots$ будет сходиться к случайной величине

$\xi \equiv 0$ для любого элементарного исхода ω (причем даже равномерно). Тем не менее, $F_{\xi_n}(0) = 1$ при всех n , но $F_\xi(0) = 0$. \square

Приведенный пример показывает, что $F_{\xi_n}(0)$ не стремится к $F_\xi(0)$, хотя естественно было бы ожидать сходимости $F_{\xi_n}(x)$ к $F_\xi(x)$ в любой точке x , поскольку $\xi_n(\omega) \xrightarrow{n \rightarrow \infty} \xi(\omega)$ при всех элементарных исходах ω .

Разгадка этого парадокса заключается в том, что 0 является точкой разрыва $F_\xi(x)$, а при определении слабой сходимости функций распределения сходимости в таких точках мы не требовали.

Наконец, если последовательность функций распределения $F_1(x), F_2(x), \dots, F_n(x), \dots$ сходится к некоторой функции $F(x)$ в каждой точке непрерывности последней, то это не гарантирует слабой сходимости, поскольку $F(x)$ может вообще не быть функцией распределения.

Пример 14. Пусть $\xi_n \equiv n$ для всех ω . Тогда $F_{\xi_n}(x) \xrightarrow{n \rightarrow \infty} F(x) \equiv 0$ при каждом x . Но $F(x)$ не является функцией распределения, так как $F(\infty) = 0 \neq 1$. \square

Значит, при определении слабой сходимости обязательно нужно требовать, чтобы предельная функция являлась функцией распределения.

Теперь мы можем привести (без доказательства) формулировку теоремы непрерывности.

Теорема непрерывности. Для того чтобы последовательность функций распределения $F_1(x), F_2(x), \dots, F_n(x), \dots$ слабо сходилась к функции распределения $F(x)$, необходимо и достаточно, чтобы последовательность характеристических функций $f_1(t), f_2(t), \dots, f_n(t), \dots$ сходилась к характеристической функции $f(t)$ равномерно на любом отрезке $[-T, T]$.

Теорема непрерывности является тем краеугольным камнем, который позволяет свести задачу изучения предельного поведения распределений сумм независимых случайных величин к задаче изучения предельного поведения характеристических функций этих сумм. Отметим, что в разных учебниках приведены различные эквивалентные формулировки теоремы непрерывности; данная здесь формулировка наиболее естественна для наших дальнейших рассуждений.

Пример 15. Пусть $\xi_1, \xi_2, \dots, \xi_n, \dots$ — последовательность независимых одинаково распределенных случайных величин, имеющих характеристические функции $f_{\xi_n}(t) = f_1(t)$. Рассмотрим последовательность средних арифметических $\eta_n = \sum_{i=1}^n \xi_i/n$ и найдем в терминах характеристических функций необходимое и достаточное условие для выполнения (слабого) закона больших чисел. Используя определение, можно показать, что закон больших чисел эквивалентен слабой сходимости последовательности функций распределения $F_{\eta_1}(x), F_{\eta_2}(x), \dots, F_{\eta_n}(x), \dots$ к предельной $F(x)$, представляющей собой функцию распределения случайной величины η , принимающей единственное значение a . Но η имеет характеристическую функцию $f(t) = e^{iat}$, и в силу

теоремы непрерывности необходимым и достаточным условием для выполнения закона больших чисел является сходимость $f_{\eta_n}(t)$ к $f(t)$ равномерно на любом интервале $[-T, T]$. В свою очередь, используя свойства характеристических функций, имеем

$$f_{\eta_n}(t) = \left[f_1\left(\frac{t}{n}\right) \right]^n.$$

Логарифмируя $f_{\eta_n}(t)$ (это можно сделать, поскольку $f_{\eta_n}(t) \xrightarrow[n \rightarrow \infty]{} f(t)$ равномерно на любом интервале $[-T, T]$, а $f(t)$ не обращается в нуль), получаем следующее необходимое и достаточное условие: $n \ln f_1(t/n) \xrightarrow[n \rightarrow \infty]{} iat$ равномерно на любом интервале $[-T, T]$. Последнее условие, в свою очередь, эквивалентно условию

$$\ln f_1\left(\frac{t}{n}\right) \approx iat \frac{t}{n}.$$

Разлагая логарифм в ряд Маклорена до первого члена, получаем

$$\ln f_1\left(\frac{t}{n}\right) = \ln \left(1 - \left[1 - f_1\left(\frac{t}{n}\right)\right]\right) \approx f_1\left(\frac{t}{n}\right) - 1 = f_1\left(\frac{t}{n}\right) - f_1(0) \approx iat \frac{t}{n}.$$

Наконец, полагая $h = t/n$, окончательно находим, что *необходимым и достаточным условием для выполнения закона больших чисел является существование предела*

$$\lim_{n \rightarrow \infty} \frac{f_1(h) - f_1(0)}{h} = ia$$

или, иными словами, *существование в нуле производной $f'_1(0)$ характеристической функции $f_1(t)$* , причем постоянная a , фигурирующая в определении закона больших чисел, вычисляется через эту производную по формуле $a = -if'_1(0)$.

Таким образом, с использованием аппарата характеристических функций нам удалось чрезвычайно просто найти необходимое и достаточное условие выполнения (слабого) закона больших чисел для последовательности независимых одинаково распределенных случайных величин $\xi_1, \xi_2, \dots, \xi_n, \dots$

Скажем еще несколько слов о связи слабого и усиленного законов больших чисел. Из свойств характеристических функций следует, что если случайная величина ξ_n имеет математическое ожидание $M\xi_n = a$, то производная $f'_1(0)$ существует и равна ia . Значит, для последовательности $\xi_1, \xi_2, \dots, \xi_n, \dots$ выполнен слабый закон больших чисел и не выполнен усиленный тогда и только тогда, когда $M\xi_n$ не существует, но тем не менее существует производная $f'_1(0) = ia$. \square

В заключение этого параграфа отметим, что наряду с характеристическими функциями в теории вероятностей используются также преобразования Лапласа–Стилтьеса $\tilde{f}(s) = M e^{-s\xi}$ (для неотрицательных случайных величин ξ) и производящие функции $f^*(z) = M z^\xi$ (для неотрицательных целочисленных случайных величин ξ). Ясно, что преобразование Лапласа–Стилтьеса $\tilde{f}(s)$ и производящая функция $f^*(z)$ связаны с характеристической функцией той же самой случайной величины соотношениями: $f(t) = f(-it)$, $f(t) = f^*(e^{it})$. Преобразование Лапласа–Стилтьеса и производящая функция, по сути дела, имеют те же самые свойства, что и характеристическая функция, но с ними существенно проще обращаться уже хотя бы потому, что они являются действительными функциями.

4. Центральная предельная теорема

Рассмотрим последовательность независимых одинаково распределенных случайных величин $\xi_1, \xi_2, \dots, \xi_n, \dots$, имеющих математическое ожидание $M\xi_n = m$. Предположим также, что существует дисперсия $D\xi_n = \sigma^2$. Закон больших чисел для этой последовательности можно представить в следующей форме:

$$\frac{1}{n} \sum_{i=1}^n (\xi_i - M\xi_i) = \frac{1}{n} (S_n - nm) \xrightarrow{n \rightarrow \infty} 0,$$

где сходимость можно понимать как в смысле сходимости по вероятности (слабый закон больших чисел), так и в смысле сходимости с вероятностью, равной единице (усиленный закон больших чисел).

Однако сразу же возникает вопрос: поскольку от случайных величин ξ_n мы потребовали существования не только математического ожидания, но и дисперсии, то нельзя ли получить более «тонкую» предельную теорему, позволяющую точнее описать предельное поведение распределений центрированных сумм $S_n - nm$? Такая теорема имеется и носит название центральной предельной теоремы. При этом нормировка $S_n - nm$ осуществляется величиной \sqrt{n} , а не n , т. е. последовательностью постоянных, стремящейся к бесконечности медленнее, чем n .

Центральная предельная теорема. Пусть $\xi_1, \xi_2, \dots, \xi_n, \dots$ — последовательность независимых одинаково распределенных случайных величин, $M\xi_n = m$, $D\xi_n = \sigma^2$. Тогда

$$P\left\{ \frac{S_n - nm}{\sqrt{n}\sigma} < x \right\} \xrightarrow{n \rightarrow \infty} \Phi(x)$$

(напомним, что $\Phi(x)$ — функция стандартного нормального распределения).

Доказательство. Прежде всего заметим, что поскольку функция стандартного нормального распределения $\Phi(x)$ является непрерывной, то сходимость к ней последовательности функций распределения в каждой точке представляет собой слабую сходимость и, значит, для доказательства центральной предельной теоремы можно воспользоваться теоремой непрерывности. Обозначим через $f_1(t)$ характеристическую функцию случайных величин ξ_n , а через $g_n(t)$ — характеристическую функцию случайной величины $(S_n - nm)/\sqrt{n}\sigma$. Воспользовавшись теперь свойствами 2 и 3 характеристических функций, имеем

$$g_n(t) = \left[f_1 \left(\frac{t}{\sqrt{n}\sigma} \right) e^{-\frac{int}{\sqrt{n}\sigma^2}} \right]^n.$$

Поскольку $f_1(t)$ имеет производные первых двух порядков (свойство 4 характеристических функций), то $\ln f_1(t/\sqrt{n}\sigma^2)$ можно равномерно на любом отрезке $[-T, T]$ разложить в ряд Маклорена по степеням $1/\sqrt{n}$ до второго члена:

$$\begin{aligned} \ln f_1\left(\frac{t}{\sqrt{n\sigma^2}}\right) &= \ln f_1(0) + \frac{t}{\sqrt{n\sigma^2}} \frac{f'_1(0)}{f_1(0)} + \frac{t^2}{2n\sigma^2} \left(\frac{f''_1(0)}{f_1(0)} - \frac{(f'_1(0))^2}{(f_1(0))^2}\right) + o\left(\frac{1}{n}\right) = \\ &= \frac{imt}{\sqrt{n\sigma^2}} - \frac{\sigma^2 t^2}{2n\sigma^2} + o\left(\frac{1}{n}\right) \end{aligned}$$

и, значит,

$$\ln g_n(t) \approx -\frac{t^2}{2}, \quad g_n(t) \xrightarrow[n \rightarrow \infty]{} e^{-t^2/2}.$$

Но $e^{-t^2/2}$ есть не что иное, как характеристическая функция $f(t)$ стандартного нормального распределения (см. пример 7). Тем самым завершается доказательство теоремы. \square

Центральная предельная теорема выявляет ту роль, которую играет нормальное распределение. Оно обычно возникает в явлениях, подверженных большому количеству малых случайных воздействий. Уже само название «нормальный закон» объясняется тем широким распространением, которое он находит на практике в самых различных областях научных исследований.

Пример 16. Рассмотрим n испытаний по схеме Бернулли с вероятностью успеха p и вероятностью неудачи $q = 1 - p$ в каждом. Пусть μ_j — число успехов в j -м испытании. Тогда $\mathbf{M}\mu_j = p$, $\mathbf{D}\mu_j = pq$. Обозначим через $S_n = \mu_1 + \dots + \mu_n$ суммарное число успехов в n испытаниях. В силу центральной предельной теоремы с ростом n распределения случайных величин $(S_n - np)/\sqrt{npq}$ сходятся к стандартному нормальному закону, т. е.

$$\mathbf{P}\left\{\frac{S_n - np}{\sqrt{npq}} < x\right\} \xrightarrow[n \rightarrow \infty]{} \Phi(x).$$

Это утверждение представляет собой не что иное, как интегральную теорему Муавра—Лапласа (см. параграф 3 гл. 4). \square

Пример 17. Для определения скорости v движения объекта выполняются n измерений v_1, \dots, v_n , причем j -е измерение производится со случайной ошибкой ξ_j (т. е. $v_j = v + \xi_j$). Предполагая, что ошибки измерений ξ_j независимы и одинаково распределены с математическим ожиданием $\mathbf{M}\xi_j = 0$ (отсутствуют систематические ошибки наблюдений) и дисперсией $\mathbf{D}\xi_j = \sigma^2$, оценим вероятность того, что средняя наблюденная скорость $v_{\text{ср}} = (v_1 + \dots + v_n)/n$ будет отличаться от истинной скорости v не более чем на ε . Тогда

$$\begin{aligned} \mathbf{P}\{|v_{\text{ср}} - v| < \varepsilon\} &= \mathbf{P}\left\{-\varepsilon < \frac{v_1 + \dots + v_n - nv}{n} < \varepsilon\right\} = \\ &= \mathbf{P}\left\{-\varepsilon \sqrt{\frac{n}{\sigma^2}} < \frac{v_1 + \dots + v_n - nv}{\sqrt{n\sigma^2}} < \varepsilon \sqrt{\frac{n}{\sigma^2}}\right\}. \end{aligned}$$

Считая теперь, что число n измерений велико, воспользуемся центральной предельной теоремой, по которой случайная величина $(v_1 + \dots + v_n - nv)/\sqrt{n\sigma^2}$ распределена приближенно по стандартномуциальному закону. Значит,

$$\mathbf{P}\{|v_{\text{ср}} - v| < \varepsilon\} \approx \Phi\left(\varepsilon \sqrt{\frac{n}{\sigma^2}}\right) - \Phi\left(-\varepsilon \sqrt{\frac{n}{\sigma^2}}\right) = 2\Phi_0\left(\varepsilon \sqrt{\frac{n}{\sigma^2}}\right),$$

где значение интеграла Лапласа $\Phi_0(x)$ приведено в табл. 3 приложения. \square

Список литературы

1. Гмурман В.Е. Теория вероятностей и математическая статистика. — М.: Высшая школа, 1972.
2. Вентцель Е.С. Теория вероятностей. — М.: Наука, 1969.
3. Пугачев В.С. Теория вероятностей и математическая статистика. — М.: Наука, 1979.
4. Розанов Ю.А. Лекции по теории вероятностей. — М.: Наука, 1986.
5. Башарин Г.П. Введение в теорию вероятностей. — М.: Изд-во УДН, 1990.
6. Чистяков В.П. Курс теории вероятностей. — М.: Наука, 1987.
7. Севастьянов Б.А. Курс теории вероятностей и математической статистики. — М.: Наука, 1982.
8. Гнеденко Б.В. Курс теории вероятностей. — М.: Наука, 1988.
9. Климов Г.П. Теория вероятностей и математическая статистика. — М.: Изд-во МГУ, 1983.
10. Боровков А.А. Курс теории вероятностей. — М.: Наука, 1976.
11. Прохоров Ю.В., Розанов Ю.А. Теория вероятностей. — М.: Наука, 1985.
12. Феллер В. Введение в теорию вероятностей и ее приложения. Т. 1, 2. — М.: Мир, 1984.
13. Лоэв М. Теория вероятностей. — М.: ИЛ, 1962.
14. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. — М.: Высшая школа, 1979.
15. Сборник задач по теории вероятностей, математической статистике и теории случайных функций / Под ред. А. А. Свешникова. — М.: Наука, 1970.
16. Агапов Г.И. Задачник по теории вероятностей. — М.: Высшая школа, 1986.
17. Мешалкин Л.Д. Сборник задач по теории вероятностей. — М.: Изд-во МГУ, 1963.

Часть II

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Глава 1

ОБЩИЕ СВЕДЕНИЯ

Так же как и теория вероятностей, математическая статистика имеет свои ключевые понятия, к которым относятся: генеральная совокупность, теоретическая функция распределения, выборка, эмпирическая функция распределения, статистика. Именно с определения этих понятий, а также с установления связи между ними и объектами, изучаемыми в теории вероятностей, мы начнем изложение математической статистики, предварительно дав краткое описание задач, которые собираемся решать. Кроме того, в последнем параграфе главы остановимся на некоторых распределениях, наиболее часто встречающихся в математической статистике.

1. Задачи математической статистики

Математическая статистика, являясь частью общей прикладной математической дисциплины «Теория вероятностей и математическая статистика», изучает, как и теория вероятностей, случайные явления, использует одинаковые с ней определения, понятия и методы и основана на той же самой аксиоматике А. Н. Колмогорова.

Однако задачи, решаемые математической статистикой, носят специфический характер. Теория вероятностей исследует явления, заданные *полностью* их моделью, и выявляет еще до опыта те статистические закономерности, которые будут иметь место после его проведения. В математической статистике вероятностная модель явления определена *с точностью до неизвестных параметров*. Отсутствие сведений о параметрах компенсируется тем, что нам позволено проводить «пробные» испытания и на их основе восстанавливать недостающую информацию.

Попытаемся показать различие этих двух взаимосвязанных дисциплин на простейшем примере — последовательности независимых одинаковых испытаний, или схеме Бернулли (часть 1, гл. 4). Схему Бернулли можно трактовать как подбрасывание несимметричной монеты с вероятностью выпадения «герба» (успеха) p и «цифры» (неудачи) $q = 1 - p$. В теории вероятностей p и q задаются «извне» (например, для симметричной монеты $p = q = 1/2$). Методы теории вероятностей позволяют, зная p и q , определить вероятность выпадения m «гербов» при n подбрасываниях монеты (биномиальное распределение, часть 1, гл. 4, параграф 1), найти асимптотику этой вероятности при увеличении числа подбрасываний (теоремы Пуассона и Муавра–Лапласа,

часть 1, гл. 4, параграфы 2–4) и т. д. В математической статистике значения p и q неизвестны заранее, но мы можем произвести серию подбрасываний монеты. Цель проведения испытаний как раз и заключается либо в определении p и q , либо в проверке некоторых априорных суждений относительно их значений. Таким образом, судя уже по этому простейшему примеру, задачи математической статистики являются в некотором смысле обратными задачам теории вероятностей.

В математической статистике обычно принято выделять два основных направления исследования.

Первое направление связано с *оценкой неизвестных параметров*. Возвращаясь к нашему примеру, предположим, что мы произвели n подбрасываний монеты и установили, что в μ из них выпал «герб». Тогда наиболее естественной оценкой вероятности p является наблюденная частота $p^* = \mu/n$. Как известно из закона больших чисел Бернулли (часть 1, гл. 4, параграф 5), с увеличением числа испытаний частота p^* стремится к вероятности p , т. е. p^* является *состоятельной* оценкой вероятности p . Оказывается, наряду с простотой и естественностью оценка p^* будет и наилучшей с многих точек зрения, т. е. она обладает свойством *эффективности*. Однако если нам заранее определено число n подбрасываний монеты, то сказать со 100%-й гарантией что-либо об истинном значении p мы не можем (за исключением разве что тривиальных суждений типа «если выпадет хотя бы один „герб“ то вероятность выпадения „герба“ не может равняться нулю»). Поэтому наряду с *точечными* оценками в математической статистике принято определять *интервальные* оценки или, иными словами, *доверительные интервалы*, опираясь при этом на «уровень доверия», или *доверительную вероятность*.

Второе направление в математической статистике связано с *проверкой* некоторых априорных предположений, или *статистических гипотез*. Так, до опыта мы можем предположить, что монета симметрична, т. е. высказать гипотезу о равенстве $p = q = 1/2$. Противоположное предположение, естественно, будет состоять в том, что $p \neq 1/2$, и тоже представляет собой гипотезу. Принято называть одну из этих гипотез (как правило, более важную с практической точки зрения) *основной* H_0 , а вторую — *альтернативной* или *конкурирующей* H_1 . В приведенном выше примере нужно проверить основную гипотезу $H_0: p = 1/2$ против конкурирующей гипотезы $H_1: p \neq 1/2$. Заметим, что в нашем случае основная гипотеза H_0 полностью определяет вероятностную модель подбрасывания монеты, т. е. является *простой* (состоит из одной точки), в отличие от конкурирующей гипотезы H_1 , являющейся *сложной* (состоит из более чем одной точки). Задача проверки статистических гипотез состоит в выборе правила или *критерия*, позволяющего по результатам наблюдений проверить (по возможности, наилучшим образом) справедливость этих гипотез и принять одну из них. Так же, как и при оценке неизвестных параметров, мы не застрахованы от неверного решения; в математической статистике они подразделяются

на ошибки первого и второго рода. Ошибка первого рода состоит в том, что мы принимаем конкурирующую гипотезу H_1 , в то время как справедлива основная гипотеза H_0 ; аналогично определяется ошибка второго рода. Возвращаясь к примеру с монетой, приведем следующий критерий проверки двух перечисленных гипотез: основную гипотезу H_0 ($p = 1/2$) будем принимать в том случае, если наблюденная частота p^* удовлетворяет неравенству $|p^* - 1/2| < \varepsilon$; в противном случае считаем верной конкурирующую гипотезу H_1 . Вероятность ошибки первого рода (принять симметричную монету за несимметричную) в этом случае определяется как вероятность выполнения неравенства $|p^* - 1/2| \geq \varepsilon$ в схеме Бернулли с равновероятными исходами. Вероятность ошибки второго рода (принять несимметричную монету за симметричную) также определяется из схемы Бернулли, но с неравновероятными исходами и будет зависеть от истинного значения p .

Далее мы увидим, что задача проверки статистических гипотез наиболее полно решается для случая двух простых гипотез. Можно поставить и задачу проверки нескольких гипотез (в примере с монетой можно взять, например, три гипотезы: $H_0: p = 1/2$, $H_1: p < 1/2$, $H_2: p > 1/2$), однако мы такие задачи рассматривать не будем.

Условно математическую статистику можно подразделить на исследование байесовских и небайесовских моделей.

Байесовские модели возникают тогда, когда неизвестный параметр является случайной величиной и имеется априорная информация о его распределении. При байесовском подходе на основе опытных данных априорные вероятности пересчитываются в апостериорные. Применение байесовского подхода фактически сводится к использованию формулы Байеса (см. часть 1, гл. 3, параграф 5), откуда, собственно говоря, и пошло его название. Байесовский подход нами будет применяться только как вспомогательный аппарат при доказательстве некоторых теорем.

Небайесовские модели появляются тогда, когда неизвестный параметр нельзя считать случайной величиной и все статистические выводы приходится делать, опираясь только на результаты «пробных» испытаний. Именно такие модели мы будем рассматривать в дальнейшем изложении.

В заключение этого параграфа отметим, что в математической статистике употребляют также понятия параметрических и непараметрических моделей. *Параметрические модели* возникают тогда, когда нам известна с точностью до параметра (скалярного или векторного) функция распределения наблюдаемой характеристики и необходимо по результатам испытаний определить этот параметр (задача оценки неизвестного параметра) или проверить гипотезу о принадлежности его некоторому заранее выделенному множеству значений (задача проверки статистических гипотез). Все приведенные выше примеры с подбрасыванием монеты представляют собой параметрические модели. Примеры *непараметрических моделей* мы рассмотрим позже.

2. Основные понятия математической статистики

Основными понятиями математической статистики являются: генеральная совокупность, выборка, теоретическая функция распределения.

Генеральная совокупность. Будем предполагать, что у нас имеются N объектов, каждому из которых присуще определенное значение некоторой числовой характеристики X . Характеристика X , вообще говоря, может быть и векторной (например, линейные размеры объекта), однако для простоты изложения мы ограничимся только скалярным случаем, тем более что переход к векторному случаю никаких трудностей не вызывает. Совокупность этих N объектов назовем *генеральной совокупностью*.

Поскольку все наши статистические выводы мы будем делать, основываясь только на значениях числовой характеристики X , естественно абстрагироваться от физической природы самих объектов и отождествить каждый объект с присущей ему характеристикой X . Таким образом, с точки зрения математической статистики генеральная совокупность представляет собой N чисел, среди которых, конечно, могут быть и одинаковые.

Выборка. Для того чтобы установить параметры генеральной совокупности, нам позволено произвести некоторое число n испытаний. Каждое испытание состоит в том, что мы случайным образом выбираем один объект генеральной совокупности и определяем его значение X . Полученный таким образом ряд чисел X_1, \dots, X_n будем называть (*случайной*) *выборкой* объема n , а число X_i — i -м *элементом выборки*.

Заметим, что сам процесс выбора можно осуществлять различными способами: выбрав объект и определив его значение, изымать этот объект и не допускать к последующим испытаниям (выборка без возвращения); после определения его значения объект возвращается в генеральную совокупность и может полноценно участвовать в дальнейших испытаниях (выборка с возвращением) и т. д.

Разумеется, если бы мы смогли провести сплошное обследование всех объектов генеральной совокупности, то не нужно было бы применять никакие статистические методы и саму математическую статистику можно было бы отнести к чисто теоретическим наукам. Однако такой полный контроль невозможен по следующим причинам. Во-первых, часто испытание сопровождается разрушением испытуемого объекта; в этом случае мы имеем выборку без возвращения. Во-вторых, обычно необходимо исследовать весьма большое количество объектов, что просто невозможно физически. Наконец, может возникнуть такое положение, когда многократно измеряется один и тот же объект, но каждый замер производится со случайной ошибкой, и цель последующей статистической обработки заключается именно в уточнении характеристик объекта на основе многократных наблюдений; при этом

результат каждого наблюдения надо считать новым объектом генеральной совокупности (простейшим примером такой ситуации является многократное подбрасывание монеты с целью определения вероятности выпадения «герба»). Следует помнить также, что выборка обязательно должна удовлетворять условию репрезентативности или, говоря более простым языком, давать обоснованное представление о генеральной совокупности.

С ростом объема N генеральной совокупности исчезает различие между выборками с возвращением и без возвращения. Мы, как обычно это делается в математической статистике, будем рассматривать случай бесконечно большого объема генеральной совокупности и поэтому, употребляя слово «выборка», не будем указывать, какая она — с возвращением или без него.

Теоретическая функция распределения. Пусть X_1 — выборка единичного объема из заданной генеральной совокупности. Поскольку сам процесс выбора производится случайным образом, то X_1 является случайной величиной и, как всякая случайная величина, имеет функцию распределения $F(x) = \mathbf{P}\{X_1 < x\}$. Нетрудно видеть, что если объем N генеральной совокупности конечен, то при случайному выборе объекта мы находимся в рамках схемы классической вероятности (часть 1, гл. 2, параграф 1) и значение функции распределения $F(x)$ совпадает с отношением N_x/N , где N_x — число тех объектов генеральной совокупности, значения которых меньше x .

В случае выборки X_1, \dots, X_n произвольного объема n каждый элемент X_i выборки также будет иметь функцию распределения $F(x)$, причем для выборки с возвращением наблюдения X_1, \dots, X_n будут независимы между собой (чего нельзя сказать о выборке без возвращения). Поскольку, как уже говорилось, мы будем рассматривать выборки из генеральной совокупности бесконечно большого объема, а в этом случае исчезает различие между выборками разного типа, мы приходим к интерпретации (с точки зрения теории вероятностей) выборки X_1, \dots, X_n как *n независимых одинаково распределенных с функцией распределения F(x) случайных величин* или, допуская некоторую вольность речи, как *n независимых реализаций наблюдаемой случайной величины X, имеющей функцию распределения F(x)*. Функция распределения $F(x)$ называется *теоретической функцией распределения*. Однако теоретическая функция распределения $F(x)$ либо неизвестна, либо известна не полностью, и именно относительно $F(x)$ мы будем делать наши статистические выводы. Заметим, что в соответствии с общими положениями теории вероятностей совместная функция распределения $F_{X_1, \dots, X_n}(x_1, \dots, x_n)$ выборки X_1, \dots, X_n задается формулой

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = \mathbf{P}\{X_1 < x_1, \dots, X_n < x_n\} = F(x_1) \dots F(x_n).$$

В дальнейшем, как правило, мы будем предполагать, что $F(x)$ является функцией распределения либо дискретной, либо непрерывной наблюдаемой случайной величины X . В первом случае будем оперировать рядом распределения случайной величины X , записанным в виде табл. 1, а во втором — плотностью распределения $p(x) = F'(x)$.

Таблица 1

X	b_1	b_2	...	b_L
P	$P(b_1)$	$P(b_2)$...	$P(b_L)$

3. Простейшие статистические преобразования

Прежде чем переходить к детальному анализу наблюденных статистических данных, обычно проводят их предварительную обработку. Иногда результаты такой обработки уже сами по себе дают наглядную картину исследуемого явления, в большинстве же случаев они служат исходным материалом для получения более подробных статистических выводов.

Вариационный и статистический ряды. Часто бывает удобно пользоваться не самой выборкой X_1, \dots, X_n , а некоторой ее модификацией, называемой вариационным рядом. *Вариационный ряд* X_1^*, \dots, X_n^* представляет собой ту же самую выборку X_1, \dots, X_n , но расположенную в порядке возрастания элементов: $X_1^* \leq X_2^* \leq \dots \leq X_n^*$. Такое преобразование не приводит к потере информации относительно теоретической функции распределения $F(x)$, поскольку, переставив элементы вариационного ряда X_1^*, \dots, X_n^* в случайном порядке, мы получим новый набор случайных величин X'_1, \dots, X'_n , совместная функция распределения $F_{X'_1, \dots, X'_n}(x_1, \dots, x_n)$ которых в точности совпадает с функцией распределения $F_{X_1, \dots, X_n}(x_1, \dots, x_n)$ первоначальной выборки X_1, \dots, X_n .

Для X_1^* и X_n^* употребляют название «крайние члены вариационного ряда».

Пример 1. Измерение проекции вектора скорости молекул водорода на одну из осей координат дало (с учетом направления вектора) результаты ($\times 10^4$ м/с), представленные в табл. 2.

Вариационный ряд этой выборки приведен в табл. 3. Крайними членами вариационного ряда X_1^*, \dots, X_n^* являются $X_1^* = 2,33$ и $X_{50}^* = 3,47$. \square

Если среди элементов выборки X_1, \dots, X_n (а значит, и среди элементов вариационного ряда X_1^*, \dots, X_n^*) имеются одинаковые, что происходит при наблюдении дискретной случайной величины, а также довольно часто встречается при наблюдении непрерывной случайной величины с округлением значений, то наряду с вариационным рядом используют представление выборки в виде *статистического*

Таблица 2

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
-1,04	-1,06	1,06	-0,53	-1,58	0,01	0,41	-0,79	-0,18	-0,52
X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}	X_{17}	X_{18}	X_{19}	X_{20}
-1,60	-1,29	-0,10	1,27	0,01	0,60	2,25	-0,88	0,01	0,30
X_{21}	X_{22}	X_{23}	X_{24}	X_{25}	X_{26}	X_{27}	X_{28}	X_{29}	X_{30}
-0,08	0,54	1,02	1,68	1,12	-0,01	2,15	0,96	-0,80	-0,50
X_{31}	X_{32}	X_{33}	X_{34}	X_{35}	X_{36}	X_{37}	X_{38}	X_{39}	X_{40}
-2,33	-0,72	0,14	-0,98	0,74	-1,32	-1,46	0,35	0,32	0,35
X_{41}	X_{42}	X_{43}	X_{44}	X_{45}	X_{46}	X_{47}	X_{48}	X_{49}	X_{50}
-0,05	-0,27	0,65	3,47	2,19	0,40	0,52	-0,28	-1,57	1,92

Таблица 3

X_1^*	X_2^*	X_3^*	X_4^*	X_5^*	X_6^*	X_7^*	X_8^*	X_9^*	X_{10}^*
-2,33	-1,60	-1,58	-1,57	-1,46	-1,32	-1,29	-1,06	-1,04	-0,98
X_{11}^*	X_{12}^*	X_{13}^*	X_{14}^*	X_{15}^*	X_{16}^*	X_{17}^*	X_{18}^*	X_{19}^*	X_{20}^*
-0,90	-0,88	-0,80	-0,79	-0,72	-0,53	-0,52	-0,28	-0,27	-0,18
X_{21}^*	X_{22}^*	X_{23}^*	X_{24}^*	X_{25}^*	X_{26}^*	X_{27}^*	X_{28}^*	X_{29}^*	X_{30}^*
-0,10	-0,08	-0,05	-0,01	0,01	0,01	0,01	0,14	0,30	0,32
X_{31}^*	X_{32}^*	X_{33}^*	X_{34}^*	X_{35}^*	X_{36}^*	X_{37}^*	X_{38}^*	X_{39}^*	X_{40}^*
0,35	0,35	0,40	0,41	0,52	0,54	0,60	0,65	0,74	0,96
X_{41}^*	X_{42}^*	X_{43}^*	X_{44}^*	X_{45}^*	X_{46}^*	X_{47}^*	X_{48}^*	X_{49}^*	X_{50}^*
1,02	1,06	1,12	1,27	1,68	1,92	2,15	2,19	2,25	3,47

ряда (табл. 4), в котором Z_1, \dots, Z_m представляют собой расположенные в порядке возрастания различные значения элементов выборки X_1, \dots, X_n , а n_1, \dots, n_m — числа элементов выборки, значения которых равны соответственно Z_1, \dots, Z_m .

Пример 2. В течение минуты каждую секунду регистрировалось число попавших в счетчик Гейгера частиц. Результаты наблюдений приведены в табл. 5.

Статистический ряд выборки представлен в табл. 6.

□

Таблица 4

Z_1	Z_2	\dots	Z_m
n_1	n_2	\dots	n_m

Таблица 5

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
1	4	3	5	3	5	0	5	0	4
X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}	X_{17}	X_{18}	X_{19}	X_{20}
1	1	4	2	3	1	5	3	1	4
X_{21}	X_{22}	X_{23}	X_{24}	X_{25}	X_{26}	X_{27}	X_{28}	X_{29}	X_{30}
3	0	6	1	2	2	4	2	4	3
X_{31}	X_{32}	X_{33}	X_{34}	X_{35}	X_{36}	X_{37}	X_{38}	X_{39}	X_{40}
3	1	4	3	5	4	4	5	2	1
X_{41}	X_{42}	X_{43}	X_{44}	X_{45}	X_{46}	X_{47}	X_{48}	X_{49}	X_{50}
2	0	4	1	1	4	5	2	3	2
X_{51}	X_{52}	X_{53}	X_{54}	X_{55}	X_{56}	X_{57}	X_{58}	X_{59}	X_{60}
4	5	4	1	3	2	1	3	9	5

Таблица 6

0	1	2	3	4	5	6	9
4	12	9	11	13	9	1	1

Статистики. Для получения обоснованных статистических выводов необходимо проводить достаточно большое число испытаний, т. е. иметь выборку достаточно большого объема n . Ясно, что не только использование такой выборки, но и хранение ее весьма затруднительно. Чтобы избавиться от этих трудностей, а также для других целей, полезно ввести понятие статистики, общее определение которой формулируется следующим образом. Назовем *статистикой* $S = (S_1, \dots, S_k)$ произвольную (измеримую) k -мерную функцию от выборки X_1, \dots, X_n :

$$S_1 = S_1(X_1, \dots, X_n),$$

$$\dots \dots \dots \dots \dots$$

$$S_k = S_k(X_1, \dots, X_n).$$

Как функция от случайного вектора (X_1, \dots, X_n) статистика S также будет случайным вектором (см. часть 1, гл. 6, параграф 7), и ее функция распределения

$$F_{S_1, \dots, S_k}(x_1, \dots, x_k) = \mathbf{P}\{S_1 < x_1, \dots, S_k < x_k\}$$

определяется для дискретной наблюдаемой случайной величины X формулой

$$F_{S_1, \dots, S_k}(x_1, \dots, x_k) = \sum P(y_1) \dots P(y_n)$$

и для непрерывной — формулой

$$F_{S_1, \dots, S_k}(x_1, \dots, x_k) = \int \dots \int p(y_1) \dots p(y_n) dy_1 \dots dy_n,$$

где суммирование или интегрирование производится по всем возможным значениям y_1, \dots, y_n (в дискретном случае каждое y_i принадлежит множеству $\{b_1, \dots, b_L\}$), для которых выполнена система неравенств

$$S_1(y_1, \dots, y_n) < x_1, \dots, S_k(y_1, \dots, y_n) < x_k.$$

Пример 3. Пусть выборка X_1, \dots, X_n произведена из генеральной совокупности с теоретической функцией распределения $F(x) = \Phi(x; m, \sigma^2)$, являющейся нормальной с математическим ожиданием (средним значением) m и дисперсией σ^2 . Рассмотрим двумерную статистику (S_1, S_2) , где

$$\begin{aligned} S_1 &= S_1(X_1, \dots, X_n) = \frac{1}{n} (X_1 + \dots + X_n), \\ S_2 &= S_2(X_1, \dots, X_n) = (X_1 - S_1)^2 + \dots + (X_n - S_1)^2. \end{aligned}$$

Тогда

$$F_{S_1, S_2}(x_1, x_2) = \int \dots \int (2\pi\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2}[(y_1-m)^2 + \dots + (y_n-m)^2]} dy_1 \dots dy_n.$$

$S_1 = (y_1 + \dots + y_n)/n < x_1,$
 $S_2 = (y_1 - S_1)^2 + \dots + (y_n - S_1)^2 < x_2$

Мы, однако, не будем вычислять записанный интеграл, а воспользуемся тем фактом (см. пример 29, часть 1, гл. 6, параграф 7), что любое линейное преобразование переводит нормально распределенный вектор в вектор, снова имеющий нормальное распределение, причем ортогональное преобразование переводит вектор с независимыми координатами, имеющими одинаковые дисперсии, в вектор с также независимыми и имеющими те же самые дисперсии координатами.

Из курса теории вероятностей известно, что статистика S_1 имеет нормальное распределение со средним m и дисперсией σ^2/n . Положим

$$X'_i = X_i - m, \quad S'_1 = \frac{1}{n} (X'_1 + \dots + X'_n).$$

Очевидно, что

$$S_1 = m + S'_1, \quad S_2 = (X'_1 - S'_1)^2 + \dots + (X'_n - S'_1)^2.$$

Пусть теперь A — линейное ортогональное преобразование пространства R^n , ставящее в соответствие каждому вектору $\mathbf{x} = (x_1, \dots, x_n)$ вектор $\mathbf{y} = (y_1, \dots, y_n) = A\mathbf{x}$, где $y_1 = (x_1 + \dots + x_n)/\sqrt{n}$ (как известно из курса линейной алгебры, такое преобразование всегда существует). Тогда, если $\mathbf{X}' = (X'_1, \dots, X'_n)$, то $(Y_1, \dots, Y_n) = \mathbf{Y} = A\mathbf{X}'$ будет нормально распределен

ленным случайным вектором, имеющим независимые координаты Y_i с нулевым средним и дисперсией σ^2 . Кроме того, $Y_1 = \sqrt{n} S'_1$. Далее, рассмотрим $S^2 = \sum_{i=1}^n (X'_i)^2$ — квадрат длины вектора \mathbf{X}' . Простейшие преобразования показывают, что

$$S^2 = \sum_{i=1}^n (X'_i - S'_1)^2 + n(S'_1)^2 = S_2 + n(S'_1)^2.$$

С другой стороны, в силу ортогональности преобразования A

$$S^2 = \sum_{i=1}^n Y_i^2 = n(S'_1)^2 + Y_2^2 + \dots + Y_n^2.$$

Отсюда, в частности, следует, что

$$S_2 = Y_2^2 + \dots + Y_n^2,$$

т. е. S_2/σ^2 представляет собой сумму квадратов $n - 1$ независимых случайных величин, распределенных по стандартному нормальному закону. Вспоминая теперь, что случайные величины $Y_1 = \sqrt{n} S'_1, Y_2, \dots, Y_n$ независимы, получаем окончательный ответ: статистики S_1 и S_2 независимы ($F_{S_1, S_2}(x_1, x_2) = F_{S_1}(x_1) F_{S_2}(x_2)$), статистика S_1 распределена по нормальному закону с параметрами t и σ^2/n , а случайная величина S_2/σ^2 (в том случае, когда дисперсия σ^2 неизвестна, отношение S_2/σ^2 не является статистикой, поскольку зависит от неизвестного параметра σ^2) — по закону χ^2 с $n - 1$ степенями свободы (см. также параграф 4). \square

Отметим, что проведенные рассуждения будут нами постоянно использоваться в гл. 4, посвященной статистическим задачам, связанным с нормально распределенными наблюдениями.

Важный класс статистик составляют так называемые *достаточные статистики*. Не давая пока строгого математического определения, скажем, что статистика S является достаточной, если она содержит всю ту информацию относительно теоретической функции распределения $F(x)$, что и исходная выборка X_1, \dots, X_n . В частности, вариационный ряд всегда представляет собой достаточную статистику. Более сложными примерами достаточных статистик являются число успехов в схеме Бернулли и двумерная статистика S из примера 3 для выборки из генеральной совокупности с нормальной теоретической функцией распределения. В современной математической статистике достаточные статистики играют очень важную роль.

Эмпирическая функция распределения. Пусть мы имеем выборку X_1, \dots, X_n объема n из генеральной совокупности с теоретической функцией распределения $F(x)$. Построим по выборке X_1, \dots, X_n аналог теоретической функции распределения $F(x)$. Положим

$$F^*(x) = \frac{\mu_x}{n},$$

где μ_x — число элементов выборки, значения которых X_i меньше x . Поскольку каждое X_i меньше x с вероятностью $p_x = F(x)$, а сами X_i

независимы, то μ_x является целочисленной случайной величиной, распределенной по биномиальному закону:

$$\mathbb{P}\{\mu_x = m\} = P_n(m) = \binom{n}{m} [F(x)]^m [1 - F(x)]^{n-m}.$$

Функция $F^*(x)$ носит название *эмпирической (выборочной) функции распределения*. Ясно, что при каждом x значение эмпирической функции распределения $F^*(x)$ является случайной величиной, принимающей значения $0, 1/n, \dots, 1$; если же рассматривать $F^*(x)$ как функцию от x , то $F^*(x)$ представляет собой случайный процесс.

Построение эмпирической функции распределения $F^*(x)$ удобно производить с помощью вариационного ряда X_1^*, \dots, X_n^* . Функция $F^*(x)$ постоянна на каждом интервале $(X_i^*, X_{i+1}^*]$, а в точке X_i^* увеличивается на $1/n$.

Пример 4. График эмпирической функции распределения, построенной по вариационному ряду из табл. 3, приведен на рис. 1. □

Если выборка задана статистическим рядом (см. табл. 4), то эмпирическая функция распределения также постоянна на интервалах $(Z_i, Z_{i+1}]$, но ее значение в точке Z_i увеличивается на n_i/n , а не на $1/n$.

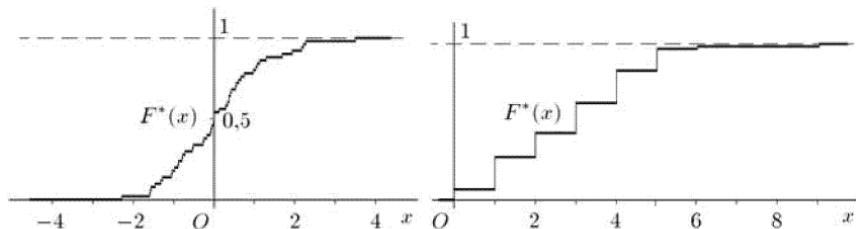


Рис. 1

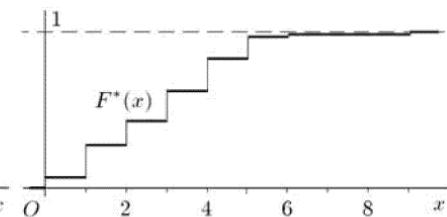


Рис. 2

Пример 5. График эмпирической функции распределения, построенной по статистическому ряду из табл. 6, приведен на рис. 2. □

Гистограмма, полигон. Для наглядности выборку иногда преобразуют следующим образом. Всю ось абсцисс делят на интервалы $[x_i, x_{i+1})$ длиной $\Delta_i = x_{i+1} - x_i$ и определяют функцию $p^*(x)$, постоянную на i -м интервале и принимающую на этом интервале значение $p^*(x) = \nu_i/(n\Delta_i)$, где ν_i — число элементов выборки, попавших в интервал $[x_i, x_{i+1})$. Функция $p^*(x)$ называется *гистограммой*.

При наблюдении дискретной случайной величины вместо гистограммы часто используют *полигон частот*. Для этого по оси абсцисс откладывают все возможные значения b_1, \dots, b_L наблюдаемой величины X , а по оси ординат, пользуясь статистическим рядом, либо числа

ν_1, \dots, ν_L элементов выборки, принявших значения b_1, \dots, b_L (полигон частот), либо соответствующие наблюденные частоты

$$p_1^* = \frac{\nu_1}{n}, \dots, p_L^* = \frac{\nu_L}{n}$$

(полигон относительных частот). Для большей наглядности соседние точки соединяются отрезками прямой.

Для непрерывной наблюдаемой случайной величины полигоном относительных частот иногда называют ломаную линию, соединяющую середины отрезков, составляющих гистограмму.

Пример 6. Построим гистограмму и полигон относительных частот выборки, представленной в табл. 2. Для этого выберем интервалы одинаковой длины $\Delta_i = 1,00$. Числа ν_i и значения $p^*(x)$ на каждом интервале приведены в табл. 7. Гистограмма выборки показана на рис. 3 сплошной линией, а полигон относительных частот — штриховой линией. \square

Таблица 7

	−4, −3	−3, −2	−2, −1	−1, 0	0, 1	1, 2	2, 3	3, 4
ν_i	0	1	8	15	16	6	3	1
$p^*(x)$	0	0,02	0,16	0,30	0,32	0,12	0,06	0,02

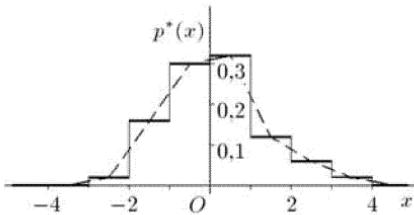


Рис. 3

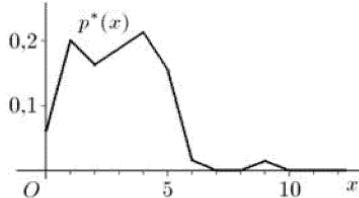


Рис. 4

Пример 7. Построим полигон относительных частот выборки, приведенной в табл. 5. Возможные значения наблюдаемой случайной величины X (числа частиц, попавших в счетчик Гейгера) представляют собой неотрицательные целые числа. Воспользовавшись статистическим рядом из табл. 6, получаем полигон относительных частот, изображенный на рис. 4. \square

Предельное поведение эмпирической функции распределения. Предположим, что по выборке X_1, \dots, X_n мы построили эмпирическую функцию распределения $F_{(n)}^*(x)$ (здесь и в дальнейшем в том случае, когда нам важна зависимость какой-то характеристики от объема выборки n), будем снабжать ее дополнительным нижним индексом (n)). Как мы уже говорили, число $\mu_x = nF_{(n)}^*(x)$ элементов выборки, принявших значение, меньшее x , распределено по биномиальному закону с вероятностью успеха $p_x = F(x)$. Тогда при $n \rightarrow \infty$ в силу усиленного закона больших чисел (часть 1, гл. 8, параграф 2) значения

эмпирических функций распределения $F_{(n)}^*(x)$ сходятся при каждом x к значению теоретической функции распределения $F(x)$. В. И. Гливенко и Ф. П. Кантелли обобщили этот факт и доказали следующую теорему.

Теорема Гливенко–Кантелли. При $n \rightarrow \infty$ с вероятностью, равной единице,

$$\sup_x |F_{(n)}^*(x) - F(x)| \rightarrow 0.$$

Смысл теоремы Гливенко–Кантелли заключается в том, что при увеличении объема выборки n у эмпирической функции распределения исчезают свойства случайности и она приближается к теоретической функции распределения.

Аналогично, если n велико, то значение гистограммы $p_{(n)}^*(x)$ в точке x приближенно равно

$$\frac{F(x_{i+1}) - F(x_i)}{\Delta_i},$$

где x_i, x_{i+1} — концы интервала, в котором находится x , а $\Delta_i = x_{i+1} - x_i$ есть длина этого интервала. Если теоретическая функция распределения имеет плотность распределения $p(x)$ и при этом длины интервалов Δ_i малы, то гистограмма $p_{(n)}^*(x)$ достаточно хорошо воспроизводит эту плотность.

Выборочные характеристики. Эмпирическая функция распределения $F^*(x)$, построенная по фиксированной выборке X_1, \dots, X_n , обладает всеми свойствами обычной функции распределения (дискретной случайной величины). В частности, по ней можно найти математическое ожидание (среднее)

$$m^* = \frac{1}{n} (X_1 + \dots + X_n),$$

второй момент

$$m_2^* = \frac{1}{n} (X_1^2 + \dots + X_n^2),$$

дисперсию

$$\sigma^{2*} = \frac{1}{n} [(X_1 - m^*)^2 + \dots + (X_n - m^*)^2] = m_2^* - (m^*)^2,$$

момент k -го порядка

$$m_k^* = \frac{1}{n} (X_1^k + \dots + X_n^k),$$

центральный момент k -го порядка

$$\mu_k^* = \frac{1}{n} [(X_1 - m^*)^k + \dots + (X_n - m^*)^k]$$

и т. д. Соответствующие характеристики называются *выборочными* (выборочное среднее, выборочный второй момент, выборочная диспер-

сия и т. п.). Ясно, что выборочные характеристики как функции от случайных величин X_1, \dots, X_n сами являются случайными величинами, причем их распределения определяются в соответствии с общими положениями теории вероятностей (см. часть 1, гл. 6, параграф 7). Так, функция распределения выборочного среднего m^* для случая дискретной наблюдаемой случайной величины определяется формулой

$$F_{m^*}(x) = \sum P(x_1) \dots P(x_n),$$

где суммирование ведется по всем x_1, \dots, x_n , принимающим значения $\{b_1, \dots, b_L\}$ и удовлетворяющим неравенству $(x_1 + \dots + x_n)/n < x$, а функция распределения выборочного второго момента m_2^* для непрерывного случая — формулой

$$F_{m_2^*}(x) = \int \dots \int_{\frac{1}{n}(x_1^2 + \dots + x_n^2) < x} p(x_1) \dots p(x_n) dx_1 \dots dx_n.$$

Наряду с выборочной дисперсией σ^{2*} часто используют и другую характеристику разброса выборки вокруг среднего:

$$s^{2*} = \frac{1}{n-1} [(X_1 - m^*)^2 + \dots + (X_n - m^*)^2] = \frac{n}{n-1} \sigma^{2*}.$$

Характеристику s^{2*} также будем называть выборочной дисперсией, а для того чтобы не путать σ^{2*} и s^{2*} , каждый раз будем указывать, о какой именно выборочной дисперсии идет речь. Выборочная дисперсия s^{2*} отличается от выборочной дисперсии σ^{2*} только лишь наличием множителя $n/(n-1)$, который с увеличением объема выборки n стремится к единице, и, казалось бы, нет смысла вводить две практически одинаковые величины. Однако, как мы увидим из дальнейшего, s^{2*} является несмещенной оценкой теоретической дисперсии σ^2 , чего нельзя сказать о выборочной дисперсии σ^{2*} , хотя стандартные методы приводят именно к σ^2 .

Пример 8. Подсчитаем выборочное среднее и выборочные дисперсии для выборки, приведенной в табл. 2:

$$m^* = \frac{1}{50} (-1,04 - 1,06 + 1,06 - \dots - 1,57 + 1,92) = 0,082,$$

$$\sigma^{2*} = \frac{1}{50} [(-1,04 - 0,082)^2 + (-1,06 - 0,082)^2 + \dots + (1,92 - 0,082)^2] = 1,34,$$

$$s^{2*} = \frac{50}{49} \sigma^{2*} = 1,37.$$

Для подсчета выборочной дисперсии σ^{2*} можно было бы воспользоваться также формулой $\sigma^{2*} = m_2^* - (m^*)^2$. \square

4. Основные распределения математической статистики

Наиболее часто в математической статистике используются: нормальное распределение, χ^2 -распределение (распределение Пирсона), t -распределение (распределение Стьюдента), F -распределение (распределение Фишера), распределение Колмогорова и ω^2 -распределение. Все эти распределения связаны с нормальным. В свою очередь, широкое распространение нормального распределения обусловлено исключительно центральной предельной теоремой (см. часть 1, гл. 8, параграф 4). Ввиду их особой важности все названные распределения заложены и содержатся в различных статистических таблицах, а также, частично, в большинстве учебников по теории вероятностей и математической статистике. Наиболее полными из известных и доступных читателю в нашей стране являются таблицы Л. Н. Большева и Н. В. Смирнова [1], на которые мы и будем ссылаться в дальнейшем.

Нормальное распределение. Одномерное стандартное нормальное распределение (стандартный нормальный закон) задается своей плотностью распределения (см. часть 1, гл. 5, параграф 4)

$$\varphi(x) = \Phi'(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Значения функции $\Phi(x)$ и плотности $\varphi(x)$ стандартного нормального распределения, а также квантилей φ_α (функции φ_α , обратной функции стандартного нормального распределения) приведены в [1], табл. 1.1–1.3 (см. также табл. 2 и 3 приложения).

Общее одномерное нормальное распределение характеризуется двумя параметрами: средним (математическим ожиданием) m и дисперсией σ^2 . Его можно трактовать как распределение случайной величины

$$\eta = m + \xi \sqrt{\sigma^2},$$

где случайная величина ξ подчинена стандартному нормальному закону. Плотность распределения и функцию распределения общего нормального закона будем обозначать через $\varphi(x; m, \sigma^2)$ и $\Phi(x; m, \sigma^2)$.

Многомерное (k -мерное) нормальное распределение (часть 1, гл. 6, параграф 4) определяется вектором средних $\mathbf{m} = (m_1, \dots, m_k)$ и матрицей ковариаций $A = (\rho_{ij})$.

χ^2 -распределение (см. часть 1, гл. 5, параграф 4, а также примеры 28 и 30, часть 1, гл. 6, параграф 7). Пусть ξ_1, \dots, ξ_n — независимые случайные величины, распределенные по стандартному нормальному закону. Распределение случайной величины

$$\chi^2 = \xi_1^2 + \dots + \xi_n^2$$

носит название χ^2 -распределения с n степенями свободы. χ^2 -распределение имеет плотность распределения

$$h(x) = H'(x) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2} \quad (x > 0),$$

где $\Gamma(\cdot)$ введено в параграфе 4 гл. 5.

Значения функции χ^2 -распределения и α -процентных точек (α -процентная точка χ^2 -распределения представляет собой $(1 - \alpha/100)$ -квантиль $\chi^2_{1-\alpha/100}$) χ^2 -распределения приведены в [1], табл. 2.1а и 2.2а.

В дальнейшем нам будет полезно следующее свойство. Пусть ξ_1, \dots, ξ_n — независимые случайные величины, распределенные по нормальному закону с одинаковыми параметрами m и σ^2 . Положим

$$\eta = \frac{1}{n} (\xi_1 + \dots + \xi_n).$$

Тогда случайная величина

$$\chi^2 = \frac{1}{\sigma^2} [(\xi_1 - \eta)^2 + \dots + (\xi_n - \eta)^2]$$

имеет χ^2 -распределение, но с $n - 1$ степенями свободы. Доказательство этого факта содержится в примере 3.

Еще одна схема, в которой появляется χ^2 -распределение — полиномиальная схема (см. часть 1, гл. 4, параграф 7). Пусть производится n независимых одинаковых испытаний, в каждом из которых с вероятностью p_l может произойти одно из событий A_l ($l = 1, \dots, L$). Обозначим через ν_l число появлений события A_l . Тогда из многомерного аналога интегральной теоремы Муавра–Лапласа следует, что случайная величина

$$\chi^2 = \frac{(\nu_1 - np_1)^2}{np_1} + \dots + \frac{(\nu_L - np_L)^2}{np_L}$$

при $n \rightarrow \infty$ асимптотически распределена по закону χ^2 с $L - 1$ степенями свободы.

t -распределение. Пусть ξ и χ^2 — независимые случайные величины, причем ξ распределена по стандартному нормальному закону, а χ^2 имеет χ^2 -распределение с n степенями свободы. Распределение случайной величины

$$\tau = \frac{\xi}{\sqrt{\chi^2/n}}$$

называется t -распределением с n степенями свободы. t -распределение имеет плотность распределения

$$t(x) = T'(x) = \frac{1}{\sqrt{\pi n}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}.$$

Значения функции t -распределения и α -процентных точек ($(1 - \alpha/100)$ -квантилей $t_{1-\alpha/100}$) t -распределения приведены в [1], табл. 3.1а и 3.2.

Далее, пусть ξ_1, \dots, ξ_n — независимые одинаково распределенные случайные величины, подчиненные нормальному закону со средним m . Положим

$$\eta = \frac{1}{n} (\xi_1 + \dots + \xi_n), \quad \zeta = (\xi_1 - \eta)^2 + \dots + (\xi_n - \eta)^2.$$

Тогда случайные величины η и ζ независимы, а случайная величина

$$\tau = \frac{\sqrt{n}(\eta - m)}{\sqrt{\frac{\zeta}{n-1}}}$$

имеет t -распределение с $n - 1$ степенями свободы (доказательство этого см. в примере 3).

F -распределение. Пусть χ_1^2 и χ_2^2 — две независимые случайные величины, имеющие χ^2 -распределения с n_1 и n_2 степенями свободы. Распределение случайной величины

$$\varkappa = \frac{n_2 \chi_1^2}{n_1 \chi_2^2}$$

носит название F -распределения с параметрами n_1 и n_2 . F -распределение имеет плотность распределения

$$\psi(x) = \Psi'(x) = \frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} n_1^{n_1/2} n_2^{n_2/2} x^{n_1/2-1} (n_2 + n_1 x)^{-(n_1+n_2)/2} \quad (x > 0).$$

Значения α -процентных точек ($(1 - \alpha/100)$ -квантилей $\varphi_{1-\alpha/100}$) F -распределения приведены в [1], табл. 3.5.

Распределение Колмогорова. Функция распределения Колмогорова имеет вид

$$K(x) = \sum_{j=-\infty}^{\infty} (-1)^j e^{-2j^2 x^2} \quad (x > 0).$$

Распределение Колмогорова является распределением случайной величины

$$\eta = \sup_{0 \leq t \leq 1} |\xi(t)|,$$

где $\xi(t)$ — броуновский мостик, т. е. винеровский процесс с закрепленными концами $\xi(0) = 0$, $\xi(1) = 0$ на отрезке $0 \leq t \leq 1$ (см. [11]).

Значения функции распределения Колмогорова приведены в [1], табл. 6.1. Квантили распределения Колмогорова будем обозначать через k_α .

ω^2 -распределение. Функция ω^2 -распределения задается формулой

$$A(x) = \frac{1}{\sqrt{2x}} \sum_{j=0}^{\infty} \frac{\Gamma\left(j + \frac{1}{2}\right) \sqrt{4j+1}}{\Gamma\left(\frac{1}{2}\right) \Gamma(j+1)} e^{-\frac{(4j+1)^2}{16x}} \times \\ \times \left[I_{-\frac{1}{4}}\left(\frac{(4j+1)^2}{16x}\right) - I_{\frac{1}{4}}\left(\frac{(4j+1)^2}{16x}\right) \right] \quad (x > 0).$$

Здесь $I_y(z)$ — модифицированная функция Бесселя. ω^2 -распределение представляет собой распределение случайной величины

$$\omega^2 = \int_0^1 \xi^2(t) dt,$$

где $\xi(t)$ — броуновский мостик.

Значения функции ω^2 -распределения приведены в [1], табл. 6.4а. Квантили ω^2 -распределения будем обозначать через a_α .

Г л а в а 2

ОЦЕНКИ НЕИЗВЕСТНЫХ ПАРАМЕТРОВ

Как уже говорилось в гл. 1, одним из двух основных направлений в математической статистике является оценивание неизвестных параметров. В этой главе мы дадим определение оценки, опишем те свойства, которые желательно требовать от оценки, и приведем основные методы построения оценок. Завершается глава изложением метода построения доверительных интервалов для неизвестных параметров.

1. Статистические оценки и их свойства

Предположим, что в результате наблюдений мы получили выборку X_1, \dots, X_n из генеральной совокупности с теоретической функцией распределения $F(x)$. Относительно $F(x)$ обычно бывает известно только, что она принадлежит определенному *параметрическому семейству* $F(x; \vartheta)$, зависящему от числового или векторного параметра ϑ . Как правило, для простоты изложения будем рассматривать случай числового параметра ϑ и лишь иногда обращаться к векторному параметру $\boldsymbol{\vartheta} = (\vartheta_1, \dots, \vartheta_k)$; в векторном случае будем использовать запись $F(x; \vartheta_1, \dots, \vartheta_k)$. Для большей наглядности будем все неизвестные параметры (за исключением теоретических моментов m_k , σ^2 и μ_k) обозначать буквой ϑ (снабжая их при необходимости индексами), хотя в теории вероятностей для них обычно приняты другие обозначения. Наша цель состоит в том, чтобы, опираясь только на выборку X_1, \dots, X_n , оценить неизвестный параметр ϑ .

Оценкой неизвестного параметра ϑ , построенной по выборке X_1, \dots, X_n , назовем произвольную функцию

$$\vartheta^* = \vartheta^*(X_1, \dots, X_n),$$

зависящую только от выборки X_1, \dots, X_n . Ясно, что как функция от случайной величины (X_1, \dots, X_n) оценка ϑ^* сама будет являться *случайной величиной* и, как всякая случайная величина, будет иметь функцию распределения $F_{\vartheta^*}(x)$, определяемую в дискретном случае формулой

$$F_{\vartheta^*}(x) = \sum P(x_1; \vartheta) \cdots P(x_n; \vartheta),$$

где суммирование ведется по всем переменным x_1, \dots, x_n , принимающим значения $\{b_1, \dots, b_L\}$ из ряда распределения наблюданной случай-

ной величины X и удовлетворяющим неравенству $\vartheta^*(x_1, \dots, x_n) < x$, и в непрерывном случае — формулой

$$F_{\vartheta^*}(x) = \int \dots \int p(x_1; \vartheta) \cdots p(x_n; \vartheta) dx_1 \cdots dx_n,$$

где интегрирование ведется по области, выделяемой неравенством $\vartheta^*(x_1, \dots, x_n) < x$. Как уже говорилось, иногда для того, чтобы подчеркнуть зависимость оценки от объема выборки n , будем наряду с обозначением ϑ^* употреблять обозначение $\vartheta^*_{(n)}$. Нужно четко представлять себе, что *зависимость оценки ϑ^* от неизвестного параметра ϑ осуществляется только через зависимость от ϑ выборки X_1, \dots, X_n* , что в свою очередь реализуется *зависимостью от ϑ функции распределения $F_{\vartheta^*}(x) = F_{\vartheta^*}(x; \vartheta)$* . Приведенное выше определение отождествляет понятие оценки ϑ^* (вектора оценок $\boldsymbol{\vartheta}^* = (\vartheta_1^*, \dots, \vartheta_k^*)$) с одномерной (k -мерной) статистикой.

Пример 1. Предположим, что проведено n испытаний в схеме Бернулли с неизвестной вероятностью успеха ϑ . В результате наблюдений получена выборка X_1, \dots, X_n , где X_i — число успехов в i -м испытании. Ряд распределения наблюдаемой величины X — числа успехов в одном испытании представлен в табл. 1.

Таблица 1

X	0	1
P	$1 - \vartheta$	ϑ

В качестве оценки ϑ^* рассмотрим наблюденную частоту успехов

$$\vartheta^* = \frac{\mu}{n} = m^*,$$

где

$$\mu = X_1 + \dots + X_n$$

представляет собой суммарное число успехов в n испытаниях Бернулли. Статистика μ распределена по биномциальному закону с параметром ϑ , поэтому ряд распределения оценки ϑ^* имеет вид, приведенный в табл. 2. \square

Таблица 2

ϑ^*	0	$1/n$	\dots	i/n	\dots	1
P	$(1 - \vartheta)^n$	$n\vartheta(1 - \vartheta)^{n-1}$	\dots	$\binom{n}{i} \vartheta^i (1 - \vartheta)^{n-i}$	\dots	ϑ^n

Пример 2. Выборка X_1, \dots, X_n произведена из генеральной совокупности с теоретической функцией распределения $F(x) = \Phi(x; \vartheta, \sigma^2)$, являющейся нормальной с неизвестным средним ϑ . В качестве оценки ϑ снова рассмотрим выборочное среднее

$$\vartheta^* = m^* = \frac{1}{n} (X_1 + \dots + X_n).$$

Функция распределения $F_{\vartheta^*}(x)$ задается формулой

$$F_{\vartheta^*}(x) = \int_{\substack{\frac{1}{n}(x_1+\dots+x_n) < x}} \dots \int \left(2\pi\sigma^2\right)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}[(x_1-\vartheta)^2 + \dots + (x_n-\vartheta)^2]} dx_1 \dots dx_n.$$

Однако вместо непосредственного вычисления написанного n -мерного интеграла заметим, что статистика

$$S = X_1 + \dots + X_n$$

распределена по нормальному закону с параметрами $n\vartheta$ (математической ожидание) и $n\sigma^2$ (дисперсия). Значит, оценка $\vartheta^* = S/n$ распределена также по нормальному закону с параметрами ϑ и σ^2/n . \square

Разумеется, на практике имеет смысл использовать далеко не любую оценку.

Пример 3. Как и в примере 1, рассмотрим испытания в схеме Бернулли. Однако теперь в качестве оценки неизвестной вероятности успеха ϑ возьмем

$$\vartheta^* = \vartheta^*(X_1, \dots, X_n) = \frac{1}{2}.$$

Такая оценка будет хороша лишь в том случае, когда истинное значение $\vartheta = 1/2$; ее качество ухудшается с увеличением отклонения ϑ от $1/2$. \square

Приведенный пример показывает, что желательно употреблять только те оценки, которые по возможности принимали бы значения, наиболее близкие к неизвестному параметру. Однако в силу случайности выборки в математической статистике мы, как правило, не застрахованы полностью от сколь угодно большой ошибки. Значит, гарантировать достаточную близость оценки ϑ^* к оцениваемому параметру ϑ можно только с некоторой вероятностью и для того, чтобы увеличить эту вероятность, приходится приносить необходимую жертву — увеличивать объем выборки n .

Опишем теперь те свойства, которые мы хотели бы видеть у оценки.

Главное свойство любой оценки, оправдывающее само название «оценка», — возможность хотя бы ценой увеличения объема выборки до бесконечности получить точное значение неизвестного параметра ϑ . Оценка ϑ^* называется *состоятельной*, если с ростом объема выборки она сходится к оцениваемому параметру ϑ . Можно рассматривать сходимость различных типов: по вероятности, с вероятностью единицы, в среднем квадратичном и т. д. Обычно рассматривается сходимость по вероятности, т. е. состоятельной называется такая оценка $\vartheta^* = \vartheta_{(n)}^*$, которая для любого $\varepsilon > 0$ при всех возможных значениях неизвестного параметра ϑ удовлетворяет соотношению

$$\mathbf{P}\{|\vartheta_{(n)}^* - \vartheta| > \varepsilon\} \xrightarrow[n \rightarrow \infty]{} 0.$$

Отметим, что правильнее было бы говорить о состоятельности *последовательности оценок* $\vartheta_{(n)}^*$, поскольку для каждого значения n объема выборки оценка $\vartheta_{(n)}^*$ может определяться по своему правилу. Однако в дальнейшем мы будем употреблять понятие состоятельности

только для оценок, построенных по определенным алгоритмам, поэтому будем говорить просто о состоятельности оценки.

Пример 4. Оценка ϑ^* из примера 1 является состоятельной оценкой неизвестной вероятности успеха ϑ . Это является прямым следствием закона больших чисел Бернуlli. \square

Пример 5. Пусть выборка X_1, \dots, X_n произведена из генеральной совокупности с неизвестной теоретической функцией распределения $F(x)$. Тогда в силу закона больших чисел выборочный момент

$$m_k^* = \frac{1}{n} (X_1^k + \dots + X_n^k)$$

сходится к теоретическому моменту $m_k = M X^k$ и, значит, представляет собой состоятельную оценку m_k . Аналогично, выборочные дисперсии σ^{2*} и s^{2*} и выборочные центральные моменты μ_k^* являются состоятельными оценками теоретической дисперсии σ^2 и теоретических центральных моментов μ_k . Отметим, что поскольку в этом примере не предполагается принадлежность теоретической функции распределения $F(x)$ какому-либо параметрическому семейству, то мы имеем дело с задачей оценки неизвестных моментов теоретической функции распределения в непараметрической модели. \square

Пример 6. Выборка X_1, \dots, X_n произведена из генеральной совокупности с теоретической функцией распределения $F(x)$, имеющей плотность распределения Коши

$$p(x) = \frac{1}{\pi [1 + (x - \vartheta)^2]}$$

с неизвестным параметром ϑ . Поскольку плотность распределения Коши симметрична относительно ϑ , то казалось бы естественным в качестве оценки ϑ^* параметра ϑ взять выборочное среднее

$$\vartheta^* = m^* = \frac{1}{n} (X_1 + \dots + X_n).$$

Однако ϑ^* , как и сама наблюдаемая случайная величина X , имеет распределение Коши с тем же параметром ϑ (это легко установить с помощью характеристических функций, см. часть 1, гл. 8, параграф 3), т. е. не сближается с параметром ϑ , а значит, не является состоятельной оценкой параметра ϑ . \square

Из курса теории вероятностей известно (см. часть 1, гл. 7, параграф 1), что мерой отклонения оценки ϑ^* от параметра ϑ служит разность $M\vartheta^* - \vartheta$. В математической статистике разность

$$M\vartheta^* - \vartheta = \delta(\vartheta)$$

называется *смещением* оценки ϑ^* . Ясно, что

$$\delta(\vartheta) = \sum [\vartheta^*(x_1, \dots, x_n) - \vartheta] P(x_1; \vartheta) \cdots P(x_n; \vartheta)$$

в дискретном случае и

$$\delta(\vartheta) = \int \dots \int [\vartheta^*(x_1, \dots, x_n) - \vartheta] p(x_1; \vartheta) \cdots p(x_n; \vartheta) dx_1 \cdots dx_n$$

в непрерывном, где суммирование или интегрирование ведется по всем возможным значениям x_1, \dots, x_n .

Оценка ϑ^* называется *несмешенной*, если

$$\delta(\vartheta) = 0$$

при всех ϑ , т. е. ее среднее значение $\mathbf{M}\vartheta^*$ совпадает с оцениваемым параметром ϑ .

Пример 7. Оценка ϑ^* неизвестной вероятности успеха ϑ из примера 1 является несмешенной. Действительно,

$$\mathbf{M}\vartheta^* = \frac{1}{n} \mathbf{M}(X_1 + \dots + X_n) = \frac{1}{n} (\mathbf{M}X_1 + \dots + \mathbf{M}X_n) = \mathbf{M}X = \vartheta. \quad \square$$

Пример 8. Выборочные моменты m_k^* являются несмешенными оценками теоретических моментов m_k , поскольку

$$\mathbf{M}m_k^* = \frac{1}{n} \mathbf{M}(X_1^k + \dots + X_n^k) = \frac{1}{n} (\mathbf{M}X_1^k + \dots + \mathbf{M}X_n^k) = \mathbf{M}X^k = m_k.$$

Вычислим теперь математическое ожидание выборочной дисперсии σ^{2*} :

$$\begin{aligned} \mathbf{M}\sigma^{2*} &= \mathbf{M}\left(\frac{1}{n}\left[(X_1 - m^*)^2 + \dots + (X_n - m^*)^2\right]\right) = \\ &= \frac{1}{n} \sum_{i=1}^n \mathbf{M}X_i^2 - \frac{2}{n} \sum_{i=1}^n \mathbf{M}(m^* X_i) + \mathbf{M}(m^*)^2 = \\ &= m_2 - \frac{2}{n^2} \sum_{i=1}^n \sum_{j=1}^n \mathbf{M}(X_i X_j) + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \mathbf{M}(X_i X_j) = \\ &= m_2 - \frac{1}{n^2} \sum_{i=1}^n \mathbf{M}X_i^2 - \frac{1}{n^2} \sum_{\substack{i,j, \\ i \neq j}} \mathbf{M}(X_i X_j) = \frac{n-1}{n} m_2 - \frac{1}{n^2} \sum_{\substack{i,j, \\ i \neq j}} \mathbf{M}(X_i) \mathbf{M}(X_j) = \\ &= \frac{n-1}{n} m_2 - \frac{n-1}{n} m^2 = \frac{n-1}{n} \sigma^2. \end{aligned}$$

Таким образом, σ^{2*} является смешенной (хотя и состоятельной, см. пример 5) оценкой дисперсии σ^2 . Поскольку

$$s^{2*} = \frac{n}{n-1} \sigma^{2*},$$

то

$$\mathbf{M}s^{2*} = \sigma^2$$

и s^{2*} представляет собой уже несмешенную оценку σ^2 . Можно показать также, что выборочные центральные моменты μ_k^* ($k \geq 3$) являются смешенными оценками теоретических центральных моментов μ_k . \square

Пример 9. Пусть X_1, \dots, X_n — выборка из генеральной совокупности с теоретической функцией распределения $F(x) = \Phi(x; \vartheta, \sigma^2)$, являющейся нормальной с неизвестным средним ϑ . Поскольку $\mathbf{M}X = \vartheta$, то оценка

$$\vartheta^* = \vartheta^*(X_1, \dots, X_n) = X_1$$

является несмешенной. Очевидно, однако, что она не является состоятельной. \square

Примеры 8 и 9 показывают, что состоятельная оценка может быть смешенной и, наоборот, несмешенная оценка не обязана быть состоятельной.

Рассматривая несколько оценок неизвестного параметра ϑ , мы, разумеется, хотели бы выбрать из них ту, которая имела бы наименьший разброс, причем при любом значении неизвестного параметра ϑ . Мерой разброса оценки ϑ^* , как и всякой случайной величины, является дисперсия

$$\mathbf{D}\vartheta^* = \mathbf{M}(\vartheta^* - \mathbf{M}\vartheta^*)^2$$

(дисперсия, как и распределение оценки, зависит от неизвестного параметра ϑ). Однако для смещенной оценки ϑ^* дисперсия служит мерой близости не к оцениваемому параметру ϑ , а к математическому ожиданию $\mathbf{M}\vartheta^*$. Поэтому естественно искать оценки с наименьшей дисперсией не среди всех оценок, а только среди несмешанных, что мы и будем делать в дальнейшем. Для несмешанных оценок дисперсия определяется также формулой

$$\mathbf{D}\vartheta^* = \mathbf{M}(\vartheta^* - \vartheta)^2.$$

Имеется несколько подходов к нахождению несмешанных оценок с минимальной дисперсией. Это связано с тем, что такие оценки существуют не всегда, а найти их бывает чрезвычайно сложно. Здесь мы изложим понятие эффективности оценки, основанное на неравенстве Рао–Крамера.

Теорема 1 (неравенство Рао–Крамера). Пусть $\vartheta_{(n)}^*$ — несмещенная оценка неизвестного параметра ϑ , построенная по выборке объема n . Тогда (при некоторых дополнительных условиях регулярности, наложенных на семейство $F(x; \vartheta)$),

$$\mathbf{D}\vartheta_{(n)}^* \geqslant \frac{1}{nI},$$

где $I = I(\vartheta)$ — информация Фишера, определяемая в дискретном случае формулой

$$I = \mathbf{M}[(\ln P(X; \vartheta))'_\vartheta]^2 = \sum_{l=1}^L \left[\frac{P'_\vartheta(b_l; \vartheta)}{P(b_l; \vartheta)} \right]^2 P(b_l; \vartheta).$$

а в непрерывном — формулой

$$I = \mathbf{M}[(\ln p(X; \vartheta))'_\vartheta]^2 = \int_{-\infty}^{\infty} \left[\frac{p'_\vartheta(x; \vartheta)}{p(x; \vartheta)} \right]^2 p(x; \vartheta) dx.$$

Прежде чем переходить к доказательству теоремы, заметим, что по неравенству Рао–Крамера дисперсия любой несмешанной оценки не может быть меньше $1/(nI)$. Назовем эффективностью $e = e(\vartheta)$ несмешанной оценки ϑ^* величину

$$e = \frac{1}{nI \mathbf{D}\vartheta^*}.$$

Ясно, что эффективность любой оценки ϑ^* при каждом ϑ заключена между нулем и единицей, причем чем она ближе к единице при каком-либо ϑ , тем лучше оценка ϑ^* при этом значении неизвестного параметра.

Несмешенная оценка ϑ^* называется *эффективной* (по Рао–Крамеру), если $e(\vartheta) = 1$ при любом ϑ .

Доказательство теоремы 1. Доказательство этой и всех остальных теорем будем проводить (если не сделано специальной оговорки) для непрерывного случая. Это связано с тем, что непрерывный случай, как правило, более сложен, и читатель, усвоивший доказательство для непрерывного случая, легко проведет его для дискретного.

Как мы увидим из хода доказательства, условия регулярности семейства $F(x; \vartheta)$, упомянутые в формулировке теоремы, есть не что иное, как условия, гарантирующие законность дифференцирования под знаком интеграла в формулах (1) и (3). В разных книгах сформулированы различные достаточные условия. Мы упомянем одно из них, приведенное в [11]:

функция $p(x; \vartheta)$ для всех (точнее, для почти всех) x непрерывно дифференцируема по ϑ , информация Фишера $I(\vartheta)$ конечна, положительна и непрерывна по ϑ .

Приступим теперь к собственно доказательству теоремы. Заметим прежде всего, что, дифференцируя тождество

$$\int_{-\infty}^{\infty} p(x; \vartheta) dx = 1 \quad (1)$$

(в силу сформулированного условия это можно делать), получаем

$$\left(\int_{-\infty}^{\infty} p(x; \vartheta) dx \right)'_{\vartheta} = \int_{-\infty}^{\infty} \frac{p'_{\vartheta}(x; \vartheta)}{p(x; \vartheta)} p(x; \vartheta) dx = 0. \quad (2)$$

Далее, в силу несмешенности оценки $\vartheta_{(n)}^* = \vartheta_{(n)}^*(X_1, \dots, X_n)$ имеем

$$\mathbf{M}\vartheta_{(n)}^* = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \vartheta_{(n)}^*(x_1, \dots, x_n) p(x_1; \vartheta) \dots p(x_n; \vartheta) dx_1 \dots dx_n = \vartheta. \quad (3)$$

Дифференцируя это равенство по ϑ и учитывая очевидное тождество

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \vartheta p(x_1; \vartheta) \dots p'_{\vartheta}(x_i; \vartheta) \dots p(x_n; \vartheta) dx_1 \dots dx_n = 0,$$

полученное из (1) и (2), находим

$$\begin{aligned} & \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \left(\sum_{i=1}^n [\vartheta_{(n)}^*(x_1, \dots, x_n) - \vartheta] p(x_1; \vartheta) \dots p'_{\vartheta}(x_i; \vartheta) \dots p(x_n; \vartheta) \right) dx_1 \dots dx_n = \\ & = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} [\vartheta_{(n)}^*(x_1, \dots, x_n) - \vartheta] \left[\sum_{i=1}^n \frac{p'_{\vartheta}(x_i; \vartheta)}{p(x_i; \vartheta)} \right] p(x_1; \vartheta) \dots p(x_n; \vartheta) dx_1 \dots dx_n = 1. \end{aligned} \quad (4)$$

Воспользовавшись неравенством Коши–Буняковского

$$\begin{aligned} & \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} a(x_1, \dots, x_n) b(x_1, \dots, x_n) p(x_1; \vartheta) \cdots p(x_n; \vartheta) dx_1 \cdots dx_n \leq \\ & \leq \left[\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} a^2(x_1, \dots, x_n) p(x_1; \vartheta) \cdots p(x_n; \vartheta) dx_1 \cdots dx_n \right]^{1/2} \times \\ & \quad \times \left[\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} b^2(x_1, \dots, x_n) p(x_1; \vartheta) \cdots p(x_n; \vartheta) dx_1 \cdots dx_n \right]^{1/2} \end{aligned}$$

при

$$a(x_1, \dots, x_n) = \vartheta_{(n)}^*(x_1, \dots, x_n) - \vartheta, \quad b(x_1, \dots, x_n) = \sum_{i=1}^n \frac{p'_{\vartheta}(x_i; \vartheta)}{p(x_i; \vartheta)},$$

имеем

$$1 \leq \left[\mathbf{M}(\vartheta_{(n)}^* - \vartheta)^2 \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\sum_{i=1}^n \frac{p'_{\vartheta}(x_i; \vartheta)}{p(x_i; \vartheta)} \right)^2 p(x_1; \vartheta) \cdots p(x_n; \vartheta) dx_1 \cdots dx_n \right]^{1/2}. \quad (5)$$

Заметим теперь, что в силу тождества (2)

$$\begin{aligned} & \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[\sum_{i=1}^n \frac{p'_{\vartheta}(x_i; \vartheta)}{p(x_i; \vartheta)} \right]^2 p(x_1; \vartheta) \cdots p(x_n; \vartheta) dx_1 \cdots dx_n = \\ & = \sum_{i=1}^n \int_{-\infty}^{\infty} \left[\frac{p'_{\vartheta}(x_i; \vartheta)}{p(x_i; \vartheta)} \right]^2 p(x_i; \vartheta) dx_i = nI. \end{aligned}$$

Тогда неравенство (5) можно переписать в виде $1 \leq nI \mathbf{D}\vartheta_{(n)}^*$, откуда и следует неравенство Рао–Крамера. \square

З а м е ч а н и е 1 к теореме 1. Для превращения используемого при доказательстве теоремы 1 неравенства Коши–Буняковского, в равенство необходимо и достаточно существование таких функций $\vartheta^*(x)$ аргумента x и $c(\vartheta)$ аргумента ϑ , что

$$\frac{p'_{\vartheta}(x_i; \vartheta)}{p(x_i; \vartheta)} = [\vartheta^*(x_i) - \vartheta] c(\vartheta). \quad (6)$$

При этом оценка $\vartheta_{(n)}^*(x_1, \dots, x_n)$ должна иметь вид

$$\vartheta_{(n)}^*(x_1, \dots, x_n) = \frac{1}{n} [\vartheta^*(x_1) + \dots + \vartheta^*(x_n)].$$

Обозначая

$$\vartheta^*(x) = S(x), \quad A(\vartheta) = \int c(\vartheta) d\vartheta, \quad B(\vartheta) = - \int \vartheta c(\vartheta) d\vartheta$$

и интегрируя уравнение (6), получаем, что необходимым условием существования эффективной оценки является возможность представления плотности распределения $p(x; \vartheta)$ в виде

$$p(x; \vartheta) = h(x) e^{S(x) A(\vartheta) + B(\vartheta)}, \quad (7)$$

где $h(x)$ и $S(x)$ — функции, зависящие только от x , а $A(\vartheta)$ и $B(\vartheta)$ — функции, зависящие только от ϑ .

Аналогичное представление для ряда распределения $P(x; \vartheta)$ должно иметь место и в дискретном случае. Семейство плотностей или рядов распределения такого вида носит название экспоненциального.

Экспоненциальные семейства играют в математической статистике важную роль. В частности, как мы показали, только для этих семейств могут существовать эффективные оценки, которые к тому же определяются формулой

$$\vartheta_{(n)}^* = \frac{1}{gn} [S(X_1) + \dots + S(X_n)] \quad (8)$$

(появление множителя $1/g$ связано с неоднозначностью определения функций $h(x)$, $A(\vartheta)$, $B(\vartheta)$ и $S(x)$ в представлении (7)). Однако следует помнить, что не для всякого экспоненциального семейства существует эффективная оценка (в принятом нами смысле), поскольку эффективная оценка по определению должна быть несмешенной, что, вообще говоря, нельзя сказать об оценке (8) в случае произвольного экспоненциального семейства. Впрочем, из тождества (1) вытекает весьма простой способ проверки несмешенности (8) непосредственно по $A(\vartheta)$ и $B(\vartheta)$, заключающийся в выполнении равенства $B'(\vartheta) = -g\vartheta A'(\vartheta)$.

Замечание 2 к теореме 1. Неравенство Рао–Крамера можно обобщить на случай смешенных оценок:

$$\mathbf{D}\vartheta_{(n)}^* \geq \frac{1}{nI} [1 + \delta'(\vartheta)]^2.$$

И в этом случае неравенство превращается в равенство только тогда, когда семейство распределений экспоненциально.

Пример 10. Рассмотрим оценку ϑ^* неизвестной вероятности успеха ϑ в схеме Бернулли из примера 1. Как показано в примере 7, эта оценка несмешенная. Дисперсия ϑ^* имеет вид

$$\mathbf{D}\vartheta^* = \mathbf{D}\left[\frac{1}{n}(X_1 + \dots + X_n)\right] = \frac{1}{n^2} [\mathbf{D}(X_1 + \dots + X_n)] = \frac{\mathbf{D}X}{n} = \frac{\vartheta(1-\vartheta)}{n}.$$

Найдем информацию Фишера (напомним, что в данном случае наблюдаемая величина X принимает всего два значения 0 и 1 с вероятностями $P(0; \vartheta) = 1 - \vartheta$ и $P(1; \vartheta) = \vartheta$ соответственно):

$$\begin{aligned} I &= \left[\frac{P'_\vartheta(0; \vartheta)}{P(0; \vartheta)}\right]^2 P(0; \vartheta) + \left[\frac{P'_\vartheta(1; \vartheta)}{P(1; \vartheta)}\right]^2 P(1; \vartheta) = \\ &= \left(\frac{-1}{1-\vartheta}\right)^2 (1-\vartheta) + \left(\frac{1}{\vartheta}\right)^2 \vartheta = \frac{1}{1-\vartheta} + \frac{1}{\vartheta} = \frac{1}{\vartheta(1-\vartheta)}. \end{aligned}$$

Таким образом, $e(\vartheta) = 1$ и, значит, оценка ϑ^* эффективная. \square

Пример 11. Рассмотрим оценку ϑ^* неизвестного среднего нормального закона из примера 2. Поскольку эта оценка представляет собой выборочное среднее, то в соответствии с результатами, полученными в примере 8, она является несмешенной. Найдем ее эффективность. Для этого прежде всего заметим, что

$$\mathbf{D}\vartheta^* = \mathbf{D}\left[\frac{1}{n}(X_1 + \dots + X_n)\right] = \frac{\mathbf{D}X}{n} = \frac{\sigma^2}{n}.$$

Далее,

$$\ln p(X; \vartheta) = \ln \varphi(X; \vartheta, \sigma^2) = -\ln \sqrt{2\pi\sigma^2} - \frac{(X-\vartheta)^2}{2\sigma^2},$$

$$\begin{aligned} (\ln p(X; \vartheta))'_{\vartheta} &= \frac{X - \vartheta}{\sigma^2}, & [(\ln p(X; \vartheta))'_{\vartheta}]^2 &= \frac{(X - \vartheta)^2}{(\sigma^2)^2}, \\ I = \mathbf{M} [(\ln p(X; \vartheta))'_{\vartheta}]^2 &= \frac{\mathbf{M}(X - \vartheta)^2}{(\sigma^2)^2} = \frac{\mathbf{D}X}{(\sigma^2)^2} = \frac{1}{\sigma^2}, & e(\vartheta^*) &= 1. \end{aligned}$$

И в этом примере оценка ϑ^* является эффективной. \square

Пример 12. Оценим неизвестную дисперсию $\vartheta = \mathbf{D}X$ нормального закона при известном среднем m . Плотность нормального распределения представима в виде

$$p(x; \vartheta) = e^{S(x) A(\vartheta) + B(\vartheta)},$$

где

$$S(x) = (x - m)^2, \quad A(\vartheta) = -\frac{1}{2\vartheta}, \quad B(\vartheta) = -\ln \sqrt{2\pi\vartheta},$$

т. е. по отношению к неизвестной дисперсии ϑ принадлежит экспоненциальному семейству. Поэтому эффективная оценка ϑ^* дисперсии ϑ должна по формуле (8) иметь вид

$$\vartheta^* = \frac{1}{gn} [S(X_1) + \dots + S(X_n)] = \frac{1}{gn} \sum_{i=1}^n (X_i - m)^2.$$

С другой стороны, нетрудно видеть, что $\vartheta A'(\vartheta) = -B'(\vartheta)$, откуда следует несмещенность оценки

$$\vartheta^* = \frac{1}{n} \sum_{i=1}^n (X_i - m)^2$$

и, значит, ее эффективность. Впрочем, эффективность оценки ϑ^* легко установить и на основе неравенства Рао–Крамера.

Пусть теперь мы оцениваем не дисперсию, а среднее квадратичное отклонение $\vartheta = \sqrt{\mathbf{D}X}$. И в этом случае имеет место представление (7), только теперь

$$A(\vartheta) = -\frac{1}{2\vartheta^2}, \quad B(\vartheta) = -\ln \sqrt{2\pi\vartheta^2}.$$

Поэтому равенство $g\vartheta A'(\vartheta) = -B'(\vartheta)$ не превращается в тождество ни при каком выборе g , и, значит, эффективной (в смысле Рао–Крамера) оценки среднего квадратичного отклонения нормального закона не существует. Рассмотрим оценку

$$\vartheta^* = \sqrt{\frac{n}{2}} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n+1}{2}\right)} \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - m)^2},$$

равную корню квадратному из оценки дисперсии с точностью до постоянного множителя $\sqrt{n/2} \Gamma(n/2)/\Gamma((n+1)/2)$. Читателю предлагается проверить, что оценка ϑ^* несмещенная. Кроме того, в следующем параграфе будет показано, что среди всех несмешанных оценок среднего квадратичного отклонения ϑ она имеет минимальную дисперсию (хотя и не является эффективной). \square

Пример 13. Пусть выборка X_1, \dots, X_n произведена из генеральной совокупности с равномерным на интервале $(0, \vartheta)$ теоретическим распределением. Оценим неизвестный параметр ϑ . Обозначим через X_n^* максимальный член вариационного ряда. В качестве оценки параметра ϑ возьмем

$$\vartheta^* = \frac{n+1}{n} X_n^*.$$

Функция распределения $F_{X_n^*}(x)$ статистики X_n^* задается формулой

$$F_{X_n^*}(x) = \mathbb{P}\{X_n^* < x\} = \mathbb{P}\{X_1 < x, \dots, X_n < x\} =$$

$$= \mathbb{P}\{X_1 < x\} \cdots \mathbb{P}\{X_n < x\} = \left(\frac{x}{\vartheta}\right)^n \quad (0 \leq x \leq \vartheta).$$

Тогда

$$\mathbf{M}\vartheta^* = \frac{n+1}{n} \int_0^\vartheta nx \frac{x^{n-1}}{\vartheta^n} dx = \vartheta.$$

Значит, оценка ϑ^* несмешенная. Далее,

$$\begin{aligned} \mathbf{M}(\vartheta^*)^2 &= \left(\frac{n+1}{n}\right)^2 \int_0^\vartheta nx^2 \frac{x^{n-1}}{\vartheta^n} dx = \frac{(n+1)^2}{n(n+2)} \vartheta, \\ \mathbf{D}\vartheta^* &= \mathbf{M}(\vartheta^*)^2 - (\mathbf{M}\vartheta^*)^2 = \frac{\vartheta^2}{n(n+2)}. \end{aligned}$$

Мы видим, что дисперсия оценки ϑ^* при $n \rightarrow \infty$ убывает, как $1/n^2$. Такая оценка оказалась более эффективной, поскольку дисперсия эффективной оценки убывает только, как $1/n$. Разгадка парадокса чрезвычайно проста: для данного семейства не выполнены условия регулярности, необходимые при доказательстве неравенства Рао–Крамера. Используя понятие достаточной статистики, в следующем параграфе мы докажем минимальность дисперсии данной оценки. \square

В заключение этого параграфа отметим, что эффективные по Рао–Крамеру оценки существуют крайне редко. Правда, как мы увидим в параграфе 4, эффективность по Рао–Крамеру играет существенную роль в асимптотическом анализе оценок, получаемых методом максимального правдоподобия. Кроме того, существуют обобщения неравенства Рао–Крамера (например, неравенство Бхаттачария [7]), позволяющие доказывать оптимальность более широкого класса оценок.

В следующем параграфе мы рассмотрим другой подход к определению оценок с минимальной дисперсией, базирующийся на достаточных статистиках.

Наиболее распространенные методы нахождения оценок приводятся в параграфах 3–6.

Наконец, в параграфе 7 описан подход к построению доверительных интервалов для неизвестных параметров.

2. Достаточные оценки

Первый шаг в поисках другого (не основанного на неравенстве Рао–Крамера) принципа построения оценок с минимальной дисперсией состоит во введении понятия достаточной статистики (отметим, что достаточные статистики играют в современной математической статистике весьма важную роль, причем как при оценке неизвестных

параметров, так и при проверке статистических гипотез). Назовем k -мерную статистику

$$S = (S_1, \dots, S_k) = (S_1(X_1, \dots, X_n), \dots, S_k(X_1, \dots, X_n))$$

достаточной для параметра ϑ , если условное распределение $F_{X_1, \dots, X_n}(x_1, \dots, x_n | S = s)$ выборки X_1, \dots, X_n при условии $S = s$ не зависит от параметра ϑ .

Пример 14. Пусть X_i — число успехов в i -м испытании Бернулли (см. пример 1). Рассмотрим статистику

$$S = X_1 + \dots + X_n$$

— общее число успехов в n испытаниях Бернулли. Покажем, что она является достаточной для вероятности успеха ϑ . Для этого найдем условное распределение $P\{X_1 = x_1, \dots, X_n = x_n | S = s\}$. Воспользовавшись определением условной вероятности, получаем

$$P\{X_1 = x_1, \dots, X_n = x_n | S = s\} = \frac{P\{X_1 = x_1, \dots, X_n = x_n, S = s\}}{P\{S = s\}}. \quad (9)$$

Если $x_1 + \dots + x_n = s$, то вероятность $P\{X_1 = x_1, \dots, X_n = x_n, S = s\}$ совпадает с вероятностью $P\{X_1 = x_1, \dots, X_n = x_n\}$, т. е.

$P\{X_1 = x_1, \dots, X_n = x_n, S = s\} = P\{X_1 = x_1, \dots, X_n = x_n\} = \vartheta^s(1 - \vartheta)^{n-s}$ (напомним еще раз, что каждое x_i может принимать здесь только значение 0 или 1, причем $x_1 + \dots + x_n = s$). Поскольку вероятность $P\{S = s\}$ определяется формулой Бернулли

$$P\{S = s\} = \binom{n}{s} \vartheta^s (1 - \vartheta)^{n-s},$$

то из (9) получаем, что

$$P\{X_1 = x_1, \dots, X_n = x_n | S = s\} = \frac{\vartheta^s(1 - \vartheta)^{n-s}}{\binom{n}{s} \vartheta^s (1 - \vartheta)^{n-s}} = \frac{1}{\binom{n}{s}},$$

т. е. не зависит от ϑ . Если же $x_1 + \dots + x_n \neq s$, то

$$P\{X_1 = x_1, \dots, X_n = x_n, S = s\} = 0,$$

откуда

$$P\{X_1 = x_1, \dots, X_n = x_n | S = s\} = 0,$$

т. е. опять-таки не зависит от ϑ . Таким образом, S — достаточная статистика. \square

Очевидно, что использовать приведенное выше определение для проверки достаточности конкретных статистик весьма сложно, особенно в непрерывном случае. Простой критерий достаточности задается следующей теоремой.

Теорема 2 (факторизационная теорема Неймана–Фишера). Для того чтобы статистика $S = S(X_1, \dots, X_n)$ была достаточной для параметра ϑ , необходимо и достаточно, чтобы ряд распределения

$$P(x_1, \dots, x_n; \vartheta) = P(x_1; \vartheta) \cdots P(x_n; \vartheta)$$

в дискретном случае или плотность распределения

$$p(x_1, \dots, x_n; \vartheta) = p(x_1; \vartheta) \cdots p(x_n; \vartheta)$$

в непрерывном случае выборки X_1, \dots, X_n были представимы в виде

$$\begin{aligned} P(x_1, \dots, x_n; \vartheta) &= A(x_1, \dots, x_n) B(S; \vartheta), \\ p(x_1, \dots, x_n; \vartheta) &= A(x_1, \dots, x_n) B(S; \vartheta), \end{aligned} \quad (10)$$

где функция $A(x_1, \dots, x_n)$ зависит только от x_1, \dots, x_n , а функция $B(S; \vartheta)$ — только от S и ϑ .

Доказательство. Для простоты изложения ограничимся только дискретным случаем. По определению условной вероятности,

$$\mathbb{P}\{X_1 = x_1, \dots, X_n = x_n \mid S = s\} = \frac{\mathbb{P}\{X_1 = x_1, \dots, X_n = x_n, S = s\}}{\mathbb{P}\{S = s\}}. \quad (11)$$

Очевидно, что числитель в правой части (11) совпадает с вероятностью $\mathbb{P}\{X_1 = x_1, \dots, X_n = x_n\}$ в том случае, когда $s = S(x_1, \dots, x_n)$, и равен нулю в противном. Поскольку событиями нулевой вероятности можно пренебречь, то ограничимся случаем $s = S(x_1, \dots, x_n)$ и запишем (11) в виде

$$\mathbb{P}\{X_1 = x_1, \dots, X_n = x_n \mid S = s\} = \frac{\mathbb{P}\{X_1 = x_1, \dots, X_n = x_n\}}{\mathbb{P}\{S = s\}}. \quad (12)$$

Теперь, если S — достаточная статистика, то левая часть (12) не зависит от ϑ . Обозначая ее через $A(x_1, \dots, x_n)$, а $\mathbb{P}\{S = s\}$ — через $B(s; \vartheta)$, приходим к (10), что доказывает необходимость (10).

И наоборот, пусть выполнено (10). Тогда

$$\begin{aligned} \mathbb{P}\{S = s\} &= \sum_{x_1, \dots, x_n : S(x_1, \dots, x_n) = s} P(x_1, \dots, x_n; \vartheta) = \\ &= B(s; \vartheta) \sum_{x_1, \dots, x_n : S(x_1, \dots, x_n) = s} A(x_1, \dots, x_n). \end{aligned}$$

Подставляя последнее равенство в (12), имеем

$$\begin{aligned} \mathbb{P}\{X_1 = x_1, \dots, X_n = x_n \mid S = s\} &= \\ &= A(x_1, \dots, x_n) \left[\sum_{y_1, \dots, y_n : S(y_1, \dots, y_n) = s} A(y_1, \dots, y_n) \right]^{-1}, \end{aligned}$$

т. е. не зависит от ϑ , а значит, статистика S является достаточной. \square

Замечание к теореме 2. Очевидно, что представление (10) справедливо с точностью до функции $g = g(S)$, зависящей только от $S = S(x_1, \dots, x_n)$.

Пример 15. Пусть X_1, \dots, X_n — выборка из генеральной совокупности с теоретической функцией распределения, являющейся нормальной со средним ϑ_1 и дисперсией ϑ_2 . Покажем, что (двумерная) статистика $S = (S_1, S_2)$, где

$$S_1 = \frac{1}{n}(X_1 + \dots + X_n), \quad S_2 = (X_1 - S_1)^2 + \dots + (X_n - S_1)^2,$$

является достаточной для (двумерного) параметра $\vartheta = (\vartheta_1, \vartheta_2)$ (см. также пример 3 из гл. 1). Действительно, плотность распределения $p(x_1, \dots, x_n; \vartheta)$ выборки X_1, \dots, X_n представима в виде

$$p(x_1, \dots, x_n; \vartheta) = (2\pi\vartheta_2)^{-n/2} e^{-\frac{S_2+n(S_1-\vartheta_1)^2}{2\vartheta_2}},$$

т. е. имеет вид (10), где

$$A(x_1, \dots, x_n) = 1, \quad B(S; \vartheta) = (2\pi\vartheta_2)^{-n/2} e^{-\frac{S_2+n(S_1-\vartheta_1)^2}{2\vartheta_2}}. \quad \square$$

Пример 16. Пусть X_1, \dots, X_n — выборка из генеральной совокупности с равномерным на интервале $(0, \vartheta)$ теоретическим распределением (см. пример 13). Покажем, что максимальный член вариационного ряда

$$S = S(X_1, \dots, X_n) = X_n^*$$

является (одномерной) достаточной статистикой для ϑ . Действительно, вспоминая, что плотность $p(x; \vartheta)$ равномерно распределенной на интервале $(0, \vartheta)$ величины равна $1/\vartheta$ при $x \in (0, \vartheta)$ и нулю в противном случае, получаем для плотности распределения выборки X_1, \dots, X_n выражение

$$p(x_1, \dots, x_n; \vartheta) = \begin{cases} \left(\frac{1}{\vartheta}\right)^n, & \text{если } 0 < x_i < \vartheta \text{ для всех } i; \\ 0 & \text{в противном случае.} \end{cases}$$

В частности, область изменения каждого аргумента x_i при отличной от нуля плотности распределения зависит от параметра ϑ . Рассмотрим функцию

$$S = S(x_1, \dots, x_n) = \max_{1 \leq i \leq n} x_i$$

и положим

$$B(S; \vartheta) = \begin{cases} \left(\frac{1}{\vartheta}\right)^n, & \text{если } 0 < S < \vartheta; \\ 0 & \text{в противном случае,} \end{cases}$$

$$A(x_1, \dots, x_n) = \begin{cases} 1, & \text{если } x_i > 0 \text{ для всех } i; \\ 0 & \text{в противном случае.} \end{cases}$$

С учетом введенных функций

$$p(x_1, \dots, x_n; \vartheta) = A(x_1, \dots, x_n) B(S; \vartheta).$$

Здесь уже при определении функции $A(x_1, \dots, x_n)$ на x_i сверху не наложено никаких ограничений, поскольку они автоматически ограничены своим максимальным значением S , которое в свою очередь не превосходит ϑ . Но это означает, что функция $A(x_1, \dots, x_n)$ не зависит от параметра ϑ и в соответствии с теоремой 2 статистика

$$S = S(X_1, \dots, X_n) = X_n^*$$

является достаточной для параметра ϑ . \square

Пример 17. Покажем, что для экспоненциального семейства (7) существует одномерная достаточная статистика. Этот факт легко установить, если подставить выражение (7) в формулу для плотности распределения выборки X_1, \dots, X_n :

$$\begin{aligned} p(x_1, \dots, x_n; \vartheta) &= p(x_1; \vartheta) \cdots p(x_n; \vartheta) = \\ &= h(x_1) \cdots h(x_n) e^{[S(x_1) + \dots + S(x_n)] A(\vartheta) + nB(\vartheta)}. \end{aligned}$$

Полагая теперь

$$A(x_1, \dots, x_n) = h(x_1) \cdots h(x_n),$$

$$S = S(x_1, \dots, x_n) = S(x_1) + \dots + S(x_n),$$

$$B(S; \vartheta) = e^{SA(\vartheta) + nB(\vartheta)},$$

видим, что одномерная статистика

$$S(X_1, \dots, X_n) = S(X_1) + \dots + S(X_n)$$

является достаточной для параметра ϑ . \square

Как уже говорилось в гл. 1, смысл достаточной статистики S заключается в том, что она включает в себя всю ту информацию о неизвестном параметре ϑ , которая содержится в исходной выборке X_1, \dots, X_n . Интуиция подсказывает нам: оценка с наименьшей дисперсией (если она существует) должна зависеть только от достаточной статистики S . И действительно, следующий наш шаг будет заключаться в переходе от произвольной оценки ϑ^* к оценке ϑ_S^* , зависящей только от достаточной статистики S , причем этот переход совершится таким образом, чтобы дисперсия оценки ϑ_S^* не превосходила дисперсии исходной оценки ϑ^* .

Начиная с этого момента и до конца параграфа будем для простоты предполагать, что неизвестный параметр ϑ является одномерным.

Пусть имеется некоторая оценка ϑ^* этого параметра, а также (произвольная) статистика S . Рассмотрим условное математическое ожидание $\mathbf{M}(\vartheta^* | S)$ случайной величины ϑ^* при условии S (см. часть 1, гл. 7, параграф 5). Следующее утверждение, играющее основную роль в наших рассуждениях, было получено независимо Д. Блекузелом, М. М. Рао и А. Н. Колмогоровым.

Теорема 3 (улучшение оценки по достаточной статистике). Пусть S — достаточная статистика, а ϑ^* — несмещенная оценка параметра ϑ . Тогда условное математическое ожидание $\vartheta_S^* = \mathbf{M}(\vartheta^* | S)$ является несмещенной оценкой параметра ϑ , зависящей только от достаточной статистики S и удовлетворяющей неравенству

$$\mathbf{D}\vartheta_S^* \leq \mathbf{D}\vartheta^* \quad (13)$$

при всех ϑ .

Доказательство. В силу достаточности статистики S условное распределение, а значит, и условное математическое ожидание оценки ϑ^* при условии S не зависит от неизвестного параметра ϑ (для произвольной статистики S функция ϑ_S^* , вообще говоря, может зависеть от ϑ), т. е. ϑ_S^* представляет собой оценку параметра ϑ , причем зависящую только от S .

Далее, из равенства

$$\mathbf{M}\vartheta_S^* = \mathbf{M}[\mathbf{M}(\vartheta^* | S)] = \mathbf{M}\vartheta^*$$

для условного математического ожидания немедленно следует несмещенность оценки ϑ_S^* .

Наконец,

$$\begin{aligned} \mathbf{D}\vartheta^* &= \mathbf{M}(\vartheta^* - \vartheta)^2 = \mathbf{M}(\vartheta^* - \vartheta_S^* + \vartheta_S^* - \vartheta)^2 = \\ &= \mathbf{M}(\vartheta_S^* - \vartheta)^2 + 2\mathbf{M}[(\vartheta_S^* - \vartheta)(\vartheta^* - \vartheta_S^*)] + \mathbf{M}(\vartheta^* - \vartheta_S^*)^2. \end{aligned}$$

Используя опять свойство условного математического ожидания, получаем

$$\begin{aligned} \mathbf{M}[(\vartheta_S^* - \vartheta)(\vartheta^* - \vartheta_S^*)] &= \mathbf{M}(\mathbf{M}[(\vartheta_S^* - \vartheta)(\vartheta^* - \vartheta_S^*) | S]) = \\ &= \mathbf{M}((\vartheta_S^* - \vartheta)\mathbf{M}[(\vartheta^* - \vartheta_S^*) | S]) = 0. \end{aligned}$$

Поэтому

$$\mathbf{D}\vartheta^* = \mathbf{D}\vartheta_S^* + \mathbf{M}(\vartheta^* - \vartheta_S^*)^2 \geq \mathbf{D}\vartheta_S^*. \quad \square$$

Замечание 1 к теореме 3. Неравенство (13) превращается для некоторого ϑ в равенство тогда и только тогда, когда $\vartheta^* = \vartheta_S^*$ (почти всюду по мере $F(x; \vartheta)$).

Замечание 2 к теореме 3. Утверждение теоремы остается в силе и для смещенной оценки ϑ^* . В частности, $\mathbf{M}\vartheta^* = \mathbf{M}\vartheta_S^*$.

Смысл теоремы 3 заключается в том, что взятие условного математического ожидания, т. е. переход к оценке ϑ_S^* , зависящей только от достаточной статистики S , не ухудшает любую оценку ϑ^* при всех значениях неизвестного параметра ϑ .

Пример 18. Пусть X_1, \dots, X_n — выборка из нормально распределенной генеральной совокупности с неизвестным средним ϑ и известной дисперсией σ^2 . В примере 9 было показано, что оценка $\vartheta^* = X_1$ даже не является состоятельной оценкой ϑ , хотя она и несмещенная. Рассмотрим статистику

$$S = X_1 + \dots + X_n.$$

Нетрудно показать, что статистика S является достаточной для параметра ϑ . Поэтому мы можем определить новую оценку $\vartheta_S^* = \mathbf{M}(\vartheta^* | S)$. Для ее вычисления заметим, что величины $\vartheta^* = X_1$ и $S = X_1 + \dots + X_n$ имеют двумерное нормальное распределение со средними $\mathbf{M}\vartheta^* = \vartheta$ и $\mathbf{M}S = n\vartheta$, дисперсиями $\mathbf{D}\vartheta^* = \sigma^2$ и $\mathbf{D}S = n\sigma^2$ и ковариацией $\mathbf{cov}(\vartheta^*, S) = \mathbf{M}[(\vartheta^* - \vartheta)(S - n\vartheta)] = \sigma^2$. Но тогда, как известно из курса теории вероятностей, условное распределение ϑ^* при условии $S = s$ также является нормальным со средним значением $\vartheta + \rho\sqrt{\sigma}(s - n\vartheta)/\sqrt{n\sigma^2}$, как раз и представляющим собой значение $\mathbf{M}(\vartheta^* | S)$ при $S = s$. Поскольку коэффициент корреляции $\rho = \mathbf{cov}(\vartheta^*, S)/\sqrt{\mathbf{D}\vartheta^*\mathbf{D}S} = 1/\sqrt{n}$, то среднее значение условного распределения ϑ^* совпадает с s/n и окончательно получаем

$$\vartheta_S^* = \frac{S}{n} = \frac{1}{n} (X_1 + \dots + X_n) = m^*.$$

Иными словами, мы из совсем плохой оценки $\vartheta^* = X_1$ получили эффективную (см. пример 11) оценку $\vartheta_S^* = m^*$. \square

Рассмотренный пример приоткрывает нам те возможности, которые несет с собой теорема 3. Однако, прежде чем сделать последний шаг, введем еще одно определение. Назовем статистику $S = S(X_1, \dots, X_n)$ полной для семейства распределений $F(x; \vartheta)$, если из того, что

$$\mathbf{M}g(S) = \mathbf{M}g(S(X_1, \dots, X_n)) =$$

$$= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(S(x_1, \dots, x_n)) p(x_1; \vartheta) \cdots p(x_n; \vartheta) dx_1 \cdots dx_n = 0$$

при всех ϑ (мы для простоты предположили существование плотности распределения $p(x; \vartheta)$), следует, что функция $g(s)$ тождественно равна нулю. Теперь мы в состоянии сформулировать окончательный итог наших поисков.

Теорема 4 (минимальность дисперсии оценки, зависящей от полной достаточной статистики). Пусть S — полная достаточная статистика, а ϑ^* — несмешенная оценка неизвестного параметра ϑ . Тогда

$$\vartheta_S^* = \mathbf{M}(\vartheta^* | S)$$

является единственной несмешенной оценкой с минимальной дисперсией.

Доказательство теоремы немедленно вытекает из предыдущих результатов. Действительно, в силу теоремы 3 оценка с минимальной дисперсией обязательно должна находиться среди оценок, зависящих только от достаточной статистики S ; в противном случае ее можно было бы улучшить с помощью условного математического ожидания. Но среди оценок, зависящих только от S , может быть максимум одна несмешенная. В самом деле, если таких оценок две: $\vartheta_1^* = \vartheta_1^*(S)$ и $\vartheta_2^* = \vartheta_2^*(S)$, то функция

$$g = g(S) = \vartheta_1^* - \vartheta_2^*$$

имеет при всех значениях ϑ математическое ожидание

$$\mathbf{M}g(S) = \mathbf{M}\vartheta_1^* - \mathbf{M}\vartheta_2^* = \vartheta - \vartheta = 0,$$

что в силу полноты статистики S влечет за собой равенство $g(s)$ нулю. Само же существование несмешенной оценки $\vartheta_S^* = \mathbf{M}(\vartheta^* | S)$, зависящей только от S , гарантируется существованием просто несмешенной оценки ϑ^* . \square

Перейдем к обсуждению полученных результатов.

Условие полноты статистики S , как мы видим, сводится к единственности несмешенной оценки ϑ^* , зависящей только от статистики S . Нам не известно общих теорем, которые давали бы простые правила проверки полноты произвольной статистики S . Однако, как мы увидим из примеров, в конкретных случаях кустарные способы обычно дают хорошие результаты.

Сравнение размерностей полной статистики S и оцениваемого параметра ϑ дает право говорить, что, как правило, статистика S должна иметь ту же размерность, что и ϑ , а поскольку мы ограничились одномерным параметром ϑ , то S также должна быть одномерной. Это приводит к следующим полезным определениям. Оценка ϑ^* называется *достаточной*, если она является достаточной как одномерная статистика. Аналогично, назовем оценку ϑ^* *полной*, если она является полной статистикой.

Сформулируем очевидное следствие из теоремы 4, которое удобно применять во многих частных случаях.

Следствие из теоремы 4. Если оценка ϑ^* несмешенная и зависит только от полной достаточной статистики S , то она имеет минимальную дисперсию.

Пример 19. Пусть X_1, \dots, X_n — выборка из генеральной совокупности, распределенной по нормальному закону с известным средним m и неизвестным средним квадратичным отклонением ϑ ($\vartheta > 0$). Нетрудно показать, что статистика

$$S = (X_1 - m)^2 + \dots + (X_n - m)^2$$

является достаточной для параметра ϑ . Покажем, что она также полная. Для этого вспомним (см. параграф 4 гл. 1), что случайная величина $\chi^2 = S/\vartheta^2$ имеет χ^2 -распределение с n степенями свободы, а значит, статистика $S = \vartheta^2 \chi^2$ имеет плотность распределения

$$p_S(x) = \frac{x^{n/2-1}}{2^{n/2} \vartheta^n \Gamma\left(\frac{n}{2}\right)} e^{-x/(2\vartheta^2)}.$$

Пусть теперь $g(x)$ — такая функция, что $\mathbf{M}g(S) = 0$ при всех ϑ . Положим

$$g_1(x) = \frac{2^{-n/2} x^{n/2-1}}{\Gamma\left(\frac{n}{2}\right)} g(x).$$

Тогда

$$\mathbf{M}g(S) = \frac{1}{\vartheta^n} \int_0^\infty g_1(x) e^{-x/(2\vartheta^2)} dx = \frac{1}{\vartheta^n} \tilde{g}_1\left(\frac{1}{2\vartheta^2}\right) = 0,$$

где $\tilde{g}_1(\omega)$ — преобразование Лапласа функции $g_1(x)$. Отсюда, в частности, следует, что $\tilde{g}_1(\omega) = 0$ для всех $\omega > 0$. Но из теории преобразований Лапласа известно, что в этом случае оригинал $g_1(x)$, а значит, и функция $g(x)$ также должны тождественно равняться нулю, что и доказывает полноту статистики S .

Рассмотрим теперь оценку

$$\vartheta^* = \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n+1}{2}\right)} \sqrt{\frac{S}{2}}$$

(см. пример 12) неизвестного среднего квадратичного отклонения ϑ . Эта оценка несмещенная и зависит только от полной достаточной статистики S . Поэтому по следствию из теоремы 4 она имеет минимальную дисперсию, хотя, как было показано в примере 12, и не является эффективной по Рао–Крамеру. \square

Пример 20. Рассмотрим оценку

$$\vartheta_n^* = \frac{n+1}{n} X_n^*$$

параметра ϑ равномерного на интервале $(0, \vartheta)$ распределения (см. пример 13). В примере 13 показано, что эта оценка несмещенная. Статистика X_n^* является достаточной (см. пример 16). Покажем, наконец, что X_n^* — полная статистика. Действительно, для любой функции $g(x)$

$$\mathbf{M}g(S) = \int_0^\vartheta g(x) p_{X_n^*}(x) dx = \frac{n}{\vartheta^n} \int_0^\vartheta g(x) x^{n-1} dx.$$

Отсюда, в частности, следует, что если $\mathbf{M}g(S) = 0$ при всех ϑ , то

$$x^{n-1} g(x) = \left(\int_0^x g(y) y^{n-1} dy \right)' = 0$$

при всех x . Поэтому $g(x) \equiv 0$ и статистика X_n^* полная.

Таким образом, в силу следствия из теоремы 4 и в этом примере оценка ϑ^* имеет минимальную дисперсию. \square

3. Метод моментов

Пусть мы имеем выборку X_1, \dots, X_n из генеральной совокупности с теоретической функцией распределения $F(x)$, принадлежащей k -параметрическому семейству $F(x; \vartheta_1, \dots, \vartheta_k)$ с неизвестными параметрами $\vartheta_1, \dots, \vartheta_k$, которые нужно оценить. Поскольку нам известен вид теоретической функции распределения, мы можем вычислить первые k теоретических моментов. Эти моменты, разумеется, будут зависеть от k неизвестных параметров $\vartheta_1, \dots, \vartheta_k$:

$$\begin{aligned} m_1 &= \mathbf{M}X = m_1(\vartheta_1, \dots, \vartheta_k), \\ m_2 &= \mathbf{M}X^2 = m_2(\vartheta_1, \dots, \vartheta_k), \\ &\dots \\ m_k &= \mathbf{M}X^k = m_k(\vartheta_1, \dots, \vartheta_k). \end{aligned}$$

Суть метода моментов заключается в следующем: так как выборочные моменты являются состоятельными оценками теоретических моментов (см. пример 8), мы можем в написанной системе равенств при большом объеме выборки n теоретические моменты m_1, \dots, m_k заменить на выборочные m_1^*, \dots, m_k^* , а затем, решая эту систему относительно $\vartheta_1, \dots, \vartheta_k$, найти оценки неизвестных параметров. Таким образом, в *методе моментов* оценки $\vartheta_1^*, \dots, \vartheta_k^*$ неизвестных параметров $\vartheta_1, \dots, \vartheta_k$ определяются из системы уравнений

$$\begin{aligned} m_1^* &= m_1(\vartheta_1^*, \dots, \vartheta_k^*), \\ m_2^* &= m_2(\vartheta_1^*, \dots, \vartheta_k^*), \\ &\dots \\ m_k^* &= m_k(\vartheta_1^*, \dots, \vartheta_k^*). \end{aligned}$$

Можно показать, что при условии непрерывной зависимости решения этой системы от начальных условий m_1^*, \dots, m_k^* оценки, полученные методом моментов, будут состоятельными. Более того, справедлива следующая теорема.

Теорема 5 (асимптотическая нормальность оценок, полученных методом моментов). *При некоторых условиях, наложенных на семейство $F(x; \vartheta_1, \dots, \vartheta_k)$, совместное распределение случайных величин*

$$\sqrt{n} (\vartheta_1^* - \vartheta_1), \dots, \sqrt{n} (\vartheta_k^* - \vartheta_k)$$

при $n \rightarrow \infty$ сходится к (многомерному) нормальному закону с нулевыми средними и матрицей ковариаций, зависящей от теоретических моментов m_1, \dots, m_{2k} и матрицы $(\partial m_i / \partial \vartheta_j)$.

Доказательство. Будем полагать, что выполнены следующие условия:

- а) параметры $\vartheta_1, \dots, \vartheta_k$ однозначно определяются своими моментами m_1, \dots, m_k ;

б) существует теоретический момент m_{2k} порядка $2k$ (это эквивалентно существованию дисперсий у выборочных моментов m_1^*, \dots, m_k^*);

в) функция

$$\mathbf{m}(\vartheta) = (m_1(\vartheta_1, \dots, \vartheta_k), \dots, m_k(\vartheta_1, \dots, \vartheta_k))$$

дифференцируема по $\boldsymbol{\vartheta} = (\vartheta_1, \dots, \vartheta_k)$ с отличным от нуля якобианом $|\partial m_i / \partial \vartheta_j|$.

Доказательство теоремы проведем для одномерного случая, предоставляемый общий случай читателю. Оно является комбинацией следующих результатов: теоремы о дифференцируемости обратного отображения и центральной предельной теоремы. Действительно, поскольку существует дисперсия $\mathbf{D}X$, то при каждом истинном значении $\hat{\vartheta}$ параметра ϑ в силу центральной предельной теоремы выборочное среднее

$$m^* = \frac{1}{n} (X_1 + \dots + X_n)$$

асимптотически при $n \rightarrow \infty$ распределено по нормальному закону с параметрами $\hat{m} = m_1(\hat{\vartheta})$ и $\hat{\sigma}^2/n = [m_2(\hat{\vartheta}) - (m_1(\hat{\vartheta}))^2]/n$. С другой стороны, сама оценка ϑ^* записывается в виде

$$\vartheta^* = m_1^{(-1)}(m^*),$$

где $m_1^{(-1)}(x)$ — обратная к $m_1(\vartheta)$ функция. В силу сделанных предположений обратное отображение $m_1^{(-1)}(x)$ в окрестности точки \hat{m} приближенно представляет собой линейную функцию

$$m_1^{(-1)}(x) = a(x - \hat{m}) + \hat{\vartheta},$$

причем $a = 1/m'(\hat{\vartheta})$. Но тогда и случайная величина ϑ^* как приближенно линейное преобразование приближенно нормальней случайной величины m^* распределена приближенно поциальному закону со средним $\hat{\vartheta}$ и дисперсией $a^2 \hat{\sigma}^2/n$. Это доказывает утверждение теоремы. \square

Пример 21. Найдем методом моментов оценку неизвестной вероятности успеха ϑ в схеме Бернулли. Поскольку в схеме Бернулли только один неизвестный параметр, для его определения необходимо приравнять теоретическое математическое ожидание числа успехов в одном испытании $m = m(\vartheta) = \vartheta$ выборочному среднему $m^* = (X_1 + \dots + X_n)/n$:

$$m^* = \vartheta^*.$$

Итак, оценка ϑ^* , полученная методом моментов, представляет собой наблюденную частоту успехов. Свойства этой оценки были нами достаточно полно исследованы в примерах 1, 4, 7 и 10. \square

Пример 22. Выборка X_1, \dots, X_n произведена из генеральной совокупности с теоретической функцией распределения, имеющей гамма-плотность

$$p(x) = p(x; \vartheta_1, \vartheta_2) = \frac{\vartheta_2^{\vartheta_1} x^{\vartheta_1-1}}{\Gamma(\vartheta_1)} e^{-\vartheta_2 x} \quad (x > 0)$$

с двумя неизвестными параметрами ϑ_1 и ϑ_2 . Первые два момента случайной величины X , имеющей гамма-распределение, задаются формулами:

$$m_1 = \mathbf{M}X = \frac{\vartheta_1}{\vartheta_2}, \quad m_2 = \mathbf{M}X^2 = \frac{\vartheta_1(\vartheta_1+1)}{\vartheta_2^2}.$$

Отсюда для определения оценок ϑ_1^* и ϑ_2^* неизвестных параметров ϑ_1 и ϑ_2 получаем систему двух уравнений:

$$m^* = \frac{\vartheta_1^*}{\vartheta_2^*}, \quad \sigma^{2*} = \frac{\vartheta_1^*}{(\vartheta_2^*)^2},$$

решение которой имеет вид

$$\vartheta_1^* = \frac{(m^*)^2}{\sigma^{2*}}, \quad \vartheta_2^* = \frac{m^*}{\sigma^{2*}}.$$
□

Вообще говоря, в методе моментов не обязательно использовать первые k моментов. Более того, можно рассматривать моменты не обязательно целого порядка. Иногда для использования в методе моментов привлекают более или менее произвольные функции $g_1(x), \dots, g_k(x)$, сравнивая выборочные средние

$$m_{11}^* = \frac{1}{n} [g_1(X_1) + \dots + g_1(X_n)], \dots, m_{1k}^* = \frac{1}{n} [g_k(X_1) + \dots + g_k(X_n)]$$

функций $g_1(X), \dots, g_k(X)$ с теоретическими средними

$$m_{11} = \mathbf{M}g_1(X), \dots, m_{1k} = \mathbf{M}g_k(X).$$

Пример 23. Пусть выборка X_1, \dots, X_n произведена из нормальной генеральной совокупности с известным средним m и неизвестной дисперсией ϑ . Попробуем для оценивания ϑ применить метод моментов, взяв выборочное среднее m_1^* . Но теоретическое среднее $m_1 = m$ не зависит от параметра ϑ . Это означает, что использование выборочного среднего для оценивания неизвестной дисперсии неправомочно и нужно привлекать моменты других порядков. В частности, применяя второй выборочный момент m_2^* и вспоминая, что $m_2 = \vartheta + m^2$, получаем оценку

$$\vartheta^* = m_2^* - m^2.$$
□

Следует отметить, что оценки, полученные методом моментов, обычно имеют эффективность существенно меньше единицы и даже являются смещенными. Иногда из-за своей простоты они используются в качестве начального приближения для нахождения более эффективных оценок.

4. Метод максимального правдоподобия

Метод максимального правдоподобия является наиболее распространенным методом нахождения оценок. Пусть по-прежнему выборка X_1, \dots, X_n произведена из генеральной совокупности с неизвестной теоретической функцией распределения $F(x)$, принадлежащей известному однопараметрическому семейству $F(x; \vartheta)$. Функция

$$L(X_1, \dots, X_n) = L(X_1, \dots, X_n; \vartheta) = P(X_1; \vartheta) \cdots P(X_n; \vartheta)$$

в дискретном случае и

$$L(X_1, \dots, X_n) = L(X_1, \dots, X_n; \vartheta) = p(x_1; \vartheta) \cdots p(x_n; \vartheta)$$

в непрерывном называется *функцией правдоподобия*. Отметим, что в функции правдоподобия $L(X_1, \dots, X_n; \vartheta)$ элементы выборки X_1, \dots, X_n являются фиксированными параметрами, а ϑ — аргументом (а не истинным значением неизвестного параметра). Функция правдоподобия по своей сути представляет собой не что иное, как вероятность (в непрерывном случае плотность распределения) получить именно ту выборку X_1, \dots, X_n , которую мы реально имеем, если бы значение неизвестного параметра равнялось ϑ . Естественно поэтому в качестве оценки неизвестного параметра ϑ выбрать ϑ^* , доставляющее наибольшее значение функции правдоподобия $L(X_1, \dots, X_n; \vartheta)$. *Оценкой максимального правдоподобия* называется такое значение ϑ^* , для которого

$$L(X_1, \dots, X_n; \vartheta^*) = \max_{\vartheta} L(X_1, \dots, X_n; \vartheta).$$

При практической реализации метода максимального правдоподобия удобно пользоваться не самой функцией правдоподобия, а ее логарифмом.

Уравнением правдоподобия называется уравнение

$$\frac{\partial}{\partial \vartheta} \ln L(X_1, \dots, X_n; \vartheta) = 0.$$

Если функция правдоподобия дифференцируема по ϑ в каждой точке, то оценку максимального правдоподобия ϑ^* следует искать среди значений ϑ , удовлетворяющих уравнению правдоподобия или принадлежащих границе области допустимых значений ϑ . Для наиболее важных семейств $F(x; \vartheta)$ уравнение правдоподобия имеет единственное решение ϑ^* , которое и является оценкой максимального правдоподобия.

Пример 24. Найдем оценку неизвестной вероятности успеха ϑ в схеме Бернулли, но теперь уже в отличие от примера 21 методом максимального правдоподобия. Поскольку $P(X; \vartheta) = \vartheta$, если $X = 1$, и $P(X; \vartheta) = 1 - \vartheta$, если $X = 0$, то функцию правдоподобия можно записать так:

$$L(X_1, \dots, X_n; \vartheta) = P(X_1; \vartheta) \cdots P(X_n; \vartheta) = \vartheta^\mu (1 - \vartheta)^{n - \mu},$$

где $\mu = X_1 + \dots + X_n$ — суммарное число успехов в n испытаниях. Тогда уравнение правдоподобия принимает вид

$$(\ln L(X_1, \dots, X_n; \vartheta))'_{\vartheta} = (\mu \ln \vartheta + (n - \mu) \ln (1 - \vartheta))'_{\vartheta} = \frac{\mu}{\vartheta} - \frac{n - \mu}{1 - \vartheta} = 0.$$

Решая это уравнение, имеем

$$\vartheta^* = \frac{\mu}{n} = m^*.$$

Поскольку

$$(\ln L(X_1, \dots, X_n; \vartheta))''_{\vartheta \vartheta} = -\left(\frac{\mu}{\vartheta^2} + \frac{n - \mu}{(1 - \vartheta)^2}\right) > 0,$$

то $\ln L(X_1, \dots, X_n; \vartheta)$ представляет собой выпуклую вверх функцию ϑ . Значит, ϑ^* доставляет максимум функции правдоподобия $L(X_1, \dots, X_n; \vartheta)$, т. е. является оценкой максимального правдоподобия. Эта оценка представляет собой, как и в примере 21, наблюденную частоту успехов. \square

Оказывается, имеется тесная связь между эффективными оценками и оценками, полученными методом максимального правдоподобия. А именно, справедлива следующая теорема.

Теорема 6 (совпадение эффективной оценки с оценкой максимального правдоподобия). Если (естественно, при условиях регулярности теоремы 1) существует эффективная оценка $\vartheta_{\text{эфф}}^*$, то она является оценкой максимального правдоподобия $\vartheta_{\text{пп}}^*$.

Доказательство теоремы 6 представляет собой дальнейшее уточнение доказательства теоремы 1. Действительно, как следует из замечания 1 к теореме 1, из существования эффективной оценки $\vartheta_{\text{эфф}}^*$ вытекает (6) и (8) (с $g = 1$). Отсюда и из (4) следует равенство

$$1 = c(\vartheta) \mathbf{M}(\vartheta_{\text{эфф}}^* - \vartheta)^2 = \frac{c(\vartheta)}{nI}.$$

Поэтому из условия строгой положительности информации I вытекает строгая положительность $c(\vartheta)$, которая в свою очередь влечет за собой единственность решения

$$\vartheta_{\text{пп}}^* = \frac{1}{n} [\vartheta^*(X_1) + \dots + \vartheta^*(X_n)]$$

уравнения правдоподобия

$$\frac{\partial}{\partial \vartheta} \ln L(X_1, \dots, X_n; \vartheta) = \sum_{i=1}^n \frac{p'_{\vartheta}(X_i; \vartheta)}{p(X_i; \vartheta)} = c(\vartheta) \sum_{i=1}^n [\vartheta^*(X_i) - \vartheta] = 0.$$

Это решение совпадает с эффективной оценкой $\vartheta_{\text{эфф}}^*$ и задает единственный максимум функции правдоподобия $L(X_1, \dots, X_n; \vartheta)$. \square

В общем случае оценка максимального правдоподобия может быть не только неэффективной, но и смещенной. Тем не менее она обладает свойством асимптотической эффективности в следующем смысле.

Теорема 7 (асимптотическая эффективность оценки максимального правдоподобия). При некоторых условиях на семейство $F(x; \vartheta)$ уравнение правдоподобия имеет решение, при $n \rightarrow \infty$ асимптотически распределенное по нормальному закону со средним ϑ и дисперсией $1/(nI)$, где I — информация Фишера.

Доказательство. Сначала сформулируем условия теоремы (см. [9]), которые, как мы увидим далее, гарантируют возможность дифференцируемости под знаком интеграла и разложения $\partial \ln p(x; \vartheta) / \partial \vartheta$ по ϑ в ряд Тейлора до первого члена:

а) для (почти) всех x существуют производные

$$\frac{\partial \ln p(x; \vartheta)}{\partial \vartheta}, \quad \frac{\partial^2 \ln p(x; \vartheta)}{\partial \vartheta^2}, \quad \frac{\partial^3 \ln p(x; \vartheta)}{\partial \vartheta^3};$$

б) при всех ϑ справедливы неравенства

$$\left| \frac{\partial p(x; \vartheta)}{\partial \vartheta} \right| < G_1(x), \quad \left| \frac{\partial^2 p(x; \vartheta)}{\partial \vartheta^2} \right| < G_2(x), \quad \left| \frac{\partial^3 p(x; \vartheta)}{\partial \vartheta^3} \right| < H(x),$$

где функции $G_1(x)$ и $G_2(x)$ интегрируемы на $(-\infty, \infty)$ и $\int_{-\infty}^{\infty} H(x)p(x; \vartheta)dx < M$, причем M не зависит от ϑ ;

в) информация I конечна и положительна для всех ϑ .

Обозначим через $\widehat{\vartheta}$ истинное значение неизвестного параметра ϑ . В силу условий теоремы справедливо следующее разложение $(\ln p(x; \vartheta))'_{\vartheta}$ в окрестности $\widehat{\vartheta}$:

$$\frac{\partial}{\partial \vartheta} \ln p(x; \vartheta) = \frac{\partial}{\partial \vartheta} \ln p(x; \vartheta) \Big|_{\vartheta=\widehat{\vartheta}} + (\vartheta - \widehat{\vartheta}) \frac{\partial^2}{\partial \vartheta^2} \ln p(x; \vartheta) \Big|_{\vartheta=\widehat{\vartheta}} + \frac{1}{2} \gamma(\vartheta - \widehat{\vartheta})^2 H(x),$$

причем $|\gamma| < 1$. Тогда после умножения на $1/n$ уравнение правдоподобия можно записать в виде

$$\frac{1}{n} \frac{\partial}{\partial \vartheta} \ln L(X_1, \dots, X_n; \vartheta) = \xi_0 + (\vartheta - \widehat{\vartheta}) \xi_1 + \frac{1}{2} \gamma (\vartheta - \widehat{\vartheta})^2 \xi_2 = 0, \quad (14)$$

где случайные величины ξ_0 , ξ_1 и ξ_2 определяются выражениями

$$\xi_0 = \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \vartheta} \ln p(X_i; \vartheta) \Big|_{\vartheta=\widehat{\vartheta}},$$

$$\xi_1 = \frac{1}{n} \sum_{i=1}^n \frac{\partial^2}{\partial \vartheta^2} \ln p(X_i; \vartheta) \Big|_{\vartheta=\widehat{\vartheta}},$$

$$\xi_2 = \frac{1}{n} \sum_{i=1}^n H(X_i).$$

Рассмотрим поведение ξ_0 , ξ_1 , ξ_2 при больших n . Дифференцируя (1) по ϑ , получаем

$$\int_{-\infty}^{\infty} \frac{\partial}{\partial \vartheta} p(x; \vartheta) dx = \int_{-\infty}^{\infty} \frac{\partial^2}{\partial \vartheta^2} p(x; \vartheta) dx = 0.$$

Поэтому

$$\mathbf{M} \left[\frac{\partial}{\partial \vartheta} \ln p(X; \vartheta) \Big|_{\vartheta=\widehat{\vartheta}} \right] = \int_{-\infty}^{\infty} \left[\frac{1}{p(x; \vartheta)} \cdot \frac{\partial}{\partial \vartheta} p(x; \vartheta) \right] \Big|_{\vartheta=\widehat{\vartheta}} p(x; \vartheta) dx = 0,$$

$$\begin{aligned} \mathbf{M} \left[\frac{\partial^2}{\partial \vartheta^2} \ln p(X; \vartheta) \Big|_{\vartheta=\widehat{\vartheta}} \right] &= \\ &= \int_{-\infty}^{\infty} \left[\frac{1}{p(x; \vartheta)} \cdot \frac{\partial^2}{\partial \vartheta^2} p(x; \vartheta) - \left(\frac{1}{p(x; \vartheta)} \cdot \frac{\partial}{\partial \vartheta} p(x; \vartheta) \right)^2 \right] \Big|_{\vartheta=\widehat{\vartheta}} p(x; \vartheta) dx = \\ &= -\mathbf{M} \left[\left(\frac{\partial}{\partial \vartheta} \ln p(x; \vartheta) \right)^2 \Big|_{\vartheta=\widehat{\vartheta}} \right] = -I = -I(\widehat{\vartheta}), \end{aligned}$$

$$\mathbf{M} H(X) = \int_{-\infty}^{\infty} H(x) p(x; \widehat{\vartheta}) dx < M.$$

Вернемся к уравнению (14) и воспользуемся сначала тем фактом, что при $n \rightarrow \infty$ в силу закона больших чисел $\xi_0 \rightarrow 0$, $\xi_1 \rightarrow -I$ и $\xi_2 \rightarrow \mathbf{M} H(x)$, причем, согласно условиям теоремы, $I > 0$. Тогда можно показать, что уравнение (14) будет в некоторой окрестности $\widehat{\vartheta}$ иметь асимптотически единственное решение ϑ^* , которое к тому же определяется приближенной формулой

$$\vartheta^* \approx \widehat{\vartheta} + \frac{\xi_0}{I}.$$

Величина ξ_0 , по центральной предельной теореме, при $n \rightarrow \infty$ имеет асимптотически нормальное распределение с нулевым средним и дисперсией I/n .

Поэтому оценка ϑ^* также асимптотически распределена по нормальному закону с параметрами $\hat{\vartheta}$ и $1/(nI)$. \square

Замечание к теореме 7. Доказанная теорема гарантирует, что среди всех решений уравнения правдоподобия существует по крайней мере одно ϑ^* , обладающее свойством асимптотической эффективности в указанном смысле. Более того, такое решение асимптотически единственны в некоторой окрестности точки $\hat{\vartheta}$ (т. е. вероятность того, что в этой окрестности имеется другое решение уравнения правдоподобия, с ростом n стремится к нулю) и именно оно доставляет локальный максимум функции правдоподобия в этой окрестности. Но с самого начала мы назвали оценкой максимального правдоподобия оценку, доставляющую глобальный максимум функции правдоподобия. Такая оценка, вообще говоря, может не совпадать с ϑ^* и даже быть неединственной. Однако если семейство распределений $F(x; \vartheta)$ удовлетворяет естественному свойству разделимости, смысл которого сводится к тому, что для достаточно удаленных друг от друга ϑ и $\hat{\vartheta}$ распределения $F(x; \vartheta)$ и $F(x; \hat{\vartheta})$ также достаточно хорошо отличаются друг от друга, то любая оценка максимального правдоподобия будет состоятельной, т. е. стремиться к оцениваемому параметру. Вкупе с доказанной теоремой это означает асимптотическую единственность оценки максимального правдоподобия и совпадение ее с ϑ^* , что позволяет при асимптотическом анализе свойств оценки максимального правдоподобия говорить не об одном из решений уравнения правдоподобия или даже не об одной из оценок максимального правдоподобия, а просто об оценке максимального правдоподобия ϑ^* . Детальный разбор этого явления можно найти в [11]. Там же показано, что для оценки близости распределений удобно использовать расстояние Кульбака–Лейблера

$$\rho(F(x; \vartheta), F(x; \hat{\vartheta})) = \int_{-\infty}^{\infty} \left[\ln \frac{p(x; \hat{\vartheta})}{p(x; \vartheta)} \right] p(x; \hat{\vartheta}) dx,$$

поскольку в силу закона больших чисел именно к расстоянию Кульбака–Лейблера при $n \rightarrow \infty$ сходится с точностью до знака, постоянной $\int_{-\infty}^{\infty} \ln [p(x; \hat{\vartheta})] p(x; \hat{\vartheta}) dx$ и множителя $1/n$ логарифм функции правдоподобия

$$\ln \Lambda(X_1, \dots, X_n; \vartheta) = \ln p(X_1; \vartheta) + \dots + \ln p(X_n; \vartheta);$$

здесь ϑ — аргумент функции правдоподобия, а $\hat{\vartheta}$ — истинное значение неизвестного параметра.

В случае, когда семейство $F(x; \vartheta_1, \dots, \vartheta_k)$ зависит от нескольких неизвестных параметров $\vartheta_1, \dots, \vartheta_k$, при использовании метода максимального правдоподобия нужно искать максимум функции правдоподобия или ее логарифма по k аргументам $\vartheta_1, \dots, \vartheta_k$. Уравнение правдоподобия превращается в систему уравнений

$$\frac{\partial}{\partial \vartheta_1} \ln L(X_1, \dots, X_n; \vartheta_1, \dots, \vartheta_k) = 0,$$

...

$$\frac{\partial}{\partial \vartheta_k} \ln L(X_1, \dots, X_n; \vartheta_1, \dots, \vartheta_k) = 0.$$

Пример 25. Выборка X_1, \dots, X_n произведена из нормальной генеральной совокупности с неизвестными параметрами ϑ_1 (среднее) и ϑ_2 (дисперсия). Найдем их оценки ϑ_1^* и ϑ_2^* методом максимального правдоподобия. Логарифм функции правдоподобия задается формулой

$$\ln L(X_1, \dots, X_n; \vartheta_1, \vartheta_2) = -\frac{n}{2} \ln(2\pi\vartheta_2) - \frac{1}{2\vartheta_2} \left[(X_1 - \vartheta_1)^2 + \dots + (X_n - \vartheta_1)^2 \right].$$

Система уравнений правдоподобия имеет вид

$$(\ln L(X_1, \dots, X_n; \vartheta_1, \vartheta_2))'_{\vartheta_1} = \frac{1}{\vartheta_2} (X_1 + \dots + X_n - n\vartheta_1) = 0,$$

$$(\ln L(X_1, \dots, X_n; \vartheta_1, \vartheta_2))'_{\vartheta_2} = -\frac{n}{2\vartheta_2} + \frac{1}{2\vartheta_2^2} \left[(X_1 - \vartheta_1)^2 + \dots + (X_n - \vartheta_1)^2 \right] = 0$$

или

$$X_1 + \dots + X_n - n\vartheta_1 = 0, \quad (X_1 - \vartheta_1)^2 + \dots + (X_n - \vartheta_1)^2 - n\vartheta_2 = 0.$$

Таким образом,

$$\begin{aligned} \vartheta_1^* &= \frac{1}{n} (X_1 + \dots + X_n) = m^*, \\ \vartheta_2^* &= \frac{1}{n} \left[(X_1 - m^*)^2 + \dots + (X_n - m^*)^2 \right] = \sigma^{2*}. \end{aligned}$$

Читателю предлагается самостоятельно показать, что ϑ_1^* и ϑ_2^* доставляют максимум функции правдоподобия $L(X_1, \dots, X_n; \vartheta_1, \vartheta_2)$.

Оценки ϑ_1^* и ϑ_2^* параметров ϑ_1 и ϑ_2 совпадают с выборочным средним m^* и выборочной дисперсией σ^{2*} . Отметим, что оценка ϑ_1^* неизвестного математического ожидания ϑ_1 является эффективной (см. пример 11), чего нельзя сказать об оценке ϑ_2^* неизвестной дисперсии ϑ_2 , которая, как мы знаем, является даже смещенной.

Оказывается, однако, что если мы в качестве оценки параметра ϑ_2 рассмотрим выборочную дисперсию s^{2*} , то эта оценка будет уже не только несмещенной, но и иметь минимальную дисперсию среди всех несмешанных оценок параметра ϑ_2 . Последний факт вытекает из неравенства Бхаттачария [7], обобщающего неравенство Рао–Крамера, а также может быть установлен из свойств многомерных достаточных оценок [11]. \square

5. Метод минимального расстояния

Суть этого метода заключается в следующем. Предположим, что любым двум функциям распределения $F_1(x)$ и $F_2(x)$ поставлено в соответствие число

$$\rho = \rho(F_1(x), F_2(x)) \geq 0,$$

называемое расстоянием, причем $\rho(F(x), F(x)) = 0$. Пусть теперь, как обычно, задана выборка X_1, \dots, X_n из генеральной совокупности с теоретической функцией распределения $F(x)$, принадлежащей параметрическому семейству $F(x; \vartheta)$. Вычислим расстояние между эмпирической функцией распределения $F^*(x)$ и функциями распределения $F(x; \vartheta)$ из данного семейства. Оценкой, полученной методом минимального расстояния, называется такое значение ϑ^* , для которого

$$\rho(F^*(x), F(x; \vartheta^*)) = \min_{\vartheta} \rho(F^*(x), F(x; \vartheta)),$$

т. е. такое значение ϑ^* , которое определяет ближайшую к $F^*(x)$ в смысле расстояния ρ функцию распределения из семейства $F(x; \vartheta)$.

Приведем примеры некоторых наиболее часто встречающихся в математической статистике расстояний.

Равномерное расстояние (расстояние Колмогорова) определяется формулой

$$\rho_R(F_1(x), F_2(x)) = \sup_x |F_1(x) - F_2(x)|.$$

Расстояние ω^2 имеет вид

$$\rho_{\omega^2}(F_1(x), F_2(x)) = \int_{-\infty}^{\infty} [F_1(x) - F_2(x)]^2 dF_2(x).$$

Расстояние χ^2 употребляется для функций распределения $F_1(x)$ и $F_2(x)$ дискретных случайных величин ξ_1 и ξ_2 , принимающих одинаковые значения b_1, \dots, b_L , и задается выражением

$$\rho_{\chi^2}(F_1(x), F_2(x)) = \sum_{i=1}^L \frac{[P_1(b_i) - P_2(b_i)]^2}{P_2(b_i)},$$

где вероятности $P_1(b_i) = \mathbf{P}\{\xi_1 = b_i\}$ и $P_2(b_i) = \mathbf{P}\{\xi_2 = b_i\}$ определяются рядами распределения случайных величин ξ_1 и ξ_2 .

Использование приведенных выше расстояний для получения оценок весьма сложно в вычислительном плане, и поэтому они употребляются крайне редко. Здесь мы упомянули об этих расстояниях только потому, что применение оценок, полученных с их помощью, позволяет упростить вычисление уровней значимости критериев при проверке сложных непараметрических статистических гипотез, поскольку такие оценки естественным образом связаны с соответствующими критериями (см. параграф 5 гл. 3).

6. Метод номограмм

Еще одним методом, позволяющим, пользуясь только номограммами (специальным образом разлинованными листами бумаги, которые в математической статистике носят название вероятностной бумаги), весьма просто и быстро оценить неизвестные параметры, является метод номограмм. Его сущность состоит в следующем. Пусть мы имеем выборку X_1, \dots, X_n из генеральной совокупности с неизвестной теоретической функцией распределения, принадлежащей двухпараметрическому семейству $F(x; \vartheta_1, \vartheta_2)$. Предположим теперь, что каким-то чрезвычайно простым способом удалось построить функцию распределения $F(x; \vartheta_1^*, \vartheta_2^*)$ из семейства $F(x; \vartheta_1, \vartheta_2)$, достаточно хорошо приближающую эмпирическую функцию распределения $F^*(x)$. Тогда ϑ_1^* и ϑ_2^* будут являться оценками неизвестных параметров ϑ_1 и ϑ_2 , при-

чем в силу теоремы Гливенко–Кантелли состоятельными при весьма слабых условиях, накладываемых на семейство $F(x; \vartheta_1, \vartheta_2)$.

Казалось бы, мы пришли к не менее сложной задаче: найти «чрезвычайно простой» способ приближения эмпирической функции распределения функцией распределения из семейства $F(x; \vartheta_1, \vartheta_2)$. Оказывается, однако, что графики функций распределения тех семейств $F(x; \vartheta_1, \vartheta_2)$, в которых ϑ_1 и ϑ_2 , по сути дела, связаны с параметрами «сдвига» и «масштаба» (к таким семействам относятся, например, нормальное, логнормальное и т. д.), можно с помощью некоторых нелинейных преобразований координат превратить в семейство прямых линий. Тогда, построив в этих новых координатах график эмпирической функции распределения $F^*(x)$, нетрудно визуально провести прямую, которая достаточно хорошо приближает $F^*(x)$, а затем уже по коэффициентам проведенной прямой найти оценки ϑ_1^* и ϑ_2^* неизвестных параметров ϑ_1 и ϑ_2 .

Практическая реализация метода номограмм происходит следующим образом. Сначала выборку X_1, \dots, X_n преобразуют в вариационный ряд X_1^*, \dots, X_n^* и на номограмме для соответствующего семейства $F(x; \vartheta_1, \vartheta_2)$ откладывают точки A_i ($i = 1, \dots, n$) с координатами $(X_i^*, (2i - 1)/(2n))$, абсциссы которых X_i^* представляют собой точки скачков эмпирической функции распределения $F^*(x)$, а ординаты $(2i - 1)/(2n)$ — середины этих скачков. Затем «на глаз» проводят прямую линию, проходящую как можно ближе ко всем точкам A_i . Наконец, с помощью пояснений к номограмме по коэффициентам прямой находят оценки ϑ_1^* и ϑ_2^* неизвестных параметров ϑ_1 и ϑ_2 .

Пример 26. Предполагая в примере 1 из гл. 1, что проекция вектора скорости молекул водорода распределена по нормальному закону, оценим с помощью метода номограмм неизвестное математическое ожидание ϑ_1 и дисперсию ϑ_2 . Воспользовавшись вариационным рядом выборки, найдем координаты точек A_i (табл. 3). Отложим точки A_i на номограмме для нормального распределения (на нормальной вероятностной бумаге) и проведем «на глаз» прямую A , задаваемую уравнением $y = kx + a$ (рис. 1).

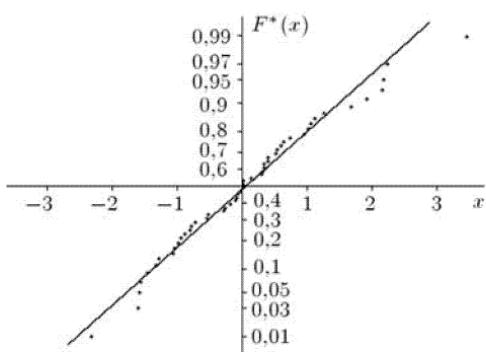


Рис. 1

Оценка ϑ_1^* математического ожидания ϑ_1 совпадает с точкой пересечения прямой A с осью абсцисс, т. е. $\vartheta_1^* = a = 0,06$. Для того чтобы найти оценку ϑ_2^* дисперсии ϑ_2 , определим значение коэффициента k : $k = 0,90$. Тогда $\vartheta_2^* = 1/k^2 = 1,23$. Для сравнения приведем значения оценок этих же параметров, полученные методом максималь-

Таблица 3

A_i	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}
$x_i = X_i^*$	-2,33	-1,60	-1,58	-1,57	-1,36	-1,32	-1,29	-1,06	-1,04	-0,98
$y_i = \frac{2i-1}{2n}$	0,01	0,03	0,05	0,07	0,09	0,11	0,13	0,15	0,17	0,19
A_i	A_{11}	A_{12}	A_{13}	A_{14}	A_{15}	A_{16}	A_{17}	A_{18}	A_{19}	A_{20}
$x_i = X_i^*$	-0,95	-0,88	-0,80	-0,79	-0,72	-0,53	-0,52	-0,28	-0,27	-0,18
$y_i = \frac{2i-1}{2n}$	0,21	0,23	1,25	0,27	0,29	0,31	0,33	0,35	0,37	0,39
A_i	A_{21}	A_{22}	A_{23}	A_{24}	A_{25}	A_{26}	A_{27}	A_{28}	A_{29}	A_{30}
$x_i = X_i^*$	-0,10	-0,08	-0,05	-0,01	0,01	0,01	0,01	0,14	0,30	0,32
$y_i = \frac{2i-1}{2n}$	0,41	0,43	0,45	0,47	0,49	0,51	0,53	0,55	0,57	0,59
A_i	A_{31}	A_{32}	A_{33}	A_{34}	A_{35}	A_{36}	A_{37}	A_{38}	A_{39}	A_{40}
$x_i = X_i^*$	0,35	0,35	0,40	0,41	0,52	0,54	0,60	0,65	0,74	0,96
$y_i = \frac{2i-1}{2n}$	0,61	0,63	0,65	0,67	0,69	0,71	0,73	0,75	0,77	0,79
A_i	A_{41}	A_{42}	A_{43}	A_{44}	A_{45}	A_{46}	A_{47}	A_{48}	A_{49}	A_{50}
$x_i = X_i^*$	1,02	1,06	1,12	1,27	1,68	1,92	2,15	2,19	2,25	3,47
$y_i = \frac{2i-1}{2n}$	0,81	0,83	0,85	0,87	0,89	0,91	0,93	0,95	0,97	0,99

ного правдоподобия (см. пример 18, а также пример 8 из гл. 1): $\vartheta_1^* = 0,082$, $\vartheta_2^* = 1,34$. Как видим, оценки весьма близки. \square

Следует отметить, что с помощью метода номограмм можно судить также о правильности выбора семейства $F(x; \vartheta_1, \vartheta_2)$. Действительно, по множеству точек A_i сразу видно, группируются они вокруг некоторой прямой или нет. Если нет, то возникают серьезные сомнения в принадлежности теоретического распределения $F(x)$ семейству $F(x; \vartheta_1, \vartheta_2)$.

7. Доверительные интервалы

Полученные в предыдущих параграфах оценки неизвестных параметров естественно называть *точечными*, поскольку они оценивают неизвестный параметр одним числом или точкой. Однако, как мы знаем, точечная оценка не совпадает с оцениваемым параметром и более разумно было бы указывать те *допустимые границы*, в которых может находиться неизвестный параметр ϑ при наблюденной выборке

X_1, \dots, X_n . К сожалению, в подавляющем большинстве важных для практики случаев при любой выборке X_1, \dots, X_n достоверная область, в которой может находиться неизвестный параметр ϑ , совпадает со всей возможной областью изменения этого параметра, поскольку такую выборку мы можем получить с ненулевой вероятностью (или плотностью распределения) при каждом значении ϑ . Поэтому приходится ограничиваться нахождением границ изменения неизвестного параметра с некоторой наперед заданной степенью доверия или доверительной вероятностью.

Доверительной вероятностью назовем такую вероятность α , что событие вероятности $1 - \alpha$ можно считать невозможным. Разумеется, выбор доверительной вероятности полностью зависит от исследователя, причем во внимание принимаются не только его личные наклонности, но и физическая суть рассматриваемого явления. Так, степень доверия авиапассажира к надежности самолета, несомненно, должна быть выше степени доверия покупателя к надежности электрической лампочки. В математической статистике обычно используют значения доверительной вероятности 0,9, 0,95, 0,99, реже 0,999, 0,9999 и т. д.

Задавшись доверительной вероятностью α , мы уже можем по выборке X_1, \dots, X_n определить интервал $[\vartheta' = \vartheta'(X_1, \dots, X_n), \vartheta'' = \vartheta''(X_1, \dots, X_n)]$, в котором будет находиться неизвестный параметр ϑ . Такой интервал называется *доверительным интервалом* (иногда также говорят «*интервальная оценка*») доверительной вероятности α для неизвестного параметра ϑ . Отметим, что доверительная вероятность α ни в коей мере не является вероятностью неизвестному параметру ϑ принадлежать доверительному интервалу $[\vartheta', \vartheta'']$, поскольку, как мы предположили с самого начала, априорные сведения о параметре ϑ , в частности о его распределении, отсутствуют. Когда говорят, что неизвестный параметр ϑ не может выйти за границу доверительного интервала $[\vartheta', \vartheta'']$, констатируют только, что если при любом истинном значении ϑ в результате эксперимента получена выборка X_1, \dots, X_n , а затем по ней построен доверительный интервал $[\vartheta', \vartheta'']$, то этот интервал с вероятностью α накроет значение ϑ .

Доверительные интервалы определим, следуя Ю. Нейману, опираясь на точечные оценки. По заданной оценке ϑ^* доверительные интервалы доверительной вероятности α можно построить различными способами. На практике обычно используют два типа доверительных интервалов: *симметричные* и *односторонние*. Ограничимся описанием процедуры построения симметричных доверительных интервалов. Односторонние доверительные интервалы находятся совершенно аналогично.

Итак, пусть у нас имеется выборка X_1, \dots, X_n из генеральной совокупности с неизвестной теоретической функцией распределения $F(x)$, принадлежащей однопараметрическому семейству $F(x; \vartheta)$. Предположим также, что нами выбрана некоторая оценка ϑ^* , по которой мы хотим построить симметричный доверительный интервал доверительной

вероятности α . Для этого возьмем произвольное значение ϑ и найдем функцию распределения $F_{\vartheta^*}(x; \vartheta)$ оценки ϑ^* . Определим $x_1 = x_1(\vartheta)$ и $x_2 = x_2(\vartheta)$ из решения уравнений (см. рис. 2):

$$F_{\vartheta^*}(x_1; \vartheta) = \frac{1+\alpha}{2}, \quad F_{\vartheta^*}(x_2; \vartheta) = \frac{1-\alpha}{2}$$

(напомним, что x_1 и x_2 носят название $(1+\alpha)/2$ - и $(1-\alpha)/2$ -квантилей функции распределения $F_{\vartheta^*}(x; \vartheta)$). Таким образом, при заданном ϑ оценка ϑ^* будет с вероятностью α заключена в интервале $[x_2, x_1]$, причем вероятность попадания ϑ^* как влево, так и вправо от интервала $[x_2, x_1]$ имеет одно и то же значение $(1-\alpha)/2$ (отсюда происходит название «симметричный»). Откладывая теперь на графике рис. 3 по оси абсцисс значение параметра ϑ , а по оси ординат — соответствующие ему значения x_1 и x_2 , получим кривые $x_1(\vartheta)$ и $x_2(\vartheta)$. В силу принципа невозможности события, происходящего с вероятностью $1-\alpha$, заключаем, что все возможные пары (ϑ, ϑ^*) могут находиться только внутри области G между кривыми $x_1(\vartheta)$ и $x_2(\vartheta)$. Для окончания построения доверительного интервала остается заметить, что, получив по выборке X_1, \dots, X_n оценку ϑ^* , мы вправе сделать вывод: неизвестный параметр ϑ обязан лежать внутри интервала $[\vartheta', \vartheta'']$, где ϑ' и ϑ'' определяются из решения уравнений

$$\vartheta^* = x_1(\vartheta'), \quad \vartheta^* = x_2(\vartheta'').$$

Именно интервал $[\vartheta', \vartheta'']$ и является симметричным доверительным интервалом доверительной вероятности α .

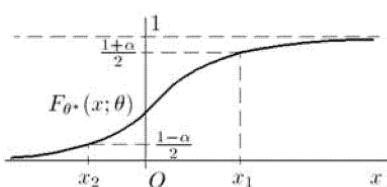


Рис. 2

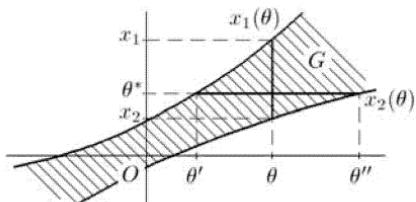


Рис. 3

Пример 27. Построим симметричный доверительный интервал доверительной вероятности α для неизвестной вероятности успеха ϑ в схеме Бернулли. Естественно в качестве оценки ϑ^* взять наблюденную частоту

$$\vartheta^* = \frac{\mu}{n},$$

где μ — суммарное наблюденное число успехов (см. пример 24).

При малом объеме выборки n процедура построения доверительных интервалов трудоемка, поскольку она практически сводится к перебору значений неизвестного параметра. Поэтому существуют специальные таблицы (см. [1], табл. 5.2), которые по наблюденным значениям числа успехов μ и числа неудач $n - \mu$ дают границы доверительного интервала доверительной вероятности α .

При больших объемах выборки n пользуются тем фактом, что в силу интегральной теоремы Муавра–Лапласа оценка ϑ^* распределена приближенно по нормальному закону со средним ϑ и дисперсией $\vartheta(1-\vartheta)/n$. Тогда решения уравнений

$$F_{\vartheta^*}(x_1; \vartheta) = \frac{1+\alpha}{2}, \quad F_{\vartheta^*}(x_2; \vartheta) = \frac{1-\alpha}{2}$$

связаны с $(1+\alpha)/2$ - и $(1-\alpha)/2$ -квантилями $\varphi_{(1+\alpha)/2}$ и $\varphi_{(1-\alpha)/2}$ (см. [1], табл. 1.3) стандартного нормального закона формулами

$$x_1 = x_1(\vartheta) \approx \vartheta + \varphi_{\frac{1+\alpha}{2}} \sqrt{\frac{\vartheta(1-\vartheta)}{n}}, \quad x_2 = x_2(\vartheta) \approx \vartheta + \varphi_{\frac{1-\alpha}{2}} \sqrt{\frac{\vartheta(1-\vartheta)}{n}}.$$

Учитывая, что $\varphi_{(1+\alpha)/2} = -\varphi_{(1-\alpha)/2}$, уравнения кривых $x_1(\vartheta)$ и $x_2(\vartheta)$ можно записать в единой эквивалентной форме

$$(x - \vartheta)^2 = \frac{\vartheta(1-\vartheta)}{n} \varphi_{\frac{1+\alpha}{2}}^2.$$

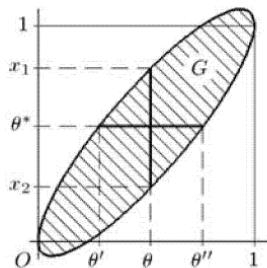


Рис. 4

Последнее уравнение, как нетрудно видеть, представляет собой уравнение эллипса (рис. 4) (физически непонятный выход эллипса за полосу $0 \leq x \leq 1$ связан с тем, что при ϑ , близких к нулю или единице, необходимо в соответствии с теоремой Пуассона использовать не нормальную, а пуассоновскую аппроксимацию оценки ϑ^*). Уравнение для определения границ ϑ' и ϑ'' доверительного интервала

имеет вид

$$(\vartheta^* - \vartheta)^2 = \frac{\vartheta(1-\vartheta)}{n} \varphi_{\frac{1+\alpha}{2}}^2,$$

откуда окончательно получаем

$$\vartheta'^{\prime, \prime \prime} = \frac{\vartheta^* + \frac{1}{2n} \varphi_{\frac{1+\alpha}{2}}^2 \mp \varphi_{\frac{1+\alpha}{2}} \sqrt{\frac{1}{n} \vartheta^*(1-\vartheta^*) + \frac{1}{4n^2} \varphi_{\frac{1+\alpha}{2}}^2}}{1 + \frac{1}{n} \varphi_{\frac{1+\alpha}{2}}^2}. \quad \square$$

Пример 28. Построим симметричный доверительный интервал доверительной вероятности α для неизвестного среднего ϑ нормального закона при известной дисперсии σ^2 . Эффективной оценкой ϑ^* параметра ϑ , как мы знаем (пример 18), является выборочное среднее

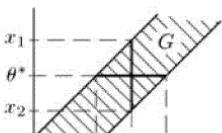
$$\vartheta^* = m^* = \frac{1}{n} (X_1 + \dots + X_n).$$

Оценка ϑ^* также распределена по нормальному закону с параметрами ϑ и σ^2/n . Поэтому

$$x_1 = x_1(\vartheta) = \vartheta + \varphi_{\frac{1+\alpha}{2}} \sqrt{\frac{\sigma^2}{n}}, \quad x_2 = x_2(\vartheta) = \vartheta + \varphi_{\frac{1-\alpha}{2}} \sqrt{\frac{\sigma^2}{n}},$$

т. е. $x_1(\vartheta)$ и $x_2(\vartheta)$ представляют собой уравнения двух параллельных прямых (рис. 5). Решая уравнения

$$\vartheta^* = x_1(\vartheta'), \quad \vartheta^* = x_2(\vartheta'),$$



получаем границы доверительного интервала

$$\vartheta' = \vartheta^* - \varphi_{\frac{1+\alpha}{2}} \sqrt{\frac{\sigma^2}{n}}, \quad \vartheta'' = \vartheta^* - \varphi_{\frac{1-\alpha}{2}} \sqrt{\frac{\sigma^2}{n}},$$

или, учитывая, что $\varphi_{(1+\alpha)/2} = -\varphi_{(1-\alpha)/2}$,

$$\vartheta'^* = \vartheta^* \mp \varphi_{\frac{1+\alpha}{2}} \sqrt{\frac{\sigma^2}{n}}. \quad \square$$

Пример 29. Как и в предыдущем примере, предположим, что выборка X_1, \dots, X_n произведена из нормальной генеральной совокупности, но с неизвестной дисперсией ϑ , а среднее известно и равно m . В качестве оценки ϑ^* неизвестной дисперсии ϑ возьмем выборочную дисперсию

$$\vartheta^* = \frac{1}{n} \left[(X_1 - m)^2 + \dots + (X_n - m)^2 \right].$$

Тогда случайная величина $\chi^2 = n\vartheta^*/\vartheta$ будет иметь χ^2 -распределение с n степенями свободы, а значит, решения уравнений

$$F_{\vartheta^*}(x_1; \vartheta) = \frac{1+\alpha}{2}, \quad F_{\vartheta^*}(x_2; \vartheta) = \frac{1-\alpha}{2}$$

определяются формулами

$$x_1 = \frac{1}{n} h_{\frac{1+\alpha}{2}} \vartheta, \quad x_2 = \frac{1}{n} h_{\frac{1-\alpha}{2}} \vartheta,$$

где h_α — α -квантиль χ^2 -распределения с n степенями свободы (см. [1], табл. 2.2б). Уравнения

$$x_1 = x_1(\vartheta) = \frac{1}{n} h_{\frac{1+\alpha}{2}} \vartheta,$$

$$x_2 = x_2(\vartheta) = \frac{1}{n} h_{\frac{1-\alpha}{2}} \vartheta$$

представляют собой уравнения двух лучей, исходящих из начала координат (рис. 6), и, значит, границы симметричного доверительного интервала доверительной вероятности α для неизвестной дисперсии ϑ задаются формулами

$$\vartheta' = \frac{n\vartheta^*}{h_{\frac{1+\alpha}{2}}}, \quad \vartheta'' = \frac{n\vartheta^*}{h_{\frac{1-\alpha}{2}}}. \quad \square$$

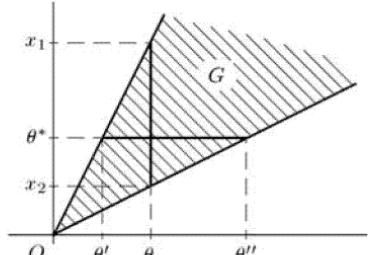


Рис. 6

Пример 30. Рассмотрим, наконец, случай, когда в выборке из нормальной генеральной совокупности неизвестны оба параметра: среднее ϑ_1 и дисперсия ϑ_2 . В качестве их оценок воспользуемся выборочным средним

$$\vartheta_1^* = m^* = \frac{1}{n} (X_1 + \dots + X_n)$$

и выборочной дисперсией

$$\vartheta_2^* = s^{2*} = \frac{1}{n-1} \left[(X_1 - m)^2 + \dots + (X_n - m)^2 \right]$$

(см. пример 25).

Построение доверительного интервала $[\vartheta'_1, \vartheta''_1]$ для неизвестного среднего ϑ_1 начнем с определения случайной величины

$$\tau = (\vartheta_1^* - \vartheta_1) \sqrt{\frac{n}{\vartheta_2^*}},$$

которая, как говорилось в параграфе 4 гл. 1, имеет t -распределение с $n - 1$ степенями свободы. Обозначим через $t_{(1+\alpha)/2}$ и $t_{(1-\alpha)/2}$ $(1 + \alpha)/2$ - и $(1 - \alpha)/2$ -квантили t -распределения (см. [1], табл. 3.2). Тогда значение оценки среднего ϑ_1^* с вероятностью α будет лежать в пределах

$$\text{от } x_2 = \vartheta_1 + t_{\frac{1-\alpha}{2}} \sqrt{\frac{\vartheta_2^*}{n}} \quad \text{до} \quad x_1 = \vartheta_1 + t_{\frac{1+\alpha}{2}} \sqrt{\frac{\vartheta_2^*}{n}}.$$

Продолжая рассуждения, как и в случае известной дисперсии, и учитывая равенство $t_{(1+\alpha)/2} = -t_{(1-\alpha)/2}$, получаем окончательные выражения для границ ϑ'_1 и ϑ''_1 симметричного доверительного интервала доверительной вероятности α :

$$\vartheta'_1 = \vartheta_1^* - t_{\frac{1+\alpha}{2}} \sqrt{\frac{\vartheta_2^*}{n}}, \quad \vartheta''_1 = \vartheta_1^* + t_{\frac{1+\alpha}{2}} \sqrt{\frac{\vartheta_2^*}{n}}.$$

Доверительный интервал $[\vartheta'_1, \vartheta''_1]$ доверительной вероятности α для неизвестной дисперсии ϑ_2 строится точно так же, как и в примере 29:

$$\vartheta'_2 = \frac{(n-1)\vartheta_2^*}{h_{\frac{1+\alpha}{2}}}, \quad \vartheta''_2 = \frac{(n-1)\vartheta_2^*}{h_{\frac{1-\alpha}{2}}}.$$

При этом нужно учитывать, что квантили $h_{(1+\alpha)/2}$ и $h_{(1-\alpha)/2}$ берутся для χ^2 -распределения с $n - 1$ степенями свободы, поскольку одна степень свободы уходит на определение неизвестного среднего ϑ_1 . \square

В заключение отметим, что в современной математической статистике доверительные интервалы строят так же, основываясь на критериях значимости.

Глава 3

ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

В этой главе мы обратимся ко второму направлению в математической статистике — проверке статистических гипотез. Сначала определим понятия статистической гипотезы и критерия, а затем рассмотрим некоторые наиболее часто встречающиеся на практике гипотезы и приведены критерии для их проверки.

1. Статистическая гипотеза. Критерий

Снова предположим, что в результате эксперимента мы получили выборку X_1, \dots, X_n из генеральной совокупности с неизвестной теоретической функцией распределения $F(x)$. *Статистической гипотезой* (в дальнейшем для краткости вместо «статистическая гипотеза» будем говорить просто «гипотеза») назовем любое предположение о виде теоретической функции распределения $F(x)$. Так, в схеме Бернуlli гипотезами будут являться следующие предположения: «вероятность успеха равна $1/2$ »; «вероятность успеха больше $1/3$ »; «вероятность успеха заключена между $0,4$ и $0,7$ » и т. д. С нормальным распределением можно связать такие гипотезы: «теоретическая функция распределения нормальна со средним, равным нулю»; «теоретическая функция распределения нормальна с дисперсией, не превосходящей квадрата среднего значения», и т. д. Все перечисленные выше гипотезы являются *параметрическими*, поскольку в них предположения делаются относительно области изменения неизвестного параметра (или нескольких параметров) для заданных параметрических семейств функций распределения. Примерами *непараметрических* гипотез служат высказывания: «теоретическая функция распределения является нормальной»; «теоретическая функция распределения не является нормальной»; «теоретическая функция распределения имеет положительное математическое ожидание». Гипотезы будем обозначать буквой H , снабжая при необходимости индексами.

Всюду в дальнейшем будем предполагать, что у нас имеются две непересекающиеся гипотезы: H_0 и H_1 . Гипотезу H_0 будем называть *основной*, а гипотезу H_1 — *конкурирующей* или *альтернативной*. Выбор названия условен, но, как правило, удобно основной гипотезой H_0 называть более конкретное предположение о виде теоретической функции распределения или предположение, влекущее за собой более важные практические последствия. *Задача проверки статисти-*

ческих гипотез состоит в том, чтобы на основе выборки X_1, \dots, X_n принять (т. е. считать справедливой) либо основную гипотезу H_0 , либо конкурирующую гипотезу H_1 .

Различают *простую* и *сложную* гипотезы. Простая гипотеза полностью определяет теоретическую функцию распределения $F(x)$. Так, простыми будут гипотезы: «вероятность успеха в схеме Бернулли равна $1/2$ »; «теоретическая функция распределения является нормальной с нулевым средним и единичной дисперсией». Гипотеза, не являющаяся простой, носит название *сложной*. Примерами сложных гипотез будут: «вероятность успеха в схеме Бернулли заключена между $0,4$ и $0,7$ »; «теоретическая функция распределения является нормальной с нулевым средним, но произвольной дисперсией»; «теоретическая функция распределения не является нормальной». Сложная гипотеза среди возможных функций распределения выделяет некоторое подмножество \mathcal{F}_0 , содержащее более одной функции распределения. При этом если мы имеем сложную параметрическую гипотезу, то заранее в силу каких-то уже проверенных соображений ограничиваемся рассмотрением некоторого параметрического семейства функций распределения $F(x; \vartheta)$ с неизвестным параметром (или параметрами) ϑ , а сама гипотеза выделяет среди всех функций распределения данного семейства те, у которых $\vartheta \in \Theta_0$, где Θ_0 в свою очередь — некоторое подмножество области Θ всех возможных значений неизвестного параметра ϑ .

Статистическим критерием (или просто критерием) называется правило, позволяющее, основываясь только на выборке X_1, \dots, X_n , принять либо основную гипотезу H_0 , либо конкурирующую H_1 . Каждый критерий характеризуется *допустимой областью* W_d , т. е. областью в n -мерном пространстве R^n , попадание в которую выборки X_1, \dots, X_n влечет за собой принятие основной гипотезы H_0 . Дополнительная область $W_k = R^n \setminus W_d$, попадание в которую выборки X_1, \dots, X_n приводит к принятию конкурирующей гипотезы H_1 , носит название *критической области*.

Предположим теперь, что у нас имеется две гипотезы H_0 и H_1 , т. е. в множестве всех функций распределения выделены два непересекающихся подмножества \mathcal{F}_0 и \mathcal{F}_1 ; при этом основная гипотеза H_0 заключается в том, что выборка X_1, \dots, X_n произведена из генеральной совокупности с теоретической функцией распределения $F(x)$, принадлежащей подмножеству \mathcal{F}_0 , а конкурирующая гипотеза H_1 — с теоретической функцией распределения $F(x)$, принадлежащей подмножеству \mathcal{F}_1 . Пусть также задан критерий для проверки этих гипотез, т. е. разбиение n -мерного пространства R^n на две области: допустимую W_d и критическую $W_k = R^n \setminus W_d$. В силу случайности выборки какой бы критерий мы не взяли, обязательно возможно появление ошибок двух родов.

Ошибка первого рода возникает тогда, когда имеет место основная гипотеза H_0 , но выборка X_1, \dots, X_n попадает в критическую область W_k и мы принимаем конкурирующую гипотезу H_1 . Вероят-

нность α ошибки первого рода называется *уровнем значимости критерия* и определяется формулой

$$\alpha = \mathbf{P}\{(X_1, \dots, X_n) \in W_{\kappa}\} = \sum P(x_1) \cdots P(x_n)$$

в дискретном случае и

$$\alpha = \mathbf{P}\{(X_1, \dots, X_n) \in W_{\kappa}\} = \int \cdots \int p(x_1) \cdots p(x_n) dx_1 \cdots dx_n$$

в непрерывном, где $P(x)$ или $p(x)$ — ряд распределения или плотность распределения наблюдаемой случайной величины X при условии справедливости основной гипотезы H_0 , а суммирование или интегрирование, как обычно, ведется по всем точкам $(x_1, \dots, x_n) \in W_{\kappa}$ (в дискретном случае каждое x_i может принимать только значения b_1, \dots, b_L). В случае, когда гипотеза H_0 сложная, уровень значимости $\alpha = \alpha(F(x))$, естественно, будет зависеть от реальной теоретической функции распределения $F(x)$ из подмножества \mathcal{F}_0 . Кроме того, если гипотеза H_0 параметрическая, т. е. подмножество \mathcal{F}_0 представляет собой параметрическое семейство функций распределения $F(x; \vartheta)$, зависящее от неизвестного параметра ϑ с областью изменения Θ_0 , являющейся подобластью области Θ всех возможных значений параметра ϑ , то будем вместо записи $\alpha(F(x))$ употреблять запись $\alpha(\vartheta)$, предполагая при этом, что $\vartheta \in \Theta_0$.

Пусть теперь справедлива конкурирующая гипотеза H_1 , но выборка X_1, \dots, X_n попала в допустимую область W_d и мы приняли основную гипотезу H_0 . Тогда мы имеем дело с *ошибкой второго рода*. Вероятность ошибки второго рода β' носит название *оперативной характеристики критерия*. Однако обычно в статистике предпочитают иметь дело с *мощностью критерия* $\beta = 1 - \beta'$ (т. е. вероятностью того, что при справедливой конкурирующей гипотезе H_1 мы ее примем), задаваемой формулой

$$\beta = \mathbf{P}\{(X_1, \dots, X_n) \in W_{\kappa}\} = \sum P(x_1) \cdots P(x_n)$$

в дискретном случае и

$$\beta = \mathbf{P}\{(X_1, \dots, X_n) \in W_{\kappa}\} = \int \cdots \int p(x_1) \cdots p(x_n) dx_1 \cdots dx_n$$

в непрерывном, где так же, как и при определении уровня значимости, суммирование или интегрирование ведется по всем $(x_1, \dots, x_n) \in W_{\kappa}$, однако ряд распределения $P(x)$ или плотность распределения $p(x)$ берутся при условии справедливости конкурирующей гипотезы H_1 . Разумеется, в случае сложной гипотезы H_1 мощность $\beta = \beta(F(x))$ будет зависеть от реального теоретического распределения $F(x)$ из подмножества \mathcal{F}_1 . Если конкурирующая гипотеза H_0 параметрическая, то вместо $\beta(F(x))$ будем писать $\beta(\vartheta)$, считая при этом, что $\vartheta \in \Theta_1$, где Θ_1 — область изменения неизвестного параметра ϑ при условии справедливости гипотезы H_1 .

Таким образом, и уровень значимости, и мощность критерия задаются одной и той же формулой и их различие состоит в том, что уровень значимости $\alpha(F(x))$ определяется только для теоретических функций распределения $F(x)$, принадлежащих подмножеству \mathcal{F}_0 , а мощность $\beta(F(x))$ — подмножеству \mathcal{F}_1 . Впрочем, иногда эти два понятия объединяют в одно, называя *функцией мощности критерия* $\beta = \beta(F(x))$ величину, равную уровню значимости $\alpha(F(x))$ при $F(x) \in \mathcal{F}_0$ и мощности $\beta(F(x))$ при $F(x) \in \mathcal{F}_1$. Отметим, что уровень значимости и оперативная характеристика критерия могут иметь совершенно разную физическую природу. Так, пропуск партии бракованных изделий влечет за собой, как правило, более тяжелые последствия, чем выбраковка партии годных изделий.

Естественное желание каждого исследователя состоит в предоставлении ему такого критерия, который позволил бы как можно реже делать ошибки и первого и второго рода (в идеале — совсем не ошибаться!), т. е. минимизировал бы и уровень значимости α , и оперативную характеристику β' . Но такое желание невыполнимо, поскольку требование делать реже ошибку первого рода влечет за собой увеличение допустимой области W_d , в то время как требование реже делать ошибку второго рода предписывает увеличить критическую область W_k . Поэтому обычно поступают следующим образом: фиксируют уровень значимости (как более важный с практической точки зрения) и среди нескольких критериев, имеющих заданный уровень значимости, предпочтение отдают более мощному.

Остановимся на этом несколько подробнее. Назовем *размером критерия* α_0 максимальное значение вероятности ошибки первого рода при использовании данного критерия, т. е.

$$\alpha_0 = \sup_{F(x) \in \mathcal{F}_0} \alpha(F(x)).$$

Отметим, что в дальнейшем нам довольно часто будут встречаться критерии, уровень значимости $\alpha(F(x))$ которых не зависит от конкретной функции распределения $F(x)$ (из подмножества \mathcal{F}_0) и, естественно, совпадает с размером критерия α_0 . В таких случаях мы будем говорить просто об уровне значимости, не связывая его с конкретным распределением $F(x)$, а в скобках писать «размер».

Равномерно наиболее мощным критерием заданного размера α_0 будем называть критерий, имеющий среди всех критериев размера α_0 наибольшую мощность $\beta = \beta(F(x))$ при любом распределении $F(x) \in \mathcal{F}_1$. Равномерно наиболее мощные критерии существуют в крайне редких случаях, наиболее известными из которых являются случай простых гипотез H_0 и H_1 и случай односторонней и двусторонней параметрических гипотез для некоторых однопараметрических семейств $F(x; \vartheta)$ (см. параграфы 2 и 3 данной главы).

В ряде задач, хотя и не существует равномерно наиболее мощный критерий, можно построить равномерно наиболее мощный несме-

щенный критерий. Критерий называется *несмешенным*, если его уровень значимости ни в коем случае не превосходит мощности, т. е. $\beta(F(x)) \geq \alpha_0$ для любых $F(x) \in \mathcal{F}_1$ (невыполнение требования несмешенности ведет к парадоксальной ситуации: в некоторых случаях мы будем чаще браковать партии годных изделий, чем негодных). Критерий, имеющий при любом $F(x) \in \mathcal{F}_1$ наибольшую мощность среди всех несмешенных критериев размера α_0 , носит название *равномерно наиболее мощного несмешенного критерия* размера α_0 . Естественно, равномерно наиболее мощный критерий является также равномерно наиболее мощным несмешенным, что вытекает из сравнения его с триадальным рандомизированным критерием (о понятии рандомизированного критерия см. ниже), который независимо от выборки X_1, \dots, X_n предписывает случайным образом с вероятностью α_0 принять конкурирующую гипотезу H_1 ; очевидно, и мощность, и уровень значимости такого критерия тождественно равны α_0 .

Наконец, еще один класс критериев представляют так называемые инвариантные критерии. Критерий называется *инвариантным* относительно группы преобразований G выборки X_1, \dots, X_n , если он не зависит от преобразований из этой группы. В частности, от любых критериев, проверяющих гипотезы о теоретической функции распределения $F(x)$, разумно потребовать инвариантность относительно всех перестановок элементов выборки X_1, \dots, X_n , иначе сам критерий зависел бы от того, в каком порядке проводились наблюдения. Обычно группа преобразований G естественным образом определяется рассматриваемыми гипотезами H_0 и H_1 , и поэтому говорят просто об инвариантных критериях. Критерий, имеющий при заданном размере α_0 наибольшую мощность среди всех инвариантных критериев, называется *равномерно наиболее мощным инвариантным критерием*. Поскольку понятие инвариантного критерия является сложным, мы в дальнейшем в основном тексте не будем его использовать, отсылая читателя к специальной литературе (например, [11, 12]), хотя это понятие и прояснило бы смысл построения некоторых критериев.

В общем случае, когда не существует ни равномерно наиболее мощного, ни даже равномерно наиболее мощного несмешенного или равномерно наиболее мощного инвариантного критерия, критерии обычно строят, опираясь на интуитивные соображения разумности. К таким соображениям относятся: простота и наглядность критерия, независимость уровня значимости критерия от вида теоретической функции распределения, асимптотическая эффективность критерия и т. д.

Как правило, критерии строят, основываясь на *статистике критерия*. Пусть $S = S(X_1, \dots, X_n)$ — произвольная (одномерная) статистика. Тогда неравенство $S > C$ определяет критерий, критическая область которого W_k состоит из всех точек (x_1, \dots, x_n) n -мерного пространства R^n , для которых

$$S(x_1, \dots, x_n) > C.$$

Заставляя критическое значение пробегать все числа от $-\infty$ до ∞ , получим семейство критериев, имеющих различные уровни значимости и мощности. Теперь, если, например, мы хотим получить критерий заданного размера α_0 , то должны выбрать такое значение C , которое давало бы нам критерий требуемого размера α_0 . Обычно критерии строят таким образом, чтобы статистика S при условии справедливости основной гипотезы H_0 имела (хотя бы асимптотически при $n \rightarrow \infty$) одно из распределений, описанных в параграфе 4 гл. 1. Тогда критическое значение C определяется как $(1 - \alpha_0)$ -квантиль соответствующего распределения. Разумеется, как уже говорилось ранее, статистика S должна по возможности обеспечивать меньший уровень значимости и большую мощность построенного на ее основе критерия.

В заключение этого параграфа скажем несколько слов о рандомизированных критериях. *Рандомизированные критерии* возникают (обычно в случае дискретной наблюдаемой случайной величины X) тогда, когда критерий, определяемый неравенством

$$S(X_1, \dots, X_n) \geq C,$$

имеет размер больше требуемого, а неравенством

$$S(X_1, \dots, X_n) > C$$

— уже меньше требуемого. В этом случае наряду с критической W_k и допустимой W_d областями вводят *область рандомизации* W_p ; при попадании выборки X_1, \dots, X_n в область W_p производят дополнительное испытание типа подбрасывания несимметричной монеты и по его исходу принимают либо основную гипотезу H_0 , либо конкурирующую H_1 . Здесь в основном тексте мы не будем рассматривать рандомизированные критерии.

2. Простые гипотезы

Изучение статистических критериев начнем со случая двух простых гипотез. Пусть выборка X_1, \dots, X_n произведена из генеральной совокупности с теоретической функцией распределения $F(x)$, относительно которой имеются две простые гипотезы: основная H_0 : $F(x) = F_0(x)$ и конкурирующая H_1 : $F(x) = F_1(x)$, где $F_0(x)$ и $F_1(x)$ — известные функции распределения. Поскольку гипотезы H_0 и H_1 простые, уровень значимости и мощность каждого критерия для проверки этих гипотез будут представлять собой два числа: α и β . Оказывается, в данном случае существует наиболее мощный критерий (при двух простых гипотезах вместо «равномерно наиболее мощный критерий» говорят просто «наиболее мощный критерий»), т. е. критерий, имеющий при заданном уровне значимости α наибольшую мощность β . Этот критерий называется *критерием отношения правдоподобия* и описывается следующим образом.

Введем статистику

$$\Lambda = \Lambda(X_1, \dots, X_n) = \frac{L_1(X_1, \dots, X_n)}{L_0(X_1, \dots, X_n)},$$

где

$$L_0(X_1, \dots, X_n) = P_0(X_1) \cdots P_0(X_n)$$

в случае дискретной наблюдаемой величины X и

$$L_0(X_1, \dots, X_n) = p_0(X_1) \cdots p_0(X_n)$$

в случае непрерывной X представляет собой функцию правдоподобия при условии справедливости гипотезы H_0 , а

$$L_1(X_1, \dots, X_n) = P_1(X_1) \cdots P_1(X_n)$$

в дискретном случае и

$$L_1(X_1, \dots, X_n) = p_1(X_1) \cdots p_1(X_n)$$

в непрерывном — ту же самую функцию правдоподобия, но при условии справедливости гипотезы H_1 . Статистика Λ носит название *отношения правдоподобия* и является отношением вероятностей (или плотностей распределения) получить выборку X_1, \dots, X_n при условиях справедливости гипотез H_1 и H_0 . Естественно предположить (и это подтверждается леммой Неймана–Пирсона, которую мы докажем ниже), что чем больше отношение правдоподобия, тем большее предпочтение мы должны оказать гипотезе H_1 . Таким образом, *критическая область W_k критерия отношения правдоподобия состоит из всех тех точек (x_1, \dots, x_n) , для которых $\Lambda(x_1, \dots, x_n)$ больше критического значения C .* Критерий отношения правдоподобия подобен рачительной хозяйке, которая всегда на имеющиеся деньги старается купить как можно больше товаров.

Теорема 1 (лемма Неймана–Пирсона). Среди всех критериев заданного уровня значимости α , проверяющих две простые гипотезы H_0 и H_1 , критерий отношения правдоподобия является наиболее мощным.

Доказательство. Пусть критерий отношения правдоподобия уровня значимости α для проверки H_0 и H_1 задается критической областью W_k . Рассмотрим любой другой критерий того же уровня значимости для проверки тех же гипотез и обозначим через W'_k его критическую область. Тогда при попадании выборки X_1, \dots, X_n в область $V = W_k \setminus W'_k$ мы должны принять гипотезу H_1 по критерию отношения правдоподобия, но отвергнуть в соответствии со вторым критерием, а при попадании в область $V' = W'_k \setminus W_k$ — наоборот, отвергнуть по критерию отношения правдоподобия, но принять в соответствии со вторым критерием (рис. 1). Тогда, поскольку оба критерия имеют одинаковый уровень значимости,

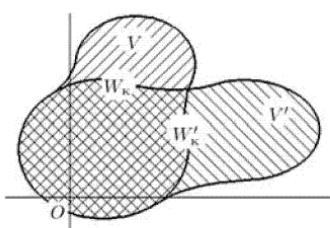


Рис. 1

то вероятности попадания выборки X_1, \dots, X_n в области V и V' при условии справедливости основной гипотезы H_0 равны, т. е.

$$\begin{aligned} \mathsf{P}\{(X_1, \dots, X_n) \in V \mid H_0\} &= \int_V \dots \int p_0(x_1) \dots p_0(x_n) dx_1 \dots dx_n = \\ &= \mathsf{P}\{(X_1, \dots, X_n) \in V' \mid H_0\} = \int_{V'} \dots \int p_0(x_1) \dots p_0(x_n) dx_1 \dots dx_n = \gamma. \end{aligned} \quad (1)$$

Далее, мощность β критерия отношения правдоподобия задается как суммарная вероятность попадания выборки в пересечение областей W_k и W'_k и область V при условии справедливости конкурирующей гипотезы H_1 :

$$\beta = \mathsf{P}\{(X_1, \dots, X_n) \in W_k \cap W'_k \mid H_1\} + \int_V \dots \int p_1(x_1) \dots p_1(x_n) dx_1 \dots dx_n. \quad (2)$$

Аналогично определяется мощность второго критерия:

$$\beta' = \mathsf{P}\{(X_1, \dots, X_n) \in W_k \cap W'_k \mid H_1\} + \int_{V'} \dots \int p_1(x_1) \dots p_1(x_n) dx_1 \dots dx_n. \quad (3)$$

Вспоминая теперь, что по построению критерия отношения правдоподобия отношение правдоподобия

$$\Lambda = \frac{p_1(X_1) \dots p_1(X_n)}{p_0(X_1) \dots p_0(X_n)}$$

в области V больше C , а в области V' не превосходит C , получаем из (1)–(3):

$$\beta \geq \mathsf{P}\{X_1, \dots, X_n \in W_k \cap W'_k \mid H_1\} + C\gamma \geq \beta'.$$

Значит, мощность второго критерия не больше мощности критерия отношения правдоподобия. \square

Замечание 1 к теореме 1. Нетрудно видеть, что мощности критерия отношения правдоподобия и второго критерия совпадают тогда и только тогда, когда $\gamma = 0$, т. е. вероятности попадания в области V и V' при условии справедливости как основной, так и конкурирующей гипотез равны нулю. Поэтому критерий отношения правдоподобия единственен (с точностью до множества, вероятность попадания в которое равна нулю).

Замечание 2 к теореме 1. Мы рассмотрели критерий отношения правдоподобия, критическая область которого задается неравенством $\Lambda > C$. Аналогично можно было бы ввести критерий отношения правдоподобия с критической областью $\Lambda \geq C$, имеющий то же самое свойство оптимальности.

Пусть нам теперь нужно построить наиболее мощный критерий заданного уровня значимости α . Может случиться так, что, определив критическое значение C , мы придем к следующей ситуации: критерий, задаваемый неравенством $\Lambda > C$, будет иметь уровень значимости α' меньше α ($\alpha' < \alpha$), а задаваемый неравенством $\Lambda \geq C$ – уровень значимости α'' уже больше α ($\alpha'' > \alpha$).

Возможный способ устраниТЬ возникшее затруднение – добавить к критической области W_k , задаваемой неравенством $\Lambda > C$, некоторую «лишнюю» подобласть W_l области W_p , определяемой равенством $\Lambda = C$, с таким расчетом, чтобы вероятность попадания в W_l при справедливой H_0 равнялась $\alpha - \alpha'$. Нетрудно видеть, что если это удастся сделать, то построенный критерий будет наиболее мощным критерием уровня значимости α .

Но, с одной стороны, подобласть W_l , вообще говоря, можно выбрать не единственным способом, поэтому могут появляться различные наиболее мощ-

ные критерии, отличающиеся друг от друга только при попадании в область W_p . С другой стороны, в дискретном случае этого, как правило, вообще нельзя сделать, поскольку область W_p состоит из конечного числа точек. Поэтому обычно идут более простым путем: при попадании выборки X_1, \dots, X_n в область W_p , которую называют *областью рандомизации*, проводят дополнительный независимый эксперимент с двумя исходами типа подбрасывания несимметричной монеты с вероятностью выпадения «цифры» $(\alpha - \alpha')/(\alpha'' - \alpha')$. При этом если выпадает «герб», то принимают гипотезу H_0 , «цифра» — гипотезу H_1 . Такой критерий называется *рандомизированным*. В соответствии с формулой полной вероятности уровень значимости рандомизированного критерия равен $\alpha' + (\alpha'' - \alpha')(\alpha - \alpha')/(\alpha'' - \alpha') = \alpha$. Очевидно также, что рандомизированный критерий отношения правдоподобия будет наиболее мощным. Естественно, в наше время при дополнительном эксперименте не бросают монету, а моделируют на ЭВМ псевдослучайное число.

В дальнейшем, как уже говорилось, мы не будем рассматривать рандомизированные критерии, хотя это и вызовет у нас определенные трудности при формулировке и доказательстве некоторых результатов.

При практической реализации критерия отношения правдоподобия обычно удобно пользоваться не отношением правдоподобия Λ , а его логарифмом $\lambda = \ln \Lambda$. В этом случае мы должны принять гипотезу H_0 , если $\lambda = \lambda(X_1, \dots, X_n) \leq c = \ln C$, и отвергнуть ее (принять гипотезу H_1), если $\lambda > c$. В соответствии с общим правилом уровень значимости α и мощность β критерия отношения правдоподобия в зависимости от критического значения c определяются формулами

$$\alpha = \alpha(c) = \mathbf{P}\{\lambda(X_1, \dots, X_n) > c \mid H_0\} = \sum P_0(x_1) \cdots P_0(x_n),$$

$$\beta = \beta(c) = \mathbf{P}\{\lambda(X_1, \dots, X_n) > c \mid H_1\} = \sum P_1(x_1) \cdots P_1(x_n)$$

в дискретном случае, где сумма берется по всем значениям x_1, \dots, x_n , принадлежащим множеству $\{b_1, \dots, b_L\}$ и удовлетворяющим условию $\lambda(x_1, \dots, x_n) > c$, и

$$\alpha = \alpha(c) = \mathbf{P}\{\lambda(X_1, \dots, X_n) > c \mid H_0\} =$$

$$= \int_{\lambda(x_1, \dots, x_n) > c} \cdots \int p_0(x_1) \cdots p_0(x_n) dx_1 \cdots dx_n,$$

$$\beta = \beta(c) = \mathbf{P}\{\lambda(X_1, \dots, X_n) > c \mid H_1\} =$$

$$= \int_{\lambda(x_1, \dots, x_n) > c} \cdots \int p_1(x_1) \cdots p_1(x_n) dx_1 \cdots dx_n$$

в непрерывном. В этих формулах запись $\mathbf{P}\{S(X_1, \dots, X_n) > C \mid H_i\}$ ($i = 0, 1$) обозначает вероятность статистике $S(X_1, \dots, X_n)$ принять значение больше C при условии справедливости гипотезы H_i .

Если задан уровень значимости α , то критическое значение c определяется из решения относительно c уравнения $\alpha = \alpha(c)$.

Аналогично поступают и в том случае, когда задана мощность β .

Наконец, встречаются задачи, в которых одновременно заданы и уровень значимости α , и мощность β , а требуется определить минимальный объем выборки n , при котором можно разделить гипотезы H_0 и H_1 с такими значениями α и β . В этом случае заметим, что функции $\alpha(c) = \alpha(c, n)$ и $\beta(c) = \beta(c, n)$ зависят, разумеется, и от объема выборки n . Необходимый объем выборки определяется как минимальное значение n , при котором существуют решения неравенств $\alpha \geq \alpha(c, n)$ и $\beta \leq \beta(c, n)$.

Пример 1. Пусть выборка X_1, \dots, X_n произведена из нормальной генеральной совокупности с известной дисперсией σ^2 и неизвестным средним ϑ , относительно которого имеются две гипотезы: основная $H_0: \vartheta = \vartheta_0$ и конкурирующая $H_1: \vartheta = \vartheta_1$ (для определенности пусть $\vartheta_1 > \vartheta_0$).

Построим критерий отношения правдоподобия для проверки гипотез H_0 и H_1 . Для этого выпишем функции правдоподобия $L_0(X_1, \dots, X_n)$ и $L_1(X_1, \dots, X_n)$:

$$L_0(X_1, \dots, X_n) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n e^{-\frac{1}{2\sigma^2}[(X_1 - \vartheta_0)^2 + \dots + (X_n - \vartheta_0)^2]},$$

$$L_1(X_1, \dots, X_n) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n e^{-\frac{1}{2\sigma^2}[(X_1 - \vartheta_1)^2 + \dots + (X_n - \vartheta_1)^2]}$$

и определим логарифм отношения правдоподобия

$$\lambda(X_1, \dots, X_n) = \ln \frac{L_1(X_1, \dots, X_n)}{L_0(X_1, \dots, X_n)} = \frac{1}{2\sigma^2} \left[(X_1 - \vartheta_0)^2 + \dots + (X_n - \vartheta_0)^2 - (X_1 - \vartheta_1)^2 - \dots - (X_n - \vartheta_1)^2 \right] = \frac{\vartheta_1 - \vartheta_0}{\sigma^2} \left(X_1 + \dots + X_n - n \frac{\vartheta_0 + \vartheta_1}{2} \right).$$

Таким образом, мы должны принять гипотезу H_0 , если

$$\frac{\vartheta_1 - \vartheta_0}{\sigma^2} \left(X_1 + \dots + X_n - n \frac{\vartheta_0 + \vartheta_1}{2} \right) < c,$$

или, вводя обозначение

$$c^* = \frac{c\sigma^2}{\vartheta_1 - \vartheta_0} + \frac{n(\vartheta_0 + \vartheta_1)}{2},$$

если $X_1 + \dots + X_n < c^*$, и отвергнуть в противном случае.

Для того чтобы при заданном c^* найти уровень значимости критерия α , заметим, что при условии справедливости гипотезы H_0 статистика $X_1 + \dots + X_n$ распределена по нормальному закону с параметрами $n\vartheta_0$ и $n\sigma^2$. Поэтому

$$\begin{aligned} \alpha &= P\{X_1 + \dots + X_n > c^* | H_0\} = 1 - \Phi(c^*; n\vartheta_0, n\sigma^2) = \\ &= 1 - \Phi\left(\frac{c^* - n\vartheta_0}{\sqrt{n\sigma^2}}\right) = \Phi\left(\frac{n\vartheta_0 - c^*}{\sqrt{n\sigma^2}}\right), \end{aligned}$$

где, как обычно, $\Phi(x)$ — функция стандартного нормального распределения. Аналогично определяется мощность критерия β

$$\begin{aligned} \beta &= P\{X_1 + \dots + X_n > c^* | H_1\} = 1 - \Phi(c^*; n\vartheta_1, n\sigma^2) = \\ &= 1 - \Phi\left(\frac{c^* - n\vartheta_1}{\sqrt{n\sigma^2}}\right) = \Phi\left(\frac{n\vartheta_1 - c^*}{\sqrt{n\sigma^2}}\right). \end{aligned}$$

Если задан уровень значимости α , то критическое значение c^* определяется формулой

$$c^* = n\vartheta_0 - \varphi_\alpha \sqrt{n\sigma^2},$$

где φ_α — α -квантиль стандартного нормального закона.

Наконец, если заданы уровень значимости α и мощность β и требуется найти минимальный объем выборки n , позволяющий разделить гипотезы H_0 и H_1 с такими α и β , то c^* должно удовлетворять двум равенствам:

$$c^* = n\vartheta_0 - \varphi_\alpha \sqrt{n\sigma^2}, \quad c^* = n\vartheta_1 - \varphi_\beta \sqrt{n\sigma^2}.$$

Вычитая второе равенство из первого, получаем

$$n(\vartheta_1 - \vartheta_0) = (\varphi_\beta - \varphi_\alpha) \sqrt{n\sigma^2} \quad \text{или} \quad n = \frac{(\varphi_\beta - \varphi_\alpha)^2 \sigma^2}{(\vartheta_1 - \vartheta_0)^2}.$$

Разумеется, реальный объем выборки должен быть ближайшим к n сверху целым числом. \square

Зададим вопрос: а можно ли для проверки двух простых гипотез H_0 и H_1 построить критерий с заданным уровнем значимости α и мощностью β , который потребовал бы меньшего объема выборки, чем критерий отношения правдоподобия? Очевидно, среди критериев с фиксированным объемом выборки такого нет. Однако можно рассматривать *последовательные критерии*, в которых испытания проводятся последовательно, и после каждого испытания мы вправе либо прекратить испытания и принять одну из гипотез H_0 или H_1 , либо продолжить наблюдения. При последовательном проведении испытаний после первого же испытания могут появиться столь серьезные основания в пользу той или иной гипотезы, что дальнейшие наблюдения становятся просто бессмысленными. Так, хороший врач по одному ярко выраженому признаку может поставить точный диагноз пациенту. Ясно, что у последовательного критерия число наблюдений случайно и поэтому вместо объема выборки следует оперировать средним числом наблюдений, которое, естественно, будет различным при условиях справедливости гипотез H_0 и H_1 . Оказывается, среди последовательных критериев также существует наилучший, который мы сейчас и опишем.

Последовательный критерий отношения правдоподобия (критерий Вальда) строят, опираясь на логарифм отношения правдоподобия (и это естественно, поскольку отношение правдоподобия — наиболее объективная мера различия гипотез). Сама процедура принятия решения реализуется следующим образом. Задаются критические значения c_0 и c_1 ($c_0 < c_1$). Проводят первое испытание и по его результату X_1 определяют логарифм отношения правдоподобия

$$\lambda(X_1) = \ln \frac{P_1(X_1)}{P_0(X_1)}$$

в дискретном случае или

$$\lambda(X_1) = \ln \frac{p_1(X_1)}{p_0(X_1)}$$

в непрерывном. Если $\lambda(X_1) \geq c_1$, то принимают гипотезу H_1 , если $\lambda(X_1) \leq c_0$ — гипотезу H_0 , если же $c_0 < \lambda(X_1) < c_1$, то проводят второе испытание. Снова определяют логарифм отношения правдоподобия

$$\lambda(X_1, X_2) = \lambda(X_1) + \lambda(X_2)$$

и сравнивают его с c_1 и c_0 . Если $\lambda(X_1, X_2) \geq c_1$, то принимают гипотезу H_1 , если $\lambda(X_1, X_2) \leq c_0$ — гипотезу H_0 , если $c_0 < \lambda(X_1, X_2) < c_1$, то переходят к третьему испытанию, по результатам которого определяют

$$\lambda(X_1, X_2, X_3) = \lambda(X_1) + \lambda(X_2) + \lambda(X_3)$$

и т. д. Графическая схема проведения испытаний приведена на рис. 2. Нахождение уровня значимости α и мощности β последовательного

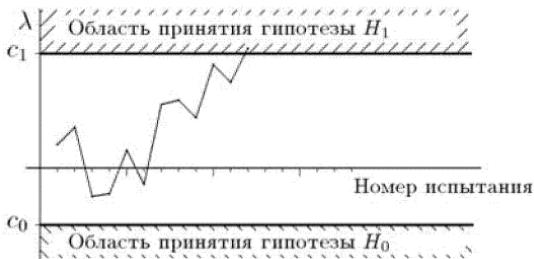


Рис. 2

критерия отношения правдоподобия представляет собой весьма сложную задачу (задача выхода случайного блуждания из полосы), все имеющиеся решения которой (за исключением некоторых частных случаев) нужно признать неудовлетворительными в смысле методов вычисления. Обычно используют следующие приближенные оценки:

$$\alpha \approx \frac{1 - e^{c_0}}{e^{c_1} - e^{c_0}}, \quad \beta \approx \frac{e^{c_1}(1 - e^{c_0})}{e^{c_1} - e^{c_0}},$$

точность которых, как правило, увеличивается с ростом c_1 и уменьшением c_0 . Из этих оценок можно при заданных α и β найти приближенные выражения для критических значений c_0 и c_1 :

$$c_0 \approx \ln \frac{1 - \beta}{1 - \alpha}, \quad c_1 \approx \ln \frac{\beta}{\alpha}.$$

Другими важными характеристиками последовательного критерия отношения правдоподобия являются средние числа наблюдений N_0 и N_1 при условиях справедливости гипотез H_0 и H_1 , которые можно приближенно вычислить по формулам

$$N_0 \approx \frac{\alpha c_1 + (1 - \alpha)c_0}{\mathbf{M}[\lambda(X) | H_0]}, \quad N_1 \approx \frac{\beta c_1 + (1 - \beta)c_0}{\mathbf{M}[\lambda(X) | H_1]},$$

где

$$\mathbf{M}[\lambda(X) | H_0] = \sum_{l=1}^L \ln \frac{P_1(b_l)}{P_0(b_l)} P_0(b_l),$$

$$\mathbf{M}[\lambda(X) | H_1] = \sum_{l=1}^L \ln \frac{P_1(b_l)}{P_0(b_l)} P_1(b_l)$$

в дискретном случае и

$$\mathbf{M}[\lambda(X) | H_0] = \int_{-\infty}^{\infty} \ln \frac{p_1(x)}{p_0(x)} p_0(x) dx,$$

$$\mathbf{M}[\lambda(X) | H_1] = \int_{-\infty}^{\infty} \ln \frac{p_1(x)}{p_0(x)} p_1(x) dx$$

в непрерывном.

Оптимальные свойства последовательного критерия отношения правдоподобия задаются следующей теоремой.

Теорема 2 (Вальда). Среди всех критериев (последовательных или нет) с заданными уровнем значимости α и мощностью β и конечными средними числами наблюдений N_0 при условии справедливости гипотезы H_0 и N_1 при условии справедливости гипотезы H_1 последовательный критерий отношения правдоподобия минимизирует как N_0 , так и N_1 .

Доказательство теоремы наиболее просто получается с использованием байесовского подхода. Пусть гипотезы H_0 и H_1 появляются случайным образом с известными вероятностями π появления гипотезы H_0 и $1 - \pi$ гипотезы H_1 . Пусть также задана стоимость d каждого наблюдения и при неправильном отклонении гипотезы H_0 (H_1) мы несем потери w_0 (w_1). Тогда для любого (в том числе последовательного) критерия K общий байесовский риск (общие средние потери) $R(\pi, K)$ определяются формулой

$$R(\pi, K) = \pi(\alpha_0 w_0 + dN_0) + (1 - \pi)(\alpha_1 w_1 + dN_1), \quad (4)$$

где α_0 (α_1) — вероятность неправильного отклонения гипотезы H_0 (H_1), а N_0 (N_1) — среднее число наблюдений при условии справедливости гипотезы H_0 (H_1).

Доказательство состоит из двух основных частей. Первая часть, представленная леммой 2.1, определяет байесовскую процедуру, минимизирующую (4). Отметим, что интерпретация (4) как байесовского риска помогает лучше понять доказательство и приводит к задачам, представляющим самостоятельный интерес.

Прежде чем сформулировать лемму 2.1, произведем некоторые дополнительные рассмотрения. Обозначим через $\rho(\pi)$ минимальный байесовский риск для всех критериев, требующих хотя бы одного наблюдения (класс таких

критериев обозначим через \mathcal{K}). Тогда в силу линейности байесовского риска $R(\pi, K)$ по π для $K \in \mathcal{K}$ и любого γ ($0 < \gamma < 1$) справедливо соотношение

$$\begin{aligned} \rho(\gamma\pi_1 + (1 - \gamma)\pi_2) &= \inf_{K \in \mathcal{K}} R(\gamma\pi_1 + (1 - \gamma)\pi_2, K) = \\ &= \inf_{K \in \mathcal{K}} [\gamma R(\pi_1, K) + (1 - \gamma)R(\pi_2, K)] \geqslant \\ &\geqslant \gamma \inf_{K \in \mathcal{K}} R(\pi_1, K) + (1 - \gamma) \inf_{K \in \mathcal{K}} R(\pi_2, K) = \gamma\rho(\pi_1) + (1 - \gamma)\rho(\pi_2). \end{aligned}$$

Следовательно, $\rho(\pi)$ — выпуклая (вверх) функция (рис. 3).

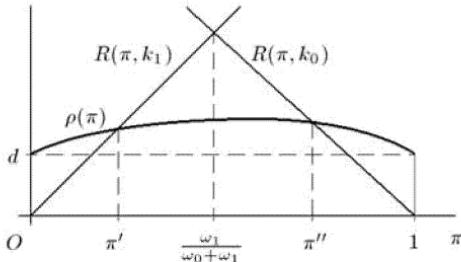


Рис. 3

Рассмотрим критерий K_0 (K_1), отвергающий H_0 (H_1) без проведения испытаний. Имеем

$$\begin{aligned} R(\pi, K_0) &= \pi w_0 \\ (R(\pi, K_1) &= (1 - \pi)w_1). \end{aligned}$$

Если

$$\rho\left(\frac{w_1}{w_0 + w_1}\right) < \frac{w_0 w_1}{w_0 + w_1},$$

то определим π' и π'' из решения следующих уравнений (см. рис. 3):

$$R(\pi', K_0) = \rho(\pi'), \quad R(\pi'', K_1) = \rho(\pi''); \tag{5}$$

в противном случае положим

$$\pi' = \pi'' = \frac{w_1}{w_0 + w_1}. \tag{6}$$

В силу свойств выпуклости и положительности функции $\rho(\pi)$ числа π' и π'' , удовлетворяющие уравнениям (5) или (6), определяются единственным образом.

Лемма 2.1. *Если $0 < \pi' < \pi'' < 1$, то при всех π ($\pi' < \pi < \pi''$) байесовский риск (4) минимизирует последовательный критерий отношения правдоподобия с критическими значениями*

$$c_0 = \ln C_0 = \ln \left(\frac{\pi}{1 - \pi} \cdot \frac{1 - \pi''}{\pi''} \right), \quad c_1 = \ln C_1 = \ln \left(\frac{\pi}{1 - \pi} \cdot \frac{1 - \pi'}{\pi'} \right). \tag{7}$$

Доказательство леммы 2.1. Заметим прежде всего, что если $\pi \leqslant \pi'$ или $\pi \geqslant \pi''$, то вообще не имеет смысла проводить испытания. Поэтому оптимальный критерий состоит в следующем поведении на первом шаге: мы принимаем гипотезу H_0 , если $\pi \leqslant \pi'$, отвергаем, если $\pi \geqslant \pi''$, и производим первое наблюдение, если $\pi' < \pi < \pi''$.

Доказательство леммы 2.1 завершается теперь по индукции. Действительно, если уже сделано n наблюдений X_1, \dots, X_n , то мы попадаем в ту же ситуацию, что и перед первым наблюдением: либо не производить дополнительные наблюдения и принять H_0 или H_1 с потерями w_0 или w_1 в случае неправильного решения, либо произвести $(n+1)$ -е наблюдение. Наличие предыдущих наблюдений в силу аддитивности потерь никак не влияет на решение задачи, поскольку их вклад в общие потери последующими испытаниями уже нельзя ни увеличить, ни уменьшить. Итак, процедура принятия решения прежняя: если $\pi \leq \pi'$ или $\pi \geq \pi''$, то испытания прекращаются и принимается H_0 или H_1 , а если $\pi' < \pi < \pi''$, то производится дополнительное наблюдение. Однако при этом мы должны оперировать уже не априорной вероятностью π , а апостериорной вероятностью $\pi(X_1, \dots, X_n)$ справедливости гипотезы H_0 при условии выборки X_1, \dots, X_n , определяемой формулой Байеса

$$\pi(X_1, \dots, X_n) = \frac{\pi p_0(X_1) \cdots p_0(X_n)}{\pi p_0(X_1) \cdots p_0(X_n) + (1-\pi) p_1(X_1) \cdots p_1(X_n)}.$$

Оптимальный байесовский критерий, таким образом, требует продолжения наблюдений только в случае $\pi' < \pi(X_1, \dots, X_n) < \pi''$ или после проведения элементарных преобразований в случае

$$C_0 = \frac{\pi}{1-\pi} \cdot \frac{1-\pi''}{\pi''} < \frac{p_1(X_1) \cdots p_1(X_n)}{p_0(X_1) \cdots p_0(X_n)} < \frac{\pi}{1-\pi} \cdot \frac{1-\pi'}{\pi'} = C_1. \quad \square$$

Вторая часть доказательства теоремы 2 устанавливает связь между параметрами основной и байесовской задач и состоит в доказательстве следующей леммы.

Лемма 2.2. Для любых π' и π'' ($0 < \pi' < \pi'' < 1$) найдутся числа w ($0 < w < 1$) и $d > 0$, такие, что решение байесовской задачи с потерями $w_0 = 1-w$, $w_1 = w$, стоимостью одного испытания d и априорной вероятностью π появления гипотезы H_0 задается последовательным критерием отношения правдоподобия с границами

$$C_0 = \frac{\pi}{1-\pi} \cdot \frac{1-\pi''}{\pi''}, \quad C_1 = \frac{\pi}{1-\pi} \cdot \frac{1-\pi'}{\pi'}.$$

Доказательство леммы 2.2 представляет, по сути дела, доказательство существования обратного отображения для двумерного вектора (π', π'') как функции от w и d . Оно носит чисто аналитический характер и основывается на таких почти очевидных свойствах π' и π'' , как монотонность, непрерывность и т. д. Поэтому мы его здесь не приводим, отправляя заинтересованного читателя к [12]. Отметим, что требование леммы $w_0 + w_1 = 1$ нисколько не ограничивает общности задачи, поскольку этого всегда можно добиться линейной заменой «масштаба цен» («денежной реформой»). \square

Для окончания доказательства теоремы 2 рассмотрим последовательный критерий отношения правдоподобия с границами C_0 и C_1 ($C_0 < 1 < C_1$). Для любого числа π положим

$$\pi' = \frac{\pi}{C_1(1-\pi) + \pi}, \quad \pi'' = \frac{\pi}{C_0(1-\pi) + \pi}.$$

Эти значения удовлетворяют равенствам (7) и неравенствам $0 < \pi' < \pi'' < 1$. Поэтому, по леммам 2.1 и 2.2, найдутся такие w ($0 < w < 1$) и $d > 0$, что рассматриваемый критерий будет являться решением байесовской задачи с априорной вероятностью π , потерями $w_0 = 1-w$ и $w_1 = w$ и стоимостью одного наблюдения d . Обозначим через α_0 , α_1 , N_0 и N_1 вероятности ошибок

и средние числа наблюдений данного последовательного критерия отношения правдоподобия. Кроме того, рассмотрим любой другой критерий для проверки гипотез H_0 и H_1 с вероятностями ошибок $\tilde{\alpha}_0 \leq \alpha_0$ и $\tilde{\alpha}_1 \leq \alpha_1$ и средними числами наблюдений \tilde{N}_0 и \tilde{N}_1 . Поскольку последовательный критерий отношения правдоподобия минимизирует байесовский риск, то применяя к байесовской задаче и второй критерий, имеем

$$\begin{aligned} \pi[(1-w)\alpha_0 + dN_0] + (1-\pi)(w\alpha_1 + dN_1) &\leq \\ &\leq \pi[(1-w)\tilde{\alpha}_0 + d\tilde{N}_0] + (1-\pi)(w\tilde{\alpha}_1 + d\tilde{N}_1), \end{aligned}$$

откуда получаем

$$\pi N_0 + (1-\pi)N_1 \leq \pi \tilde{N}_0 + (1-\pi)\tilde{N}_1.$$

Из справедливости последнего неравенства при всех π ($0 < \pi < 1$) вытекают, в частности, соотношения $N_0 \leq \tilde{N}_0$ и $N_1 \leq \tilde{N}_1$, что и доказывает теорему. \square

Замечание к теореме 2. Анализируя доказательство леммы 2.1, нетрудно заметить, что если на первом шаге $\pi = \pi'$ ($\pi = \pi''$), то у нас имеются две (а с учетом рандомизации и больше) идентичные с точки зрения байесовского риска возможности: принять H_0 (H_1) или продолжить испытания; такая же ситуация может возникнуть и после любого наблюдения. В том случае, когда отношение правдоподобия

$$\Lambda(X_1, \dots, X_n) = \frac{p_1(X_1) \dots p_1(X_n)}{p_0(X_1) \dots p_0(X_n)}$$

может равняться C_0 (C_1) только с нулевой вероятностью, последнее обстоятельство не влияет на последовательный критерий отношения правдоподобия. Если же эта вероятность не равна нулю, то можно предложить, по крайней мере, две модификации критерия, одна из которых будет предписывать прекращение испытаний при равенстве отношения правдоподобия соответствующей константе, а другая, наоборот, продолжение наблюдений. Обе эти модификации, как уже говорилось, имеют одинаковый байесовский риск. Однако другие характеристики для них будут различными; в частности, вторая модификация требует большего числа наблюдений. Впрочем, никакого противоречия с изложенной теорией здесь нет, поскольку увеличение количества наблюдений компенсируется уменьшением уровня значимости и увеличением мощности. Поэтому для построения последовательного критерия отношения правдоподобия, имеющего заданные уровень значимости и мощность, необходимо, вообще говоря, привлекать рандомизацию. Подробнее на этом мы останавливаться не будем.

Пример 2. Пусть наблюдается нормально распределенная случайная величина X с известной дисперсией σ^2 и неизвестным средним ϑ , относительно которого имеются две простые гипотезы: $H_0: \vartheta = \vartheta_0$ и $H_1: \vartheta = \vartheta_1$ ($\vartheta_0 < \vartheta_1$). Как показано в примере 1, отношение правдоподобия имеет вид

$$\lambda(X_1, \dots, X_n) = \frac{\vartheta_1 - \vartheta_0}{\sigma^2} \left[X_1 + \dots + X_n - \frac{n(\vartheta_0 + \vartheta_1)}{2} \right].$$

При применении последовательного критерия отношения правдоподобия наблюдения продолжаются, пока $c_0 < \lambda(X_1, \dots, X_n) < c_1$; в противном случае принимается либо гипотеза H_0 (если $\lambda(X_1, \dots, X_n) < c_0$), либо гипотеза H_1 (если $\lambda(X_1, \dots, X_n) > c_1$).

Вычислим $\mathbf{M}[\lambda(X) | H_0]$ и $\mathbf{M}[\lambda(X) | H_1]$. Поскольку $\mathbf{M}(X | H_0) = \vartheta_0$, а $\mathbf{M}(X | H_1) = \vartheta_1$, то

$$\begin{aligned}\mathbf{M}[\lambda(X) | H_0] &= \mathbf{M}\left[\frac{\vartheta_1 - \vartheta_0}{\sigma^2} \left(X - \frac{\vartheta_1 + \vartheta_0}{2}\right) | H_0\right] = \\ &= \frac{\vartheta_1 - \vartheta_0}{\sigma^2} \left(\vartheta_0 - \frac{\vartheta_1 + \vartheta_0}{2}\right) = -\frac{(\vartheta_1 - \vartheta_0)^2}{2\sigma^2}, \\ \mathbf{M}[\lambda(X) | H_1] &= \mathbf{M}\left[\frac{\vartheta_1 - \vartheta_0}{\sigma^2} \left(X - \frac{\vartheta_1 + \vartheta_0}{2}\right) | H_1\right] = \\ &= \frac{\vartheta_1 - \vartheta_0}{\sigma^2} \left(\vartheta_1 - \frac{\vartheta_1 + \vartheta_0}{2}\right) = \frac{(\vartheta_1 - \vartheta_0)^2}{2\sigma^2}.\end{aligned}$$

Предположим теперь, что заданы конкретные значения: $\alpha = 0,05$, $\beta = 0,9$, $\sigma^2 = 1$, $\vartheta_0 = -0,5$ и $\vartheta_1 = 0,5$. Тогда $c_0 = -2,25$, $c_1 = 2,89$, $\mathbf{M}[\lambda(X) | H_0] = -0,5$, $\mathbf{M}[\lambda(X) | H_1] = 0,5$, откуда определяем средние числа наблюдений $N_0 \approx 4$ и $N_1 \approx 4,75$. Для сравнения, подставив в примере 1 значения $\varphi_{0,9} = 1,281552$ и $\varphi_{0,05} = -1,644854$ ([1, с. 136]), находим, что при использовании обычного критерия отношения правдоподобия для разделения гипотез H_0 и H_1 с уровнем значимости $\alpha = 0,05$ и мощностью $\beta = 0,9$ необходимо иметь выборку объема $n = 9$. Значит, применение последовательного критерия отношения правдоподобия в случае справедливости гипотезы H_0 позволяет в среднем сократить число наблюдений более чем в 2 раза, а в случае справедливости гипотезы H_1 — почти в 2 раза. \square

У последовательного критерия отношения правдоподобия можно отметить два основных недостатка.

Первым является невозможность одновременного проведения нескольких испытаний.

Второй связан с тем, что если произошла ошибка в определении гипотез H_0 и H_1 и истинная теоретическая функция распределения $F(x)$ заключена между гипотетическими $F_0(x)$ и $F_1(x)$, то потребуется очень большое число наблюдений, поскольку логарифм отношения правдоподобия будет колебаться вокруг нуля, не выходя ни на одну из границ c_0 или c_1 . Для того чтобы компенсировать этот второй недостаток, на практике часто принудительно ограничивают число наблюдений.

3. Однопараметрические гипотезы. Равномерно наилучшие критерии

Пусть выборка X_1, \dots, X_n извлечена из генеральной совокупности с теоретической функцией распределения $F(x)$, принадлежащей однопараметрическому семейству $F(X; \vartheta)$ с неизвестным параметром ϑ , область возможных значений которого будем обозначать через Θ (обычно Θ представляет собой либо всю прямую $(-\infty, \infty)$, либо полупрямую (a, ∞) или $(-\infty, a)$, либо отрезок $[a, b]$). Предположим, что в Θ выделено некоторое подмножество Θ_0 ; дополнение к нему в Θ обозначим через $\Theta_1 = \Theta \setminus \Theta_0$. Относительно неизвестного параметра ϑ имеются две гипотезы: основная $H_0: \vartheta \in \Theta_0$ и конкурирующая $H_1: \vartheta \in \Theta_1$. Зада-

ча состоит в построении критерия для проверки двух (вообще говоря, сложных) однопараметрических гипотез H_0 и H_1 . Отметим, что если подмножество Θ_0 состоит всего из одной точки ϑ_0 , то критерии для проверки гипотез H_0 и H_1 называют обычно критериями значимости. В этом параграфе мы рассмотрим те случаи, в которых существуют равномерно наиболее мощные несмещенные критерии. Построение критериев в остальных случаях можно проводить по общему рецепту, приведенному для (много) параметрических гипотез в следующем параграфе.

Односторонние гипотезы. Пусть подмножество Θ_0 состоит из всех $\vartheta \leq \vartheta_0$; соответственно подмножество Θ_1 будет содержать все $\vartheta > \vartheta_0$. Таким образом, мы имеем две односторонние гипотезы: основную H_0 : $\vartheta \leq \vartheta_0$ и конкурирующую H_1 : $\vartheta > \vartheta_0$. Возьмем произвольные $\vartheta' < \vartheta''$ и составим отношение правдоподобия

$$\Lambda(X_1, \dots, X_n; \vartheta', \vartheta'') = \frac{L(X_1, \dots, X_n; \vartheta'')}{L(X_1, \dots, X_n; \vartheta')},$$

где

$$L(X_1, \dots, X_n; \vartheta) = P(X_1; \vartheta) \cdots P(X_n; \vartheta)$$

в дискретном случае и

$$L(X_1, \dots, X_n; \vartheta) = p(X_1; \vartheta) \cdots p(X_n; \vartheta)$$

в непрерывном — функция правдоподобия.

Теорема 3 (односторонние гипотезы). Предположим, что существует (одномерная) статистика $S = S(X_1, \dots, X_n)$, такая, что для любых $\vartheta' < \vartheta''$ отношение правдоподобия можно представить в виде

$$\Lambda(X_1, \dots, X_n; \vartheta', \vartheta'') = \Lambda(S(X_1, \dots, X_n), \vartheta', \vartheta''),$$

причем $\Lambda(S, \vartheta', \vartheta'')$ — неубывающая функция от S . Тогда существует равномерно наиболее мощный критерий для проверки гипотез H_0 и H_1 , критическая область W_k которого состоит из всех точек (x_1, \dots, x_n) , удовлетворяющих неравенству $S(x_1, \dots, x_n) > C$, а уровень значимости $\alpha = \alpha(\vartheta)$ и мощность $\beta = \beta(\vartheta)$ являются неубывающими функциями от ϑ и, как обычно, определяются формулами

$$\begin{aligned} \alpha(\vartheta) &= \mathbf{P}\{S(X_1, \dots, X_n) > C \mid \vartheta\} = \sum_{S(x_1, \dots, x_n) > C} P(x_1; \vartheta) \cdots P(x_n; \vartheta) \\ &\quad (\vartheta \leq \vartheta_0), \end{aligned}$$

$$\begin{aligned} \beta(\vartheta) &= \mathbf{P}\{S(X_1, \dots, X_n) > C \mid \vartheta\} = \sum_{S(x_1, \dots, x_n) > C} P(x_1; \vartheta) \cdots P(x_n; \vartheta) \\ &\quad (\vartheta > \vartheta_0) \end{aligned}$$

в дискретном случае и

$$\begin{aligned}\alpha(\vartheta) &= \mathbf{P}\{S(X_1, \dots, X_n) > C \mid \vartheta\} = \\ &= \int_{S(x_1, \dots, x_n) > C} \dots \int p(x_1; \vartheta) \cdots p(x_n; \vartheta) dx_1 \cdots dx_n \quad (\vartheta \leq \vartheta_0),\end{aligned}$$

$$\begin{aligned}\beta(\vartheta) &= \mathbf{P}\{S(X_1, \dots, X_n) > C \mid \vartheta\} = \\ &= \int_{S(x_1, \dots, x_n) > C} \dots \int p(x_1; \vartheta) \cdots p(x_n; \vartheta) dx_1 \cdots dx_n \quad (\vartheta > \vartheta_0)\end{aligned}$$

в непрерывном. Здесь $\mathbf{P}\{S > C \mid \vartheta\}$ — вероятность статистике $S = S(X_1, \dots, X_n)$ принять значение больше C при условии, что истинное значение неизвестного параметра равно ϑ .

Доказательство. Используем тот факт, что предложенный критерий является наиболее мощным для проверки простых гипотез H'_0 : $\vartheta = \vartheta'$ и H'_1 : $\vartheta = \vartheta''$ при всех $\vartheta' < \vartheta''$. Действительно, для гипотез H'_0 и H'_1 критерий отношения правдоподобия определяется неравенством

$$\Lambda(X_1, \dots, X_n; \vartheta', \vartheta'') > C_0$$

или, что в силу предположений теоремы то же самое, неравенством

$$S(X_1, \dots, X_n) > C = \Lambda^{(-1)}(C_0),$$

где $\Lambda^{(-1)}(y)$ — обратная к $\Lambda(S) = \Lambda(S; \vartheta', \vartheta'')$ функция, причем поскольку критерий отношения правдоподобия является несмешанным, то уровень значимости α' и мощность β' для этого критерия удовлетворяют неравенству $\alpha' = \beta(\vartheta') \leq \beta' = \beta(\vartheta'')$, где $\beta(\vartheta)$ — функция мощности (т. е. обобщенная запись $\alpha(\vartheta)$ и $\beta(\vartheta)$, см. параграф 1). Таким образом, мы доказали, что $\alpha(\vartheta)$ и $\beta(\vartheta)$ являются неубывающими функциями от ϑ , откуда, в частности, следует: $\alpha_0 = \alpha(\vartheta_0)$.

Далее, из сказанного также вытекает, что рассматриваемый критерий является наиболее мощным уровня значимости α_0 для проверки простых гипотез \tilde{H}_0 : $\vartheta = \vartheta_0$ и \tilde{H}_1 : $\vartheta = \vartheta'$, где ϑ' — любое число, большее ϑ_0 . Но тогда он является равномерно наиболее мощным критерием размера α_0 для проверки изначальных гипотез H_0 : $\vartheta \leq \vartheta_0$ и H_1 : $\vartheta > \vartheta_0$. Действительно, любой другой критерий размера α_0 должен в точке ϑ_0 (как и в любой другой точке $\vartheta < \vartheta_0$) иметь уровень значимости $\alpha'(\vartheta_0) \leq \alpha_0$. Значит, и мощность $\beta'(\vartheta)$ этого другого критерия при любом $\vartheta' > \vartheta_0$ не должна превосходить мощности наиболее мощного критерия уровня значимости α_0 для проверки простых гипотез \tilde{H}_0 : $\vartheta = \vartheta_0$ и \tilde{H}_1 : $\vartheta = \vartheta'$, т. е. должна удовлетворять неравенству $\beta'(\vartheta) \leq \beta(\vartheta)$. Последнее неравенство и доказывает теорему. \square

Замечание 1 к теореме 3. Как обычно, чтобы соблюсти необходимую строгость приведенного доказательства, нужно привлекать рандомизацию. Кроме того, если потребовать, чтобы функции распределения $F(x; \vartheta')$ и $F(x; \vartheta'')$ были различными для разных ϑ' и ϑ'' , то утверждение теоремы о монотонности функций $\alpha(\vartheta)$ и $\beta(\vartheta)$ можно усилить до строгой монотонности. Наконец, заметим, что семейство распределений $F(x; \vartheta)$ с указанным в условии

ях теоремы свойством называют обычно семейством с монотонным отношением правдоподобия.

Замечание 2 к теореме 3. Нетрудно видеть, что при заданном размере $\alpha_0 = \alpha(\vartheta_0)$ рассматриваемый критерий наряду с максимизацией мощности $\beta(\vartheta)$ при каждом $\vartheta > \vartheta_0$ минимизирует также уровень значимости $\alpha(\vartheta)$ при любом $\vartheta < \vartheta_0$. Поэтому обращенный критерий $S(X_1, \dots, X_n) < C$, используемый для проверки основной гипотезы $H_0: \vartheta \geq \vartheta_0$ против конкурирующей гипотезы $H_1: \vartheta < \vartheta_0$, будет также равномерно наиболее мощным.

Пример 3. Пусть выборка X_1, \dots, X_n произведена из нормальной генеральной совокупности с известной дисперсией σ^2 и неизвестным средним ϑ , относительно которого имеются две гипотезы: основная $H_0: \vartheta \leq \vartheta_0$ и конкурирующая $H_1: \vartheta > \vartheta_0$. Определим отношение правдоподобия

$$\Lambda(X_1, \dots, X_n; \vartheta', \vartheta'') = e^{\frac{\vartheta'' - \vartheta'}{\sigma^2} (X_1 + \dots + X_n) + \frac{n[(\vartheta')^2 - (\vartheta'')^2]}{2\sigma^2}}.$$

Вводя статистику

$$S = S(X_1, \dots, X_n) = X_1 + \dots + X_n,$$

видим, что

$$\Lambda(X_1, \dots, X_n; \vartheta', \vartheta'') = e^{\frac{\vartheta'' - \vartheta'}{\sigma^2} S + \frac{n[(\vartheta')^2 - (\vartheta'')^2]}{2\sigma^2}} = \Lambda(S, \vartheta', \vartheta''),$$

причем при $\vartheta' < \vartheta''$ функция $\Lambda(S, \vartheta', \vartheta'')$ является неубывающей по S . Значит, существует равномерно наиболее мощный критерий для проверки гипотез H_0 и H_1 , критическая область которого W_k задается неравенством $x_1 + \dots + x_n > C$.

Для того чтобы определить уровень значимости $\alpha = \alpha(\vartheta)$ и мощность $\beta = \beta(\vartheta)$ этого критерия, заметим, что статистика $S = X_1 + \dots + X_n$ распределена по нормальному закону с параметрами $n\vartheta$ и $n\sigma^2$. Тогда

$$\begin{aligned} \alpha(\vartheta) &= P\{X_1 + \dots + X_n > C \mid \vartheta\} = 1 - \Phi(C; n\vartheta, n\sigma^2) = \\ &= 1 - \Phi\left(\frac{C - n\vartheta}{\sqrt{n\sigma^2}}\right) = \Phi\left(\frac{n\vartheta - C}{\sqrt{n\sigma^2}}\right) \quad (\vartheta \leq \vartheta_0), \end{aligned}$$

$$\begin{aligned} \beta(\vartheta) &= P\{X_1 + \dots + X_n > C \mid \vartheta\} = 1 - \Phi(C; n\vartheta, n\sigma^2) = \\ &= 1 - \Phi\left(\frac{C - n\vartheta}{\sqrt{n\sigma^2}}\right) = \Phi\left(\frac{n\vartheta - C}{\sqrt{n\sigma^2}}\right) \quad (\vartheta > \vartheta_0). \end{aligned}$$

Поскольку $\alpha(\vartheta)$ — возрастающая функция от ϑ , то размер критерия

$$\alpha_0 = \alpha(\vartheta_0) = \Phi\left(\frac{n\vartheta_0 - C}{\sqrt{n\sigma^2}}\right).$$

Теперь, если, наоборот, нужно построить критерий заданного размера α_0 , то критическое значение C определяется выражением

$$C = n\vartheta_0 - \varphi_{\alpha_0} n\sigma^2,$$

где φ_α — α -квантиль стандартного нормального закона. \square

Двусторонняя основная гипотеза. Рассмотрим двустороннюю основную гипотезу $H_0: \vartheta \leq \vartheta_0$ или $\vartheta \geq \vartheta_1$ ($\vartheta_0 < \vartheta_1$). Соответственно конкурирующая гипотеза H_0 имеет вид: $\vartheta_0 < \vartheta < \vartheta_1$. Пусть существуют возрастающая функция $A(\vartheta)$ и функции $B(x)$, $S(x)$, такие,

что плотность распределения $p(x; \vartheta)$ любой функции распределения из семейства $F(x; \vartheta)$ (чтобы не рассматривать рандомизированные критерии, мы ограничимся здесь только непрерывным случаем) представима в виде

$$p(x; \vartheta) = h(x) e^{A(\vartheta) S(x) + B(\vartheta)}$$

(принадлежит экспоненциальному семейству, см. гл. 2, параграф 1). Введем статистику

$$S = S(X_1, \dots, X_n) = S(X_1) + \dots + S(X_n).$$

Определим для любых двух чисел C_1 и C_2 ($C_1 < C_2$) вероятности

$$q_0 = q_0(C_1, C_2) = \mathbf{P}\{C_1 < S(X_1, \dots, X_n) < C_2 \mid \vartheta_0\} =$$

$$= \int_{C_1 < S(x_1, \dots, x_n) < C_2} \dots \int p(x_1; \vartheta_0) \cdots p(x_n; \vartheta_0) dx_1 \cdots dx_n,$$

$$q_1 = q_1(C_1, C_2) = \mathbf{P}\{C_1 < S(X_1, \dots, X_n) < C_2 \mid \vartheta_1\} =$$

$$= \int_{C_1 < S(x_1, \dots, x_n) < C_2} \dots \int p(x_1; \vartheta_1) \cdots p(x_n; \vartheta_1) dx_1 \cdots dx_n.$$

Теорема 4 (двусторонняя основная гипотеза). При сделанных предположениях существует равномерно наиболее мощный критерий разме-ра α_0 для проверки гипотез H_0 и H_1 , критическая область W_k которого задается неравенствами

$$C_1 < S(x_1, \dots, x_n) < C_2,$$

где C_1 и C_2 определяются из уравнения

$$q_0(C_1, C_2) = q_1(C_1, C_2) = \alpha_0. \quad (8)$$

Уровень значимости $\alpha = \alpha(\vartheta)$ и мощность $\beta = \beta(\vartheta)$ задаются формулами:

$$\alpha(\vartheta) = \mathbf{P}\{C_1 < S(X_1, \dots, X_n) < C_2 \mid \vartheta\} =$$

$$= \int_{C_1 < S(x_1, \dots, x_n) < C_2} \dots \int p(x_1; \vartheta) \cdots p(x_n; \vartheta) dx_1 \cdots dx_n \quad (\vartheta \leq \vartheta_0 \text{ или } \vartheta \geq \vartheta_1),$$

$$\beta(\vartheta) = \mathbf{P}\{C_1 < S(X_1, \dots, X_n) < C_2 \mid \vartheta\} =$$

$$= \int_{C_1 < S(x_1, \dots, x_n) < C_2} \dots \int p(x_1; \vartheta) \cdots p(x_n; \vartheta) dx_1 \cdots dx_n \quad (\vartheta_0 < \vartheta < \vartheta_1).$$

Доказательство. Для доказательства теоремы нам понадобится следующая лемма.

Лемма 4.1. Система уравнений (8) при любом α_0 ($0 < \alpha_0 < 1$) имеет решение.

Доказательство леммы 4.1, как и доказательство леммы 2.2, носит аналитический характер, и мы его здесь не приводим (см. [11]).

Представим функцию правдоподобия в виде

$$L(X_1, \dots, X_n; \vartheta) = p(X_1; \vartheta) \cdots p(X_n; \vartheta) = c(\vartheta) e^{A(\vartheta) S} h(X_1, \dots, X_n), \quad (9)$$

где

$$S = S(X_1, \dots, X_n) = S(X_1) + \dots + S(X_n).$$

Рассмотрим теперь следующую байесовскую постановку задачи. Пусть основная «смешанная» гипотеза H'_0 состоит в том, что выборка X_1, \dots, X_n произведена из генеральной совокупности с теоретической плотностью распределения $p(x; \vartheta_0)$, появившейся с вероятностью q , или с теоретической плотностью распределения $p(x; \vartheta_1)$, появившейся с вероятностью $1 - q$, а конкурирующая гипотеза H'_1 — из генеральной совокупности с теоретической плотностью распределения $p(x; \vartheta')$, где ϑ' — произвольное, но фиксированное число, $\vartheta_0 < \vartheta' < \vartheta_1$. Соответственно гипотезы H'_0 и H'_1 имеют априорные вероятности π и $1 - \pi$. Нетрудно видеть, что мы фактически имеем дело с двумя простыми гипотезами, причем по формуле полной вероятности гипотеза H'_0 приписывает выборке X_1, \dots, X_n плотность распределения

$$p_0(x_1, \dots, x_n) = qp(x_1; \vartheta_0) \cdots p(x_n; \vartheta_0) + (1 - q)p(x_1; \vartheta_1) \cdots p(x_n; \vartheta_1),$$

а гипотеза H'_1 — плотность распределения

$$p_1(x_1, \dots, x_n) = p(x_1; \vartheta') \cdots p(x_n; \vartheta').$$

Байесовский риск R (вероятность принятия неправильной гипотезы) имеет вид

$$R = \pi \int_{W'_k} \dots \int p_0(x_1, \dots, x_n) dx_1 \cdots dx_n + (1 - \pi) \int_{W'_\Delta} \dots \int p_1(x_1, \dots, x_n) dx_1 \cdots dx_n,$$

где W'_k и W'_Δ — критическая и допустимая области принятия H'_0 . Можно показать, что по аналогии с критерием отношения правдоподобия оптимальный байесовский критерий предписывает принять H'_0 при

$$\Lambda'(X_1, \dots, X_n) = \frac{p_1(X_1, \dots, X_n)}{p_0(X_1, \dots, X_n)} < \frac{\pi}{(1 - \pi)}$$

и отвергнуть в противном случае. В силу (9) неравенство для попадания выборки в критическую область можно переписать в виде

$$q \frac{c(\vartheta_0)}{c(\vartheta')} e^{[A(\vartheta_0) - A(\vartheta')]S} + (1 - q) \frac{c(\vartheta_0)}{c(\vartheta')} e^{[A(\vartheta_1) - A(\vartheta')]S} < \frac{1 - \pi}{\pi}.$$

Нетрудно видеть, что из-за монотонности $A(\vartheta)$ левая часть этого неравенства — выпуклая вниз функция, неограниченно возрастающая при $S \rightarrow \pm\infty$. Поэтому его можно переписать в виде $c_1 < S < c_2$, где

$$c_1 = c_1(q, \pi), \quad c_2 = c_2(q, \pi),$$

причем при $c_1 < c_2$ существует обратное преобразование

$$q = q(c_1, c_2), \quad \pi = \pi(c_1, c_2). \quad \square$$

В соответствии с леммой 4.1 для любого α_0 существует решение уравнения (8), а значит, найдутся такие $q = q(\alpha_0)$ и $\pi = \pi(\alpha_0)$, при которых оптимальный байесовский критерий для проверки гипотез H'_0 и H'_1 имеет при ϑ_0 и ϑ_1 равные вероятности ошибки (уровни значимости) $\alpha(\vartheta_0) = \alpha(\vartheta_1) = \alpha_0$.

Но этот же критерий можно применить и в небайесовской модели для проверки основной сложной гипотезы $\tilde{H}_0: \vartheta = \vartheta_0$ или $\vartheta = \vartheta_1$ против простой гипотезы $\tilde{H}_1: \vartheta = \vartheta'$. Ясно, что полученный критерий представляет собой именно тот критерий, о котором говорится в утверждении теоремы. Покажем, что он является равномерно наиболее мощным размера α_0 для проверки гипотез \tilde{H}_0 и \tilde{H}_1 . Для этого рассмотрим любой другой критерий размера не больше α_0 и обозначим через $\tilde{\alpha}(\vartheta_0)$, $\tilde{\alpha}(\vartheta_1)$ и $\tilde{\beta}(\vartheta')$ его уровни значимости в точках ϑ_0 и ϑ_1 и мощность соответственно. Тогда

$$\tilde{\alpha}(\vartheta_0) \leq \alpha_0 = \alpha(\vartheta_0), \quad \tilde{\alpha}(\vartheta_1) \leq \alpha_0 = \alpha(\vartheta_1). \quad (10)$$

Снова считая, что оба критерия (основной и только что введенный) являются байесовскими, вычислим для них байесовские риски R и \tilde{R} :

$$\begin{aligned} R = \pi [q \alpha(\vartheta_0) + (1 - q) \alpha(\vartheta_1)] + (1 - \pi) [1 - \beta(\vartheta')] = \\ = \pi \alpha_0 + (1 - \pi) [1 - \beta(\vartheta')], \end{aligned} \quad (11)$$

$$\tilde{R} = \pi [q \tilde{\alpha}(\vartheta_0) + (1 - q) \tilde{\alpha}(\vartheta_1)] + (1 - \pi) [1 - \tilde{\beta}(\vartheta')]. \quad (12)$$

Вспоминая теперь, что при $q = q(\alpha_0)$, $\pi = \pi(\alpha_0)$ первый из этих критериев является оптимальным байесовским, получаем неравенство $R \leq \tilde{R}$, или с учетом (10)–(12) $\beta(\vartheta') \geq \tilde{\beta}(\vartheta')$.

В силу произвольности выбора ϑ' ($\vartheta_0 < \vartheta' < \vartheta_1$) предложенный критерий является равномерно наиболее мощным размера α_0 для проверки сложных гипотез $\tilde{H}_0: \vartheta = \vartheta_0$ или $\vartheta = \vartheta_1$ и $H_1: \vartheta_0 < \vartheta < \vartheta_1$.

Совершенно аналогично показывается, что построенный критерий является при $\vartheta' < \vartheta_0$ ($\vartheta' > \vartheta_1$) наиболее мощным для проверки гипотезы $H''_0: \vartheta = \vartheta'$ против гипотезы $H''_1: \vartheta = \vartheta_0$ или $\vartheta = \vartheta_1$ среди всех критериев, для которых

$$\begin{aligned} \beta''(\vartheta_0) = \alpha(\vartheta_0) \geq \alpha_0, \quad \beta''(\vartheta_1) = \alpha(\vartheta_1) \leq \alpha_0, \\ (\beta''(\vartheta_0) = \alpha(\vartheta_0) \leq \alpha_0, \quad \beta''(\vartheta_1) = \alpha(\vartheta_1) \geq \alpha_0), \end{aligned}$$

откуда, в частности, следует, что $\alpha(\vartheta) \leq \alpha_0$ для любого $\vartheta < \vartheta_0$ ($\vartheta > \vartheta_1$). Значит, размер построенного критерия в точности равен α_0 .

Для окончания доказательства теоремы осталось заметить, что поскольку построенный критерий имеет размер α_0 при проверке изначальных гипотез $H_0: \vartheta \leq \vartheta_0$ или $\vartheta \geq \vartheta_1$ и $H_1: \vartheta_0 < \vartheta < \vartheta_1$ является равномерно наиболее мощным размера α_0 при замене основной гипотезы H_0 на «упрощенную» $\tilde{H}_0: \vartheta = \vartheta_0$ или $\vartheta = \vartheta_1$, то он (см. окончание доказательства теоремы 3) является также равномерно наиболее мощным для проверки гипотез H_0 и H_1 . \square

Замечание 1 к теореме 4. Вообще говоря, уравнение (8) имеет решение только в тех случаях, когда вероятности событий

$$\{S = S(X_1) + \dots + S(X_n) = C_1\}$$

и

$$\{S = S(X_1) + \dots + S(X_n) = C_2\}$$

при истинных значениях ϑ_0 и ϑ_1 параметра ϑ равны нулю. В общем случае (в частности, при дискретных наблюдениях), как обычно, нужно использовать рандомизацию, т. е. выбирать наряду с C_1 и C_2 вероятности p_1 и p_2 принятия гипотезы H_1 при условиях $S = C_1$ и $S = C_2$.

Замечание 2 к теореме 4. Фактически при доказательстве теоремы было показано больше, чем требовалось. А именно, рассматриваемый критерий

минимизирует уровень значимости $\alpha(\vartheta)$ среди всех критериев, удовлетворяющих соотношениям

$$\alpha(\vartheta_0) = \alpha(\vartheta_1) = \alpha_0.$$

Это замечание нам понадобится при рассмотрении двусторонней конкурирующей гипотезы.

Замечание 3 к теореме 4. Можно показать, что (при некотором несущественном условии, которое мы не приводим) функция мощности $\beta(\vartheta)$ (т. е. мощность $\beta(\vartheta)$ и уровень значимости $\alpha(\vartheta)$) построенного критерия будет иметь вид кривой 1 на рис. 4: существует такое значение $\tilde{\vartheta}$, что при $\vartheta < \tilde{\vartheta}$

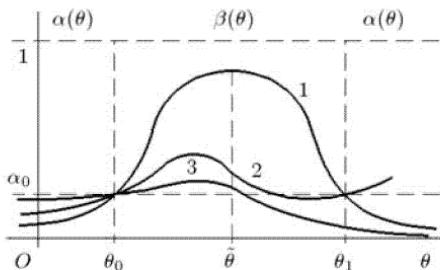


Рис. 4

функция мощности строго возрастает, а при $\vartheta > \tilde{\vartheta}$, наоборот, строго убывает. В силу утверждения теоремы 4 любой другой критерий для проверки H_0 и H_1 , удовлетворяющий условию

$$\alpha'(\vartheta_0) = \alpha'(\vartheta_1) = \alpha_0,$$

будет иметь функцию мощности, задаваемую кривой типа 2, т. е. лежащей не ниже кривой 1 при $\vartheta < \vartheta_0$ и $\vartheta > \vartheta_1$ и не выше кривой 1 при $\vartheta_0 < \vartheta < \vartheta_1$. Наконец, если потребовать, чтобы просто размер критерия равнялся α_0 , то для некоторых ϑ уровень значимости может оказаться меньше $\alpha(\vartheta)$ (кривая 3).

Пример 4. Выборка X_1, \dots, X_n произведена из нормальной генеральной совокупности с известной дисперсией σ^2 и неизвестным средним ϑ , относительно которого имеются две гипотезы: двусторонняя основная гипотеза H_0 : $\vartheta \leq \vartheta_0$ или $\vartheta \geq \vartheta_1$ ($\vartheta_0 < \vartheta_1$) и конкурирующая гипотеза H_1 : $\vartheta_0 < \vartheta < \vartheta_1$. Полагая

$$h(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}, \quad A(\vartheta) = \frac{\vartheta}{\sigma^2}, \quad S(x) = x, \quad B(\vartheta) = -\frac{\vartheta^2}{2\sigma^2},$$

видим, что плотность нормального распределения представима в виде

$$p(x; \vartheta) = \varphi(x; \vartheta, \sigma^2) = h(x) e^{A(\vartheta) S(x) + B(\vartheta)},$$

причем $A(\vartheta)$ — возрастающая функция от ϑ . Поскольку статистика

$$S(X_1, \dots, X_n) = S(X_1) + \dots + S(X_n) = X_1 + \dots + X_n$$

распределена по нормальному закону с параметрами $n\vartheta$ и $n\sigma^2$, то вероятности $q_0(C_1, C_2)$ и $q_1(C_1, C_2)$ определяются выражениями:

$$q_0(C_1, C_2) = P\{C_1 < X_1 + \dots + X_n < C_2 \mid \vartheta_0\} =$$

$$= \Phi(C_2; n\vartheta_0, n\sigma^2) - \Phi(C_1; n\vartheta_0, n\sigma^2) = \Phi\left(\frac{C_2 - n\vartheta_0}{\sqrt{n\sigma^2}}\right) - \Phi\left(\frac{C_1 - n\vartheta_0}{\sqrt{n\sigma^2}}\right),$$

$$q_1(C_1, C_2) = \mathbf{P}\{C_1 < X_1 + \dots + X_n < C_2 \mid \vartheta_1\} =$$

$$= \Phi(C_2; n\vartheta_1, n\sigma^2) - \Phi(C_1; n\vartheta_1, n\sigma^2) = \Phi\left(\frac{C_2 - n\vartheta_1}{\sqrt{n\sigma^2}}\right) - \Phi\left(\frac{C_1 - n\vartheta_1}{\sqrt{n\sigma^2}}\right).$$

Рассмотрим уравнение

$$\Phi\left(\frac{n(\vartheta_1 - \vartheta_0)/2 + C}{\sqrt{n\sigma^2}}\right) - \Phi\left(\frac{n(\vartheta_1 - \vartheta_0)/2 - C}{\sqrt{n\sigma^2}}\right) = \alpha_0$$

относительно неизвестного C . Это уравнение численно можно решить, например, методом последовательных приближений. Полагая теперь

$$C_1 = \frac{n(\vartheta_1 + \vartheta_0)}{2} - C, \quad C_2 = \frac{n(\vartheta_1 + \vartheta_0)}{2} + C$$

и вспоминая тождество $\Phi(x) = 1 - \Phi(-x)$, убеждаемся, что C_1 и C_2 удовлетворяют равенствам

$$q_0(C_1, C_2) = q_1(C_1, C_2) = \alpha_0.$$

Таким образом, равномерно наиболее мощный критерий размера α_0 предписывает нам отвергнуть гипотезу H_0 , если $C_1 < X_1 + \dots + X_n < C_2$, и принять ее в противном случае. Уровень значимости $\alpha(\vartheta)$ и мощность $\beta(\vartheta)$ критерия задаются формулами

$$\alpha(\vartheta) = \mathbf{P}\{C_1 < X_1 + \dots + X_n < C_2 \mid \vartheta\} = \Phi\left(\frac{C_2 - n\vartheta}{\sqrt{n\sigma^2}}\right) - \Phi\left(\frac{C_1 - n\vartheta}{\sqrt{n\sigma^2}}\right) \\ (\vartheta \leq \vartheta_0 \text{ или } \vartheta \geq \vartheta_1),$$

$$\beta(\vartheta) = \mathbf{P}\{C_1 < X_1 + \dots + X_n < C_2 \mid \vartheta\} = \Phi\left(\frac{C_2 - n\vartheta}{\sqrt{n\sigma^2}}\right) - \Phi\left(\frac{C_1 - n\vartheta}{\sqrt{n\sigma^2}}\right) \\ (\vartheta_0 < \vartheta < \vartheta_1). \quad \square$$

Двусторонняя конкурирующая гипотеза. Пусть теперь двусторонней является конкурирующая гипотеза H_1 : $\vartheta < \vartheta_0$ или $\vartheta > \vartheta_1$ ($\vartheta_0 < \vartheta_1$), а основная гипотеза имеет вид H_0 : $\vartheta_0 \leq \vartheta \leq \vartheta_1$. Предположим также, что все допущения относительно семейства функций распределения $F(x; \vartheta)$, принятые при рассмотрении двусторонней основной гипотезы, остаются в силе. Таким образом, мы только поменяли местами основную и конкурирующую гипотезы. Казалось бы, от такой замены ничего не должно измениться и равномерно наиболее мощный критерий для проверки основной двусторонней гипотезы будет таковым и для проверки двусторонней конкурирующей гипотезы. Однако это не так. Более того, для двусторонней конкурирующей гипотезы вообще не существует равномерно наиболее мощного критерия. Причина такого «неравноправия» кроется в том, что в определение равномерно наиболее мощного критерия уровень значимости и мощность входят несимметрично: от уровня значимости требуется только, чтобы он при каждом $\vartheta \in \Theta_0$ не превосходил размера α_0 , в то время как мощность при каждом $\vartheta \in \Theta_1$ должна быть максимальна. Тем не менее, имеет место следующая теорема.

Теорема 5 (двусторонняя конкурирующая гипотеза). Пусть дополнительно к условиям теоремы 4 функция $A(\vartheta)$ является непрерывной.

Тогда для проверки гипотез H_0 и H_1 существует равномерно наибольший мощный несмешанный критерий, который в точности совпадает с равномерно наиболее мощным критерием для проверки двусторонней основной гипотезы, за исключением того, что меняются местами критическая и допустимая области.

Доказательство. Можно показать, что из условия непрерывности $A(\vartheta)$ следует непрерывность функции мощности любого критерия, которая в свою очередь для несмешенного критерия влечет за собой равенство уровня значимости в точках ϑ_0 и ϑ_1 размеру α_0 . Таким образом, равномерно наиболее мощный несмешанный критерий можно искать только среди критериев, для которых

$$\alpha(\vartheta_0) = \alpha(\vartheta_1) = \alpha_0. \quad (13)$$

С другой стороны, как вытекает из замечания 2 к теореме 4, «обращенный» критерий теоремы 5 является наиболее мощным для проверки основной гипотезы H_0 : $\vartheta = \vartheta_0$ или $\vartheta = \vartheta_1$ против конкурирующей гипотезы H_1 : $\vartheta < \vartheta_0$ или $\vartheta > \vartheta_1$ среди всех критериев, удовлетворяющих (13). Но из самого утверждения теоремы 4 следует, что рассматриваемый критерий при каждом ϑ ($\vartheta_0 \leq \vartheta \leq \vartheta_1$) имеет уровень значимости $\alpha(\vartheta) \leq \alpha_0$, т. е. является несмешенным, а значит, он будет также равномерно наиболее мощным и среди всех несмешенных критериев для проверки гипотез H_0 и H_1 , поскольку, как уже показывалось при доказательстве теорем 3 и 4, наличие дополнительных ограничений может привести только к уменьшению мощности критерия. \square

На рис. 5 (кривая 1) приведен типичный график функции мощности равномерно наиболее мощного несмешенного критерия размера α_0 для

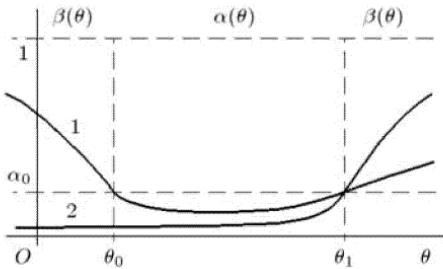


Рис. 5

проверки двусторонней конкурирующей гипотезы. Кривая 2 изображает функцию мощности другого критерия, имеющего тот же размер α_0 , а при $\vartheta > \vartheta_1$ — мощность, большую, чем равномерно наиболее мощный несмешанный критерий. Однако этот другой критерий является смешанным, поскольку при $\vartheta < \vartheta_0$ его мощность меньше размера α_0 .

4. Многопараметрические гипотезы

Пусть выборка X_1, \dots, X_n произведена из генеральной совокупности с теоретической функцией распределения $F(x)$, принадлежащей параметрическому семейству $F(x; \boldsymbol{\theta}) = F(x; \vartheta_1, \dots, \vartheta_k)$, зависящему от

неизвестного векторного параметра $\boldsymbol{\theta} = (\vartheta_1, \dots, \vartheta_k)$. Множество возможных значений параметра $\boldsymbol{\theta}$ будем обозначать через Θ . Не вдаваясь в подробное описание Θ , скажем только, что Θ представляет собой либо все k -мерное пространство R^k , либо достаточно «большую» его часть (например, в случае двух неизвестных параметров ϑ_1 и ϑ_2 множество Θ может быть полуплоскостью, полосой, прямоугольником и т. д.). Предположим теперь, что в множестве Θ выделено некоторое подмножество Θ_0 . Как обычно, дополнение к Θ_0 в Θ будем обозначать через $\Theta_1 = \Theta \setminus \Theta_0$. Нам нужно проверить две сложные параметрические гипотезы: основную $H_0: \boldsymbol{\theta} \in \Theta_0$ против конкурирующей $H_1: \boldsymbol{\theta} \in \Theta_1$.

Основным методом для проверки таких гипотез является *метод отношения правдоподобия*, представляющий естественное обобщение критерия отношения правдоподобия (см. параграф 2). Этот метод заключается в следующем. Рассмотрим функцию правдоподобия

$$L(X_1, \dots, X_n; \vartheta_1, \dots, \vartheta_k) = P(X_1; \vartheta_1, \dots, \vartheta_k) \cdots P(X_n; \vartheta_1, \dots, \vartheta_k)$$

в дискретном случае или

$$L(X_1, \dots, X_n; \vartheta_1, \dots, \vartheta_k) = p(X_1; \vartheta_1, \dots, \vartheta_k) \cdots p(X_n; \vartheta_1, \dots, \vartheta_k)$$

в непрерывном и определим два ее максимальных значения как функции от аргументов $\vartheta_1, \dots, \vartheta_k$:

$$L_0^*(X_1, \dots, X_n) = \sup_{\boldsymbol{\theta} \in \Theta_0} L(X_1, \dots, X_n; \vartheta_1, \dots, \vartheta_k)$$

и

$$L^*(X_1, \dots, X_n) = \sup_{\boldsymbol{\theta} \in \Theta} L(X_1, \dots, X_n; \vartheta_1, \dots, \vartheta_k)$$

(очевидно, что $L_0^*(X_1, \dots, X_n)$ не превосходит $L^*(X_1, \dots, X_n)$).

Так же, как и в методе максимального правдоподобия (см. параграф 4 гл. 2), для нахождения значений $\boldsymbol{\theta}_0^* = (\vartheta_1^{*(0)}, \dots, \vartheta_k^{*(0)})$ и $\boldsymbol{\theta}^* = (\vartheta_1^*, \dots, \vartheta_k^*)$, доставляющих максимум функции правдоподобия $L(X_1, \dots, X_n; \vartheta_1, \dots, \vartheta_k)$ в подмножестве Θ_0 и множестве Θ , обычно используют систему уравнений правдоподобия

$$\frac{\partial}{\partial \vartheta_i} (\ln L(X_1, \dots, X_n; \vartheta_1, \dots, \vartheta_k)) = 0$$

с соблюдением соответствующих ограничений.

Определим теперь отношение правдоподобия

$$\Lambda = \Lambda(X_1, \dots, X_n) = \frac{L^*(X_1, \dots, X_n)}{L_0^*(X_1, \dots, X_n)}.$$

Интуитивно ясно, что если Λ мало отличается от единицы, то это говорит в пользу основной гипотезы H_0 . Полагая $\lambda = 2 \ln \Lambda$, получим следующий критерий для проверки двух сложных параметрических гипотез: мы должны принять гипотезу H_0 , если $\lambda = \lambda(X_1, \dots, X_n) < C$, и отвергнуть в противном случае, где C — критическое значение критерия.

Хотя уровень значимости α и мощность β полученного критерия и определяются, как обычно, формулами

$$\begin{aligned}\alpha &= \alpha(\vartheta_1, \dots, \vartheta_k) = \mathbf{P}\{\lambda(X_1, \dots, X_n) > C \mid \vartheta_1, \dots, \vartheta_k\} \\ &\quad ((\vartheta_1, \dots, \vartheta_k) \in \Theta_0),\end{aligned}$$

$$\begin{aligned}\beta &= \beta(\vartheta_1, \dots, \vartheta_k) = \mathbf{P}\{\lambda(X_1, \dots, X_n) > C \mid \vartheta_1, \dots, \vartheta_k\} \\ &\quad ((\vartheta_1, \dots, \vartheta_k) \in \Theta_1),\end{aligned}$$

их нахождение для конкретных семейств $F(x; \vartheta_1, \dots, \vartheta_k)$ представляет, как правило, сложную в вычислительном плане задачу. Поэтому ограничимся выписыванием приближенного значения уровня значимости α , справедливого при большом объеме выборки n .

Теорема 6 (асимптотическое свойство метода отношения правдоподобия). Пусть Θ_0 есть m -мерное (не обязательно линейное) подпространство R^k ($m < k$). Тогда (при некоторых дополнительных предположениях относительно семейства $F(X; \vartheta_1, \dots, \vartheta_k)$) статистика $\lambda(X_1, \dots, X_k)$ при условии, что истинное значение параметра $\boldsymbol{\theta} = (\vartheta_1, \dots, \vartheta_k) \in \Theta_0$, асимптотически при $n \rightarrow \infty$ имеет χ^2 -распределение с $k - m$ степенями свободы.

Доказательство теоремы 6, которое мы здесь не приводим, в идейном плане состоит из двух частей. Первая часть устанавливает асимптотическую нормальность оценок $\boldsymbol{\theta}^*$ и $\boldsymbol{\theta}_0^*$ и является многомерным аналогом теоремы 7 гл. 2. Вторая часть заключается в подстановке полученных в первой части асимптотически нормальных оценок в квадратичную форму, приближенно описывающую отношение правдоподобия вблизи истинного значения параметра $\boldsymbol{\theta}$. \square

Таким образом, уровень значимости критерия определяется приближенной формулой $\alpha = \mathbf{P}\{\chi^2 > C\}$, где χ^2 — случайная величина, имеющая χ^2 -распределение с $k - m$ степенями свободы (см. [1, табл. 2.1а]), и, что следует отметить особо, уровень значимости при большом объеме выборки практически не зависит от истинного значения параметра $\boldsymbol{\theta}$. Если же, наоборот, задан уровень значимости α , то критическое значение C приближенно совпадает с $h_{1-\alpha} = (1 - \alpha)$ -квантилью χ^2 -распределения с $k - m$ степенями свободы [1, табл. 2.2а].

Пример 5. Выборка X_1, \dots, X_n произведена из нормальной генеральной совокупности с неизвестными средним ϑ_1 и дисперсией ϑ_2 . Требуется построить критерий уровня значимости α для проверки сложной основной гипотезы $H_0: \vartheta_1 = m_0$ против сложной конкурирующей гипотезы $H_1: \vartheta_1 \neq m_0$. В этом примере множество всех возможных значений параметров ϑ_1 и ϑ_2 представляет полуплоскость $\vartheta_2 > 0$, а гипотеза H_0 выделяет в этой полуплоскости полупрямую $\vartheta_1 = m_0$. Функция правдоподобия имеет вид

$$L(X_1, \dots, X_n; \vartheta_1, \vartheta_2) = \left(\frac{1}{\sqrt{2\pi\vartheta_2}} \right)^n e^{-\frac{1}{2\vartheta_2}[(X_1 - \vartheta_1)^2 + \dots + (X_n - \vartheta_1)^2]}.$$

Для определения значения $\boldsymbol{\theta}^* = (\vartheta_1^*, \vartheta_2^*)$, доставляющего максимум функции правдоподобия в множестве Θ , обратимся к примеру 25 из гл. 2. Тогда

$$\vartheta_1^* = m^*, \quad \vartheta_2^* = \sigma^{2*},$$

где m^* и σ^{2*} — выборочные среднее и дисперсия, и, значит, само максимальное значение

$$L^*(X_1, \dots, X_n) =$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^{2*}}} \right)^n e^{-\frac{1}{2\sigma^{2*}}[(X_1 - m^*)^2 + \dots + (X_n - m^*)^2]} = \left(\frac{1}{\sqrt{2\pi\sigma^{2*}}} \right)^n e^{-n/2}.$$

Найдем теперь $\vartheta_0^* = (\vartheta_1^{*(0)}, \vartheta_2^{*(0)})$, максимизирующую функцию правдоподобия в подмножестве Θ_0 . Для этого заметим, что поскольку $\vartheta_1 = m_0$, то система уравнений правдоподобия превращается в одно уравнение

$$\begin{aligned} (\ln L(X_1, \dots, X_n; m_0, \vartheta_2))'_{\vartheta_2} &= \\ &= \left(-\frac{n}{2} \ln(2\pi\vartheta_2) - \frac{1}{2\vartheta_2} [(X_1 - m_0)^2 + \dots + (X_n - m_0)^2] \right)'_{\vartheta_2} = \\ &= -\frac{n}{2\vartheta_2} + \frac{1}{2\vartheta_2^2} [(X_1 - m_0)^2 + \dots + (X_n - m_0)^2] = 0, \end{aligned}$$

решая которое, получаем

$$\vartheta_1^{*(0)} = m_0, \quad \vartheta_2^{*(0)} = \frac{1}{n} [(X_1 - m_0)^2 + \dots + (X_n - m_0)^2] = \bar{\sigma}^{2*}.$$

Таким образом,

$$L_0^*(X_1, \dots, X_n) =$$

$$= \left(\frac{1}{\sqrt{2\pi\bar{\sigma}^{2*}}} \right)^n e^{-\frac{1}{2\bar{\sigma}^{2*}}[(X_1 - m_0)^2 + \dots + (X_n - m_0)^2]} = \left(\frac{1}{\sqrt{2\pi\bar{\sigma}^{2*}}} \right)^n e^{-n/2}.$$

Отношение правдоподобия имеет вид

$$\Lambda = \frac{L^*(X_1, \dots, X_n)}{L_0^*(X_1, \dots, X_n)} = \left(\frac{\bar{\sigma}^{2*}}{\sigma^{2*}} \right)^{n/2},$$

а сам критерий предписывает нам принять гипотезу H_0 , если $\lambda = 2 \ln \Lambda < C$, и отвергнуть ее, если $\lambda > C$. Поскольку множество Θ имеет размерность $k = 2$, а подмножество Θ_0 — размерность $m = 1$, то критическое значение $C \approx h_{1-\alpha}$, где h_α — α -квантиль χ^2 -распределения с одной степенью свободы.

Полученный критерий удобно записать в несколько ином виде. Действительно, производя элементарные преобразования, имеем

$$\bar{\sigma}^{2*} = \sigma^{2*} + (m^* - m_0)^2.$$

Используя теперь монотонность функции $f(x) = \ln x$, видим, что неравенство $2\lambda = n \ln(\bar{\sigma}^{2*}/\sigma^{2*}) < C$ эквивалентно неравенству $n(m^* - m_0)^2 < C_1^2 s^{2*}$, где $C_1^2 = (n-1)(e^{C/n} - 1)$, или, что то же самое, неравенству $\sqrt{n} |m^* - m_0| < C_1 \sqrt{s^{2*}}$.

Иными словами, мы пришли к естественному критерию: принять гипотезу H_0 , если $\frac{|m^* - m_0|}{\sqrt{s^{2*}/n}} < C_1$, и отвергнуть в противном случае.

Поскольку статистика

$$t = \frac{\sqrt{n} (m^* - m_0)}{\sqrt{s^{2*}}}$$

имеет t -распределение (см. гл. 1, параграф 4), то критическое значение C_1 при размере критерия α_0 представляет собой $t_{1-\alpha_0/2}$ — $(1 - \alpha_0/2)$ -квантиль t -распределения с $n - 1$ степенями свободы.

Отметим [12], что построенный критерий является равномерно наиболее мощным несмешанным для проверки гипотез H_0 и H_1 . \square

Пример 6. Предположим, мы произвели опыт, состоящий из n испытаний, а результат каждого испытания характеризуется двумя случайными факторами (показателями), причем первый фактор может принимать значения (уровни действия фактора) $1, 2, \dots, m_1$, а второй — $1, 2, \dots, m_2$. Результаты опыта можно представить в виде табл. 1, где ν_{ij} — число испытаний, в которых первый фактор подействовал на уровне i , а второй — на уровне j . Наша задача — проверить, действуют ли эти факторы независимо (гипотеза H_0) или между ними существует зависимость (гипотеза H_1).

Таблица 1

Уровень первого фактора	Уровень второго фактора				
	1	...	i	...	m_1
1	ν_{11}	...	ν_{i1}	...	ν_{m_11}
...
j	ν_{1j}	...	ν_{ij}	...	ν_{m_1j}
...
m_2	ν_{1m_2}	...	ν_{im_2}	...	$\nu_{m_1m_2}$

Здесь мы имеем дело с так называемой двухфакторной (m_1, m_2) -уровневой моделью. Опишем эту модель. Прежде всего, если не делать предположения о независимости факторов, то имеется $m_1 m_2$ неизвестных параметров ϑ_{ij} — вероятностей того, что первый фактор подействует на уровне i , а второй — на уровне j . Значит, множество Θ представляет собой $(m_1 m_2 - 1)$ -мерное подпространство пространства $R^{m_1 m_2}$, выделяемое соотношениями $\vartheta_{ij} > 0$, $\sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \vartheta_{ij} = 1$ (в силу последнего равенства размерность пространства Θ равна $m_1 m_2 - 1$, а не $m_1 m_2$). В свою очередь, в силу независимости факторов подмножество Θ_0 является $[(m_1 - 1) + (m_2 - 1)]$ -мерным подпространством пространства $R^{m_1 m_2}$, задаваемым ограничениями

$$\vartheta_{ij} = \vartheta_i^{(1)} \vartheta_j^{(2)}, \quad \vartheta_i^{(1)} > 0, \quad \vartheta_j^{(2)} > 0, \quad \sum_{i=1}^{m_1} \vartheta_i^{(1)} = 1, \quad \sum_{j=1}^{m_2} \vartheta_j^{(2)} = 1.$$

Выпишем логарифм функции правдоподобия

$$\ell = \ln L = \ln \left(\prod_{i=1}^{m_1} \prod_{j=1}^{m_2} \vartheta_{ij}^{\nu_{ij}} \right) = \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \nu_{ij} \ln \vartheta_{ij}.$$

Поскольку ϑ_{ij} удовлетворяют уравнению связи $\sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \vartheta_{ij} = 1$, то для вычисления максимального значения L^* определим лагранжиан

$$G = \ell - \gamma \left(\sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \vartheta_{ij} - 1 \right),$$

и в соответствии с общими правилами нахождения максимума будем искать ϑ_{ij}^* из системы уравнений

$$\frac{\partial}{\partial \vartheta_{ij}} G = 0, \quad \frac{\partial}{\partial \gamma} G = 0.$$

Несложные подсчеты показывают, что $\vartheta_{ij}^* = \nu_{ij}/n$. Таким образом,

$$L^* = \prod_{i=1}^{m_1} \prod_{j=1}^{m_2} \left(\frac{\nu_{ij}}{n} \right)^{\nu_{ij}}.$$

Аналогично поступим, если $\boldsymbol{\theta} = (\vartheta_{ij}) \in \Theta_0$. Тогда

$$\begin{aligned} \ell &= \ln L = \ln \left(\prod_{i=1}^{m_1} \prod_{j=1}^{m_2} (\vartheta_i^{(1)} \vartheta_j^{(2)})^{\nu_{ij}} \right) = \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \nu_{ij} (\ln \vartheta_i^{(1)} + \ln \vartheta_j^{(2)}), \\ G &= \ell - \gamma_1 \left(\sum_{i=1}^{m_1} \vartheta_i^{(1)} - 1 \right) - \gamma_2 \left(\sum_{j=1}^{m_2} \vartheta_j^{(2)} - 1 \right) \end{aligned}$$

и, значит,

$$\vartheta_i^{(1)*} = \frac{\nu_i^{(1)}}{n}, \quad \vartheta_j^{(2)*} = \frac{\nu_j^{(2)}}{n}, \quad L_0^* = \prod_{i=1}^{m_1} \prod_{j=1}^{m_2} \left(\frac{\nu_i^{(1)} \nu_j^{(2)}}{n^2} \right)^{\nu_{ij}},$$

где $\nu_i^{(1)} = \sum_{j=1}^{m_2} \nu_{ij}$ — суммарное число наблюдений, в которых первый фактор

подействовал на уровне i , а $\nu_j^{(2)} = \sum_{i=1}^{m_1} \nu_{ij}$ — суммарное число наблюдений, в которых второй фактор подействовал на уровне j .

Выписывая отношение правдоподобия

$$\Lambda = \frac{L^*}{L_0^*} = \prod_{i=1}^{m_1} \prod_{j=1}^{m_2} \left(\frac{n \nu_{ij}}{\nu_i^{(1)} \nu_j^{(2)}} \right)^{\nu_{ij}},$$

получаем окончательно, что мы должны принять гипотезу H_0 , если $\lambda = -2 \ln \Lambda < C$, и отвергнуть в противном случае. Критическое значение C приближенно совпадает с $(1 - \alpha)$ -квантилью $h_{1-\alpha}$ χ^2 -распределения с $m_1 m_2 - m_1 - m_2 + 1 = (m_1 - 1)(m_2 - 1)$ степенями свободы, где α — уровень значимости.

Отметим, что при практической реализации описанного критерия число наблюдений n должно быть существенно больше $m_1 m_2$ и, более того, достаточно большими должны быть все ν_{ij} (обычно рекомендуется не менее 5). \square

5. Критерии согласия

Предположим, что выборка X_1, \dots, X_n произведена из генеральной совокупности с неизвестной теоретической функцией распределения, относительно которой имеются две непараметрические гипотезы: простая основная $H_0: F(x) = F_0(x)$ и сложная конкурирующая $H_1: F(x) \neq F_0(x)$, где $F_0(x)$ — известная функция распределения. Иными словами, мы хотим проверить, *согласуются эмпирические данные с нашим гипотетическим предположением относительно теоретической функции распределения или нет*. Поэтому критерии для проверки гипотез H_0 и H_1 носят название *критерии согласия*. Приведем три наиболее часто употребляемых критерия согласия.

Критерий Колмогорова. Уже говорилось (параграф 3 гл. 1), что в силу теоремы Гливенко–Кантелли эмпирическая функция распределения $F^*(x)$ представляет собой состоятельную оценку теоретической функции распределения $F(x)$. Поэтому можно сравнить эмпирическую функцию распределения $F^*(x)$ с гипотетической $F_0(x)$ и, если мера расхождения между ними мала, то считать справедливой гипотезу H_0 .

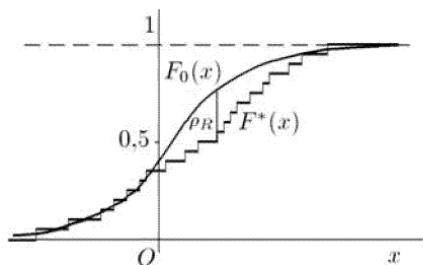


Рис. 6

Наиболее естественной и простой из таких мер (будем предполагать, что $F_0(x)$ — непрерывная функция) является равномерное расстояние

$$\rho_R = \rho_R(F^*(x), F_0(x)) = \sup_{-\infty < x < \infty} |F^*(x) - F_0(x)|$$

(рис. 6 и параграф 5 гл. 2). Однако при построении критерия

Колмогорова более удобно пользоваться нормированным расстоянием $\rho = \sqrt{n} \rho_R$.

Итак, рассмотрим статистику

$$\rho = \rho(X_1, \dots, X_n) = \sqrt{n} \sup_{-\infty < x < \infty} |F^*(x) - F_0(x)|.$$

Критерий Колмогорова предписывает принять гипотезу H_0 , если $\rho < C$, и отвергнуть в противном случае, где C — критическое значение критерия.

Если гипотеза H_0 справедлива, то распределение статистики ρ не зависит от гипотетической функции распределения $F_0(x)$ (доказательство этого факта следует из инвариантности статистики критерия Колмогорова относительно монотонных преобразований, в частности преобразования $g(x) = F_0^{(-1)}(x)$, где $F_0^{(-1)}(x)$ — обратная к $F_0(x)$ функция; преобразование $g(x)$ приводит выборку X_1, \dots, X_n к равномерно распределенной на отрезке $(0,1)$). Поэтому можно рассчитать таблицы,

которые по заданному объему выборки n и критическому значению C позволяют определить уровень значимости критерия α . Поскольку на практике обычно, наоборот, считают известными уровень значимости α и объем выборки n , а затем по ним определяют критическое значение C , то именно такая таблица приведена в [1, табл. 6.2].

При $n \rightarrow \infty$ распределение статистики ρ сходится к распределению Колмогорова [1, табл. 6.1], и критическое значение C при большом объеме выборки практически совпадает с $(1 - \alpha)$ -квантилем $k_{1-\alpha}$ распределения Колмогорова.

При практической реализации критерия Колмогорова сначала по выборке X_1, \dots, X_n составляют вариационный ряд X_1^*, \dots, X_n^* . Затем находят $F_0(X_i^*)$ и определяют значения статистики ρ по формуле

$$\rho = \sqrt{n} \left[\max_{1 \leq i \leq n} \left| F_0(X_i^*) - \frac{2i-1}{2n} \right| + \frac{1}{2n} \right].$$

Наконец, сравнивают полученное значение ρ с критическим значением C для заданного уровня значимости α и принимают или отвергают гипотезу H_0 .

Критерий ω^2 . Пусть $F_1(x)$ — некоторая функция распределения, не совпадающая с $F_0(x)$. Критерий Колмогорова хорошо разделяет выборки (имеет большую мощность) из генеральных совокупностей с теоретическими функциями распределения $F_0(x)$ и $F_1(x)$, если $|F_0(x) - F_1(x)|$ достаточно велико хотя бы на малом интервале изменения x . Встречается и обратная ситуация, когда $|F_0(x) - F_1(x)|$ мало, но постоянно на достаточно большом интервале изменения x . В этом случае для разделения гипотез H_0 и H_1 естественно пользоваться каким-либо интегральным расстоянием, например расстоянием ω^2 (см. параграф 5 гл. 2).

Статистика ω^2 критерия ω^2 задается выражением

$$\omega^2 = \omega^2(X_1, \dots, X_n) = n \int_{-\infty}^{\infty} [F^*(x) - F_0(x)]^2 p_0(x) dx$$

(мы предполагаем, что гипотетическая функция распределения $F_0(x)$ имеет плотность распределения $p_0(x)$), а критическая область W_k состоит из всех тех точек (x_1, \dots, x_n) , для которых $\omega^2 > C$, где C — критическое значение критерия. Используя вариационный ряд X_1^*, \dots, X_n^* , статистику ω^2 можно записать в более удобном для практических расчетов виде

$$\omega^2 = \sum_{i=1}^n \left[F_0(X_i^*) - \frac{2i-1}{2n} \right]^2 + \frac{1}{12n}.$$

Распределение статистики ω^2 при условии справедливости гипотезы H_0 также не зависит от гипотетической функции распределения $F_0(x)$ (это доказывается точно так же, как и инвариантность распре-

деления статистики критерия Колмогорова) и при увеличении объема выборки сходится к ω^2 -распределению. Поэтому уровень значимости критерия α определяется по критическому значению C приближенной формулой

$$\alpha \approx 1 - A(C),$$

где $A(x)$ — функция ω^2 -распределения [1, табл. 6.4а]. Если же задан уровень значимости α критерия, то критическое значение C практически совпадает с $(1 - \alpha)$ -квантилью $a_{1-\alpha}$ ω^2 -распределения.

Практическая реализация критерия ω^2 происходит в той же последовательности, что и критерия Колмогорова: сначала по выборке X_1, \dots, X_n определяется вариационный ряд X_1^*, \dots, X_n^* , затем находятся $F_0(X_i^*)$ и вычисляется значение статистики ω^2 и, наконец, полученное значение ω^2 сравнивается с критическим значением C и либо принимается, либо отвергается гипотеза H_0 .

В литературе иногда критериями ω^2 называют целое семейство критериев, основанных на интегральных расстояниях с различными весовыми функциями.

Критерий χ^2 (Пирсона). Критерий χ^2 является аналогом критерия ω^2 для дискретной наблюдаемой величины X , хотя и применяется как в дискретном, так и в непрерывном случае.

Начнем с дискретного случая. Пусть наблюдаемая случайная величина X может принимать только значения $\{b_1, \dots, b_L\}$ с неизвестными вероятностями $\vartheta_1, \dots, \vartheta_L$. Основная гипотеза H_0 выделяет среди всех распределений случайных величин, принимающих значения $\{b_1, \dots, b_L\}$, одно фиксированное распределение, для которого значения вероятностей ϑ_l известны и равны P_l . Обозначим через ν_l ($l = 1, \dots, L$) число тех элементов выборки X_1, \dots, X_n , которые приняли значение b_l . Поскольку в силу закона больших чисел наблюденная частота $\vartheta_l^* = \nu_l/n$ с ростом объема выборки n стремится к вероятности ϑ_l , мы должны признать гипотезу H_0 справедливой, если все ϑ_l^* мало отличаются от P_l . Введем теперь статистику

$$\chi^2 = \chi^2(X_1, \dots, X_n) = \sum_{l=1}^L \frac{(\nu_l - nP_l)^2}{nP_l}.$$

С одной стороны, эта статистика является мерой равномерной близости всех ϑ_l^* к P_l , с другой стороны, как говорилось в параграфе 4 гл. 1, она асимптотически при $n \rightarrow \infty$ независимо от гипотетических вероятностей P_l имеет χ^2 -распределение с $L - 1$ степенями свободы. Таким образом, критерий χ^2 предписывает принять гипотезу H_0 , если $\chi^2 < C$, и отвергнуть, если $\chi^2 \geqslant C$, где C — критическое значение критерия.

Из сказанного выше следует, что при заданном C уровень значимости α критерия χ^2 определяется приближенной формулой

$$\alpha \approx 1 - H(C),$$

где $H(x)$ — функция χ^2 -распределения с $L - 1$ степенями свободы [1, табл. 2.1а]. Наоборот, если задан уровень значимости α , то критическое значение C примерно совпадает с $(1 - \alpha)$ -квантилью $h_{1-\alpha}$ χ^2 -распределения [1, табл. 2.2а].

При практической реализации критерия χ^2 нужно следить за тем, чтобы объем выборки был велик, иначе неправомочна аппроксимация χ^2 -распределением распределения статистики χ^2 . Обычно считается, что достаточным условием для этого является выполнение неравенств $\nu_l \geq 5$ при всех l ; в противном случае маловероятные значения b_l объединяются в одно или присоединяются к другим значениям, причем объединенному значению приписывается суммарная вероятность (разумеется, уменьшается число степеней свободы при определении уровня значимости или критического значения C).

Следует отметить, что при $n \rightarrow \infty$ критерий χ^2 асимптотически совпадает с параметрическим критерием для проверки основной гипотезы $H_0: \boldsymbol{\theta} = \mathbf{P}$ ($\boldsymbol{\theta} = (\vartheta_1, \dots, \vartheta_L)$, $\mathbf{P} = (P_1, \dots, P_L)$) против сложной конкурирующей гипотезы $H_1: \boldsymbol{\theta} \neq \mathbf{P}$, построенным по методу отношения правдоподобия, хотя эти критерии и основаны на совершенно различных идеях.

В общем случае (не обязательно дискретной наблюдаемой величины X) поступают следующим образом. Сначала всю прямую разбивают на L непересекающихся интервалов $(-\infty, d_1), [d_1, d_2], \dots, [d_{L-1}, \infty)$. Затем определяют гипотетические вероятности $P_l = F_0(d_l) - F_0(d_{l-1})$ попадания в интервал $[d_{l-1}, d_l]$ и числа ν_l элементов выборки, попавших в эти интервалы. Наконец, вычисляют значение статистики

$$\chi^2 = \sum_{l=1}^L \frac{(\nu_l - nP_l)^2}{nP_l}$$

и сравнивают его с критическим значением C . Как и в дискретном случае, маловероятные интервалы объединяют. Разумеется, для того чтобы улучшить качество критерия χ^2 (увеличить его мощность), необходимо уменьшать интервалы разбиения, однако этому препятствует ограничение на числа попавших в каждый интервал наблюдений.

При применении критерия χ^2 удобно пользоваться полигоном частот или гистограммой (см. параграф 3 гл. 1).

Часто требуется проверить не совпадение теоретической функции распределения $F(x)$ с известной функцией распределения $F_0(x)$, а принадлежность $F(x)$ заданному параметрическому семейству $F(x; \boldsymbol{\theta}) = F(x; \vartheta_1, \dots, \vartheta_k)$ функций распределения, зависящему от k -мерного неизвестного параметра $\boldsymbol{\theta} = (\vartheta_1, \dots, \vartheta_k)$, т. е. разделить сложные непараметрические гипотезы $H_0: F(x) \in \{F(x; \boldsymbol{\theta})\}$ и $H_1: F(x) \notin \{F(x; \boldsymbol{\theta})\}$. Для того чтобы воспользоваться вышеописанными критериями, нужно из семейства $F(x; \boldsymbol{\theta})$ выделить ту функцию распределения $F_0(x) = F(x; \boldsymbol{\theta}_0)$, с которой уже и будет производиться сравнение эмпирической функции распределения $F(x)$ выборки X_1, \dots, X_n .

Поэтому сначала, предполагая, что верна основная гипотеза H_0 , находят оценку $\boldsymbol{\theta}^* = (\vartheta_1^*, \dots, \vartheta_k^*)$ неизвестного векторного параметра $\boldsymbol{\theta}$, а затем, полагая $\boldsymbol{\theta}_0 = \boldsymbol{\theta}^*$, с помощью выбранного критерия согласия проверяют простую основную гипотезу $H_0: F(x) = F_0(x) = F(x; \boldsymbol{\theta}^*)$ против сложной конкурирующей гипотезы $H_1: F(x) \neq F_0(x) = F(x; \boldsymbol{\theta}^*)$. Ясно, что в качестве оценки $\boldsymbol{\theta}^*$ лучше всего брать такое значение параметра $\boldsymbol{\theta}$, которое доставляло бы минимальное значение статистике соответствующего критерия (см. параграф 5 гл. 2). Однако эти естественные оценки, как правило, весьма сложны в вычислительном плане, и поэтому обычно пользуются более простыми оценками (полученными методами моментов или максимального правдоподобия).

Скажем еще несколько слов об уровне значимости критериев согласия при проверке сложной гипотезы $H_0: F(x) \in \{F(x; \boldsymbol{\theta})\}$. Вообще говоря, даже асимптотически при $n \rightarrow \infty$ уровень значимости критерия будет зависеть и от семейства $F(x; \boldsymbol{\theta})$, и от выбранной оценки $\boldsymbol{\theta}^*$, и даже от истинного значения неизвестного параметра $\boldsymbol{\theta}$. Обычно на практике для критериев Колмогорова и χ^2 считают уровень значимости таким же, как и в случае простой гипотезы H_0 . При использовании достаточно «хороших» оценок (например, оценки максимального правдоподобия) истинный уровень значимости, как правило, будет даже меньше подсчитанного таким образом. Что касается критерия χ^2 , то для него при определении уровня значимости просто уменьшают число степеней свободы χ^2 -распределения на число неизвестных параметров k . Здесь мы делаем обратную ошибку: объявляем уровень значимости меньшим, чем он есть на самом деле, правда, обычно несущественно.

Отметим, что в последнее время в специальной литературе появились работы, в которых выводятся асимптотические разложения уровней значимости (и даже мощностей при «близких» гипотезах) критериев по степеням $1/\sqrt{n}$, причем для некоторых критериев и типов распределений вычисляются также значения первых коэффициентов этих разложений.

Пример 7. Проверим с помощью критерия Колмогорова гипотезу H_0 о том, что проекция X вектора скорости молекулы водорода на ось координат (см. пример 1 из гл. 1) распределена по нормальному закону. Проверку произведем для уровня значимости $\alpha = 0.05$. Параметры нормального закона не заданы, значит, мы имеем дело со сложной гипотезой H_0 и сначала должны оценить среднее ϑ_1 и дисперсию ϑ_2 . Поскольку мы будем пользоваться критерием Колмогорова, хотелось бы оценки ϑ_1^* и ϑ_2^* неизвестных параметров ϑ_1 и ϑ_2 выбрать таким образом, чтобы они доставляли минимальное значение статистики критерия Колмогорова

$$\rho = \sqrt{n} \left[\max_{1 \leq i \leq n} \left| \Phi(X_i^*; \vartheta_1, \vartheta_2) - \frac{2i-1}{2n} \right| + \frac{1}{2n} \right],$$

где X_1^*, \dots, X_n^* — вариационный ряд выборки X_1, \dots, X_n , приведенный в табл. 3 гл. 1, а $\Phi(x; m, \sigma^2) = \Phi((x-m)/\sigma^2)$ — функция распределения нормального закона с параметрами m, σ^2 . Однако искать минимум ρ как функции от ϑ_1 и ϑ_2 — весьма сложная в вычислительном плане задача, так как $\Phi(x)$ даже не выражается в элементарных функциях. Поэтому в качестве оценок ϑ_1^* и ϑ_2^* используем оценки максимального

правдоподобия $\vartheta_1^* = m^* = 0,082$ и $\vartheta_2^* = \sigma^{*2} = 1,34$ (см. примеры 8 из гл. 1 и 15 из гл. 2). Теперь с помощью критерия Колмогорова будем проверять простую гипотезу $H_0 : F(x) = F_0(x) = \Phi(x; \vartheta_1^*, \vartheta_2^*) = \Phi(x; 0,082, 1,34)$. Вычислив сначала $Y_i^* = (X_i^* - \vartheta_1^*)/\sqrt{\vartheta_2^*}$ и воспользовавшись равенством $\Phi(X_i^*; \vartheta_1^*, \vartheta_2^*) = \Phi((X_i^* - \vartheta_1^*)/\sqrt{\vartheta_2^*})$, последовательно находим затем значения $\Phi(Y_i^*)$ [1, с. 112–117], $F_i^* = (2i-1)/(2n)$ и $S_i = |\Phi(Y_i^*) - F_i^*|$ (табл. 2). Наконец, определяя значение статистики критерия Колмогорова

$$\rho = \sqrt{n} \left[\max_{1 \leq i \leq n} \left| \Phi(X_i^*; \vartheta_1, \vartheta_2) - \frac{2i-1}{2n} \right| + \frac{1}{2n} \right] = 0,07\sqrt{50} = 0,49$$

(максимальное значение S_i равно 0,06) и сравнивая его с 0,95-квантилью распределения Колмогорова $k_{0,95} = 1,36$ ([1], с. 346), видим, что $\rho < k_{0,95}$. Значит, мы должны принять гипотезу H_0 и считать распределение проекции вектора скорости молекулы водорода нормальным. \square

Пример 8. Проверим ту же гипотезу H_0 о нормальности проекции вектора скорости молекулы водорода с тем же уровнем значимости $\alpha = 0,05$, но теперь для проверки H_0 воспользуемся критерием ω^2 . Поскольку и в этом случае мы будем пользоваться оценками максимального правдоподобия неизвестных среднего ϑ_1 и дисперсии ϑ_2 нормального закона, то все этапы вычислений, вплоть до нахождения S_i , для критерииев Колмогорова и ω^2 полностью совпадают. Остается только определить значение статистики ω^2 (см. табл. 2):

$$\omega^2 = \sum_{i=1}^n S_i^2 + \frac{1}{12n} = 0,01^2 + 0,04^2 + \dots + 0,01^2 + \frac{1}{600} = 0,05$$

и сравнить найденное значение $\omega^2 = 0,05$ с 0,95-квантилью ω^2 -распределения $a_{0,95} = 0,46$ [1, с. 348]. Таким образом, критерий ω^2 также подтверждает справедливость гипотезы H_0 . \square

Пример 9. Воспользовавшись выборкой из примера 2 гл. 1, проверим с уровнем значимости $\alpha = 0,1$ гипотезу H_0 о том, что число X регистрируемых ежесекундно счетчиком Гейгера частиц имеет распределение Пуассона. Случайная величина X , распределенная по закону Пуассона, принимает значения $l = 0, 1, \dots$ с вероятностями

$$P_l = P(l; \vartheta) = \mathbb{P}\{X = l \mid \vartheta\} = \frac{\vartheta^l}{l!} e^{-\vartheta},$$

где ϑ — математическое ожидание X (см. часть 1, гл. 5, параграф 3). Ввиду дискретности случайной величины X для проверки гипотезы H_0 следует использовать критерий χ^2 . Поскольку значение параметра ϑ неизвестно, мы имеем дело со сложной гипотезой H_0 и должны сначала оценить ϑ . Воспользовавшись методом моментов, получаем оценку

$$\vartheta^* = m^* = \frac{1}{n} (X_1 + \dots + X_n) = \frac{1}{60} (1 + 4 + \dots + 5) = 2,92$$

(читателю предоставляется возможность убедиться, что оценка максимального правдоподобия также будет совпадать с выборочным средним). В табл. 3 выписаны гипотетические вероятности

$$P_l = P(l; \vartheta) = \frac{(2,92)^l}{l!} e^{-2,92} \quad (l = 0, 1, \dots),$$

а также числа ν_l элементов выборки, принявших значение l , взятые из статистического ряда выборки (табл. 6 гл. 1). Из табл. 3 видно, что числа элементов выборки, принявших значения 0, 6, 7, 8 и т. д., меньше 5. Поэтому объединим

Таблица 2

X_i^*	-2,33	-1,60	-1,58	-1,57	-1,46	-1,32	-1,29	-1,06	-1,04	-0,98
Y_i^*	-2,08	-1,45	-1,43	-1,43	-1,33	-1,21	-1,18	-0,98	-0,97	-0,92
$\Phi(Y_i^*)$	0,02	0,07	0,08	0,08	0,09	0,11	0,12	0,16	0,17	0,18
F_i^*	0,01	0,03	0,05	0,07	0,09	0,11	0,13	0,15	0,17	0,19
S_i	0,01	0,04	0,03	0,01	0,00	0,00	0,01	0,01	0,00	0,01
X_i^*	-0,90	-0,88	-0,80	-0,79	-0,72	-0,53	-0,52	-0,28	-0,27	-0,18
Y_i^*	-0,85	-0,83	-0,76	-0,75	-0,69	-0,53	-0,52	-0,31	-0,30	-0,22
$\Phi(Y_i^*)$	0,20	0,20	0,22	0,23	0,25	0,30	0,30	0,38	0,38	0,41
F_i^*	0,21	0,23	0,25	0,27	0,29	0,31	0,33	0,35	0,37	0,39
S_i	0,01	0,03	0,03	0,04	0,04	0,01	0,03	0,05	0,01	0,02
X_i^*	-0,10	-0,08	-0,05	-0,01	0,01	0,01	0,01	0,14	0,30	0,32
Y_i^*	-0,16	-0,14	-0,11	-0,08	-0,06	-0,06	-0,06	0,05	0,19	0,21
$\Phi(Y_i^*)$	0,44	0,44	0,46	0,47	0,48	0,48	0,48	0,52	0,58	0,58
F_i^*	0,41	0,43	0,45	0,47	0,49	0,51	0,53	0,55	0,57	0,59
S_i	0,03	0,01	0,01	0,00	0,01	0,03	0,05	0,03	0,01	0,01
X_i^*	0,35	0,35	0,40	0,41	0,52	0,54	0,60	0,65	0,74	0,96
Y_i^*	0,23	0,23	0,28	0,29	0,38	0,40	0,45	0,49	0,57	0,76
$\Phi(Y_i^*)$	0,59	0,59	0,61	0,61	0,65	0,66	0,67	0,69	0,72	0,78
F_i^*	0,61	0,63	0,65	0,67	0,69	0,71	0,73	0,75	0,77	0,79
S_i	0,02	0,04	0,04	0,06	0,04	0,05	0,06	0,06	0,05	0,01
X_i^*	1,02	1,06	1,12	1,27	1,68	1,92	2,15	2,19	2,25	3,47
Y_i^*	0,81	0,85	0,90	1,03	1,38	1,59	1,79	1,82	1,87	2,93
$\Phi(Y_i^*)$	0,79	0,80	0,82	0,85	0,92	0,94	0,96	0,97	0,97	1,00
F_i^*	0,81	0,83	0,85	0,87	0,89	0,91	0,93	0,95	0,97	0,99
S_i	0,02	0,03	0,03	0,02	0,03	0,03	0,03	0,02	0,00	0,01

нулевой и первый столбцы, а шестой и последующие столбцы присоединим к пятому (табл. 4). Производя последовательно вычисления, представленные в табл. 4, определяем значение статистики

$$\chi^2 = \sum_{l=1}^L \frac{(\nu_l - nP_l)^2}{nP_l} = 0,88 + 1,67 + 0,44 + 106 + 0,04 = 4,09.$$

Так как число столбцов $L = 5$, а число неизвестных параметров $k = 1$, то χ^2 -распределение, используемое для приближенного нахождения критического значения C , имеет $L - k - 1 = 3$ степени свободы. В [1, с. 167] находим 0,9-квантиль χ^2 -распределения с тремя степенями свободы $h_{0,9} = 6,251$. Сравнивая значение $\chi^2 = 4,09$ с $h_{0,9} = 6,251$, констатируем, что следует признать справедливость гипотезы H_0 о пуассоновости распределения регистрируемых счетчиком Гейгера частиц. \square

Таблица 3

l	0	1	2	3	4	5	6	7	8	9	10	11	...
ν_l	4	12	9	11	13	9	1	0	0	1	0	0	...
P_l	0,054	0,157	0,230	0,224	0,163	0,095	0,046	0,019	0,007	0,0023	0,0007	0,0002	...

Таблица 4

l	≤ 1	2	3	4	≥ 5
P_l	0,211	0,230	0,224	0,163	0,172
ν_l	16	9	11	13	11
nP_l	12,66	13,8	13,44	9,78	10,32
$(\nu_l - nP_l)^2$	11,16	23,04	5,95	10,37	0,46
$(\nu_l - nP_l)^2 / (nP_l)$	0,88	1,67	0,44	1,06	0,04

Пример 10. Еще раз обратимся к проверке гипотезы H_0 о нормальности проекции X вектора скорости молекулы водорода (см. примеры 7 и 8). Воспользуемся критерием χ^2 . Для этого разобьем всю прямую на 8 интервалов: $(-\infty, -3)$, $(-3, -2)$, $(-2, -1)$, $(-1, 0)$, $(0, 1)$, $(1, 2)$, $(2, 3)$ и $(3, \infty)$. Однако поскольку в первые два интервала попало всего одно наблюдение (см. табл. 7 гл. 1), объединим их с третьим интервалом. Аналогично седьмой и восьмой интервалы присоединим к шестому. Окончательно получим 4 интервала с числами ν_l попавших в них наблюдений, приведенными в табл. 5. Поскольку оценки $\vartheta_1^* = 0,082$ и $\vartheta_2^* = 1,34$ неизвестных среднего ϑ_1 и дисперсии ϑ_2 нами уже получены, найдем, воспользовавшись [1, с. 112–113], гипотетические вероятности P_l попадания наблюданной величины X в рассматриваемые интервалы:

$$P_l = \Phi(-1; \vartheta_1^*, \vartheta_2^*) = \Phi\left(\frac{-1 - \vartheta_1^*}{\sqrt{\vartheta_2^*}}\right) = \Phi\left(\frac{-1 - 0,082}{\sqrt{1,34}}\right) = \Phi(-0,935) = 0,175,$$

$$\begin{aligned} P_2 &= \Phi(0; \vartheta_1^*, \vartheta_2^*) - \Phi(-1; \vartheta_1^*, \vartheta_2^*) = \\ &= \Phi(-0,071) - \Phi(-0,935) = 0,472 - 0,175 = 0,297, \\ P_3 &= \Phi(1; \vartheta_1^*, \vartheta_2^*) - \Phi(0; \vartheta_1^*, \vartheta_2^*) = \\ &= \Phi(0,793) - \Phi(-0,071) = 0,786 - 0,472 = 0,314, \\ P_4 &= 1 - \Phi(1; \vartheta_1^*, \vartheta_2^*) = 1 - \Phi(0,793) = 1 - 0,786 = 0,214. \end{aligned}$$

Таблица 5

Интервал	$(-\infty, -1)$	$(-1, 0)$	$(0, 1)$	$(1, \infty)$
P_l	0,175	0,297	0,314	0,214
ν_l	9	15	16	10
nP_l	8,7	14,8	15,7	10,7
$(\nu_l - nP_l)^2$	0,09	0,04	0,09	0,49
$(\nu_l - nP_l)^2 / (nP_l)$	0,001	0,003	0,006	0,046

Теперь определим значение статистики

$$\chi^2 = \sum_{l=1}^4 \frac{(\nu_l - nP_l)^2}{nP_l} = 0,001 + 0,003 + 0,006 + 0,046 = 0,056.$$

Число степеней свободы χ^2 -распределения равно единице (число интервалов наблюдения $L = 4$, число неизвестных параметров $k = 2$). Сравнивая полученное значение $\chi^2 = 0,056$ с 0,95-квантилем χ^2 -распределения с одной степенью свободы $h_{0,95} = 3,841$ [1, с. 167], видим, что и критерий χ^2 подтверждает гипотезу H_0 . \square

6. Критерии однородности двух выборок

В этом параграфе мы обратимся к постановке задачи, несколько отличной от изучавшихся ранее. А именно, будем рассматривать две выборки и проверять гипотезу о том, что эти *выборки извлечены из одной и той же генеральной совокупности*. Итак, пусть мы имеем независимые выборки: X_1, \dots, X_n , произведенную из генеральной совокупности с неизвестной теоретической функцией распределения $F_1(x)$, и Y_1, \dots, Y_m , произведенную из генеральной совокупности с неизвестной теоретической функцией распределения $F_2(x)$. Проверяются две сложные непараметрические гипотезы: основная $H_0: F_1(x) = F_2(x)$ и конкурирующая $H_1: F_1(x) \neq F_2(x)$. Будем предполагать, что функции $F_1(x)$ и $F_2(x)$ непрерывны. Поскольку справедливость гипотезы H_0 , по сути дела, означает, что выборки X_1, \dots, X_n и Y_1, \dots, Y_m произведены из одной и той же генеральной совокупности, критерии для проверки гипотез H_0 и H_1 называются *критериями однородности двух выборок*. Приведем два таких критерия.

Критерий Смирнова. Критерий Смирнова использует ту же идею, что и критерий Колмогорова, но только если в критерии Колмогорова эмпирическая функция распределения сравнивается с гипотетической, то в критерии Смирнова между собой сравниваются две эмпирические функции распределения. Статистика критерия Смирнова задается выражением

$$\rho = \sqrt{\frac{nm}{n+m}} \sup_{-\infty < x < \infty} |F_1^*(x) - F_2^*(x)|,$$

где $F_1^*(x)$ и $F_2^*(x)$ — эмпирические функции распределения, построенные по выборкам X_1, \dots, X_n и Y_1, \dots, Y_m соответственно. Критерий Смирнова предписывает принять гипотезу H_0 , если $\rho < C$, и отвергнуть в противном случае.

При условии справедливости гипотезы H_0 распределение статистики ρ (а значит, и уровень значимости α) не зависит от распределения $F_1(x) = F_2(x)$ (доказательство этого факта слово в слово повторяет доказательство инвариантности статистики критерия Колмогорова). При малых объемах выборок ($n, m \leq 20$) критические значения C для заданных уровней значимости (размеров) критерия α приведены в табл. 6.5а [1]. При $n, m \rightarrow \infty$ распределение статистики ρ сходится к распределению Колмогорова $K(x)$, что позволяет приближенно вычислять уровни значимости критерия Смирнова по формуле $\alpha = 1 - K(C)$ (распределение Колмогорова $K(x)$ приведено в табл. 6.1 [1]) и, наоборот, определять критическое значение C при заданном уровне значимости α как $(1 - \alpha)$ -квантиль распределения Колмогорова $K(x)$.

Пример 11. На двух реакторах были проведены сходные эксперименты, в результате которых возникли новые частицы. Для анализа экспериментальных данных были замерены энергии $n = 631$ частицы, полученной на первом реакторе, и $m = 839$ частиц, полученных на втором реакторе, и построены эмпирические функции распределения энергии частиц $F_1^*(x)$ и $F_2^*(x)$. Когда $F_1^*(x)$ и $F_2^*(x)$ сравнили, оказалось, что $\sup_{-\infty < x < \infty} |F_1^*(x) - F_2^*(x)| = 0,042$.

Проверим с помощью критерия Смирнова уровня значимости (размера) $\alpha = 0,2$ гипотезу H_0 о том, что на обоих реакторах возникали одни и те же частицы. Для этого вычислим значение статистики критерия

$$\rho = \sqrt{\frac{nm}{n+m}} \sup_{-\infty < x < \infty} |F_1^*(x) - F_2^*(x)| = \sqrt{\frac{631 \cdot 839}{631 + 839}} 0,042 = 0,80$$

и сравним полученное значение ρ с 0,8-квантилем распределения Колмогорова $k_{0,8} = 1,07$ [1, с. 346]. Поскольку $\rho < k_{0,8}$, то у нас есть основания считать гипотезу H_0 справедливой. \square

Критерий Вилкоксона. Образуем из выборок X_1, \dots, X_n и Y_1, \dots, Y_m один общий вариационный ряд (табл. 6) и отметим последовательные порядковые номера (ранги) r_1, r_2, \dots, r_m элементов выборки Y_1, \dots, Y_m в общем вариационном ряду (в табл. 6 рангами r_1, r_2, \dots, r_m будут $2, 3, \dots, n + m - 1$).

Таблица 6

X_1^*	Y_1^*	Y_2^*	X_2^*	X_3^*	X_4^*	...	Y_m^*	X_n^*
1	2	3	4	5	6	...	$n + m - 1$	$n + m$
s_1	s_2	s_3	s_4	s_5	s_6	...	s_{n+m-1}	s_{n+m}

Критерии, позволяющие только на основе рангов r_1, \dots, r_m принимать или отвергать гипотезу H_0 , называются *ранговыми критериями*. Их достоинством является чрезвычайная простота.

Поскольку при условии справедливости гипотезы H_0 все возможные комбинации рангов r_1, \dots, r_m равновероятны (всего таких комбинаций $\binom{n+m}{n}$), то уровень значимости (размер) рангового критерия не зависит от распределения $F_1(x) = F_2(x)$.

Обычно в качестве статистики рангового критерия используют сумму $f(r_1) + \dots + f(r_m)$, где $f(r)$ — некоторая функция, определенная для всех $r = 1, 2, \dots, n+m$. Мы рассмотрим один тип ранговых критериев — критерий Вилкоксона.

Пусть $(s_1, s_2, \dots, s_{n+m})$ — одна из $(n+m)!$ возможных перестановок чисел $1, 2, \dots, n+m$ (т. е. расположенные в произвольном порядке числа $1, 2, \dots, n+m$). Положим $f(r) = s_r$ (см. табл. 6). Статистика критерия Вилкоксона задается формулой

$$w = s_{r_1} + \dots + s_{r_m}.$$

Односторонний критерий Вилкоксона предписывает принять гипотезу H_0 , если $w > C$, и отвергнуть, если $w \leqslant C$, где C — критическое значение одностороннего критерия Вилкоксона.

При использовании двустороннего критерия Вилкоксона мы должны принять гипотезу H_0 , если $C_1 < w < C_2$ ($C_1 < C_2$), и отвергнуть ее, если либо $w \geqslant C_2$, либо $w \leqslant C_1$. Нижнее C_1 и верхнее C_2 критические значения двустороннего критерия Вилкоксона связаны между собой соотношением $C_1 + C_2 = N$, где $N = m(m+n+1)$.

Выбор перестановки $(s_1, s_2, \dots, s_{n+m})$ осуществляется до опыта таким образом, чтобы по возможности наилучшим образом разделить выборки X_1, \dots, X_n и Y_1, \dots, Y_m при наименее благоприятном соотношении между теоретическими функциями $F_1(x)$ и $F_2(x)$ или, иными словами, чтобы при заданном соотношении между $F_1(x)$ и $F_2(x)$ мощность критерия была бы максимальна. Так, если к наиболее опасным последствиям ведет отождествление наблюдаемых величин X и Y в случае, когда Y систематически меньше X (т. е. $F_1(x) < F_2(x)$ при всех x), то естественно положить $s_r = r$ и воспользоваться односторонним критерием Вилкоксона. Если же одинаково пагубными представляются и случай X систематически меньше Y , и случай Y систематически меньше X (т. е. одновременно для всех x либо $F_1(x) < F_2(x)$, либо $F_2(x) < F_1(x)$), то опять-таки нужно взять $s_r = r$, но использовать двусторонний критерий Вилкоксона. Или еще пример: из каких-то соображений стало известно, что наблюдаемые величины X и Y в среднем приблизительно одинаковы ($\mathbf{M}X = \mathbf{M}Y$), и нужно проверить основную гипотезу H_0 : разброс случайных величин X и Y одинаков против конкурирующей гипотезы H_1 : разброс Y больше разброса X . При выполнении гипотезы H_1 наблюдаемые значения величины Y (выборка Y_1, \dots, Y_m) будут в основном сосредоточиваться в начале и в конце

общего вариационного ряда и весьма разумным представляется выбор перестановки $s_1 = 1, s_{n+m} = 2, s_{n+m-1} = 3, s_2 = 4, s_3 = 5, s_{n+m-2} = 6$ и т. д.

Если верна основная гипотеза $H_0: F_1(x) = F_2(x)$, то распределение статистики критерия Вилкоксона зависит лишь от объемов выборок n и m и не зависит от конкретно используемой перестановки $(s_1, s_2, \dots, s_{n+m})$. Поэтому в соответствии с принципом классической вероятности уровень значимости (размер) α одностороннего критерия Вилкоксона для критического значения C определяется как число тех сочетаний (r_1, \dots, r_m) из $n + m$ элементов по m , для которых $r_1 + \dots + r_m \leq C$, отнесенного к общему числу сочетаний $\binom{n+m}{n}$. Поскольку обычно, наоборот, по уровню значимости α определяют критическое значение C , то именно такая таблица приведена в [1] (табл. 6.8).

Уровень значимости двустороннего критерия Вилкоксона находится как удвоенный уровень значимости одностороннего критерия с $C = C_1$. Для вычисления критических значений C_1 и C_2 двустороннего критерия с уровнем значимости α мы должны по табл. 6.8 определить критическое значение C одностороннего критерия с уровнем значимости $\alpha/2$, а затем положить $C_1 = C$ и $C_2 = N - C_1$ ($N = m(n + m + 1)$).

Наконец, если объем хотя бы одной из выборок X_1, \dots, X_n или Y_1, \dots, Y_m велик, можно воспользоваться асимптотической нормальностью статистики Вилкоксона w со средним $N/2$ и дисперсией $mn(m+n+1)/12$; в этом случае при заданном уровне значимости α следует положить для одностороннего критерия

$$C_1 = \frac{N}{2} + \frac{mn(m+n+1)}{12} \varphi_\alpha,$$

а для двустороннего —

$$C_{1,2} = \frac{N}{2} \pm \frac{mn(m+n+1)}{12} \varphi_{\frac{\alpha}{2}},$$

где φ_α — α -квантиль стандартного нормального закона ([1], табл. 1.3).

Пример 12. Для сравнительного анализа надежности крепежных болтов, выпускаемых двумя заводами, были проверены на разрыв $n = 24$ изделия первого завода и $m = 20$ изделий второго. Силы натяжения ($\times 10^5$ Н), при которых произошли разрывы изделий первого и второго заводов, приведены в табл. 7 и 8.

Таблица 7

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}
2,49	2,81	3,09	2,50	2,20	2,90	2,89	2,17	2,63	1,72	2,45	3,31
X_{13}	X_{14}	X_{15}	X_{16}	X_{17}	X_{18}	X_{19}	X_{20}	X_{21}	X_{22}	X_{23}	X_{24}
1,01	2,60	4,60	1,18	3,12	3,58	4,48	1,26	1,17	1,71	3,75	3,23

Таблица 8

Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9	Y_{10}
3,37	3,08	4,79	3,71	3,06	3,86	3,77	2,80	3,70	4,83
Y_{11}	Y_{12}	Y_{13}	Y_{14}	Y_{15}	Y_{16}	Y_{17}	Y_{18}	Y_{19}	Y_{20}
2,12	3,57	1,60	5,48	5,76	4,43	4,59	4,11	2,36	2,33

Проверим с помощью критерия Вилкоксона уровня значимости (размера) $\alpha = 0,01$ гипотезу H_0 о том, что надежность изделий обоих заводов одинакова. Для того чтобы воспользоваться критерием Вилкоксона, нужно сначала задать перестановку (s_1, \dots, s_{n+m}) . Анализируя условия задачи, видим, что наименее благоприятным будет случай, когда надежность болтов, выпускаемых одним заводом, систематически меньше надежности аналогичных изделий другого завода, и, значит, в качестве перестановки (s_1, \dots, s_{44}) естественно выбрать перестановку $(1, 2, \dots, 44)$, причем из-за отсутствия априорных предпосылок предпочтение изделия какого-либо завода мы должны воспользоваться двусторонним критерием Вилкоксона. Образуем теперь общий вариационный ряд выборок X_1, \dots, X_{24} и Y_1, \dots, Y_{20} (табл. 9) и определим значение статистики w критерия Вилкоксона

$$w = 5 + 8 + 11 + \dots + 43 + 44 = 571.$$

Так как мы используем двусторонний критерий Вилкоксона, то нижнее критическое значение C_1 при уровне значимости $\alpha = 0,01$ совпадает с критическим значением $C = 341$ одностороннего критерия Вилкоксона, имеющего уровень

Таблица 9

X_{13}	X_{21}	X_{16}	X_{20}	Y_{13}	X_{22}	X_{10}	Y_{11}	X_8	X_5	Y_{20}
1,01	1,17	1,18	1,26	1,60	1,71	1,72	2,12	2,17	2,20	2,33
1	2	3	4	5	6	7	8	9	10	11
Y_{19}	X_{11}	X_1	X_4	X_{14}	X_9	Y_8	X_2	X_7	X_6	X_5
2,36	2,45	2,49	2,50	2,60	2,63	2,80	2,81	2,89	2,90	3,06
12	13	14	15	16	17	18	19	20	21	22
X_2	X_3	Y_{17}	X_{24}	X_{12}	Y_1	X_{12}	X_{18}	Y_4	X_{23}	Y_9
3,08	3,09	3,12	3,23	3,31	3,37	3,57	3,58	3,71	3,75	3,76
23	24	25	26	27	28	29	30	31	32	33
Y_7	Y_6	Y_{18}	Y_{16}	X_{19}	Y_{17}	X_{15}	Y_3	Y_{10}	Y_{14}	Y_{15}
3,77	3,86	4,11	4,43	4,48	4,59	4,60	4,79	4,83	5,48	5,76
34	35	36	37	38	39	40	41	42	43	44

значимости 0,005 [1, с. 360], а верхнее критическое значение C_2 определяется формулой

$$C_2 = N - C_1 = 20(24 + 20 + 1) - 341 = 900 - 341 = 559.$$

Сравнивая значение статистики $w = 571$ с критическими значениями C_1 и C_2 , видим, что $w > C_2$. Таким образом, гипотезу H_0 об одинаковой надежности крепежных болтов, выпускаемых обоими заводами, нужно признать не соответствующей результатам проверки, а для практических потребностей рекомендовать изделия второго завода как более надежные. \square

Г л а в а 4

НЕКОТОРЫЕ ЗАДАЧИ, СВЯЗАННЫЕ С НОРМАЛЬНЫМИ ВЫБОРКАМИ

Мы уже не раз указывали на ту важную роль, которую играет в теории вероятностей и ее приложениях нормальный закон. Естественным отражением этой роли в математической статистике является наличие *специальных методов, ориентированных на нормально распределенные выборки*. Некоторые из этих методов и будут рассмотрены в настоящей главе. Однако сразу же условимся, что сюда не будут включены такие задачи, как оценка неизвестных параметров нормального закона и построение для них доверительных интервалов, проверка гипотез о параметрах нормального закона и т. п., поскольку они решаются общими методами, изложенными в предыдущих главах.

1. Общая характеристика задач

Прежде чем приступить к описанию задач, рассматриваемых в данной главе, заметим, что хотя всюду в этой главе и будет предполагаться нормальность всех выборок, предлагаемые методы с разной степенью обоснованности можно применить и к выборкам, отличающимся от нормальных. В частности, полученные по методу наименьших квадратов оценки параметров линейной и регрессионной моделей, для нормального случая являющиеся эффективными, в случае произвольно распределенной выборки также будут иметь минимальные дисперсии, но в классе линейных (т. е. выражаемых линейно через элементы выборки) оценок. Однако применение t - и χ^2 -распределений для построения доверительных интервалов и проверки гипотез в тех же моделях ведет к существенным ошибкам в выводах, причем при отклонениях теоретического распределения от нормального использование t -распределения, как правило, приводит к меньшим ошибкам, чем использование χ^2 -распределения. Здесь мы сталкиваемся с важной проблемой чувствительности (или, как сейчас модно говорить, робастости) статистических процедур к воздействиям определенного типа. Интерес к этой тематике усиливает еще и то, что в реальной жизни вряд ли можно встретить «100 %-чистое» нормальное распределение. Саму проблему робастости мы здесь не рассматриваем, отсылая заинтересованного читателя к специальной литературе.

Материалы настоящей главы можно условно разбить на две части.

В первой части (параграфы 2–4) рассматриваются некоторые специальные критерии согласия (параграф 2), критерии равенства дисперсий нескольких выборок (параграф 3) и задачи, связанные с выборочным

коэффициентом корреляции (параграф 4). Изложенные здесь результаты можно применять как вспомогательные при анализе моделей второй части, но они представляют и самостоятельный интерес.

Вторая часть главы (параграфы 5–8) предназначена для первоначального знакомства читателя с общей линейной моделью (иногда говорят «гипотезой», однако мы предпочитаем употреблять слово «гипотеза» только в первоначальном смысле), регрессионной моделью и задачами дисперсионного анализа. По сути дела, эти три раздела столь близки между собой, что задачи, формулируемые в одном из них, легко переформулировать в терминах другого; кроме того, объединяющим их стержнем является метод наименьших квадратов. Собственно говоря, параграфы 5–7 дают тройной пересказ практически одинакового материала. Однако мы сознательно позволяем себе такую роскошь, поскольку, во-первых, читатель должен научиться ориентироваться в проблемах математической статистики и ему нужно дать хотя бы минимальные сведения о всех наиболее важных возможных подходах, а во-вторых, каждый подход обладает своей «изюминкой». В частности, общая линейная модель позволяет геометрически наглядно выяснить причину появления метода наименьших квадратов, регрессионная модель обладает простым физическим смыслом, а дисперсионный анализ проясняет ту роль, которую играет правильная группировка составляющих выборочной дисперсии при проверке гипотез об отсутствии действия различных факторов. Наконец, последний параграф главы дает некоторое представление о планировании эксперимента в дисперсионном анализе, т. е. о методах, позволяющих при минимальном числе наблюдений получить по возможности наиболее обоснованный ответ на вопрос о существовании действия факторов на рассматриваемое явление. В этой связи отметим важную роль, которую играет в планировании эксперимента теория Галуа. Однако поскольку не предполагается знакомство читателя с этой теорией, мы ограничились здесь общей постановкой задачи и выписыванием простейших планов.

Последнее замечание касается принятых в этой главе обозначений. Мы вернемся к использованным нами в курсе теории вероятностей обозначениям t и σ^2 параметров нормального закона, поскольку из постановки задачи всегда будет видно, значение какого из этих параметров известно, а какого — нет.

2. Критерии согласия

Рассматриваемые в этом параграфе *критерии согласия* проверяют сложную основную гипотезу H_0 : теоретическая функция распределения $F(x)$ является нормальной (т. е. выборка X_1, \dots, X_n произведена из нормальной генеральной совокупности с неизвестными параметрами t и σ^2) против сложной альтернативной гипотезы H_1 : теоретическая функция распределения не является нормальной. Разумеется, приведенные в гл. 3 критерии согласия (Колмогорова, ω^2 и χ^2) применимы

и в этом случае, однако на практике проверку на нормальность часто начинают с использования более простых, хотя, вообще говоря, менее мощных критериев, основанных (кроме критерия исключения резко выделяющихся наблюдений) на сравнении эмпирических и теоретических моментов выборки. Естественно, поскольку мы не интересуемся конкретными значениями среднего m и дисперсии σ^2 , а они являются определяющими параметрами нормального закона, во всех предложенных ниже критериях будут использоваться их оценки: выборочное среднее $m^* = \sum_{i=1}^n \frac{X_i}{n}$ и выборочная дисперсия $s^{2*} = \sum_{i=1}^n \frac{(X_i - m^*)^2}{n-1}$.

Выборочный момент первого порядка (выборочное среднее) $m^* = m_1^*$ для построения критерия согласия мы использовать не можем, поскольку он уже применен в качестве оценки среднего нормального закона (по аналогичной причине не используется выборочный второй момент), однако мы можем обратиться к выборочному *абсолютному центральному моменту первого порядка*. Рассмотрим статистику

$$\mu^* = \frac{1}{n\sqrt{s^{2*}}} \sum_{i=1}^n |X_i - m^*|.$$

Нетрудно показать, что при условии справедливости гипотезы H_0 распределение статистики μ^* зависит только от объема выборки n и не зависит от параметров m и σ^2 *нормального* закона, а в силу закона больших чисел и состоятельности $\sqrt{s^{2*}}$ как оценки среднего квадратичного отклонения сама статистика μ^* при $n \rightarrow \infty$ сходится к первому абсолютному моменту μ стандартного нормального распределения (равному примерно 0,8). Естественно, если значение статистики μ^* сильно отличается от μ , следует отклонить гипотезу H_0 . Значит, гипотеза H_0 принимается, если $C_1 < \mu^* < C_2$. При заданном уровне значимости (размере) α обычно для симметрии полагают $C_1 = \mu_{\alpha/2}$ и $C_2 = \mu_{1-\alpha/2}$, где μ_α — α -квантиль статистики μ^* , построенной по выборке объема n при условии справедливости гипотезы H_0 . Значения μ_α для некоторых n и α можно найти в табл. 4.7а из [1].

Следующий критерий согласия основан на выборочном *коэффициенте асимметрии*

$$a^* = \frac{1}{n(s^{2*})^{3/2}} \sum_{i=1}^n (X_i - m^*)^3.$$

Поскольку нормальный закон имеет симметричную относительно среднего m плотность распределения, то при условии справедливости гипотезы H_0 плотность распределения выборочного коэффициента асимметрии симметрична относительно нуля, и, значит, естественно использовать критерий, инвариантный относительно замены знака a^* на противоположный. Кроме того, при условии справедливости гипотезы H_0 распределение выборочного коэффициента асимметрии зависит только

от объема выборки n и не зависит от m и σ^2 . Сравнивая выборочный коеффициент асимметрии a^* с нулем, получаем с учетом вышесказанного следующий критерий: принимаем гипотезу H_0 , если $-C < a^* < C$, где при уровне значимости (размере) α нужно положить $C = a_{1-\alpha/2}$, а a_α — α -квантиль статистики a^* при условии, что выборка объема n произведена из нормальной генеральной совокупности. Значения a_α для некоторых n и α приведены в табл. 4.7б из [1].

Еще один критерий согласия проверяет близость выборочного эксцесса

$$\gamma^* = \frac{1}{n(s^{2*})^2} \sum_{i=1}^n (X_i - m^*)^4$$

к теоретическому эксцессу $\gamma = \mathbf{M}(X - \mathbf{M}X)^4 / (\mathbf{D}X)^2$, равному 3 для нормального закона. По-прежнему, критерий уровня значимости α предписывает принять H_0 , если $\gamma_{\alpha/2} < \gamma^* < \gamma_{1-\alpha/2}$, где γ_α — α -квантиль статистики γ^* , построенной по нормальной выборке объема n . Значения γ_α для некоторых n и α содержатся в табл. 4.7в из [1].

Наконец, последний приводимый здесь критерий согласия основан на минимальном и максимальном значениях элементов выборки (крайних членах вариационного ряда). Часто этот критерий называют *критерием исключения резко выделяющихся наблюдений*, поскольку он обычно применяется, когда в выборке наряду с элементами нормальной генеральной совокупности присутствуют отдельные «засоряющие» элементы, не принадлежащие этой генеральной совокупности, и их необходимо исключить по результатам анализа выборки. Критерий основан на том, что плотность нормального распределения весьма быстро стремится к нулю при удалении от среднего. Следовательно, наличие очень больших или очень маленьких значений наблюдений в выборке говорит о том, что эта выборка не может принадлежать нормальной генеральной совокупности. Статистика критерия имеет вид

$$x^* = \frac{1}{\sqrt{s^{2*}}} \max_{1 \leq i \leq n} |X_i - m^*|.$$

Гипотеза H_0 принимается с уровнем значимости α , если $x^* < x_{1-\alpha}$, где x_α — α -квантиль статистики x^* при условии нормальности выборки; значения x_α для некоторых n и α приведены в табл. 4.8в из [1].

Пример 1. В примерах 7, 8 и 10 гл. 3 мы уже проверяли гипотезу H_0 о нормальности проекции X вектора скорости молекулы водорода (выборка X_1, \dots, X_{50} приведена в табл. 2 гл. 1). Вновь вернемся к этой гипотезе и проверим ее при уровне значимости $\alpha = 0,02$ с помощью описанных выше критериев. Вычисления дают следующие значения соответствующих статистик (значения m^* и s^{2*} найдены в примере 8 гл. 1): $\mu^* = 0,79$, $a^* = 0,57$, $\gamma^* = 3,14$, $x^* = 2,90$. Сравнивая эти значения с критическими значениями $\mu_{0,01} = 0,7291$, $\mu_{0,99} = 0,8648$, $a_{0,99} = 0,787$ ([1, с. 258]; значения $\mu_{0,01}$ и $\mu_{0,99}$ взяты для объема выборки $n = 51$ как ближайшего к 50), $\gamma_{0,01} = 1,95$, $\gamma_{0,99} = 4,92$ (там же, с. 259), $x_{0,98} = 3,370$ (там же, с. 262), убеждаемся, что и проверка по этим четырем критериям подтверждает гипотезу H_0 . \square

Предостережение. Итак, гипотезу о нормальности проекции вектора скорости молекулы водорода мы проверили с помощью семи (!) критериев. Собственно говоря, мы действовали как сверхсторожный исследователь, придерживающийся принципа «семь раз отмерь, один раз отрежь». Такой исследователь, испытав все известные ему критерии (допустим для определенности, что их k), принимает основную гипотезу H_0 только тогда, когда все они дают положительный результат, или, по сути дела, использует некий «обобщенный» критерий. Предполагая, что уровни значимости отдельных критериев равны $\alpha_1, \dots, \alpha_k$, вычислим уровень значимости α «обобщенного» критерия. Если статистики отдельных критериев независимы, то по формуле умножения вероятностей

$$\alpha = 1 - (1 - \alpha_1) \cdots (1 - \alpha_k).$$

В том случае, когда статистики зависимы, общий уровень значимости может достигать даже величины $\alpha_1 + \dots + \alpha_k$. Это нужно учитывать при использовании «обобщенного» критерия и задавать для каждого отдельного критерия уровень значимости, меньший, чем при использовании только одного критерия. Разумеется, это предостережение относится не только к критериям проверки выборки на нормальность, но и к критериям, проверяющим любую статистическую гипотезу.

У читателя, ознакомившегося с материалами настоящего параграфа, может возникнуть естественное желание использовать подобного вида критерии согласия, основанные на простейших статистиках типа выборочных моментов, для проверки гипотез о принадлежности теоретической функции распределения другим параметрическим семействам, например, семейству гамма-распределений. Следует сразу же предупредить читателя о возможных трудностях, которые могут встретиться на этом пути. Во-первых, для большинства семейств распределения простейших статистик зависят от истинных значений неизвестных параметров, что делает невозможным построение универсальных для всего семейства критических областей. Во-вторых, хотя мы и говорим о «простейших» статистиках, нахождение их распределений представляет собой, как правило, весьма сложную в вычислительном плане задачу. Наконец, для физического обоснования принадлежности теоретического распределения определенному семейству исследователю обычно приходится затратить такие усилия, что экономия на статистической проверке правильности выбранной модели и использование маломощных критериев просто бессмысленна.

3. Критерии равенства дисперсий

В моделях, исследуемых в дальнейшем в этой главе, предполагается, что имеется несколько нормальных выборок с возможно различными средними, но одинаковыми дисперсиями. Естественно, факт *равенства дисперсий* необходимо обосновать. Для этой цели обычно используют критерии Бартлетта, Кокрена и Фишера.

Итак, пусть имеется m независимых выборок

$$X_{11}, \dots, X_{1n_1},$$

$$\dots \dots \dots$$

$$X_{m1}, \dots, X_{mn_m}$$

объемов n_1, \dots, n_m , произведенных из различных нормальных генеральных совокупностей с неизвестными средними m_1, \dots, m_m и также неизвестными дисперсиями $\sigma_1^2, \dots, \sigma_m^2$. Основная гипотеза H_0 состоит в том, что дисперсии во всех генеральных совокупностях равны: $\sigma_1^2 = \dots = \sigma_m^2$; естественно, конкурирующая гипотеза H_1 : некоторые дисперсии могут быть различными.

Статистика критерия Бартлетта задается выражением

$$b = N \ln \left(\frac{1}{N} \sum_{i=1}^m (n_i - 1) s_i^{2*} \right) - \sum_{i=1}^m (n_i - 1) \ln s_i^{2*},$$

где

$$N = \sum_{i=1}^m (n_i - 1), \quad s_i^{2*} = \frac{1}{n_i - 1} \sum_{i=1}^{n_i} (X_{ij} - m_i^*)^2, \quad m^* = \frac{1}{n_i} \sum_{i=1}^{n_i} X_{ij}$$

(смысл введения статистики b заключается в том, что она позволяет привести задачу проверки гипотезы о равенстве дисперсий нормальных выборок к задаче проверки гипотезы о равенстве средних приближенно нормальных выборок; для строгого нормальных выборок эта задача будет решена в параграфе 7). Если гипотеза H_0 верна и все $n_i \geq 6$, то отношение

$$B = b \left[1 + \frac{1}{3(m-1)} \left(\sum_{i=1}^m \frac{1}{n_i - 1} - \frac{1}{N} \right) \right]^{-1}$$

распределено приближенно по закону χ^2 с $m-1$ степенями свободы. Поэтому критерий Бартлетта, имеющий уровень значимости (размер), примерно равный α , предписывает принять гипотезу H_0 , если $B < \chi_{1-\alpha}^2$, и отклонить в противном случае. Здесь, как обычно, через χ_α^2 обозначена α -квантиль χ^2 -распределения с $m-1$ степенями свободы (см. [1, табл. 2.2а]).

Уточнение критерия Бартлетта на случай, когда хотя бы одно из n_i может быть меньше 6, можно найти в [1]. Там же содержатся таблицы (табл. 4.3), необходимые для использования уточненного критерия.

Если все выборки имеют одинаковый объем ($n_1 = \dots = n_m = n$), то для проверки гипотезы H_0 против гипотезы H_1 можно воспользоваться, вообще говоря, менее мощным, но зато более простым *критерием Кокрена*. Статистика критерия Кокрена задается формулой

$$G = \frac{s_{\max}^{2*}}{s_1^{2*} + \dots + s_m^{2*}},$$

где $s_{\max}^{2*} = \max_{1 \leq i \leq m} s_i^{2*}$. Критерий Кокрена предписывает принять H_0 с уровнем значимости α , если $G < G_\alpha$, где G_α — α -квантиль распределения статистики G при условии справедливости основной гипотезы H_0 ; приближенные значения G_α для некоторых α , m и $\nu = n - 1$ приведены в [1, табл. 4.3б].

Наконец, если $m = 2$ (n_1 и n_2 могут быть различными), то для проверки гипотезы H_0 о равенстве дисперсий двух выборок лучше всего использовать *критерий Фишера*. Статистика критерия Фишера определяется формулой

$$\chi = \frac{s_2^{2*}}{s_1^{2*}},$$

где s_1^{2*} и s_2^{2*} — выборочные дисперсии первой и второй выборок. Поскольку при условии справедливости H_0 статистики $(n_1 - 1)s_1^{2*}$ и $(n_2 - 1)s_2^{2*}$ с точностью до одного и того же сомножителя $\sigma_1^2 = \sigma_2^2 = \sigma^2$ имеют χ^2 -распределения с $n_1 - 1$ и $n_2 - 1$ степенями свободы, то статистика χ имеет F -распределение с параметрами $n_2 - 1$ и $n_1 - 1$.

Односторонний критерий Фишера проверяет основную гипотезу H_0 : $\sigma_1^2 = \sigma_2^2$ против конкурирующей гипотезы $H_1: \sigma_1^2 < \sigma_2^2$; критическая область одностороннего критерия Фишера при уровне значимости α задается неравенством $\chi > \varphi_{1-\alpha}$.

Двусторонний критерий Фишера проверяет основную гипотезу H_0 : $\sigma_1^2 = \sigma_2^2$ против конкурирующей гипотезы $H_1: \sigma_1^2 \neq \sigma_2^2$; его допустимая область определяется двумя неравенствами: $\varphi_{\alpha/2} < \chi < \varphi_{1-\alpha/2}$. Здесь φ_α — α -квантиль F -распределения с параметрами $n_2 - 1$ и $n_1 - 1$.

Можно показать, что односторонний критерий Фишера является равномерно наиболее мощным несмешанным критерием. Двусторонний критерий Фишера также будет обладать этим свойством, если критическую область выбирать в виде $C_1 < \chi < C_2$, где C_1 и C_2 — определяемые специальным образом постоянные (см. [12]). Правда, сложность подбора постоянных C_1 и C_2 обычно не оправдывает незначительного увеличения мощности критерия от замены $\varphi_{\alpha/2}$ и $\varphi_{1-\alpha/2}$ на C_1 и C_2 .

Заметим также, что с помощью распределения Фишера нетрудно определить и мощности одностороннего и двустороннего критериев Фишера; эти мощности зависят только от отношения дисперсий σ_1^2 и σ_2^2 .

В дальнейшем (параграфы 5–8) односторонний критерий Фишера станет нашим основным рабочим инструментом, хотя использоваться он будет для проверки несколько иных гипотез.

Пример 2. Прибор, измеряющий скорость элементарной частицы, был опробован на пучке летящих с одинаковой (но неизвестной) скоростью электронов и на пучке летящих с одинаковой (но также неизвестной) скоростью протонов. Результаты измерения скорости 11 электронов $X_{1,1}, \dots, X_{1,11}$ и 11 протонов $X_{2,1}, \dots, X_{2,11}$ ($\times 10^7$ м/с) приведены в табл. 1 и 2. Проверим при уровне значимости $\alpha = 0,05$ гипотезу H_0 о том, что абсолютные точности измерения скоростей как электрона, так и протона совпадают. Естественно, мы предполагаем, что ошибки измерения скоростей как электронов, так и протонов

независимы, распределены по нормальному закону с нулевым средним, но, вообще говоря, разными дисперсиями σ_1^2 и σ_2^2 . Гипотеза H_0 как раз и состоит в том, что эти дисперсии совпадают: $\sigma_1^2 = \sigma_2^2$. \square

Таблица 1

$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$	$X_{1,5}$	$X_{1,6}$	$X_{1,7}$	$X_{1,8}$	$X_{1,9}$	$X_{1,10}$	$X_{1,11}$
7,554	7,550	7,557	7,601	7,595	7,587	7,591	7,592	7,599	7,570	7,609

Таблица 2

$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$	$X_{2,5}$	$X_{2,6}$	$X_{2,7}$	$X_{2,8}$	$X_{2,9}$	$X_{2,10}$	$X_{2,11}$
0,820	0,802	0,821	0,805	0,843	0,818	0,842	0,830	0,786	0,828	0,799

Воспользуемся сначала критерием Бартлетта. Вычислим выборочные дисперсии s_1^{2*} и s_2^{2*}

$$m_1^* = \frac{1}{11} (7,554 + \dots + 7,609) = 7,582,$$

$$m_2^* = \frac{1}{11} (0,820 + \dots + 0,799) = 0,818,$$

$$s_1^{2*} = \frac{1}{10} [(7,554 - 7,582)^2 + \dots + (7,609 - 7,582)^2] = 0,000\,433,$$

$$s_2^{2*} = \frac{1}{10} [(0,820 - 0,818)^2 + \dots + (0,799 - 0,818)^2] = 0,000\,327$$

и значения статистик b и B

$$b = 20 \ln \left(\frac{1}{20} (10 \cdot 0,000\,433 + 10 \cdot 0,000\,327) \right) - (10 \ln 0,000\,433 + 10 \ln 0,000\,327) = 10 \ln 1,01984 = 0,196,$$

$$B = 0,196 \left[1 + \frac{1}{3} \left(\frac{1}{10} + \frac{1}{10} - \frac{1}{20} \right) \right]^{-1} = 0,187.$$

Сравнивая значение статистики $B = 0,187$ со значением $\chi_{0,95}^2 = 3,841$ 0,95-квантили χ^2 -распределения с одной степенью свободы [1, с. 167], делаем вывод, что гипотеза H_0 должна быть принята.

Применим для этой же цели критерий Кокрена (это можно сделать, поскольку объемы выборок одинаковы). Значение статистики критерия Кокрена

$$G = \frac{0,000\,433}{0,000\,433 + 0,000\,327} = 0,570.$$

Сравнение статистики G с 0,95-квантилью распределения статистики Кокрена $G_{0,95} = 0,7880$ ([1], с. 243) показывает, что критерий Кокрена также предписывает принять гипотезу H_0 .

Наконец, мы имеем дело с двумя выборками и, значит, можем использовать критерий Фишера, причем, поскольку не известно заранее, какая из дисперсий σ_1^2 или σ_2^2 больше, то мы должны проверять двустороннюю конкурирующую гипотезу H_1 : $\sigma_1^2 \neq \sigma_2^2$. Вычислив значение статистики

$$\varkappa = \frac{0,000\,433}{0,000\,327} = 1,324$$

и сравнив его с 0,025- и 0,975-квантилями распределения Фишера с параметрами (10,10) $\varphi_{0,025} = 0,2690$ и $\varphi_{0,975} = 3,7168$ ([1], с. 207), мы опять-таки приходим к необходимости принять гипотезу H_0 .

Интересно отметить, что в данном случае (две выборки одинакового объема) все три критерия (Бартлетта, Кокрена и Фишера) совпадают и нам достаточно было воспользоваться только одним из них.

4. Выборочная корреляция

В этом параграфе мы обратимся к постановке задачи, несколько отличной от рассматривавшихся ранее. Будем считать, что выборка $\mathbf{X}_1, \dots, \mathbf{X}_n$ является двумерной, причем элементы выборки $\mathbf{X}_i = (X_{i1}, X_{i2})$ представляют собой двумерные случайные величины, имеющие совместное нормальное распределение со средними m_1 и m_2 , дисперсиями σ_1^2 и σ_2^2 и коэффициентом корреляции ρ . Как известно из курса теории вероятностей, коэффициент корреляции изменяется в пределах от -1 до $+1$, причем в случае нормального распределения равенство коэффициента корреляции нулю эквивалентно независимости компонент выборки X_{i1} и X_{i2} .

Довольно часто встречаются задачи, в которых необходимо либо проверить гипотезу о независимости компонент $X_{(1)}$ и $X_{(2)}$ наблюдаемого вектора \mathbf{X} (т.е. равенстве нулю коэффициента корреляции), либо найти оценку и построить доверительный интервал для коэффициента корреляции. Именно эти задачи мы рассмотрим в настоящем параграфе. Отметим, что корреляционный статистический анализ, о котором пойдет сейчас речь, обычно предшествует рассмотрению задач регрессионного и дисперсионного анализа (см. далее), а иногда даже полностью заменяет регрессионный анализ. Очевидно, что наиболее естественной оценкой коэффициента корреляции ρ является *выборочный коэффициент корреляции*

$$\rho^* = \frac{\sum_{i=1}^n (X_{i1} - m_1^*)(X_{i2} - m_2^*)}{\sqrt{\sum_{i=1}^n (X_{i1} - m_1^*)^2 \sum_{j=1}^n (X_{j2} - m_2^*)^2}},$$

где

$$m_1^* = \frac{1}{n} \sum_{i=1}^n X_{i1}, \quad m_2^* = \frac{1}{n} \sum_{i=1}^n X_{i2}.$$

Можно показать, что распределение статистики ρ^* зависит только от объема выборки n и коэффициента корреляции ρ и не зависит от остальных параметров нормального вектора $\mathbf{X} = (X_{(1)}, X_{(2)})$, а сама статистика ρ^* является состоятельной асимптотически эффективной (в смысле теоремы 7 из гл. 2) оценкой коэффициента корреляции ρ .

Разумным представляется и критерий проверки гипотезы H_0 : $\rho = 0$ о независимости $X_{(1)}$ и $X_{(2)}$ против гипотезы H_1 : $\rho \neq 0$ (компоненты $X_{(1)}$ и $X_{(2)}$ зависимы), предписывающий принять гипотезу H_0 в том случае, когда $\rho_{\alpha/2} < \rho^* < \rho_{1-\alpha/2}$. Здесь ρ_α — α -квантиль выборочного коэффициента корреляции при условии справедливости гипотезы H_0 ($\rho = 0$); в силу симметрии имеет место тождество $\rho_\alpha = -\rho_{1-\alpha}$. Значения α -квантитей выборочного коэффициента корреляции приведены в табл. 4.5а из [1].

Можно показать [12], что приведенный критерий является равномерно наиболее мощным несмещенным критерием.

Пример 3. В табл. 3 приведены значения роста (см) $X_{1,1}, \dots, X_{50,1}$ и веса (кг) $X_{1,2}, \dots, X_{50,2}$ 50 выбранных случайным образом мужчин. Считая, что рост и вес мужчины подчинен двумерному нормальному закону и используя выборочный коэффициент корреляции, проверим гипотезу H_0 о независимости роста и веса. Уровень значимости (размер) критерия $\alpha = 0,05$. \square

Таблица 3

i	1	2	3	4	5	6	7	8	9	10
X_{i1}	168,1	159,1	171,3	179,6	166,9	162,9	182,8	166,3	174,4	177,8
X_{i2}	77,24	64,48	77,93	72,51	72,08	70,02	79,91	67,22	68,42	76,20
i	11	12	13	14	15	16	17	18	19	20
X_{i1}	171,1	170,6	153,9	175,3	170,6	173,8	181,4	174,2	170,4	163,0
X_{i2}	68,41	69,92	58,18	79,16	73,34	71,66	84,04	77,41	73,36	72,94
i	21	22	23	24	25	26	27	28	29	30
X_{i1}	173,8	175,5	179,8	167,8	163,7	177,4	176,4	178,1	170,2	179,2
X_{i2}	81,85	71,01	76,69	68,97	65,41	75,28	83,33	75,73	70,43	73,01
i	31	32	33	34	35	36	37	38	39	40
X_{i1}	171,8	171,7	169,2	166,2	161,7	162,7	167,7	170,5	171,7	184,5
X_{i2}	72,00	73,03	73,98	73,70	69,90	72,33	78,63	83,05	80,46	82,73
i	41	42	43	44	45	46	47	48	49	50
X_{i1}	165,8	175,7	176,0	170,5	163,0	170,4	172,1	181,0	180,2	157,4
X_{i2}	69,80	71,46	82,41	77,08	63,77	66,78	88,15	73,60	77,54	56,84

Вычислим значение выборочного коэффициента корреляции ρ^* :

$$m_1^* = \frac{1}{50} (168,1 + \dots + 171,3) = 171,3,$$

$$m_2^* = \frac{1}{59} (77,24 + \dots + 56,84) = 73,66,$$

$$\sum_{i=1}^{50} (X_{i1} - m_1^*)^2 = (168,1 - 171,3)^2 + \dots + (157,4 - 171,3)^2 = 2266,5,$$

$$\sum_{i=1}^{50} (X_{i2} - m_2^*)^2 = (77,24 - 73,66)^2 + \dots + (56,84 - 73,66)^2 = 1991,0,$$

$$\sum_{i=1}^{50} (X_{i1} - m_1^*)(X_{i2} - m_2^*) = 1391,2, \quad \rho^* = 0,655.$$

Сравнивая ρ^* с 0,975-квантилью выборочного коэффициента корреляции $\rho_{0,975} = 0,273$ [1, с. 248], видим, что гипотеза H_0 о независимости роста и веса должна быть отвергнута.

Что касается построения *доверительных интервалов* для коэффициента корреляции ρ , то Р. Фишер предложил преобразование

$$z = \operatorname{arcth} \rho^* = \frac{1}{2} \ln \frac{1 + \rho^*}{1 - \rho^*},$$

которое при $n \geq 20$ приводит к статистике z с близким к нормальному распределением. При этом

$$\mathbf{M}z = \frac{1}{2} \ln \frac{1 + \rho}{1 - \rho} + \frac{\rho}{2(n-3)} \left(1 - \frac{3 - \rho^2}{4(n-3)} + \dots \right),$$

$$\mathbf{D}z = \frac{1}{n-3} \left(1 - \frac{\rho^2}{2(n-3)} - \frac{2 - 6\rho^2 + 3\rho^4}{6(n-3)^2} - \dots \right).$$

Таким образом, величина $(z - \mathbf{M}z)/\sqrt{\mathbf{D}z}$ распределена практически по стандартному нормальному закону и при построении доверительного интервала для коэффициента корреляции можно воспользоваться стандартным методом, изложенным в параграфе 6 гл. 2. Значения самого преобразования Фишера содержатся в табл. 4.5б из [1]. Кроме того, в табл. 4.5в [1] приведены графики квантилей выборочного коэффициента корреляции ρ^* в зависимости от ρ и n . Эти графики позволяют определить доверительные границы для коэффициента корреляции без всяких вычислений.

Пример 4. В условиях предыдущего примера определим границы двустороннего симметричного доверительного интервала доверительной вероятности $\alpha = 0,95$ для коэффициента корреляции роста и веса мужчины. Как мы вычислили ранее, значение выборочного коэффициента корреляции $\rho^* = 0,655$. Воспользуемся графиком, приведенным в [1, с. 250]. Проводя вертикаль через точку $\rho^* = 0,655$ и вычисляя ординаты ее пересечения с кривыми для $n = 50$, получаем

$$\rho' = 0,45, \quad \rho'' = 0,79.$$

□

5. Общая линейная модель, метод наименьших квадратов

Пусть имеется последовательность наблюдений X_1, \dots, X_n , однако в отличие от ранее использовавшегося понятия выборки будем полагать, что хотя наблюдения X_1, \dots, X_n независимы, они не обладают свойством одинаковой распределенности. А именно, X_i распределены нормально с одинаковой (неизвестной) дисперсией σ^2 , но с различными (неизвестными) средними. Вектор средних (m_1, \dots, m_n) будем обозначать через \mathbf{m} . *Общая линейная модель* заключается в априорном предположении, что вектор \mathbf{m} лежит в некотором линейном подпространстве L размерности $l < n$ пространства R^n . Именно относительно вектора \mathbf{m} мы будем делать наши статистические выводы.

Следует отметить, что использование общей линейной модели позволяет в случаях, когда имеется нормальность или даже асимптотическая нормальность определенных характеристик, получить многие результаты предыдущих глав [12]. Однако цель этого параграфа гораздо скромнее: выяснить, какое место в исследовании линейной модели занимает метод наименьших квадратов.

Как и раньше, статистические задачи, возникающие при анализе общей линейной модели, можно отнести к двум типам: оценка неизвестных параметров и проверка статистических гипотез.

Задачу *оценки неизвестных параметров* можно поставить следующим образом. Из курса линейной алгебры известно, что любой вектор \mathbf{m} из линейного подпространства L можно представить в виде линейной комбинации

$$\mathbf{m} = \sum_{i=1}^l \vartheta_i \mathbf{d}_i$$

линейно независимых (базисных) векторов $\mathbf{d}_1, \dots, \mathbf{d}_l$. Необходимо по наблюдениям X_1, \dots, X_n оценить коэффициенты ϑ_i разложения вектора \mathbf{m} по базисным векторам \mathbf{d}_i и построить для ϑ_i доверительные интервалы.

Пример 5. Пусть зависимость некоторой переменной x от времени описывается линейной функцией $x = x(t) = \vartheta_1 + \vartheta_2 t$, причем параметры ϑ_1 и ϑ_2 неизвестны. Для их определения в моменты t_1, \dots, t_n производятся наблюдения, однако в силу случайных ошибок вместо $x(t_i)$ результатами наблюдений являются величины

$$X_i = x(t_i) + \varepsilon_i,$$

где ошибки ε_i независимы и распределены по нормальному закону с нулевым средним и неизвестной дисперсией σ^2 (здесь мы имеем дело с простой линейной регрессией; общее понятие регрессии будет введено в следующем параграфе). Требуется найти оценки ϑ_1^* и ϑ_2^* неизвестных параметров ϑ_1 и ϑ_2 и построить для ϑ_1 и ϑ_2 доверительные интервалы.

Опишем поставленную задачу в терминах общей линейной модели. Для этого заметим, что $\mathbf{m} = (m_1, \dots, m_n)$ представляет собой вектор средних значений \mathbf{m} наблюдений X_i , т. е. в силу сделанных предположений

$$m_i = \mathbf{M}X_i = x(t_i) = \vartheta_1 + \vartheta_2 t_i.$$

Но тогда, полагая

$$\mathbf{d}_1 = (1, \dots, 1), \quad \mathbf{d}_2 = (t_1, \dots, t_n),$$

получаем, что все допустимые значения вектора \mathbf{m} описываются векторами $\vartheta_1 \mathbf{d}_1 + \vartheta_2 \mathbf{d}_2$, т. е. представляют собой двумерное линейное подпространство L , натянутое на векторы \mathbf{d}_1 и \mathbf{d}_2 . Итак, мы привели рассматриваемую задачу к задаче оценки неизвестных параметров в линейной модели. \square

Задача проверки статистических гипотез предполагает, что в линейном подпространстве L содержится некоторое линейное подпространство L' , имеющее размерность $l' < l$. Проверяемая гипотеза H_0 заключается в том, что вектор \mathbf{m} лежит не только в подпространстве L , но и в подпространстве L' .

Пример 6. Для проверки гипотезы H_0 о том, что результат некоторого эксперимента не зависит от определенного фактора, была проведена серия из $n = n_1 + n_2$ наблюдений, причем сам фактор действовал только в последних n_2 наблюдениях. Считается, что результаты наблюдений $X_1, \dots, X_{n_1}, X_{n_1+1}, \dots, X_n$ представляют собой независимые случайные величины, распределенные по нормальному закону с одинаковой (неизвестной) дисперсией σ^2 и (неизвестными) средними m' при отсутствии и m'' при наличии действия фактора. Таким образом,

$$\mathbf{M}X_1 = \dots = \mathbf{M}X_{n_1} = m', \quad \mathbf{M}X_{n_1+1} = \dots = \mathbf{M}X_n = m'',$$

и гипотеза H_0 состоит в равенстве средних m' и m'' (это так называемая одноФакторная двухуровневая модель дисперсионного анализа; общее определение многофакторной многоуровневой модели будет дано в параграфе 7).

В терминах общей линейной модели вектор

$$\mathbf{m} = (m_1, \dots, m_n) = (m', \dots, m', m'', \dots, m'')$$

априори лежит в двумерном подпространстве L , порожденном векторами $\mathbf{d}_1 = (1, \dots, 1, 0, \dots, 0)$ и $\mathbf{d}_2 = (0, \dots, 0, 1, \dots, 1)$, а проверяемая гипотеза H_0 состоит в том, что \mathbf{m} лежит также в одномерном подпространстве L' , порожденном вектором $\mathbf{d}' = (1, \dots, 1)$. \square

Начнем с задачи *проверки статистических гипотез*. Выберем в пространстве R^n новый ортонормированный базис $\mathbf{e}'_1, \dots, \mathbf{e}'_n$, такой, что первые l' базисных векторов $\mathbf{e}'_1, \dots, \mathbf{e}'_{l'}$ порождают подпространство L' , следующие $l - l'$ векторов $\mathbf{e}'_{l'+1}, \dots, \mathbf{e}'_l$ дополняют подпространство L' до L и, наконец, оставшиеся $n - l$ векторов $\mathbf{e}'_{l+1}, \dots, \mathbf{e}'_n$ дополняют L до R^n . Линейное преобразование, переводящее стандартный базис $\mathbf{e}_1 = (1, 0, \dots, 0), \dots, \mathbf{e}_n = (0, \dots, 0, 1)$ в базис $\mathbf{e}'_1, \dots, \mathbf{e}'_n$, является ортогональным. Поэтому вектор $\mathbf{X}' = (X'_1, \dots, X'_n)$, представляющий собой вектор $\mathbf{X} = (X_1, \dots, X_n)$, записанный в новом базисе $\mathbf{e}'_1, \dots, \mathbf{e}'_n$, будет также состоять из независимых координат X'_1, \dots, X'_n , распределенных поциальному закону с одинаковой дисперсией σ^2 и вектором средних $\mathbf{m}' = (m'_1, \dots, m'_n)$. Однако в силу априорного предположения

о принадлежности вектора \mathbf{m} подпространству L справедливо соотношение $m'_{l+1} = \dots = m'_n = 0$, а проверяемая гипотеза H_0 о принадлежности L' заключается в том, что равны нулю также координаты вектора \mathbf{m}' с номерами $l'+1, \dots, l$, т. е. $m'_{l'+1} = \dots = m'_l = 0$.

Определим теперь статистики

$$s_0^{2*} = \frac{1}{n-l} \sum_{i=l+1}^n (X'_i)^2, \quad s_1^{2*} = \frac{1}{l-l'} \sum_{i=l'+1}^l (X'_i)^2, \quad \varkappa = \frac{s_1^{2*}}{s_0^{2*}}.$$

Статистики s_0^{2*} и s_1^{2*} являются независимыми, и, как говорилось в параграфе 4 гл. 1, случайная величина $(n-l)s_0^{2*}/\sigma^2$ имеет χ^2 -распределение с $n-l$ степенями свободы.

Кроме того, если справедлива основная гипотеза H_0 , то величина $(l-l')s_1^{2*}/\sigma^2$ также распределена по закону χ^2 с $l-l'$ степенями свободы; статистики s_0^{2*} и s_1^{2*} представляют в этом случае две независимые несмещенные оценки неизвестной дисперсии σ^2 .

Однако если справедлива конкурирующая гипотеза H_1 , то несмешенной оценкой σ^2 будет только статистика s_0^{2*} , а статистика s_1^{2*} будет систематически больше σ^2 (ее распределение представляет собой так называемое нецентральное χ^2 -распределение), т. е. иметь положительное смещение $\delta = Ms_1^{2*} - \sigma^2$, тем большее, чем больше отклонение вектора \mathbf{m} от линейного подпространства L' . Поэтому для проверки гипотезы H_0 естественно применить *односторонний критерий Фишера*, предписывающий принять H_0 , если $\varkappa < C$, и отвергнуть в противном случае. При заданном уровне значимости критерия α критическое значение C совпадает с $(1-\alpha)$ -квантилем $\varphi_{1-\alpha}$ F -распределения с параметрами $l-l'$ и $n-l$ [1, табл. 3.5].

Построенный критерий является равномерно наиболее мощным инвариантным для проверки гипотез H_0 и H_1 .

Заметим, что нецентральное F -распределение [1, табл. 4.12] можно использовать для вычисления мощности построенного критерия; это же замечание справедливо для всех решаемых далее задач настоящей главы, в которых применяется критерий Фишера.

Однако предложенный подход обладает существенным недостатком. Обычно линейные подпространства L и L' задаются неортонормированными системами базисных векторов. Поэтому, чтобы применить полученный критерий, необходимо сначала выбрать ортонормированный базис и произвести линейное преобразование вектора наблюдений \mathbf{X} , что, как правило, представляет собой весьма трудоемкую в вычислительном плане задачу. И здесь на помощь приходит метод наименьших квадратов, позволяющий при определении s_0^{2*} и s_1^{2*} заменить задачу нахождения линейного преобразования существенно более простой задачей поиска минимума квадратичного функционала.

Для теоретического обоснования метода наименьших квадратов полезна простейшая геометрическая интерпретация статистики $(n-l)s_0^{2*}$ как квадрата расстояния от точки (X_1, \dots, X_n) до подпро-

странства L , а статистики $(l - l')s_1^{2*}$ — как квадрата проекции вектора \mathbf{X} на ортогональное дополнение L' до L .

Сам метод наименьших квадратов заключается в следующем. Пусть (m_1, \dots, m_n) — произвольная точка в R^n . Рассмотрим квадрат расстояния $S^2 = \sum_{i=1}^n (X_i - m_i)^2$ от точки (X_1, \dots, X_n) до точки (m_1, \dots, m_n) и найдем минимальное значение S^2 для всех (m_1, \dots, m_n) , принадлежащих подпространству L , которое и будет совпадать с квадратом расстояния от точки (X_1, \dots, X_n) до подпространства L , т. е.

$$s_0^{2*} = \frac{1}{n-l} \min_{(m_1, \dots, m_n) \in L} S^2.$$

Аналогично, статистика

$$S_2^2 = \min_{(m_1, \dots, m_n) \in L'} S^2$$

представляет собой квадрат расстояния от точки (X_1, \dots, X_n) до подпространства L' , а значит,

$$s_1^{2*} = \frac{S_2^2 - (n-l)s_0^{2*}}{l - l'}.$$

Теперь для применения критерия Фишера, проверяющего гипотезу H_0 , осталось составить отношение

$$\varkappa = \frac{s_1^{2*}}{s_0^{2*}}$$

и сравнить его с критическим значением $C = \varphi_{1-\alpha}$.

Как уже говорилось, при практической реализации критерия подпространства L и L' задаются системой базисных векторов

$$\mathbf{d}_1 = (d_{11}, \dots, d_{1n}), \dots, \mathbf{d}_l = (d_{l1}, \dots, d_{ln})$$

и

$$\mathbf{d}'_1 = (d'_{11}, \dots, d'_{1n}), \dots, \mathbf{d}'_{l'} = (d'_{l'1}, \dots, d'_{l'n}).$$

Тогда задача нахождения s_0^{2*} сводится к задаче нахождения минимума квадратичного функционала:

$$(n-l)s_0^{2*} = \min_{\vartheta_1, \dots, \vartheta_l} \sum_{i=1}^n (X_i - \vartheta_1 d_{1i} - \dots - \vartheta_l d_{li})^2 = \min_{\vartheta_1, \dots, \vartheta_l} S^2(\vartheta_1, \dots, \vartheta_l).$$

Значение s_0^{2*} достигается при подстановке в $S^2(\vartheta_1, \dots, \vartheta_l)$ коэффициентов $\vartheta_1^*, \dots, \vartheta_l^*$, определяемых из системы линейных уравнений, которые получаются дифференцированием $S^2(\vartheta_1, \dots, \vartheta_l)$ по ϑ_i и приравниванием производных нулю:

$$\sum_{j=1}^n (\vartheta_1^* d_{1j} + \dots + \vartheta_l^* d_{lj}) d_{ij} = \sum_{j=1}^n d_{ij} X_j. \quad (1)$$

Аналогично вычисляется значение статистики S_2^2 . Для этого сначала решается система (1), в которой вместо ϑ_i^* и d_{ij} подставлены $\vartheta_i'^*$ и d_{ij}' , а затем полагается $S_2^2 = S^2(\vartheta_1'^*, \dots, \vartheta_l'^*)$.

Пример 7. Обратимся к решению поставленной в примере 6 задачи.

Как мы уже говорили, подпространство L задается двумя базисными векторами:

$$\mathbf{d}_1 = (1, \dots, 1, 0, \dots, 0), \quad \mathbf{d}_2 = (0, \dots, 0, 1, \dots, 1).$$

Система уравнений (1) в данном случае имеет вид

$$\begin{aligned} n_1 \vartheta_1^* &= \sum_{j=1}^{n_1} X_j, & \vartheta_1^* &= \frac{1}{n_1} \sum_{j=1}^{n_1} X_j = m'^*, \\ n_2 \vartheta_2^* &= \sum_{j=n_1+1}^n X_j, & \vartheta_2^* &= \frac{1}{n_2} \sum_{j=n_1+1}^n X_j = m''*. \end{aligned}$$

Поэтому

$$s_0^{2*} = \frac{1}{n-2} \min_{\vartheta_1, \vartheta_2} S^2(\vartheta_1, \vartheta_2) = \frac{1}{n-2} \left[\sum_{j=1}^{n_1} (X_j - m'^*)^2 + \sum_{j=n_1+1}^n (X_j - m''*)^2 \right].$$

Аналогично, подпространство L' порождается вектором

$$\mathbf{d}' = (1, \dots, 1),$$

а система (1) состоит из единственного уравнения

$$(n_1 + n_2) \vartheta_1'^* = \sum_{j=1}^n X_j, \quad \vartheta_1'^* = \frac{1}{n} \sum_{j=1}^n X_j = m^*.$$

Отсюда

$$S_2^2 = \min_{\vartheta'_1} S^2(\vartheta'_1) = \sum_{j=1}^n (X_j - m^*)^2, \quad s_1^{2*} = \frac{S_2^2 - (n-2)s_0^{2*}}{2-1}.$$

Простейшие подсчеты показывают, что статистику s_1^{2*} можно записать также в виде:

$$s_1^{2*} = n_1(m'^* - m^*)^2 + n_2(m''* - m^*)^2.$$

Сам критерий уровня значимости α заключается в следующем: мы должны принять гипотезу H_0 , если $\varkappa = s_1^{2*}/s_2^{2*} < \varphi_{1-\alpha}$, где φ_α — α -квантиль F -распределения с параметрами 1 и $n-2$ [1, табл. 3.5], и отвергнуть в противном случае. \square

Перейдем к задаче *оценки неизвестных параметров*. Напомним ее постановку. Пусть вектор \mathbf{m} представим в виде суммы

$$\mathbf{m} = \sum_{i=1}^l \vartheta_i \mathbf{d}_i,$$

где \mathbf{d}_i — известные базисные векторы подпространства L , а ϑ_i — неизвестные параметры. Задача заключается в оценке ϑ_i и построении для них доверительных интервалов.

И эта задача решается наиболее просто в случае, когда \mathbf{d}_i представляют собой ортонормированную (или хотя бы ортогональную) систему

векторов. Действительно, выбирая новый ортонормированный базис

$$\mathbf{e}'_1 = \mathbf{d}_1, \dots, \mathbf{e}'_l = \mathbf{d}_l, \mathbf{e}'_{l+1}, \dots, \mathbf{e}'_n$$

и вводя в рассмотрение новый вектор $\mathbf{X}' = (X'_1, \dots, X'_n)$, состоящий из координат вектора \mathbf{X} в новом базисе $\mathbf{e}'_1, \dots, \mathbf{e}'_n$, получаем, что величины X'_i являются независимыми и нормально распределенными с неизвестной одинаковой дисперсией σ^2 и средними

$$\mathbf{M}X'_1 = \vartheta_1, \dots, \mathbf{M}X'_l = \vartheta_l, \mathbf{M}X'_{l+1} = \dots = \mathbf{M}X'_n = 0.$$

Как известно (см. пример 25 в гл. 2), эффективные оценки ϑ_i^* параметров ϑ_i задаются простейшей формулой $\vartheta_i^* = X_i$. Однако для построения доверительных интервалов для ϑ_i необходимо знать дисперсию σ^2 нормального закона или хотя бы ее оценку. Но такая оценка s_0^{2*} , не зависящая от $\vartheta_1^*, \dots, \vartheta_l^*$ и с точностью до множителя $(n-l)/\sigma^2$ распределенная по закону χ^2 с $n-l$ степенями свободы, легко находится из рассмотрения координат X'_{l+1}, \dots, X'_n :

$$s_0^{2*} = \frac{1}{n-l} \sum_{j=l+1}^n (X'_j)^2.$$

Таким образом, величина

$$t = \frac{\vartheta_i^* - \vartheta_i}{\sqrt{s_0^{2*}}}$$

имеет t -распределение с $n-l$ степенями свободы, а симметричный доверительный интервал доверительной вероятности α для ϑ_i задается (см. пример 30 в гл. 2) границами:

$$\vartheta'_i = \vartheta_i^* - t_{\frac{1+\alpha}{2}} \sqrt{s_0^{2*}}, \quad \vartheta''_i = \vartheta_i^* + t_{\frac{1+\alpha}{2}} \sqrt{s_0^{2*}},$$

где t_α — α -квантиль t -распределения с $n-l$ степенями свободы.

Как и при проверке статистических гипотез, трудности возникают тогда, когда векторы $\mathbf{d}_1, \dots, \mathbf{d}_l$ не являются ортогональными, и эти трудности также легко обходятся с помощью метода наименьших квадратов. По-прежнему, рассмотрим квадратичный функционал

$$S^2(\vartheta_1, \dots, \vartheta_l) = \sum_{i=1}^n (X_i - \vartheta_1 d_{1i} - \dots - \vartheta_l d_{li})^2,$$

зависящий от некоторых (абстрактных) аргументов $\vartheta_1, \dots, \vartheta_l$. Оценки $\vartheta_1^*, \dots, \vartheta_l^*$ по методу наименьших квадратов параметров $\vartheta_1, \dots, \vartheta_l$ получаются как значения аргументов $\vartheta_1, \dots, \vartheta_l$, доставляющих минимум квадратичному функционалу $S^2(\vartheta_1, \dots, \vartheta_l)$, т. е. определяются из системы линейных уравнений (1).

Перечислим основные свойства оценок, полученных по методу наименьших квадратов.

1. Оценки ϑ_i^* являются линейными, т. е. выражаются в виде линейной комбинации

$$\vartheta_i^* = \sum_{j=1}^n c_{ij} X_j$$

от наблюдений X_1, \dots, X_n , и нормально распределенными как линейные комбинации нормально распределенных наблюдений X_1, \dots, X_n .

2. Оценки ϑ_i^* являются эффективными, в частности, несмешенными и имеющими минимальную дисперсию среди всех возможных оценок.

Свойство 2 показывает, что оценки ϑ_i^* могут быть получены методом максимального правдоподобия, что нетрудно выявить и прямыми вычислениями; впрочем, это мы фактически и делали в случае ортогонализированных векторов $\mathbf{d}_1, \dots, \mathbf{d}_l$.

Следует отметить также, что если векторы $\mathbf{d}_1, \dots, \mathbf{d}_l$ неортогональны, то оценки $\vartheta_1^*, \dots, \vartheta_l^*$ *зависимы*.

Одномерные *доверительные интервалы* для неизвестных параметров $\vartheta_1, \dots, \vartheta_l$ строятся следующим образом. Как было показано, статистика

$$s_0^{2*} = \frac{1}{n-l} \min_{\vartheta_1, \dots, \vartheta_l} S^2(\vartheta_1, \dots, \vartheta_l) = \frac{1}{n-l} S^2(\vartheta_1^*, \dots, \vartheta_l^*)$$

представляет собой несмешенную оценку неизвестной дисперсии σ^2 , не зависящую от $\vartheta_1^*, \dots, \vartheta_l^*$ и с точностью до множителя $(n-l)/\sigma^2$ имеющую χ^2 -распределение с $n-l$ степенями свободы. Далее, поскольку

$$\vartheta_i^* = \sum_{j=1}^n c_{ij} X_j,$$

то дисперсия оценки ϑ_i^* задается формулой

$$\mathbf{D}\vartheta_i^* = \sigma^2 \sum_{j=1}^n c_{ij}^2 = \sigma^2 c_i^2. \quad (2)$$

Значит, величина

$$t = \frac{\vartheta_i^* - \vartheta_i}{\sqrt{c_i^2 s_0^{2*}}}$$

имеет t -распределение с $n-l$ степенями свободы, а границы симметричного доверительного интервала доверительной вероятности α для ϑ_i имеют вид

$$\vartheta'_i = \vartheta_i^* - t_{\frac{1+\alpha}{2}} \sqrt{c_i^2 s_0^{2*}}, \quad \vartheta''_i = \vartheta_i^* + t_{\frac{1+\alpha}{2}} \sqrt{c_i^2 s_0^{2*}}, \quad (3)$$

где, как обычно, t_α — α -квантиль t -распределения с $n-l$ степенями свободы.

Пример 8. Решим сформулированную в примере 5 задачу. Найдем сначала оценки ϑ_1^* и ϑ_2^* неизвестных параметров ϑ_1 и ϑ_2 . Система уравнений (1) принимает в данном случае вид

$$n\vartheta_1^* + \vartheta_2^* \sum_{j=1}^n t_j = \sum_{j=1}^n X_j, \quad \vartheta_1^* \sum_{j=1}^n t_j + \vartheta_2^* \sum_{j=1}^n t_j^2 = \sum_{j=1}^n t_j X_j,$$

откуда

$$\vartheta_1^* = \sum_{j=1}^n c_{1j} X_j, \quad \vartheta_2^* = \sum_{j=1}^n c_{2j} X_j,$$

где для краткости введены обозначения:

$$c_{1j} = \frac{\sum_{k=1}^n t_k^2 - t_j \sum_{k=1}^n t_k}{n \sum_{k=1}^n t_k^2 - \left(\sum_{k=1}^n t_k \right)^2}, \quad c_{2j} = \frac{n t_j - \sum_{k=1}^n t_k}{n \sum_{k=1}^n t_k^2 - \left(\sum_{k=1}^n t_k \right)^2}.$$

Отметим, что вычисления оценок ϑ_1^* и ϑ_2^* существенно упрощаются, если в качестве начала отсчета выбрано время $t_0 = \sum_{j=1}^n t_j / n$.

Приступим теперь к построению доверительных интервалов доверительной вероятности α для неизвестных параметров ϑ_1 и ϑ_2 . Статистика s_0^{2*} (называемая также остаточной дисперсией) с точностью до множителя $(n-2)/\sigma^2$ имеет χ^2 -распределение с $n-2$ степенями свободы и задается формулой

$$s_0^{2*} = \frac{1}{n-2} \sum_{j=1}^n (X_j - \vartheta_1^* - \vartheta_2^* t_j)^2.$$

Значения коэффициентов c_1^2 и c_2^2 определяются из общей формулы (2):

$$c_1^2 = \sum_{j=1}^n c_{1j}^2, \quad c_2^2 = \sum_{j=1}^n c_{2j}^2,$$

а нижняя и верхняя границы симметричного доверительного интервала доверительной вероятности α для параметра ϑ_i задаются формулой (3), в которой t_α — α -квантиль распределения Стьюдента с $n-2$ степенями свободы. □

Несомненным достоинством метода наименьших квадратов является то, что он применим для оценки неизвестных параметров и в том случае, когда наблюдения X_1, \dots, X_n распределены не обязательно по нормальному закону. Более того, наблюдения X_1, \dots, X_n могут быть разнораспределенными и даже зависимыми. Необходимо только, чтобы X_i имели одинаковую дисперсию σ^2 и были некоррелированными. Естественно, процедура построения оценок остается неизменной. Свойства 1 и 2 оценок $\vartheta_1^*, \dots, \vartheta_l^*$, полученных по методу наименьших квадратов, заменяются в этом случае на следующие:

1'. Оценки ϑ_i^* являются линейными.

2'. Оценки ϑ_i^* являются несмещанными и имеют минимальную дисперсию в классе всех линейных оценок.

Разумеется, в случае наблюдений X_1, \dots, X_n , распределенных не по нормальному закону, использование F -распределения и t -распределения для проверки гипотез и построения доверительных интервалов неправомочно.

6. Регрессионный анализ

Прежде чем переходить к рассмотрению простейших статистических задач регрессионного анализа, попробуем разобраться в самом понятии «регрессия». Для этого возвратимся к терминам теории вероятностей и рассмотрим двумерную случайную величину (ξ, η) . Пусть случайная величина η приняла значение y . Что в этом случае можно сказать о значении случайной величины ξ ? Из курса теории вероятностей известно (см. часть 1, гл. 6, параграф 4), что при этом условии случайная величина ξ имеет условную плотность распределения

$$p_\xi(x | \eta = y) = \frac{p_{\xi\eta}(x, y)}{p_\eta(y)}$$

(для простоты предположим, что существует совместная плотность распределения $p_{\xi\eta}(x, y)$). Изменение этой плотности распределения как функции от аргумента y характеризует зависимость ξ от η . Однако если мы хотим указать зависимость ξ от y «в целом», то мы должны обратиться к какой-либо числовой характеристике случайной величины ξ и рассматривать зависимость этой числовой характеристики от значения y случайной величины η . Такая зависимость носит название регрессии ξ на η . Поскольку наиболее важной числовой характеристикой случайной величины является математическое ожидание, то для этой цели обычно используют именно его и под простой (т. е. зависящей от одного параметра) *регрессией* ξ на η подразумевают *зависимость условного математического ожидания*

$$m(y) = \mathbf{M}(\xi | \eta = y) = \int_{-\infty}^{\infty} x p_\xi(x | \eta = y) dx$$

от значения y случайной величины η . Для простоты изложения мы под регрессией будем понимать не только факт зависимости $m(y)$ от y , но и саму функцию $m(y)$ (иногда функцию $m(y)$ называют функцией регрессии). График регрессии $m(y)$ носит название линии регрессии (см. также часть 1, гл. 7, параграф 5).

Пример 9. Пусть (ξ, η) — двумерная случайная величина, распределенная поциальному закону с параметрами $m_1 = \mathbf{M}\xi$, $m_2 = \mathbf{M}\eta$, $\sigma_1^2 = \mathbf{D}\xi$, $\sigma_2^2 = \mathbf{D}\eta$ и $\rho = \rho(\xi, \eta) = \text{cov}(\xi, \eta)/\sqrt{\mathbf{D}\xi\mathbf{D}\eta}$. Определим регрессию $m(y)$ величины ξ на η . Из курса теории вероятностей известно, что условное распределение ξ при условии $\eta = y$ является нормальным со средним значением $m_1 + \rho(y - m_2)\sqrt{\sigma_1^2/\sigma_2^2}$. Таким образом, регрессия $m(y)$ является линейной:

$$m(y) = \vartheta_1 + \vartheta_2 y,$$

где

$$\vartheta_1 = m_1 - \rho m_2 \sqrt{\frac{\sigma_1^2}{\sigma_2^2}}, \quad \vartheta_2 = \rho \sqrt{\frac{\sigma_1^2}{\sigma_2^2}}.$$

Отметим также, что условное распределение случайной величины $\xi - m(y)$ при условии $\eta = y$ нормально с нулевым средним (поскольку мы из ξ вычли ее математическое ожидание $m(y)$) и постоянной (не зависящей от y) дисперсией $(1 - \rho^2) \sigma_1^2$. \square

Заметим теперь, что случайную величину η обычно можно трактовать как независимый параметр, значение которого y нам известно и которым в ряде случаев мы даже можем управлять. Тогда регрессия представляет собой функцию, описывающую зависимость среднего значения случайной величины ξ от независимой переменной y .

Понятие регрессии часто возникает и из следующих соображений. Пусть некоторое физическое явление описывается двумя параметрами: (неслучайным) независимым параметром y и связанным с ним функциональной зависимостью $x = x(y)$ (также неслучайным) зависимым параметром x . Однако если мы наблюдаем за поведением зависимого параметра x , то в результате действия различных случайных причин (влияние неконтролируемых параметров, ошибки измерений и т. д.) мы получаем функцию $x(y)$ со случайными ошибками. Обычно считают, что эти ошибки имеют нулевое среднее, или, как говорят, отсутствуют систематические ошибки наблюдений. Тогда функцию $x(y)$ также естественно назвать регрессией. Примером такой модели может служить определение радиолокатором координат движущейся цели (самолета, корабля и т. п.).

Все сказанное выше можно отнести и к случаю нескольких независимых переменных (тогда говорят о множественной регрессии), и к случаю векторного зависимого параметра (многомерная регрессия). Мы, однако, ограничимся случаем простой регрессии, причем для удобства будем отождествлять независимую переменную со временем.

Итак, приступим к описанию той *регрессионной модели*, которую мы будем рассматривать далее. Предположим, что некоторое физическое явление проистекает во времени и описывается параметром $x = x(t)$, изменяющимся по закону

$$x(t) = \vartheta_1 f_1(t) + \dots + \vartheta_l f_l(t),$$

причем $f_1(t), \dots, f_l(t)$ — известные функции, а $\vartheta_1, \dots, \vartheta_l$ — неизвестные параметры. Пусть в моменты t_1, \dots, t_n производятся измерения параметра $x(t)$ с ошибками $\varepsilon_1, \dots, \varepsilon_n$, относительно которых известно, что они независимы и распределены по нормальному закону с нулевым средним и одинаковой неизвестной дисперсией σ^2 . Таким образом, результаты измерений можно представить в виде

$$X_i = \vartheta_1 f_1(t_i) + \dots + \vartheta_l f_l(t_i) + \varepsilon_i.$$

Рассмотрим следующие задачи.

1. Оценка неизвестных параметров $\vartheta_1, \dots, \vartheta_l$ и построение для них (одномерных) доверительных интервалов.
2. Проверка гипотезы о равенстве некоторых ϑ_i нулю.
3. Нахождение оценки $x^*(t)$ регрессии $x(t)$ в произвольный момент t и построение для $x(t)$ доверительных интервалов.

Прежде чем переходить к решению задач 1–3, наложим на параметры регрессии некоторые естественные ограничения.

Во-первых, будем считать, что число наблюдений n больше числа l неизвестных параметров. Действительно, если $n < l$, то число определяемых параметров больше числа наблюдений; случай же $n = l$ не дает возможности учесть элементы случайности в рассматриваемой модели.

Во-вторых, векторы

$$\mathbf{f}_1 = (f_1(t_1), \dots, f_1(t_n)),$$

...

$$\mathbf{f}_l = (f_l(t_1), \dots, f_l(t_n))$$

должны быть линейно независимыми, иначе одинаковые траектории $x(t_1), \dots, x(t_n)$ можно было бы получить при разных значениях параметров $\vartheta_1, \dots, \vartheta_l$.

Решения поставленных задач используют метод наименьших квадратов. Поскольку решения задач 1 и 2 полностью повторяют решения соответствующих задач для общей линейной модели, дадим только их краткое описание со ссылкой на предыдущий параграф.

Задача 1. Рассмотрим сумму квадратов

$$S^2 = S^2(\vartheta_1, \dots, \vartheta_l) = \sum_{i=1}^n [X_i - \vartheta_1 f_1(t_i) - \dots - \vartheta_l f_l(t_i)]^2.$$

Оценки $\vartheta_1^*, \dots, \vartheta_l^*$ неизвестных параметров $\vartheta_1, \dots, \vartheta_l$ находятся из условия минимальности функции S^2 . Дифференцируя S^2 по $\vartheta_1, \dots, \vartheta_l$ и приравнивая производные нулю, получаем для определения ϑ_i^* следующую систему линейных уравнений:

$$\vartheta_1^* \sum_{j=1}^n f_i(t_j) f_1(t_j) + \dots + \vartheta_l^* \sum_{j=1}^n f_i(t_j) f_l(t_j) = \sum_{j=1}^n X_j f_i(t_j). \quad (4)$$

Оценки $\vartheta_1^*, \dots, \vartheta_l^*$ удовлетворяют свойствам 1 и 2, приведенным в параграфе 5.

Если отказаться от требования нормальности ошибок ε_i и считать только, что они некоррелированы, имеют нулевое среднее и одинаковую дисперсию σ^2 , то оценки $\vartheta_1^*, \dots, \vartheta_l^*$ будут удовлетворять свойствам 1' и 2', приведенным в параграфе 5.

Симметричные доверительные интервалы доверительной вероятности α для неизвестных параметров $\vartheta_1, \dots, \vartheta_l$ определяются своими границами (3), где

$$s_0^{2*} = \frac{1}{n-l} \min_{\vartheta_1, \dots, \vartheta_l} S^2(\vartheta_1, \dots, \vartheta_l) = \frac{S^2(\vartheta_1^*, \dots, \vartheta_l^*)}{n-l}, \quad c_i^2 = \sum_{j=1}^n c_{ij}^2,$$

c_{ij} — коэффициенты в представлении

$$\vartheta_i^* = \sum_{j=1}^n c_{ij} X_j,$$

а t_α — α -квантиль t -распределения с $n-l$ степенями свободы.

Система уравнений (4) наиболее просто решается в том случае, когда векторы

$$\mathbf{f}_1 = (f_1(t_1), \dots, f_1(t_n)),$$

...

$$\mathbf{f}_l = (f_l(t_1), \dots, f_l(t_n))$$

ортогональны, т. е.

$$\sum_{j=1}^n f_i(t_j) f_k(t_j) = 0$$

при $k \neq i$. Тогда (4) распадается на отдельные уравнения

$$\vartheta_i \sum_{j=1}^n [f_i(t_j)]^2 = \sum_{j=1}^n X_j f_i(t_j).$$

В этом случае дополнительно к свойствам 1 и 2 оценки ϑ_i^* будут также независимыми (если не предполагать нормальности ε_i , то дополнительно к 1' и 2' оценки ϑ_i^* будут некоррелированными).

Задача 2. Сформулируем задачу следующим образом. Пусть $0 \leqslant l' < l$. Проверяемая гипотеза H_0 состоит в том, что $\vartheta_{l'+1} = \dots = \vartheta_l = 0$. Физически это означает, что мы перестраховались и ввели в модель лишние воздействия, задаваемые функциями $f_{l'+1}(t), \dots, f_l(t)$, которые на самом деле на исследуемое явление не влияют. Для решения задачи 2 определим, как и раньше,

$$s_0^{2*} = \frac{1}{n-l} \min_{\vartheta_1, \dots, \vartheta_l} S^2(\vartheta_1, \dots, \vartheta_l), \quad S_2^2 = \min_{\vartheta_1, \dots, \vartheta_{l'}} S^2(\vartheta_1, \dots, \vartheta_{l'}, 0, \dots, 0).$$

Положим

$$s_1^{2*} = \frac{S_2^2 - (n-l)s_0^{2*}}{l-l'}, \quad \varkappa = \frac{s_1^{2*}}{s_0^{2*}}.$$

Гипотезу H_0 следует принять с уровнем значимости (размером) α , если $\varkappa < \varphi_{1-\alpha}$, где φ_α — α -квантиль F -распределения с параметрами $l-l'$ и $n-l$.

Задача 3. Очевидно, что оценку $x^*(t)$ регрессии $x(t)$ проще всего получить, подставляя в $x(t)$ вместо ϑ_i их оценки:

$$x^*(t) = \vartheta_1^* f_1(t) + \dots + \vartheta_l^* f_l(t).$$

Ясно, что $x^*(t)$ является линейной, нормально распределенной и несмещенной оценкой регрессии $x(t)$. Можно показать также, что она эффективна.

Для построения доверительного интервала для $x(t)$ вспомним, что ϑ_i^* представляются в виде

$$\vartheta_i^* = \sum_{j=1}^n c_{ij} X_j.$$

Значит, дисперсия $Dx^*(t)$ оценки регрессии $x(t)$ задается формулой

$$Dx^*(t) = \sigma^2 \sum_{j=1}^n \left[\sum_{i=1}^l c_{ij} f_i(t) \right]^2 = \sigma^2 c^2(t)$$

и случайная величина

$$\frac{x^*(t) - x(t)}{\sqrt{\sigma^2 c^2(t)}}$$

распределена по стандартному нормальному закону. Поскольку статистика s_0^{2*} не зависит от $x^*(t)$ и ее дисперсия равна σ^2 , то случайная величина

$$t = \frac{x^*(t) - x(t)}{\sqrt{c^2(t) s_0^{2*}}}$$

имеет t -распределение, а симметричный доверительный интервал доверительной вероятности α для регрессии $x(t)$ задается границами

$$x'(t) = x^*(t) - t_{\frac{1+\alpha}{2}} \sqrt{c^2(t) s_0^{2*}}, \quad x''(t) = x^*(t) + t_{\frac{1+\alpha}{2}} \sqrt{c^2(t) s_0^{2*}},$$

где t_α — α -квантиль t -распределения с $n-l$ степенями свободы.

Пример 10. Рассмотрим простую линейную регрессию

$$x(t) = \vartheta_1 + \vartheta_2 t.$$

Собственно говоря, исследование этой модели уже началось в примерах 5 и 8, в которых она представлялась в виде частного случая общей линейной модели. Поскольку там мы оценили неизвестные параметры ϑ_1 и ϑ_2 и построили для них доверительные интервалы, а решение этой же задачи в терминах регрессионной модели, очевидно, полностью совпадает с ее решением в терминах общей линейной модели, то здесь остановимся только на решениях задач 2 и 3.

1°. Задачу 2 сформулируем в следующем виде. Будем проверять независимость $x(t)$ от времени, т. е. гипотезу $H_0: \vartheta_2 = 0$. Значение статистики s_0^{2*} , необходимой для применения критерия Фишера, уже было вычислено в примере 8. Определим S_2^2 . В соответствии с общими принципами мы в выражении $S^2 = S^2(\vartheta_1, \vartheta_2)$ должны положить $\vartheta_2 = 0$ и искать минимум $S^2 = S^2(\vartheta_1, 0)$ как

функции от одной переменной ϑ_1 . Уравнение (4) для определения значения ϑ_1^* , доставляющего минимум S^2 , имеет вид

$$n\vartheta_1^* = \sum_{i=1}^n X_i, \quad \vartheta_1^* = \frac{1}{n} \sum_{i=1}^n X_i = m^*,$$

а значение S_2^2 задается формулой

$$S_2^2 = \sum_{i=1}^n (X_i - m^*)^2.$$

Поэтому

$$s_1^{2*} = \frac{S_2^2 - (n-2)s_0^{2*}}{2-1},$$

и, значит, мы должны принять гипотезу H_0 , если $\varkappa = s_1^{2*}/s_0^{2*} < \varphi_{1-\alpha}$, где φ_α — α -квантиль F -распределения с параметрами 1 и $n-2$.

2°. Обратимся к решению задачи 3. Оценка $x^*(t)$ регрессии $x(t)$ имеет вид

$$x^*(t) = \vartheta_1^* + \vartheta_2^* t.$$

Далее, коэффициент $c^2(t)$ определяется формулой

$$c^2(t) = \sum_{j=1}^n (c_{1j} + c_{2j}t)^2 = \sum_{j=1}^n c_{1j}^2 + 2t \sum_{j=1}^n c_{1j}c_{2j} + t^2 \sum_{j=1}^n c_{2j}^2 = b_0 + b_1t + b_2t^2,$$

где c_{1j} и c_{2j} найдены в примере 8. Таким образом, границы симметричного доверительного интервала доверительной вероятности α для линейной регрессии задаются выражениями:

$$\begin{aligned} x'(t) &= \vartheta_1^* + \vartheta_2^* t - t \frac{1+\alpha}{2} \sqrt{s_0^{2*}(b_0 + b_1t + b_2t^2)}, \\ x''(t) &= \vartheta_1^* + \vartheta_2^* t + t \frac{1+\alpha}{2} \sqrt{s_0^{2*}(b_0 + b_1t + b_2t^2)}, \end{aligned}$$

где t_α — α -квантиль t -распределения с $n-2$ степенями свободы. Отметим еще один факт, который нетрудно установить простейшими вычислениями: минимальная ширина доверительного интервала (т. е. наиболее точная оценка регрессии $x(t)$) получается при t , равном среднему арифметическому $\sum_{i=1}^n t_i/n$ всех моментов наблюдений.

3°. Применим полученные результаты к анализу регрессии веса X_2 на рост X_1 мужчины (см. пример 3). Как следует из примера 9, в силу совместной нормальности X_1 и X_2 в данном случае можно использовать линейную регрессионную модель. В качестве независимой переменной выберем рост.

Начнем с задачи оценки неизвестных параметров ϑ_1 и ϑ_2 линейной регрессии и построения для них доверительных интервалов. Используя результаты примера 8, находим

$$\vartheta_1^* = -31,48, \quad \vartheta_2^* = 0,6138.$$

Доверительные интервалы для ϑ_1 и ϑ_2 будем строить при доверительной вероятности $\alpha = 0,9$. В [1] на с. 178 находим значение $t_{0,95} = 1,6772$ 0,95-квантили t -распределения с 48 степенями свободы. Вычисления значений статистики s_0^{2*} и коэффициентов c_1^2 и c_2^2 (см. пример 8) дают:

$$s_0^{2*} = 23,69, \quad c_1^2 = 12,97, \quad c_2^2 = 0,000\,441\,2.$$

Поэтому доверительные границы для ϑ_1 и ϑ_2 имеют вид

$$\vartheta_1' = -60,88, \quad \vartheta_1'' = -2,08, \quad \vartheta_2' = 0,4423, \quad \vartheta_2'' = 0,7853.$$

Проверим гипотезу H_0 о независимости веса от роста, которую в рамках регрессионной модели сформулируем как гипотезу о равенстве нулю параметра ϑ_2 . Уровень значимости (размер) критерия α положим равным 0,05. Используя известное из примера 3 значение среднего веса $m^* = 73,66$, находим значения статистик

$$S_2^2 = 1991,0, \quad s_1^{2*} = 853,9$$

и вычисляем отношение

$$\kappa = \frac{s_1^{2*}}{s_0^{2*}} = 36,0.$$

Сравнивая полученное значение κ с 0,95-квантилью $\varphi_{0,95} = 4,05$ ([1, с. 208]; поскольку значения $\varphi_{0,95}$ для $n_2 = 48$ в таблице нет, воспользуемся линейной интерполяцией) F -распределения с параметрами 1 и 48, убеждаемся, что гипотеза H_0 должна быть отвергнута.

Наконец, найдем оценку $x^*(t)$ веса $x(t)$ при известном росте t и построим симметричный доверительный интервал для $x(t)$ доверительной вероятности $\alpha = 0,9$. Оценка $x^*(t)$ имеет вид

$$x^*(t) = -31,48 + 0,6138 t.$$

Вычисления дают следующие значения коэффициентов:

$$b_0 = 12,9606, \quad b_1 = -0,1511, \quad b_2 = 0,01045.$$

В [1] на с. 178 находим 0,95-квантиль $t_{0,95} = 1,6772$ t -распределения с 48 степенями свободы. Поэтому доверительные границы задаются формулами:

$$x'(t) = -31,48 + 0,6138 t - 1,6772 \sqrt{307,04 - 3,5796 t + 0,01045 t^2},$$

$$x''(t) = -31,48 + 0,6138 t + 1,6772 \sqrt{307,04 - 3,5796 t + 0,01045 t^2}.$$

Графики кривых $x^*(t)$, $x'(t)$ и $x''(t)$ приведены на рис. 1.

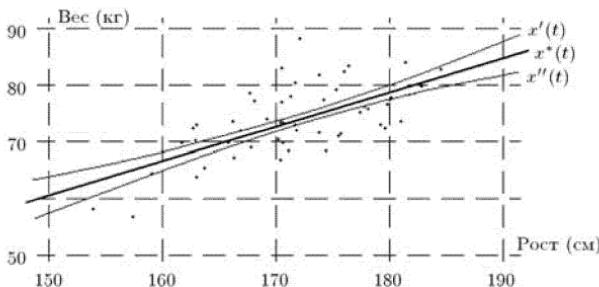


Рис. 1

4°. Интересно сравнить результаты корреляционного (см. пример 3) и регрессионного анализа. Оба метода отвергают гипотезу H_0 о независимости веса от роста. Однако в споре, какой из двух критериев лучше: критерий, основанный на выборочном коэффициенте корреляции, или критерий, основанный на регрессионной модели, — пальму первенства нужно отдать первому, как равномерно наиболее мощному; это естественно, поскольку он использует дополнительную информацию о нормальности независимой переменной X_1 — роста индивидуума. Далее, воспользовавшись результатом примера 9 о функциональной зависимости $\vartheta_2 = \rho \sqrt{\sigma_1^2 / \sigma_2^2}$ коэффициента ϑ_2 линейной регрессии от коэффициента корреляции ρ и дисперсий σ_1^2 и σ_2^2 и подставляя вместо ρ ,

σ_1^2 и σ_2^2 их оценки $\rho^* = 0,655$, $s_1^{2*} = 40,63$ и $s_2^{2*} = 46,26$ (см. пример 3), получаем оценку $\vartheta_2^* = 0,614$ параметра ϑ_2 , которая, как мы видим, совпадает с оценкой ϑ_2^* , полученной в регрессионной модели (это можно установить и прямыми вычислениями). Однако с доверительными границами для ϑ_2 и ρ такая манипуляция не проходит; этому опять-таки мешает дополнительная информация о нормальности независимой переменной в регрессионной модели. \square

В заключение этого параграфа вкратце остановимся на проблеме выбора функций $f_i(t)$. Разумеется, их выбор целиком лежит на совести исследователя и определяется физическим анализом изучаемого явления. Обычно хороший результат дает линейная регрессия

$$x(t) = \vartheta_1 + \vartheta_2 t,$$

рассматривавшаяся в примерах. Реже используется квадратичная регрессия

$$x(t) = \vartheta_1 + \vartheta_2 t + \vartheta_3 t^2,$$

еще реже — полиномиальная регрессия более высоких степеней. В ряде явлений, таких, как рост популяций, радиоактивный распад веществ и т. п., имеет смысл применять экспоненциальную регрессию вида

$$x(t) = \vartheta_1 + \vartheta_2 e^{\gamma t}.$$

Часто в этом случае считают, что и дисперсия ошибок измерения $\varepsilon(t)$ зависит от времени: $D\varepsilon(t) = \sigma^2 b(t)$, где $b(t)$ — известная функция, и применяют модифицированный метод наименьших квадратов, в котором сумма квадратов берется с весовыми коэффициентами. Наконец, при исследовании закономерностей, носящих ярко выраженный периодический характер (изменчивость в течение суток, недели, года), естественно обратиться к периодической регрессии, в которой используются разложения в ряд Фурье. Впрочем, как уже знает читатель, решение задачи проверки статистических гипотез (задачи 2) в регрессионном анализе позволяет подобрать адекватную модель и избавиться от «лишних» факторов.

7. Дисперсионный анализ

Задачи, которые рассматриваются в этом и следующем параграфах, дают некоторое представление о разделе математической статистики, называемом *дисперсионным анализом* и основанном на сравнении выборочных дисперсий.

Начнем с *простейшей задачи*. Пусть имеется l независимых выборок

$$\begin{aligned} X_{11}, \dots, X_{1n_1}; \\ \vdots \\ X_{l1}, \dots, X_{ln_l}, \end{aligned}$$

произведенных из нормальных генеральных совокупностей с неизвестными средними m_1, \dots, m_l и также неизвестными, но одинаковыми дисперсиями σ^2 . Проверяются две сложные параметрические гипотезы: основная H_0 : $m_1 = \dots = m_l$, состоящая в том, что все теоретические средние m_i равны между собой, и конкурирующая H_1 : некоторые теоретические средние m_i могут быть различными.

Построение критерия для проверки гипотез H_0 и H_1 начнем с рассмотрения отдельных внутригрупповых выборочных дисперсий

$$s_{(i)}^{2*} = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (X_{ij} - m_i^*)^2,$$

где

$$m_i^* = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}$$

— выборочное среднее выборки X_{i1}, \dots, X_{in_i} . Как известно (см. параграф 4 гл. 1), случайная величина $(n_i - 1)s_{(i)}^{2*}/\sigma^2$ имеет χ^2 -распределение с $n_i - 1$ степенями свободы. Определим статистику

$$s_0^{2*} = \frac{\sum_{i=1}^l (n_i - 1)s_{(i)}^{2*}}{\sum_{i=1}^l (n_i - 1)},$$

называемую (общей) внутригрупповой или остаточной (выборочной) дисперсией. Поскольку выборки независимы, случайная величина $s_0^{2*} \sum_{i=1}^l (n_i - 1)/\sigma^2$ также распределена по закону χ^2 , но с числом степеней свободы $\sum_{i=1}^l (n_i - 1)$, а статистика s_0^{2*} представляет собой несмешанную оценку неизвестной дисперсии σ^2 . Обозначая через

$$m^* = \frac{\sum_{i=1}^l m_i^* n_i}{\sum_{i=1}^l n_i} = \frac{\sum_{i=1}^l \sum_{j=1}^{n_i} X_{ij}}{\sum_{i=1}^l n_i}$$

общее выборочное среднее обобщенной выборки, образуем новую статистику

$$s_1^{2*} = \frac{1}{l-1} \sum_{i=1}^l (m_i^* - m^*)^2 n_i,$$

которая представляет собой межгрупповую выборочную дисперсию и не зависит от s_0^{2*} . Кроме того, при условии справедливости гипотезы H_0 случайная величина $s_1^{2*}(l-1)/\sigma^2$ имеет χ^2 -распределение с $l-1$

степенями свободы, а статистика s_1^{2*} является еще одной несмещенной оценкой дисперсии σ^2 . Однако если гипотеза H_0 не верна, то s_1^{2*} будет иметь положительное смещение $\delta = Ms_1^{2*} - \sigma^2$ (т. е. иметь нецентральное χ^2 -распределение), тем большее, чем больше расхождение между теоретическими средними m_1, \dots, m_l . Поэтому естественно применить для проверки гипотезы H_0 односторонний критерий Фишера, предписывающий принять гипотезу H_0 , если $\varkappa = s_1^{2*}/s_0^{2*} < C$. Статистика \varkappa при условии справедливости гипотезы H_0 , как уже не раз говорилось, имеет F -распределение, а значит, при заданном уровне значимости α критическое значение C совпадает с $(1 - \alpha)$ -квантилью $\varphi_{1-\alpha}$ F -распределения с параметрами $l - 1$ и $\sum_{i=1}^l (n_i - 1)$ [1, табл. 3.5].

Можно показать, что критерий Фишера является равномерно наименее мощным несмещенным критерием для проверки гипотез H_0 и H_1 .

Практическая реализация критерия происходит следующим образом. Сначала вычисляют внутригрупповые выборочные средние m_i^* и общее выборочное среднее m^* . Затем находят внутригрупповую и межгрупповую выборочные дисперсии s_0^{2*} и s_1^{2*} . Наконец, определяют значение статистики \varkappa и сравнивают его с соответствующей квантилью F -распределения; в зависимости от результатов сравнения либо принимают гипотезу H_0 , либо отвергают ее. Иногда полезно бывает ввести еще одну статистику

$$s^{2*} = \frac{\sum_{i=1}^l \sum_{j=1}^{n_i} (X_{ij} - m^*)^2}{\sum_{i=1}^l n_i - 1},$$

являющуюся выборочной дисперсией объединенной выборки. Тогда справедливо соотношение

$$s^{2*} \left(\sum_{i=1}^l n_i - 1 \right) = s_0^{2*} \sum_{i=1}^l (n_i - 1) + s_1^{2*} (l - 1).$$

Пример 11. Для определения процентного содержания вредных примесей в минерале были взяты образцы одинаковой массы из трех различных месторождений: 3 образца из первого месторождения, 2 из второго и 4 из третьего. Результаты химического анализа (процент содержания вредных примесей) приведены в табл. 4–6. Считая процентное содержание примесей в каждом образце распределенным по нормальному закону с одинаковой дисперсией, проверим при уровне значимости $\alpha = 0,05$ гипотезу H_0 о том, что среднее содержание примесей во всех трех месторождениях одинаково.

Таблица 4

X_{11}	X_{12}	X_{13}
8,35	5,40	7,16

Таблица 5

X_{21}	X_{22}
4,52	6,24

Таблица 6

X_{31}	X_{32}	X_{33}	X_{34}
8,91	7,47	9,08	9,94

Воспользуемся критерием Фишера. Вычислим внутригрупповые выборочные средние

$$m_1^* = \frac{1}{3} (8,35 + 5,40 + 7,16) = 6,97,$$

$$m_2^* = \frac{1}{2} (4,52 + 6,24) = 5,38,$$

$$m_3^* = \frac{1}{4} (8,91 + 7,47 + 9,08 + 9,94) = 8,85$$

и общее выборочное среднее

$$m^* = \frac{1}{9} (8,35 + 5,40 + \dots + 9,94) = 7,45.$$

Найдем внутригрупповую выборочную дисперсию

$$\begin{aligned} s_0^{2*} = \frac{1}{2+1+3} & [((8,35 - 6,97)^2 + (5,40 - 6,97)^2 + (7,16 - 6,97)^2) + \\ & + ((4,52 - 5,38)^2 + (6,24 - 5,38)^2) + \\ & + ((8,91 - 8,85)^2 + (7,47 - 8,85)^2 + (9,08 - 8,85)^2 + (9,94 - 8,85)^2)] = 1,50 \end{aligned}$$

и межгрупповую выборочную дисперсию

$$s_1^{2*} = \frac{1}{3-1} [3(6,97 - 7,45)^2 + 2(5,38 - 7,45)^2 + 4(8,85 - 7,45)^2] = 8,54.$$

Значение статистики $\chi = 8,54/1,50 = 5,69$. Параметры F -распределения, используемого для нахождения критического значения C , равны $3-1=2$ и $(3-1)+(2-1)+(4-1)=6$. В [1, с. 208] отыскиваем 0,95-квантиль $\varphi_{0,95} = 5,1433$ F -распределения с параметрами 2 и 6, совпадающую с критическим значением C критерия Фишера. Поскольку $\chi > \varphi_{0,95}$, есть серьезные основания отвергнуть гипотезу H_0 о равенстве среднего содержания примесей в минералах всех трех месторождений. \square

Прежде чем переходить к дальнейшему изложению, придадим рассмотренной задаче несколько иную формулировку. Будем считать, что имеется некоторый фактор, который может действовать на l уровнях. При этом действие на анализируемое явление фактора на i -м уровне определяется математическим ожиданием m_i . Такая модель в дисперсионном анализе называется однофакторной l -уровневой (предостережение: не надо путать с факторной моделью, рассмотренной в параграфе 4 гл. 3, где сам фактор являлся случайным и мог появляться с определенной вероятностью в каждом наблюдении на каком-то одном из уровней). Задача заключалась в том, чтобы проверить гипотезу H_0 об одинаковом действии фактора на всех уровнях или, по-другому, об отсутствии влияния на результат эксперимента уровня действия фактора.

Рассмотрим теперь случай действия *двух факторов*, причем первый фактор может действовать на l_1 уровнях, а второй — на l_2 уровнях (двуухфакторная (l_1, l_2) -уровневая модель дисперсионного анализа).

Предположим, что произведено n_{ij} наблюдений, в которых первый фактор действовал на i -м уровне, а второй — на j -м. Результаты наблюдений X_{ijk} представляют собой независимые нормально распределенные случайные величины с одинаковой (неизвестной) дисперсией σ^2 и, вообще говоря, с различными (также неизвестными) средними m_{ij} . Будем опять-таки проверять гипотезу H_0 об отсутствии влияния (уровней действия) первого и второго факторов на результаты наблюдений, т. е. о равенстве всех m_{ij} .

Поскольку действие двух факторов на уровнях l_1 и l_2 соответственно можно отождествить с действием одного фактора на $l_1 l_2$ уровнях, то для проверки гипотезы H_0 воспользуемся однофакторной $l_1 l_2$ -уровневой моделью и в соответствии с построенным для нее критерием вычислим значения статистик s_0^{2*} и s_1^{2*} :

$$s_0^{2*} = \frac{\sum_{i=1}^{l_1} \sum_{j=1}^{l_2} \sum_{k=1}^{n_{ij}} (X_{ijk} - m_{ij}^*)^2}{\sum_{i=1}^{l_1} \sum_{j=1}^{l_2} (n_{ij} - 1)},$$

$$s_1^{2*} = \frac{1}{l_1 l_2 - 1} \sum_{i=1}^{l_1} \sum_{j=1}^{l_2} n_{ij} (m_{ij}^* - m^*)^2,$$

где

$$m_{ij}^* = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} X_{ijk}$$

— среднее по наблюдениям, в которых первый и второй факторы действовали на уровнях i и j ,

$$m^* = \frac{\sum_{i=1}^{l_1} \sum_{j=1}^{l_2} \sum_{k=1}^{n_{ij}} X_{ijk}}{\sum_{i=1}^{l_1} \sum_{j=1}^{l_2} n_{ij}}$$

— общее среднее всех наблюдений. Теперь, если $\varkappa = s_1^{2*}/s_0^{2*} < \varphi_{1-\alpha}$, где φ_α — α -квантиль F -распределения с параметрами $\sum_{i=1}^{l_1} \sum_{j=1}^{l_2} (n_{ij} - 1)$ и $l_1 l_2 - 1$, то мы должны с уровнем значимости α принять гипотезу H_0 о полном отсутствии влияния уровней действия обоих факторов.

Однако можно произвести более детальный анализ модели. Будем предполагать, что выполнено условие $n_{ij} = n$ (общий случай различных n_{ij} исследуется несколько более сложным образом). Введем

статистики

$$\begin{aligned} m_{i(1)}^* &= \frac{1}{nl_2} \sum_{j=1}^{l_2} \sum_{k=1}^n X_{ijk}, & m_{j(2)}^* &= \frac{1}{nl_1} \sum_{i=1}^{l_1} \sum_{k=1}^n X_{ijk}, \\ s_{(1)}^{2*} &= \frac{nl_2}{l_1 - 1} \sum_{i=1}^{l_1} (m_{i(1)}^* - m^*)^2, & s_{(2)}^{2*} &= \frac{nl_1}{l_2 - 1} \sum_{j=1}^{l_2} (m_{j(2)}^* - m^*)^2, \\ s_{(12)}^{2*} &= \frac{n}{(l_1 - 1)(l_2 - 1)} \sum_{i=1}^{l_1} \sum_{j=1}^{l_2} (m_{ij}^* - m_{i(1)}^* - m_{j(2)}^* + m^*)^2. \end{aligned}$$

Эти статистики имеют весьма простой смысл: $m_{i(1)}^*$ ($m_{j(2)}^*$) — выборочное среднее при условии, что первый (второй) фактор подействовал на уровне i (j); $s_{(1)}^{2*}$ ($s_{(2)}^{2*}$) — выборочная дисперсия, связанная с действием первого (второго) фактора; $s_{(12)}^{2*}$ — выборочная дисперсия взаимодействия факторов. Статистики $s_{(1)}^{2*}$, $s_{(2)}^{2*}$ и $s_{(12)}^{2*}$ являются независимыми.

Следующая гипотеза, которую мы рассмотрим, — гипотеза H'_0 о линейности действия факторов, т. е. о возможности представления теоретического среднего в виде $m_{ij} = m_i^{(1)} + m_j^{(2)}$, где $m_i^{(1)}$ ($m_j^{(2)}$) — неизвестное среднее значение, вносимое первым (вторым) фактором при действии на уровне i (j). Естественно, гипотеза H'_0 верна далеко не всегда, поскольку довольно часто существенное влияние на результат эксперимента факторы оказывают только при совместном их действии на определенных уровнях (например, урожайность сельскохозяйственной культуры высока только при внесении необходимых удобрений в предписанных количествах; при любых отклонениях от агротехнических норм эффект от внесения удобрений существенно уменьшается). Оказывается, при условии справедливости линейной гипотезы H'_0 дисперсия взаимодействия $s_{(12)}^{2*}$ также является несмещенной оценкой σ^2 , с точностью до множителя $(l_1 - 1)(l_2 - 1)/\sigma^2$ распределенной по закону χ^2 с $(l_1 - 1)(l_2 - 1)$ степенями свободы; в противном случае эта оценка имеет положительное смещение. Поэтому критерий принятия линейной гипотезы H'_0 при заданном уровне значимости α заключается в выполнении неравенства $\varkappa = s_{(12)}^{2*}/s_0^{2*} < \varphi_{1-\alpha}$, где φ_α — α -квантиль F -распределения с параметрами $(l_1 - 1)(l_2 - 1)$ и $l_1 l_2 (n - 1)$.

Если линейная гипотеза считается выполненной (по результатам проверки или в силу априорных предположений), то можно проверять гипотезы $H_0^{(1)}$ и $H_0^{(2)}$ об отсутствии действия первого или второго факторов. В частности, если справедлива гипотеза $H_0^{(1)}$: $m_1^{(1)} = \dots = m_{l_1}^{(1)}$, то статистика $s_{(1)}^{2*}$ опять-таки будет являться несмещенной оценкой σ^2 , с точностью до множителя $(l_1 - 1)/\sigma^2$ имеющей χ^2 -распределение с $l_1 - 1$ степенями свободы. Значит, мы можем срав-

нить ее или с остаточной дисперсией s_0^{2*} , или с дисперсией взаимодействия $s_{(12)}^2$, или, еще лучше, с объединенной дисперсией

$$s_{\text{об}}^{2*} = \frac{l_1 l_2 (n - 1) s_0^{2*} + (l_1 - 1)(l_2 - 1) s_{(12)}^{2*}}{l_1 l_2 n - l_1 - l_2 + 1},$$

образованной остаточной дисперсией и дисперсией взаимодействия, и воспользоваться критерием Фишера. В частности, если мы пользуемся объединенной дисперсией $s_{\text{об}}^{2*}$, то должны принять гипотезу $H_0^{(1)}$ при выполнении неравенства $\varkappa = s_{(1)}^{2*}/s_{\text{об}}^{2*} < \varphi_{1-\alpha}$, где, как обычно, φ_α — α -квантиль F -распределения с параметрами $l_1 - 1$ и $l_1 l_2 n - l_1 - l_2 + 1$. Аналогично проверяется гипотеза $H_0^{(2)}$: $m_1^{(2)} = \dots = m_{l_2}^{(2)}$.

Отметим, что проверку гипотезы H_0 о (полном) отсутствии действия обоих факторов можно производить только в том случае, когда хотя бы одно n_{ij} больше единицы; в противном случае мы не смогли бы определить остаточную дисперсию s_0^{2*} и не имели бы «эталонной» оценки дисперсии σ^2 . Если же все n_{ij} равны единице, то возможны два пути: либо потребовать, чтобы была известна теоретическая дисперсия σ^2 наблюдений X_{ij} , либо априори считать справедливой линейную гипотезу, тогда «эталонной» оценкой дисперсии σ^2 будет дисперсия взаимодействия $s_{(12)}^{2*}$.

Совершенно аналогично рассматривается случай действия *трех и более факторов*. Однако здесь появляется новое осложнение: если имеется k факторов, причем i -й фактор действует на l_i уровнях, то для того, чтобы произвести хотя бы по одному наблюдению со всеми возможными комбинациями уровней действия факторов, необходимо в общей сложности $l_1 \dots l_k$ наблюдений. В частности, при четырех факторах и пяти уровнях действия каждого из них необходимо 625 наблюдений. Ясно, что на практике такое количество наблюдений обычно нереально. Поэтому следует так спланировать эксперимент, чтобы, с одной стороны, число наблюдений было разумным, а с другой стороны, мы бы получили достаточно аргументированные статистические выводы. Возможные методы решения этой задачи мы рассмотрим в следующем параграфе, сейчас же еще раз заметим, что в общей модели нельзя уменьшить число наблюдений с сохранением достаточной мощности критерия, поскольку влияние факторов может проявиться только при определенной комбинации уровней их действия, а так как эта комбинация нам заранее не известна, необходимо испробовать все возможные комбинации. Значит, нужно отказаться от общей постановки задачи. В дальнейшем мы будем рассматривать только линейную модель, т. е. модель, в которой $m_{i_1, \dots, i_l} = m_{i_1}^{(1)} + \dots + m_{i_l}^{(l)}$.

8. Планирование эксперимента

Прежде чем переходить к рассмотрению задач планирования эксперимента, сделаем замечание по поводу использования самого термина «планирование эксперимента». Дело в том, что разные специалисты под ним понимают различные проблемы вплоть до того, что некоторые авторы называют «планированием эксперимента» фактически всю прикладную математику, усердно сдобренную практическими рекомендациями по применению того или иного математического метода. Мы будем придерживаться традиционного определения «планирования эксперимента» как части дисперсионного анализа.

Напомним общую постановку задачи. Пусть имеется l факторов, причем i -й фактор может действовать на любом из n_i уровней, а выбор уровня действия фактора находится целиком в руках экспериментатора. Действие всех l факторов на уровнях j_1, \dots, j_l приводит к тому, что результат наблюдения является нормально распределенной случайной величиной с неизвестным средним m_{j_1, \dots, j_l} и неизвестной дисперсией σ^2 , не зависящей от уровней действия факторов. Всюду далее будем предполагать, что *факторы действуют линейно* (часто говорят независимо), т. е.

$$m_{j_1, \dots, j_l} = m_{j_1}^{(1)} + \dots + m_{j_l}^{(l)}.$$

Требуется так спланировать эксперимент, т. е. указать все возможные комбинации (j_1, \dots, j_l) уровней действия факторов, для которых нужно произвести наблюдения, чтобы, с одной стороны, количество наблюдений не было бы слишком большим, а, с другой стороны, результаты эксперимента достаточно хорошо проверяли бы гипотезы об отсутствии действия каждого фактора, т. е. гипотезы $H_0^{(i)}$: $m_1^{(i)} = \dots = m_{n_i}^{(i)}$. Естественно, *планом эксперимента* будем называть сам перечень всех наборов (j_1, \dots, j_l) уровней действия факторов, для которых необходимо произвести наблюдения. При этом будем предполагать, что для каждого набора (j_1, \dots, j_l) , входящего в план эксперимента, производится только одно наблюдение X_{j_l} . Отметим, что план эксперимента, состоящий из всех возможных комбинаций (j_1, \dots, j_l) уровней действия факторов и рассмотренный в параграфе 7, называют *(полным) факторным планом*. Однако, как уже говорилось, факторный план требует слишком большого количества наблюдений.

Разумным представляется потребовать от плана эксперимента, чтобы результаты эксперимента можно было бы достаточно просто обработать. С этим мы уже столкнулись в параграфе 7, когда в двухфакторную модель пришлось ввести условие одинакового числа наблюдений при каждом наборе уровней действия обоих факторов. Ортогональность планов или сбалансированность блоков, которые будут использоваться далее, не только упрощают расчеты, но и приводят к независимости статистик, служащих для проверки отсутствия действия

отдельных факторов, что в свою очередь избавляет от неприятной ситуации, когда принятие или отклонение гипотезы об отсутствии действия одного фактора зависит от аналогичного решения относительно остальных факторов. Кроме того, такие планы обладают некоторыми специальными свойствами оптимальности, которых мы здесь касаться не будем.

Рассмотрим способы построения планов эксперимента. Начнем со случая, когда *число уровней действия факторов одинаково* для всех факторов, т. е. $n_1 = \dots = n_l = n$. Потребуем, чтобы для любых факторов i_1 и i_2 и уровней их действия j_{i_1} и j_{i_2} в план эксперимента входил, по крайней мере, один набор, содержащий пару j_{i_1} и j_{i_2} , т. е. хотя бы в одном наблюдении фактор i_1 подействовал на уровне j_{i_1} , а фактор i_2 — на уровне j_{i_2} . С другой стороны, для минимизации числа наблюдений резонно ограничиться только планами, в которых каждая пара (j_{i_1}, j_{i_2}) уровней действия факторов i_1 и i_2 встречается не более одного раза. Суммируя эти два требования, получаем, что каждая пара (j_{i_1}, j_{i_2}) уровней действия факторов i_1 и i_2 должна входить в план эксперимента ровно один раз. Такие планы назовем ортогональными. Нетрудно видеть, что любой ортогональный план содержит ровно n^2 наблюдений и его можно представить в виде квадратной таблицы (табл. 7), в которой $j_i^{(k_1 k_2)}$ ($i = 3, \dots, l$) — числа от 1 до n . Этот план заключается в следующем. В первом наблюдении первый и второй факторы действуют на уровне 1, третий — на уровне $j_3^{(11)}, \dots, l$ -й — на уровне $j_l^{(11)}, \dots, l$ -м наблюдении первый фактор действует на уровне n , второй — на уровне 1, третий — на уровне $j_3^{(n1)}, \dots, l$ -й — на уровне $j_l^{(n1)}, \dots, l$ -м наблюдении первый и второй факторы действуют на уровне n , третий — на уровне $j_3^{(nn)}, \dots, l$ -й — на уровне $j_l^{(nn)}$. В силу требования ортогональности плана в каждой строке и в каждом столбце табл. 7 на любом месте каждое из чисел $1, \dots, n$ должно встречаться ровно один раз; для любых чисел k_1 и k_2 от 1 до n и i_1 и i_2 от 3 до l ровно один раз должны встречаться клетки, в которых на i_1 -м месте стоит k_1 , а на i_2 -м — k_2 . Пример ортогонального плана при числе действующих факторов $l = 6$ и числе уровней действия каждого фактора $n = 5$ приведен в табл. 8 (в этой и во всех дальнейших таблицах для сокращения записи первый столбец и первая строка опущены).

Обратимся к конкретным значениям числа факторов l .

При $l = 2$ ортогональный план представляет собой не что иное, как факторный план, рассмотренный в параграфе 7, в котором мы должны произвести по одному наблюдению с каждой комбинацией (j_1, j_2) уровней действия первого и второго факторов.

При $l = 3$ любой ортогональный план задается так называемым латинским квадратом n -го порядка, т. е. квадратной таблицей размера

Таблица 7

Уровень действия второго фактора	Уровень действия первого фактора		
	1	...	n
1	$j_3^{(11)} \dots j_l^{(11)}$...	$j_3^{(n1)} \dots j_l^{(n1)}$
...
n	$j_3^{(1n)} \dots j_l^{(1n)}$...	$j_3^{(nn)} \dots j_l^{(nn)}$

Таблица 8

1111	2222	3333	4444	5555
2345	3451	4512	5123	1234
3524	4135	5241	1352	2413
4253	5314	1425	2531	3142
5432	1543	2154	3215	4321

$n \times n$, в которую вписаны числа от 1 до n таким образом, чтобы в каждой строке и каждом столбце любое из этих чисел встречалось ровно один раз. Пример латинского квадрата, полученного циклическим сдвигом первой строки на единицу, приведен в табл. 9.

Критерий для проверки отсутствия действия факторов строится следующим образом. Пусть в результате эксперимента по предложенному плану мы получили наблюдения

$$X_{11}, \dots, X_{n1},$$

. . . .

$$X_{1n}, \dots, X_{nn}.$$

Обозначим

$$m^* = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n X_{ij}, \quad S^2 = \sum_{i=1}^n \sum_{j=1}^n (X_{ij} - m^*)^2.$$

Статистика S^2 представляет собой сумму квадратов отклонений наблюдений от общего выборочного среднего m^* . Используя свойство латинского квадрата, ее можно переписать в виде

1	2	3	4	5
2	3	4	5	1
3	4	5	1	2
4	5	1	2	3
5	1	2	3	4

$$\begin{aligned}
S^2 &= \sum_{i=1}^n \sum_{j=1}^n (X_{ij} - m_{i(1)}^* - m_{j(2)}^* - m_{k(3)}^* + 2m^*)^2 + \\
&+ \sum_{i=1}^n n(m_{i(1)}^* - m^*)^2 + \sum_{j=1}^n n(m_{j(2)}^* - m^*)^2 + \sum_{k=1}^n n(m_{k(3)}^* - m^*)^2 = \\
&= (n^2 - 3n + 2)s_{(123)}^{2*} + (n - 1)s_{(1)}^{2*} + (n - 1)s_{(2)}^{2*} + (n - 1)s_{(3)}^{2*},
\end{aligned}$$

где

$$m_{i(1)}^* = \frac{1}{n} \sum_{j=1}^n X_{ij}, \quad m_{j(2)}^* = \frac{1}{n} \sum_{i=1}^n X_{ij}, \quad m_{k(3)}^* = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n X_{ij}.$$

Статистики $m_{i(1)}^*$ и $m_{j(2)}^*$ являются выборочными средними при условии действия первого или второго фактора на уровне i или j . Такой же смысл имеет и статистика $m_{k(3)}^*$, поэтому при определении $m_{k(3)}^*$ суммирование должно производиться только по тем наблюдениям, в которых третий фактор действовал на k -м уровне или, в терминах латинского квадрата, по тем клеткам, в которых стоит цифра k . Заметим теперь, что в силу линейности действия факторов (которую мы предположили с самого начала) статистики $s_{(123)}^{2*}$, $s_{(1)}^{2*}$, $s_{(2)}^{2*}$ и $s_{(3)}^{2*}$ независимы, а $s_{(123)}^{2*}$ является несмещенной оценкой дисперсии σ^2 , с точностью до множителя $(n^2 - 3n + 2)/\sigma^2$ имеющей χ^2 -распределение с $n^2 - 3n + 2$ степенями свободы. Теперь, если отсутствует влияние первого фактора, то статистика $s_{(1)}^{2*}$ с точностью до множителя $(n - 1)/\sigma^2$ также распределена по закону χ^2 с $n - 1$ степенями свободы и, значит, для проверки гипотезы об отсутствии влияния первого фактора можно применить критерий Фишера, предписывающий принять эту гипотезу, если $\chi_{(1)} = s_{(1)}^{2*}/s_{(123)}^{2*} < C$. Аналогично, по статистикам $\chi_{(2)} = s_{(2)}^{2*}/s_{(123)}^{2*}$ и $\chi_{(3)} = s_{(3)}^{2*}/s_{(123)}^{2*}$ проверяются гипотезы об отсутствии действия второго и третьего факторов.

Пусть теперь $l = 4$. Назовем два латинских квадрата ортогональными, если при наложении их друг на друга каждый набор (ij) встретится ровно один раз. Пример наложения двух латинских квадратов пятого порядка приведен в табл. 10. Таким образом, задача построения ортогонального плана сводится к задаче нахождения двух ортогональных латинских квадратов. Однако ортогональные латинские квадраты существуют не для всех n . В частности, они, как известно, не существуют при $n = 6$ (задача Эйлера о 36 офицерах). Если же ортогональные латинские квадраты существуют, то ортогональный план задается их наложением. Сама процедура проверки гипотез об отсутствии действия факторов остается той же, что и в случае $l = 3$, и опирается на представление статистики S^2 в виде

$$S^2 = (n^2 - 4n + 3)s_{(1234)}^{2*} + (n - 1)s_{(1)}^{2*} +$$

Таблица 10

1	2	3	4	5
2	3	4	5	1
3	4	5	1	2
4	5	1	2	3
5	1	2	3	4

1	2	3	4	5
3	4	5	1	2
5	1	2	3	4
2	3	4	5	1
4	5	1	2	3

11	22	33	44	55
23	34	45	51	12
35	41	52	13	24
42	53	14	25	31
54	15	21	32	43

$$+ (n-1)s_{(2)}^{2*} + (n-1)s_{(3)}^{2*} + (n-1)s_{(4)}^{2*}.$$

Нетрудно теперь понять, что возможность построения ортогонального плана эксперимента при $l > 4$ сводится к возможности построения $l - 2$ ортогональных латинских квадратов. Оказывается, в любом случае нельзя построить более $n - 1$ ортогональных латинских квадратов. Система, содержащая максимальное для заданного n число ортогональных латинских квадратов, называется полной. Ясно, что любая система из $n - 1$ ортогональных латинских квадратов является полной.

Покажем, как можно построить полную систему в случае *простого* n . Первая строка всех $n - 1$ латинских квадратов состоит из записанных подряд чисел $1, \dots, n$. Вторая строка первого латинского квадрата получается циклическим сдвигом первой строки на единицу влево, третья — циклическим сдвигом второй строки на единицу влево и т. д. Вторая строка второго латинского квадрата получается циклическим сдвигом первой строки влево на два, третья — циклическим сдвигом второй строки влево на два и т. д. Полная система из четырех латинских квадратов пятого порядка приведена в табл. 11.

Таблица 11

1	2	3	4	5
2	3	4	5	1
3	4	5	1	2
4	5	1	2	3
5	1	2	3	4

1	2	3	4	5
3	4	5	1	2
5	1	2	3	4
2	3	4	5	1
4	5	1	2	3

1	2	3	4	5
4	5	1	2	3
2	3	4	5	1
5	1	2	3	4
3	4	5	1	2

1	2	3	4	5
5	1	2	3	4
4	5	1	2	3
3	4	5	1	2
2	3	4	5	1

Для читателя, знакомого с комплексными числами, опишем рассмотренный выше метод построения ортогональных латинских квадратов в терминах корней n -й степени из единицы. Пусть $z_1 = \sqrt[n]{1}$ — какой-либо (отличный от 1) корень n -й степени из 1. Положим $z_2 = z_1^2, \dots, z_n = z_1^n = 1$. Отождествим число i с z_i . Тогда первая строка любого квадрата отождествима с z_1, \dots, z_n . Каждая последующая строка первого латинского квадрата получается из предыдущей

умножением на z_1 , второго — умножением на z_2 и т. д. Таким образом, существует тесная связь приведенной процедуры с корнями n -й степени из единицы, которые в свою очередь являются одним из основополагающих понятий теории Галуа. Оказывается, использование теории Галуа позволяет построить полную систему из $n - 1$ латинских квадратов и в том случае, когда $n = p^l$, где p — простое число. Кроме того, с ее помощью можно построить k ортогональных латинских квадратов в общем случае $n = p_1^{l_1} \dots p_q^{l_q}$, причем k представляет собой минимальное из чисел $p_i^{l_i} - 1$. С соответствующей процедурой читатель может ознакомиться в специальной литературе [15, 16].

Вернемся к случаю, когда факторы могут действовать на *разных числах уровней* n_i . Рассмотрим двухфакторную модель. Если мы продолжаем настаивать на том условии, чтобы каждый уровень i_1 первого фактора обязательно хотя бы в одном наблюдении подействовал с любым уровнем i_2 второго фактора, то, как мы знаем, минимальный план эксперимента представляет собой факторный план, содержащий $n_1 n_2$ наблюдений. Однако можно уменьшить число наблюдений, если отказаться от этого требования и заменить его на более слабое условие: уровни действия i_1 и i_2 первого и второго факторов встречаются не более одного раза. Потребуем также, чтобы каждый уровень действия первого фактора встречался во всех экспериментах l_1 раз, а второго — l_2 раз. Естественно, числа n_1 , n_2 , l_1 и l_2 должны быть связаны соотношением $n_1 l_1 = n_2 l_2 = n$, где n — общее число наблюдений. План такого эксперимента можно записать в виде таблицы (табл. 12), элементы которой представляют собой числа $1, \dots, n_2$, расставленные таким образом, чтобы в каждом столбце любое из них встречалось не более одного раза, а во всей таблице — ровно l_2 раз. Отметим, что столбцы таблицы, представляющие собой те уровни, на которых в планируемом эксперименте должен подействовать второй фактор при условии, что на соответствующем уровне действует первый фактор, носят названия *блоков*.

Таблица 12

Номера наблюдений	Уровень действия первого фактора		
	1	...	n_1
$1 \div n_1$	$i_2^{(11)}$...	$i_2^{(n_1 1)}$
...
$(l_1 - 1)n_1 + 1 \div l_1 n_1$	$i_2^{(1l_1)}$...	$i_2^{(n_1 l_1)}$

Потребовав также, чтобы каждая пара i_1, i_2 , встречающаяся хотя бы в одном блоке, встречалась во всех блоках ровно k раз, получаем *неполный сбалансированный блок*. Нетрудно подсчитать, что

для неполного сбалансированного блока $l_2(l_1 - 1) = k(n_1 - 1)$. Если $n_1 = n_2$, то неполный сбалансированный блок называется *симметричным неполным сбалансированным блоком*. В табл. 13 приведен пример симметричного неполного сбалансированного блока для $n_1 = n_2 = 15$, $l_1 = l_2 = 7$, $k = 3$.

Таблица 13

1	1	1	1	1	1	1	2	2	2	2	3	3	4	5
2	2	2	3	3	8	7	6	6	7	8	4	5	5	6
5	5	5	6	6	11	4	11	11	12	3	10	11	12	7
3	7	10	4	9	9	8	4	7	3	4	5	7	6	8
8	9	15	13	12	10	11	9	10	8	9	11	9	8	9
6	12	13	10	14	12	15	13	12	13	10	12	13	13	10
11	4	14	7	15	13	14	15	14	15	14	15	14	14	15

При построении неполных сбалансированных блоков также используется теория Галуа. Описание самого метода их построения и процедуры применения плана для проверки гипотез об отсутствии действия факторов читатель может найти в [15, 16].

Случай трех и более факторов можно исследовать методом смешивания экспериментов [15, 16].

Мы рассмотрели здесь возможность применения задач планирования эксперимента к дисперсионному анализу. Однако в последнее время эти задачи находят широкое применение в множественном и многомерном регрессионном анализе для получения наилучших в некотором классе оценок параметров регрессии и для проверки гипотез о значимости отличия этих параметров от нуля. Обычно планы экспериментов выбирают таким образом, чтобы оценки или статистики критериев были независимыми, что приводит к необходимости рассматривать ортогональные планы или их обобщения. Характерной особенностью применения методов планирования эксперимента к регрессионному анализу является также тот факт, что наибольшую информацию несут наблюдения, в которых аргументы регрессии принимают крайние (минимальное и максимальное) значения. Поэтому в задачах регрессионного анализа очень часто все факторы действуют только на двух уровнях, которые для симметрии обозначаются -1 и $+1$, что, однако, не сильно упрощает сами планы из-за большого числа действующих факторов. Задачи планирования эксперимента в регрессионном анализе мы здесь не рассматриваем, а отсылаем заинтересованного читателя к специальной литературе.

Список литературы

1. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. — М.: Наука, 1983.
2. Гムрман В.Е. Теория вероятностей и математическая статистика. — М.: Высшая школа, 1972.
3. Пугачев В.С. Теория вероятностей и математическая статистика. — М.: Наука, 1979.
4. Худсон Д. Статистика для физиков. — М.: Мир, 1967.
5. Севастьянов Б.А. Курс теории вероятностей и математической статистики. — М.: Наука, 1982.
6. Климов Г.П. Теория вероятностей и математическая статистика. — М.: Изд-во МГУ, 1983.
7. Ивченко Г.И., Медведев Ю.И. Математическая статистика. — М.: Высшая школа, 1984.
8. Ван дер Варден. Математическая статистика. — М.: Мир, 1975.
9. Крамер Г. Математические методы статистики. — М.: Мир, 1975.
10. Уилкс С. Математическая статистика. — М.: Наука, 1967.
11. Боровков А.А. Математическая статистика. — М.: Наука, 1984.
12. Леман Э. Проверка статистических гипотез. — М.: Наука, 1979.
13. Вальд А. Последовательный анализ. — М.: Физматгиз, 1960.
14. Шеффе Г. Дисперсионный анализ. — М.: Физматгиз, 1963.
15. Дюoge Д. Теоретическая и прикладная статистика. — М.: Наука, 1972.
16. Ермаков С.М., Жиглявский А.А. Математическая теория оптимального эксперимента. — М.: Наука, 1987.
17. Гумрман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. — М.: Высшая школа, 1979.
18. Сборник задач по теории вероятностей, математической статистике и теории случайных функций / Под ред. А.А. Свешникова. — М.: Наука, 1970.
19. Азапов Г.И. Задачник по теории вероятностей. — М.: Высшая школа, 1986.
20. Емельянов Г.В., Скитович В.П. Задачник по теории вероятностей и математической статистике. — Л.: Изд-во ЛГУ, 1967.
21. Севастьянов Б.А., Чистяков В.П., Зубков А.М. Сборник задач по теории вероятностей. — М.: Наука, 1980.
22. Мешалкин Л.Д. Сборник задач по теории вероятностей. — М.: Изд-во МГУ, 1963.

Приложение

Таблица 1. Значения функции $P(m; \lambda) = \frac{\lambda^m}{m!} e^{-\lambda}$

Таблица 2. Значения функции $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

<i>x</i>	Сотые доли <i>x</i>									
	0	1	2	3	4	5	6	7	8	9
0,0	0,39894	39892	39886	39876	39862	39844	39822	39797	39767	39733
0,1	39695	39654	39608	39559	39505	39448	39387	39322	39253	39181
0,2	39104	39024	38940	38853	38762	38667	38568	38466	38361	38251
0,3	38139	38023	37903	37780	37654	37524	37391	37255	37115	36973
0,4	36827	36678	36526	36371	36213	36053	35889	35723	35553	35381
0,5	35207	35029	34849	34667	34482	34294	34105	33912	33718	33521
0,6	33322	33121	32918	32713	32506	32297	32086	31874	31659	31443
0,7	31225	31006	30785	30563	30339	30114	29887	29659	29431	29200
0,8	28969	28737	28504	28269	28034	27798	27562	27324	27086	26848
0,9	26609	26369	26129	25888	25647	25406	25164	24923	24681	24439
1,0	24197	23955	23713	23471	23230	22988	22747	22506	22265	22025
1,1	21785	21546	21307	21069	20831	20594	20357	20121	19886	19652
1,2	19419	19186	18954	18724	18494	18265	18037	17810	17585	17360
1,3	17137	16915	16694	16474	16256	16038	15822	15608	15395	15183
1,4	14973	14764	14556	14350	14146	13943	13742	13542	13344	13147
1,5	12952	12758	12566	12376	12188	12001	11816	11632	11450	11270
1,6	11092	10915	10741	10567	10396	10226	10059	09893	09728	09566
1,7	09405	09246	09089	08933	08780	08628	08478	08329	08183	08038
1,8	07895	07754	07614	07477	07341	07206	07074	06943	06814	06687
1,9	06562	06438	06316	06195	06077	05959	05844	05730	05618	05508
2,0	05399	05292	05186	05082	04980	04879	04780	04682	04586	04491
2,1	04398	04307	04217	04128	04041	03955	03871	03788	03706	03626
2,2	03547	03470	03394	03319	03246	03174	03103	03034	02965	02898
2,3	02833	02768	02705	02643	02582	02522	02463	02406	02349	02294
2,4	02239	02186	02134	02083	02033	01984	01936	01888	01842	01797
2,5	01753	01709	01667	01625	01585	01545	01506	01468	01431	01394
2,6	01358	01323	01289	01256	01223	01191	01160	01130	01100	01071
2,7	01042	01014	00987	00961	00935	00909	00885	00861	00837	00814
2,8	00792	00770	00748	00727	00707	00687	00668	00649	00631	00613
2,9	00595	00578	00562	00545	00530	00514	00499	00485	00470	00457
3,0	00443	00430	00417	00405	00393	00381	00370	00358	00348	00337
3,1	00327	00317	00307	00298	00288	00279	00271	00262	00254	00246
3,2	00238	00231	00224	00216	00210	00203	00196	00190	00184	00178
3,3	00172	00167	00161	00156	00151	00146	00141	00136	00132	00127
3,4	00123	00119	00115	00111	00107	00104	00100	00097	00094	00090
3,5	00087	00084	00081	00079	00076	00073	00071	00068	00066	00063
3,6	00061	00059	00057	00055	00053	00051	00049	00047	00046	00044
3,7	00042	00041	00039	00038	00037	00035	00034	00033	00031	00030
3,8	00029	00028	00027	00026	00025	00024	00023	00022	00021	00021
3,9	00020	00019	00018	00018	00017	00016	00016	00015	00014	00014
<i>x</i>	Десятые доли <i>x</i>									
	0	2	4	6					8	
4, 5,	0,000 133 8 0 000 015	0 000 589	0 000 249	0 000 101					0 000 040	

Таблица 3. Значения интеграла Лапласа $\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2} dt$

x	Сотые доли x									
	0	1	2	3	4	5	6	7	8	9
0,0	0,00000	00399	00798	01197	01595	01994	02392	02790	03188	03586
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147
1,3	40320	40490	40658	40824	40988	41149	41308	41466	41621	41774
1,4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189
1,5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408
1,6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449
1,7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327
1,8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670
2,0	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169
2,1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574
2,2	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899
2,3	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158
2,4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361
2,5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520
2,6	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643
2,7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736
2,8	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807
2,9	49813	49819	49825	49831	49836	49841	49846	49851	49856	49861
3,0	49865	49869	49874	49878	49882	49886	49889	49893	49897	49900
3,1	49903	49906	49910	49913	49916	49918	49921	49924	49926	49929
3,2	49931	49934	49936	49938	49940	49942	49944	49946	49948	49950
3,3	49952	49953	49955	49957	49958	49960	49961	49962	49964	49965
3,4	49966	49968	49969	49970	49971	49972	49973	49974	49975	49976
3,5	49977	49978	49978	49979	49980	49981	49981	49982	49983	49983
3,6	49984	49985	49985	49986	49986	49987	49987	49988	49988	49989
3,7	49989	49990	49990	49990	49991	49991	49992	49992	49992	49992
3,8	49993	49993	49993	49994	49994	49994	49994	49995	49995	49995
3,9	49995	49995	49996	49996	49996	49996	49996	49996	49997	49997
x	Десятые доли x									
	0	2	4	6	8					
4, 5,	0,499 968 3 4 999 997	4 999 867	4 999 946	4 999 979	4 999 992					

Учебное издание

*БОЧАРОВ Павел Петрович
ПЕЧИНКИН Александр Владимирович*

ТЕОРИЯ ВЕРОЯТНОСТЕЙ. МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Оригинал-макет: *В.В. Худяков*
Оформление переплета: *А.Ю. Алексина*

ЛР №071930 от 06.07.99. Подписано в печать 11.07.05.
Формат 60×90/16. Бумага офсетная. Печать офсетная.
Усл. печ. л. 18,5. Уч.-изд. л. 22,6. Заказ №

Издательская фирма «Физико-математическая литература»
МАИК «Наука/Интерпериодика»
117997, Москва, ул. Профсоюзная, 90
E-mail: fizmat@maik.ru, fmsale@maik.ru;
<http://www.fml.ru>

Отпечатано с готовых диапозитивов
в ОАО «Чебоксарская типография № 1»
428019, г. Чебоксары, пр. И. Яковleva, 15

ISBN 5-9221-0633-3



9 785922 106337