М. Я. Кельберт, Ю. М. Сухов

ВЕРОЯТНОСТЬ И СТАТИСТИКА В ПРИМЕРАХ И ЗАДАЧАХ

Том I

Основные понятия теории вероятностей и математической статистики

Москва
Издательство МЦНМО
2007

УДК 519.21
ББК 22.171
К34

*Издание осуществлено при поддержке РФФИ
(издательский проект* №05-01-14087).

Кельберт М. Я., Сухов Ю. М.
К34 Вероятность и статистика в примерах и задачах. Т. I: Основные понятия теории вероятностей и математической статистики. М.: МЦНМО, 2007. — 456 с.: ил.

ISBN 978-5-94057-253-4

Для освоения теории вероятностей и математической статистики тренировка в решении задач и выработка интуиции важны не меньше, чем изучение доказательств теорем; большое разнообразие задач по этому предмету затрудняет студентам переход от лекций к экзаменационным задачам, а от них — к практике.

Ввиду того, что предмет этой книги критически важен как для современных приложений (финансовая математика, менеджмент, телекоммуникации, обработка сигналов, биоинформатика), так и для приложений классических (актуарная математика, социология, инженерия), авторы собрали большое количество упражнений, снабженных полными решениями. Эти решения адаптированы к нуждам и умениям учащихся. Необходимые теоретические сведения приводятся по ходу изложения; кроме того, текст снабжен историческими отступлениями.

ББК 22.171

Перевод с английского В. Кноповой, Ю. Мишуры, Л. Сахно.

На обложке изображены капелла Кингс-колледжа
и улица Кингс-парад (г. Кембридж, Великобритания).

ISBN 978-5-94057-253-4 (Том I)
ISBN 978-5-94057-252-7
© Кельберт М. Я., Сухов Ю. М., 2007
© МЦНМО, 2007

Оглавление

Часть B
Основы статистики

Введение

Первоначальные побуждения к созданию данной книги были весьма личными. Один из авторов в ходе своей педагогической деятельности на факультете математики и математической статистики Кембриджского университета и Колледжа Св. Иоанна в Кембридже часто наблюдал печальную картину, когда хорошие (и даже блестящие) студенты, увлеченные математикой и ее приложениями и отлично завершившие свой первый учебный год, спотыкались или даже проваливались на экзаменах. Это вело к большому разочарованию, которое очень трудно преодолеть в последующие годы учебы. Добросовестный преподаватель всегда сочувствует таким неудачам, однако даже тщательное объяснение слабых сторон студента (если таковые имеются) не всегда помогает. Другой автор приобрел подобный опыт скорее как отец студента Кембриджского университета, чем как преподаватель.

Мы почувствовали, что книга, состоящая главным образом из заданий, предложенных на экзаменах по математике в Кембриджском университете, будет полезной большинству студентов. В соответствии с нашими собственными научными и педагогическими устремлениями мы выбрали главной темой вероятность и статистику. Отправным пунктом стало содержание курса теории вероятностей (I год обучения) и математической статистики (II год обучения). Чтобы отразить полное содержание этих курсов, потребуется еще несколько томов; так или иначе, мы решили реализовать этот проект.

Наша основная цель — представить кембриджские курсы вероятности и статистики с помощью экзаменационных заданий, предложенных в предыдущие годы. В процессе написания мы сочли необходимым время от времени включать также некоторые стандартные вопросы из списков задач, иллюстрирующих материал лекций, хотя мы и старались свести их число к минимуму.

Конечно, кембриджские экзамены никогда не были легкими. По их результатам испытуемые делятся на классы: первый, второй (он делится на две категории, 2.1 и 2.2) и третий; небольшое число испытуемых проваливается. (На самом деле составляется более подробный список всех испытуемых по ранжиру, согласно их результатам, но он не оглашается.) Официальное название экзаменов — «Математические треножники», оно происходит от трехногих табуреток, на которых испытуемые и экзаменаторы восседали (часто на протяжении многих часов) во время устных экзаменов в давние времена. Теперь все экзамены письменные. Первый год трехлетнего курса обучения именуется «часть IA», второй — «часть IB» и третий — «часть II».

В мае—июне 2003 г. студенты-математики первого года обучения сдавали 4 экзамена; каждый длился 3 часа и состоял из 12 вопросов по двум предметам. На экзамены были вынесены следующие курсы: «Алгебра и геометрия», «Числа и множества», «Анализ», «Вероятность», «Дифференциальные уравнения», «Векторное исчисление» и «Механика». Все вопросы по заданному курсу были собраны в одном экзамене, кроме курса алгебры и геометрии, по которому студенты экзаменовались дважды. В каждом экзамене 4 вопроса классифицированы как краткие (по 2 вопроса из каждого курса) и 8 как расширенные (по 4 из каждого курса). Испытуемые должны ответить на все 4 кратких вопроса и не более чем на 5 расширенных, не более чем по 3 из каждого курса; расширенный вопрос приносит в 2 раза больше очков, чем краткий, при условии верного ответа. Расчет показывает: если студент берется за все 9 положенных вопросов (так происходит чаще всего) и правильно распределяет время, то на краткий вопрос нужно ответить за 12—13 минут, а на расширенный — за 24—25 минут. Это нелегко и требует специальной практики; данная книга во многом нацелена на то, чтобы помочь в обретении необходимых «тренировочных» навыков.

Экзамены во второй год обучения частично схожи с предыдущими, но имеют свои отличия. В июне 2003 г. проводились 4 экзамена «Математических треножников» по «части IB», каждый был рассчитан на 3 часа и содержал 9 или 10 коротких и 9 или 10 расширенных вопросов по предметам, отобранным для данного экзамена. В частности, «Статистика IB» содержалась в экзаменах 1, 2 и 4, а в целом было 6 вопросов по «Статистике». Конечно, подготовка к экзаменам «части IB» отличается от подготовки к «части IA»; мы разберем эти различия в соответствующих главах.

Для типичного студента Кембриджа специальная подготовка к экзамену начинается во время пасхального семестра (с середины апреля). В идеале к работе рекомендуется приступать во время предшествующих пятинедельных каникул. (Некоторая предэкзаменационная подготовка к «части IB» и «части II», а именно вычислительные проекты, выполняется в основном на весенних (пасхальных) каникулах.) По мере приближения экзаменов атмосфера в Кембридже сгущается и становится нервной, хотя делается немало попыток ее разрядить. Многие испытуемые прилагают огромные усилия, пытаясь подсчитать как можно точнее, сколько же труда вложить в тот или иной предмет в зависимости от баллов, которые он приносит, и того, как они сами в этом предмете разбираются, с целью оптимизировать общий уровень подготовки. Можно соглашаться либо не соглашаться с таким подходом, но ясно: если студенты обладают достаточной информацией по поводу уровня и стиля вопросов к экзаменам «Математические треножники» и способны их усвоить, то у них намного больше шансов мобилизовать свои лучшие качества и способности. В настоящий момент, под гнетом нехватки времени и сил, большинство из них не в состоянии мобилизоваться, и многое отдается на волю случая. Мы будем чрезвычайно рады, если наша книга поможет изменить эту ситуацию, сняв предэкзаменационный стресс и лишив экзамены «Треножники» налета некоторой таинственности, по крайней мере в части учебных дисциплин, рассматриваемых здесь.

Таким образом, первой из причин появления этой книги было желание облегчить жизнь студентам. Однако во время работы над текстом появилась и вторая причина, и она представляет, по нашему мнению, значительный профессиональный интерес для всех, кто читает курс теории вероятностей и статистики. В 1991/2 учебном году в Кембридже произошли большие изменения, касающиеся подхода к чтению вероятностных и статистических курсов. Наиболее значимым в новом подходе оказалось то, что курсы «Вероятность IA» и «Статистика IB» были переработаны с целью сделать их привлекательными для широкой аудитории (200 студентов первого курса, изучающих «Вероятность IA», и почти такое же число студентов второго курса, изучающих «Статистику IB»). Для большого числа студентов это единственные курсы теории вероятностей и статистики, которые они слушают на протяжении учебы. Поскольку все большее число специалистов в современном мире имеют дело с теоретическими и (особенно) прикладными задачами вероятностной и статистической природы, очень важно, чтобы указанные курсы поддерживали у студентов повышенный интерес. Таким образом, основная цель переместилась с академического введения в предмет на методологический подход, который вооружил бы студентов методами, необходимыми для решения разумных практических и теоретических проблем, возникающих в реальных жизненных ситуациях.

Как следствие, акцент в курсе «Вероятность IA» сместился с сигма-алгебр, интегрирования по Лебегу и Стилтьесу и характеристических функций на прямое рассмотрение различных моделей, как дискретных, так и непрерывных, с целью подготовить студентов к будущим курсам и задачам (в частности, к курсу «Статистика IB» и «Марковские цепи IB/II»). В свою очередь, центральной частью курса «Статистика IB» стали наиболее популярные практические применения теории оценивания, проверки гипотез и регрессионного анализа. Основным показателем экзаменационной подготовки, относящейся к курсам «Вероятность IA» и «Статистика IB», стала способность студентов решать задачи без привлечения глубоких теоретических сведений, а лишь путем выбора и анализа подходящей модели и аккуратного выполнения разумного количества вычислений.

Конечно, такие изменения (и параллельное развитие других курсов) не были единогласно поддержаны на математическом факультете и вызвали большие дебаты. Однако основная масса студентов поддержала новый подход, и переработанные курсы завоевали большую популярность как в плане посещаемости, так и по числу задач данного курса, выбираемых экзаменуемыми (что является весьма важным моментом в жизни математических факультетов Кембриджа). Кроме того, с увеличением роли компьютеров студенты оказали существенное предпочтение «алгоритмическому» стилю лекций и экзаменационных вопросов (по крайней мере, по опыту авторов).

В этом отношении представляет определенный интерес следующий опыт, приобретенный одним из авторов. В течение определенного времени он опрашивал выпускников-математиков Колледжа Святого Иоанна, которые трудятся ныне в различных сферах деятельности, о том, какие разделы курса, прослушанного в Кембридже, они считают наиболее важными для их нынешней работы. Оказалось, что наибольшее влияние на большинство респондентов оказали не конкретные факты, теоремы и доказательства (надолго, однако, запоминаются шутки, прозвучавшие на лекциях). Гораздо выше ценится способность построить математическую модель, которая удачно представляет жизненную ситуацию, поставить верно задачу и решить ее аналитически или, чаще, численными методами. Этим подтверждается своевременность предложенного подхода.

Вследствие всего вышесказанного уровень и стиль «Математических треножников» претерпел изменения. Строго подразумевается (хотя это и не всегда достигается), что вопросы должны иметь четкую структуру, согласно которой испытуемые переходят от одной части задачи к следующей. Курьезным наблюдением многих экзаменаторов является то, что, как бы ни были совершенны задуманные экзаменатором решения, испытуемые следуют им очень редко. Поэтому в книге часто используются студенческие решения.

Описанная выше вторая причина внушает нам надежду, что книга будет интересна далеко за пределами Кембриджа. В связи с этим возникает естественный вопрос: каково место данной книги в длинном списке учебников по теории вероятностей и статистике? Большая часть библиографии в конце этого тома содержит список книг, опубликованных на английском языке после 1991 г., имеющих в названии термины «вероятность» и «статистика» и приобретенных как главной библиотекой Кембриджского университета, так и факультетскими библиотеками (список неполон, по нашему убеждению, и мы извиняемся за все пропуски).

Что касается основ теории вероятностей, хочется сравнить нашу книгу с тремя популярными сериями учебников и сборников задач, одна из них принадлежит С. Россу [Ros1]—[Ros5], другая — Д. Стирзакеру [St1]—[St3], третья — Г. Гриммету и Д. Стирзакеру [GriS1]—[GriS3]. Книги Росса и Стирзакера — хорошее введение в основные понятия. Стиль и уровень изложения книг Росса принят во многих американских университетах. С другой стороны, книги Гриммета и Стирзакера написаны на гораздо более высоком уровне, который можно охарактеризовать как «профессиональный». Уровень нашей книги — это нечто среднее. По нашему мнению, мы ближе к Россу и Стирзакеру, хотя и очень далеки от них в некоторых важных аспектах. Нам кажется, что уровень книг Росса и Стирзакера для Кембриджа недостаточен; его недостаточно для того, чтобы сдать кембриджские квалификационные экзамены с разрядом 2.1 и выше. Книг Гриммета и Стирзакера, конечно, более чем достаточно, однако если по ним готовиться к экзаменам, то основной проблемой будет выбор подходящих примеров из примерно тысячи предлагаемых.

Многие из этих книг — прекрасное дополнительное чтение для тех, кто хочет продвинуться в определенном направлении. Отметим лишь некоторые из них: [Chu], [Dur1], [G], [Go], [JP], [Sc] и [ChaY]. В любом случае, прошло то (ностальгическое) время, когда каждый, кто взялся изучать теорию вероятностей, должен был прилежно проштудировать (великолепный) двухтомник Феллера [Fe] (по нашему мнению, Феллера до сих пор никто не превзошел).

Что касается статистики, тут картина сложнее. Даже определение предмета статистики — до сих пор нечто спорное (см. введение к части В данного тома). Стиль чтения лекций и проведения экзаменов по основному курсу «Статистики» (и по другим курсам, относящимся к статистике) в Кембридже всегда имел отличительные особенности. Этот стиль всегда сопротивлялся намерениям сделать изложение «абсолютно строгим», несмотря на то что курс читался студентам-математикам. Меньшая часть студентов находила курс слишком сложным, однако для большинства сложностей не было. С другой стороны, уровень строгости курса достаточно высокий и требует прочных математических знаний. Из новейших изданий к кембриджскому стилю ближе всего, вероятно, [CaB]. Как пример весьма отличного подхода отметим книгу [Will] (ее стилем мы восхищаемся, но не считаем эту книгу приемлемой для первоначального знакомства с курсом или для подготовки к кембриджским экзаменам).

Отличительной чертой данной книги является наличие повторений: некоторые темы и вопросы возникают не единожды, часто в слегка измененной форме, что затрудняет ссылки на предыдущие их появления. Это типичная черта экзаменационного процесса, что становится очевидным, если читать курс лет десять или около того. Наше персональное отношение к данному вопросу выражается пословицей «повторение — мать учения», популярной (в различных формах) во многих языках. Однако мы извиняемся перед теми читателями, кому некоторые (а может, и многие) из наших повторений покажутся чрезмерными.

Книга построена следующим образом. В части А представлен материал курса «Вероятность IA» (который состоит из 24 одночасовых лекций). В этой части вопросы экзаменов «Треножники» расположены внутри или непосредственно следуют за соответствующими теоретическими сведениями. В части В представлен материал курса «Статистика IB», состоящего из 16 лекций. Здесь вопросы «Треножников» охватывают широкий класс отдельных тем, и мы решили расположить их отдельно от материала курса. Однако отдельные разделы теории всегда представлены в соответствии с той ролью, которую они играют в экзаменационных вопросах.

Задачи, представленные в книге, взяты из списков задач к экзаменам по вероятности и статистике разных лет, проводившимся в Кембриджском университете, а также в Университете Свонзи-Уэльс.

Некоторые отступления, разбросанные по тексту, позаимствованы из [Ox], [Go], [Ha], веб-сайтов Университета Святого Андрея

`http://www-history.mcs.st-andrews.ac.uk/history/`

и Массачусетского университета

`http://www.umass.edu/wsp/statistics/tales/`.

Беседы с Х. Дэниэлсом, Д. Г. Кендаллом и К. Р. Рао также привели к появлению нового материала. Однако ряд историй — это часть фольклора (большинство из них доступно в Интернете); за возможные ошибки несут ответственность авторы этой книги. Фотографии и портреты многих людей, упоминаемых в этой книге, размещены на сайтах Йоркского университета

`http://www.york.ac.uk/depts/maths/histstat/people/`

и (с биографиями) на

`http://members.aol.com/jayKplanr/images.htm`.

С появлением Всемирной Сети чрезвычайно быстро стали распространяться шутки, юмор. Мы признаем, что немало времени мы посвятили веб-сайтам, посвященным шуткам и забавным цитатам; часть из них использована в книге. Просим прощения у авторов, оставшихся безымянными (иногда мы даже меняли смысл фраз).

В процессе работы над книгой нас воодушевляла поддержка многих коллег, которые прочли весь текст или его часть. Мы благодарим Чарльза Голди, Оливера Джонсона, Джеймса Мартина, Ричарда Самворса и Аманду Тернер за плодотворные дискуссии и замечания. Особенно мы благодарны Алану Хоксу за бесконечное терпение, с которым он продирался через предварительный вариант рукописи. Мы широко использовали конспекты лекций, списки экзаменационных вопросов и другие тексты, подготовленные сотрудниками Статистической лаборатории, в частности Джеффри Гримметтом, Франком Келли, Джеймсом Норрисом, Сюзанной Питтс и Ричардом Вебером. Особая благодарность также Саре Шеи-Симондс за тщательное прочтение книги и исправление большого числа стилистических ошибок.

* * *

Английское и русское издания этой книги выходят в свет практически одновременно. Тем не менее, текст русского издания был существенно переработан и расширен (перевод был выполнен с рукописи). Авторы благодарят переводчиков — В. Кнопову, Ю. Мишуру и Л. Сахно — за добросовестное и вдумчивое отношение к тексту. Мы также признательны Джону Хейгу, указавшему на ряд неточностей в английском издании.

Часть А

Вероятность

Глава 1

Дискретные пространства элементарных исходов

§ 1.1. Равномерное распределение

Lest men suspect your tale untrue,
Keep probability in view.

Дабы люди не сочли ваши истории неправдивыми,
Учитывайте вероятность того, о чем говорите.

Дж. Гэй (1685—1732), английский поэт

В этом параграфе используется простейшая (и наиболее ранняя исторически) вероятностная модель, имеющая конечное число m возможностей (часто называемых исходами), каждый из которых имеет одну и ту же вероятность $1/m$. Множество A из k исходов при $k \leqslant m$ называется событием, и его вероятность $\mathsf{P}(A)$ равна k/m:

$$\mathsf{P}(A) = \frac{\text{число исходов в событии } A}{\text{общее число исходов}}. \tag{1.1.1}$$

Пустое множество имеет нулевую вероятность, а множество всех исходов имеет вероятность 1. Эта схема выглядит обманчиво простой; в действительности вычисление числа исходов, благоприятствующих данному событию (а иногда и общего числа исходов), может быть достаточно сложным.

Задача 1.1.1. Мы с вами подбрасываем монету: если монета падает вверх орлом, я выигрываю очко, если решкой — вы выигрываете очко. Вначале счет нулевой.

а) Какова вероятность того, что после $2n$ бросаний счет будет равным?

б) Какова вероятность того, что после $2n+1$ бросания у меня будет на 3 очка больше, чем у вас?

Решение. Исходы в п. а) — это последовательности вида $OOO\ldots O$, $POO\ldots O$, $\ldots$, $PPP\ldots P$, состоящие из $2n$ последовательных букв O или P (или, в цифрах, 0 или 1). Общее число исходов равно $m = 2^{2n}$, вероятность каждого $1/2^{2n}$. Нас интересуют исходы, для которых число букв O равно числу букв P. Число k таких исходов равно $\dfrac{(2n)!}{n!\,n!}$ (число возможностей выбрать n мест для O из всех $2n$ мест). Искомая вероятность равна $\frac{(2n)!}{n!\,n!}\cdot\frac{1}{2^{2n}}$. В п. б) исходы — это последовательности длины $2n+1$, всего их 2^{2n+1}. Вероятность равна $\frac{(2n+1)!}{(n+2)!\,(n-1)!}\cdot\frac{1}{2^{2n+1}}$. □

Задача 1.1.2. Теннисный турнир организован для 2^n игроков, проводится он в n туров, игра идет «на выбывание», последний тур называется финалом, предпоследний — полуфиналом. Двух игроков выбирают случайным образом. Найдите вероятность того, что они а) встретятся в первом или втором туре, б) встретятся в финале или полуфинале, в) вообще не встретятся.

Решение. Очень важной является фраза «двух игроков выбирают случайным образом». Например, можно считать, что выбор сделан уже после турнира, когда все результаты известны, и это совершенно другая ситуация. Итак, имеется 2^{n-1} пара игроков в первом туре, 2^{n-2} пары во втором туре, две в полуфинале, одна в финале, а всего $2^{n-1}+2^{n-2}+\ldots$ $\ldots+2+1=2^n-1$. Общее число пар, которое можно образовать, равно $C_{2^n}^2=2^{n-1}(2^n-1)$. Поэтому ответы таковы: а) $\frac{2^{n-1}+2^{n-2}}{2^{n-1}(2^n-1)}=\frac{3}{2(2^n-1)}$, б) $\frac{3}{2^{n-1}(2^n-1)}$, в) $\frac{2^{n-1}(2^n-1)-(2^n-1)}{2^{n-1}(2^n-1)}=1-\frac{1}{2^{n-1}}$. □

Задача 1.1.3. В комнате собралось n человек.

а) Какова вероятность того, что как минимум двое имеют общий день рождения? Найдите эту вероятность при $n=22$ и 23.

б) Какова вероятность того, что по крайней мере один человек родился в тот же день, что и вы? Какое значение n делает эту вероятность близкой к 1/2?

Решение. Общее число исходов равно 365^n. В п. а) число неблагоприятных исходов равно $365\cdot364\cdot\ldots\cdot(365-n+1)$. Поэтому вероятность того, что все дни рождения различны, равна $(365\cdot364\cdot\ldots\cdot(365-n+1))/365^n$, а того, что двое или более людей имеют общий день рождения, равна $1-\frac{365\cdot364\cdot\ldots\cdot(365-n+1)}{365^n}$. Для $n=22$ имеем

$$1-\frac{365}{365}\cdot\frac{364}{365}\cdot\ldots\cdot\frac{344}{365}=0{,}4927,$$

а для $n=23$ имеем

$$1-\frac{365}{365}\cdot\frac{364}{365}\cdot\ldots\cdot\frac{343}{365}=0{,}5243.$$

В п. б) число неблагоприятных исходов равно 364^n, а вероятность равна $1-\left(\frac{364}{365}\right)^n$. Если мы хотим, чтобы она была близка к 1/2, то $\left(\frac{364}{365}\right)^n\approx\frac{1}{2}$, т. е. $n=-\frac{1}{\log_2(364/365)}\approx253$. □

Задача 1.1.4. Мэри подбрасывает $n+1$ монету, а Джон подбрасывает n монет. Какова вероятность того, что у Мэри выпадет больше орлов, чем у Джона?

Решение 1. Будем предполагать, что все монеты симметричны (поскольку не сказано иного). Число исходов для Мэри равно 2^{n+1} (все возможные последовательности орлов и решек), а для Джона — 2^n; всего имеется 2^{2n+1} равновозможных исходов. Пусть $O_М$ и $P_М$ — число орлов и решек у Мэри, $O_Д$ и $P_Д$ — число их же у Джона; при этом $O_М + P_М = n+1$, $O_Д + P_Д = n$. События $\{O_М > O_Д\}$ и $\{P_М > P_Д\}$ имеют одинаковое число исходов, значит, $\mathsf{P}(O_М > O_Д) = \mathsf{P}(P_М > P_Д)$. С другой стороны, $O_М > O_Д$ тогда и только тогда, когда $n - O_М < n - O_Д$, т. е. $P_М - 1 < P_Д$, или $P_М \leqslant P_Д$. Таким образом, событие $O_М > O_Д$ совпадает с событием $P_М \leqslant P_Д$, и $\mathsf{P}(P_М \leqslant P_Д) = \mathsf{P}(O_М > O_Д)$. Но для любого (совместного) исхода либо $P_М > P_Д$, либо $P_М \leqslant P_Д$, т. е. число исходов события $\{P_М > P_Д\}$ равно разности между 2^{n+1} и числом исходов события $\{P_М \leqslant P_Д\}$. Поэтому $\mathsf{P}(P_М > P_Д) = 1 - \mathsf{P}(P_М \leqslant P_Д)$. Подведем итог: $\mathsf{P}(O_М > O_Д) = \mathsf{P}(P_М > P_Д) = 1 - \mathsf{P}(P_М \leqslant P_Д) = 1 - \mathsf{P}(O_М > O_Д)$. Отсюда $\mathsf{P}(O_М > O_Д) = 1/2$. □

Решение 2. Будем считать, что последнее бросание в игре принадлежит Мэри. Обозначим через x вероятность того, что у Джона и Мэри перед ее последним бросанием одно и то же количество орлов. Тогда в силу симметрии вероятность того, что у Мэри перед ее последним бросанием больше орлов, чем у Джона, равна $(1-x)/2$. Тем самым, вероятность того, что у Мэри будет в конце игры больше орлов, чем у Джона, равна $(1-x)/2 + x/2 = 1/2$ (Дж. Хейг). □

Решение 3. Поскольку у Мэри на единицу больше бросаний, у нее и в конце игры будет либо больше орлов, либо больше решек, чем у Джона. При этом эти два случая исключают друг друга. В силу симметрии вероятность каждого есть 1/2. □

Задача 1.1.5. Вы подбрасываете случайным образом $6n$ игральных кубиков. Покажите, что вероятность того, что каждая грань выпадет ровно n раз, равна

$$\frac{(6n)!}{(n!)^6}\left(\frac{1}{6}\right)^{6n}.$$

Решение. Общее число возможных результатов подбрасываний равно 6^{6n} (шесть для каждого кубика), каждый исход имеет вероятность $\frac{1}{6^{6n}}$. Мы хотим, чтобы на n кубиках выпала грань «1», на n кубиках — грань «2» и т. д. Чтобы подсчитать число таких исходов, зафиксируем вначале те кубики, на которых выпала грань «1»: эти кубики можно выбрать $\frac{(6n)!}{n!\,(5n)!}$

способами. Выбрав n кубиков, на которых выпала грань «1», мы зафиксируем среди оставшихся кубиков те, на которых выпала грань «2»: для этого у нас имеется $\frac{(5n)!}{n!\,(4n)!}$ способов, и т. д. Общее количество желаемых исходов равно произведению полученных чисел, что равно $\frac{(6n)!}{(n!)^6}$. Это и дает ответ. □

При решении многих задач очень важное значение имеет нахождение рекуррентных соотношений для мощностей различных событий. Например, если f_n — это число возможных способов подбрасывания монеты n раз, при которых герб никогда не выпадает два раза подряд, то имеет место соотношение $f_n = f_{n-1} + f_{n-2}$, $n \geqslant 3$ (уравнение Фибоначчи).

Задача 1.1.6. а) Определите число g_n возможных способов подбрасывания монеты n раз так, чтобы комбинация OP никогда не появилась.

б) Докажите соотношение $f_n = f_{n-1} + f_{n-2} + f_{n-3}$, $n \geqslant 3$, для числа способов подбрасывания монеты n раз, при которых никогда не выпадут три орла подряд.

Решение. а) Справедливо равенство $g_n = 1 + n$; один раз появляется последовательность $OO\ldots O$, и имеется n последовательностей вида $P\ldots PO\ldots O$ (включая и последовательность $P\ldots P$).

б) Возможными исходами являются 2^n последовательностей вида $(y_1, \ldots, y_n)$, состоящие из O и P. Пусть A_n — это событие {ни разу не выпало три орла подряд в результате n подбрасываний}, тогда f_n — это мощность $\operatorname{card} A_n$. Рассмотрим разбиение $A_n = B_n^{(1)} \cup B_n^{(2)} \cup B_n^{(3)}$, где $B_n^{(1)}$ — это событие {ни разу не выпало три орла подряд в результате n подбрасываний, и при последнем подбрасывании выпала решка}, $B_n^{(2)} =$ {ни разу не выпало три орла подряд в результате n подбрасываний, и при последних двух подбрасываниях выпало PO} и $B_n^{(3)} =$ {ни одного раза не выпало три орла подряд после n подбрасываний, и при последних трех подбрасываниях выпало POO}.

Ясно, что $B_n^{(i)} \cap B_n^{(j)} = \varnothing$, $1 \leqslant i \neq j \leqslant 3$, а также $f_n = \operatorname{card} B_n^{(1)} + \operatorname{card} B_n^{(2)} + \operatorname{card} B_n^{(3)}$.

Опустим теперь последний символ y_n: $(y_1, \ldots, y_n) \in B_n^{(1)}$ тогда и только тогда, когда $y_n = P$, $(y_1, \ldots, y_{n-1}) \in A_{n-1}$, следовательно, $\operatorname{card} B_{n-1}^{(1)} = f_{n-1}$. Далее, $(y_1, \ldots, y_n) \in B_n^{(2)}$ тогда и только тогда, когда $y_{n-1} = P$, $y_n = O$ и $(y_1, \ldots, y_{n-2}) \in A_{n-2}$. Это позволяет нам заключить, опуская два последних символа, что $\operatorname{card} B_n^{(2)} = f_{n-2}$. Аналогично $\operatorname{card} B_n^{(3)} = f_{n-3}$. Отсюда и следует требуемое равенство. □

§ 1.2. Условные вероятности. Формула Байеса. Независимые испытания

On voit par cet Essai que la théorie des probabilités
n'est au fond que le bon sens réduit au calcul...

P.-S. Laplace

Теория вероятностей есть не что иное,
как здравый смысл, сведенный к вычислениям.

П.-С. Лаплас (1749—1827)

Clockwork Omega[1]

(Из серии «Фильмы, которые не вышли на большой экран».)

Рассмотрим теперь более общую модель: пусть элементарные события имеют не обязательно равные вероятности $p_1, \ldots, p_m$, $p_i > 0$ и $p_1 + \ldots$ $\ldots + p_m = 1$.

Как и ранее, *событие* A представляет собой некоторую совокупность элементарных исходов (возможно, и пустое множество); *вероятность* $\mathsf{P}(A)$ события A теперь задается формулой

$$\mathsf{P}(A) = \sum_{\text{элементарный исход } i \in A} p_i \tag{1.2.1}$$

($\mathsf{P}(A) = 0$ для $A = \varnothing$.) Вероятность всей совокупности элементарных исходов равна единице. Всю совокупность элементарных исходов часто называют полным или достоверным событием и обозначают Ω, таким образом, $\mathsf{P}(\Omega) = 1$. Элементарный исход часто обозначают ω, и если $p(\omega)$ — это его вероятность, то

$$\mathsf{P}(A) = \sum_{\omega \in A} p(\omega). \tag{1.2.2}$$

Из этого определения следует, что вероятность объединения вычисляется по формуле

$$\mathsf{P}(A_1 \cup A_2) = \mathsf{P}(A_1) + \mathsf{P}(A_2) \tag{1.2.3}$$

для любой пары несовместных событий A_1, A_2 (т. е. таких событий, что $A_1 \cap A_2 = \varnothing$). В более общем виде

$$\mathsf{P}(A_1 \cup \ldots \cup A_n) = \mathsf{P}(A_1) + \ldots + \mathsf{P}(A_n) \tag{1.2.4}$$

для любого набора попарно несовместных событий (т. е. таких событий, что $A_j \cap A_{j'} = \varnothing$ $\forall j \neq j'$). Следовательно, а) вероятность $\mathsf{P}(A^c)$ дополнения $A^c = \Omega \setminus A$ равна $1 - \mathsf{P}(A)$, б) если $B \subseteq A$, то $\mathsf{P}(B) \leqslant \mathsf{P}(A)$

[1] Игра слов, ср. «Clockwork Orange» («Заводной апельсин», фильм С. Кубрика).

и $\mathrm{P}(A)-\mathrm{P}(B)=\mathrm{P}(A\setminus B)$, в) для любой пары событий A, B выполняется равенство $\mathrm{P}(A\setminus B)=\mathrm{P}(A\setminus(A\cap B))=\mathrm{P}(A)-\mathrm{P}(A\cap B)$.

Более того, для объединения любого числа (не обязательно несовместных) событий имеет место соотношение

$$\mathrm{P}(A_1\cup\ldots\cup A_n)\leqslant\sum_{i=1}^{n}\mathrm{P}(A_i);$$

более детально вероятность $\mathrm{P}\left(\bigcup A_i\right)$ описывается при помощи формулы включения-исключения (1.3.1).

Пусть заданы два события A и B, причем $\mathrm{P}(B)>0$. Тогда *условная вероятность* $\mathrm{P}(A\mid B)$ события A при условии, что событие B произошло, определяется как отношение

$$\mathrm{P}(A\mid B)=\frac{\mathrm{P}(A\cap B)}{\mathrm{P}(B)}. \tag{1.2.5}$$

На данном этапе условные вероятности имеют для нас важное значение благодаря двум формулам. Одна из них — это *формула полной вероятности*: если $B_1,\ldots,B_n$ являются попарно несовместными событиями, образующими разбиение множества Ω, т. е. $B_i\cap B_j=\varnothing$ для $1\leqslant i<j\leqslant n$ и $B_1\cup B_2\cup\ldots\cup B_n=\Omega$, и если, кроме того, $\mathrm{P}(B_i)>0$ для $1\leqslant i\leqslant n$, то

$$\mathrm{P}(A)=\mathrm{P}(A\mid B_1)\,\mathrm{P}(B_1)+\mathrm{P}(A\mid B_2)\,\mathrm{P}(B_2)+\ldots+\mathrm{P}(A\mid B_n)\,\mathrm{P}(B_n). \tag{1.2.6}$$

Доказательство получаем непосредственным подсчетом:

$$\mathrm{P}(A)=\sum_{1\leqslant i\leqslant n}\mathrm{P}(A\cap B_i)=\sum_{1\leqslant i\leqslant n}\frac{\mathrm{P}(A\cap B_i)}{\mathrm{P}(B_i)}\,\mathrm{P}(B_i)=\sum_{1\leqslant i\leqslant n}\mathrm{P}(A\mid B_i)\,\mathrm{P}(B_i).$$

Часто именно условные вероятности заданы, а найти необходимо безусловные вероятности; формула полной вероятности также может быть полезной для прояснения природы (безусловной) вероятности $\mathrm{P}(A)$. Несмотря на свою простоту, эта формула является мощным средством буквально во всех областях, имеющих дело с вероятностями. В частности, значительная часть теории марковских цепей основана на умелом применении этой формулы.

Представление вероятности $\mathrm{P}(A)$ в виде правой части формулы (1.2.6) называют взятием условной вероятности относительно набора событий $B_1,\ldots,B_n$.

Вторая формула — это *формула Байеса* (или *теорема Байеса*), названная в честь Т. Байеса (1702—1761), английского математика и священнослужителя. Она утверждает, что при сделанных предположениях и условии $\mathrm{P}(A)>0$ условная вероятность $\mathrm{P}(B_i\mid A)$ может быть выражена в терминах вероятностей $\mathsf{P}(B_1), \ldots, \mathsf{P}(B_n)$ и условных вероятностей $\mathsf{P}(A \mid B_1), \ldots, \mathsf{P}(A \mid B_n)$ следующим образом:

$$\mathsf{P}(B_i \mid A) = \frac{\mathsf{P}(A \mid B_i)\,\mathsf{P}(B_i)}{\sum\limits_{1 \leqslant j \leqslant n} \mathsf{P}(A \mid B_j)\,\mathsf{P}(B_j)}. \tag{1.2.7}$$

Доказательство состоит в непосредственном применении определения и формулы полной вероятности: $\mathsf{P}(B_i|A) = \dfrac{\mathsf{P}(A \cap B_i)}{\mathsf{P}(A)}$, $\mathsf{P}(A \cap B_i) = \mathsf{P}(A|B_i)\,\mathsf{P}(B_i)$ и $\mathsf{P}(A) = \sum\limits_j \mathsf{P}(A \mid B_j)\,\mathsf{P}(B_j)$.

Стандартная интерпретация равенства (1.2.7) такова: эта формула связывает *апостериорную вероятность* $\mathsf{P}(B_i \mid A)$ (условную вероятность по A) с *априорными вероятностями* $\{\mathsf{P}(B_j)\}$ (известными до того, как мы узнаем, что событие A произошло).

При жизни Байес завершил только две работы. Одна из них была по теологии. Вторая носила название «Опыт решения задачи в терминах шансов»; эта последняя содержала теорему Байеса и была опубликована спустя два года после смерти автора. Тем не менее, Байес был избран членом Королевского общества. Теория Байеса (включающая в себя указанную теорему как важную составную часть) долгое время была предметом споров. Его взгляды были полностью признаны (после значительных теоретических разъяснений) только в конце XIX в.

Задача 1.2.1. Из мышиного выводка, содержащего двух белых мышей, наугад взяты (без возвращения) четыре мыши. Вероятность того, что взяты обе белые мыши, в два раза больше, чем вероятность того, что не взята ни одна. Сколько мышей в выводке?

Решение. Пусть число мышей в выводке равно n. Введем обозначения: $\mathsf{P}(2) = \mathsf{P}$(выбраны две белые мыши) и $\mathsf{P}(0) = \mathsf{P}$(ни одна белая мышь не выбрана). Тогда

$$\mathsf{P}(2) = \frac{C_{n-2}^2}{C_n^4}.$$

Иным способом $\mathsf{P}(2)$ можно вычислить так:

$$\frac{2}{n}\cdot\frac{1}{n-1} + \frac{2}{n}\cdot\frac{n-2}{n-1}\cdot\frac{1}{n-2} + \frac{2}{n}\cdot\frac{n-2}{n-1}\cdot\frac{n-3}{n-2}\cdot\frac{1}{n-3} + \frac{n-2}{n}\cdot\frac{2}{n-1}\cdot\frac{1}{n-2} + \\ + \frac{n-2}{n}\cdot\frac{n-3}{n-1}\cdot\frac{2}{n-2}\cdot\frac{1}{n-3} + \frac{n-2}{n}\cdot\frac{2}{n-1}\cdot\frac{n-3}{n-2}\cdot\frac{1}{n-3} = \frac{12}{n(n-1)}.$$

С другой стороны,

$$\mathsf{P}(0) = \frac{C_{n-2}^4}{C_n^4}.$$

Иначе $\mathsf{P}(0)$ можно вычислить следующим образом:

$$\mathsf{P}(0) = \frac{n-2}{n}\,\frac{n-3}{n-1}\,\frac{n-4}{n-2}\,\frac{n-5}{n-3} = \frac{(n-4)(n-5)}{n(n-1)}.$$

Решая уравнение

$$\frac{12}{n(n-1)} = 2\frac{(n-4)(n-5)}{n(n-1)},$$

мы получим $n = \frac{9 \pm 5}{2}$. Решение $n = 2$ отбрасываем, поскольку $n \geqslant 6$ (в противном случае вторая вероятность равна 0). Следовательно, $n = 7$. □

Задача 1.2.2. Лорд Вайл любит выпить виски, количество выпитого за день случайно, но известно, что он может выпить за день n стаканов с вероятностью $\frac{1}{n!}e^{-1}$, $n = 0, 1, \ldots$ Вчера его жена леди Вайл, его сын Лидделл и его дворецкий решили убить лорда. Если он не пил виски в этот день, его должна была убить леди Вайл; если он выпил ровно один стакан, совершение убийства выпадало Лидделлу, в противном случае это должен был осуществить дворецкий. В два раза более вероятно, что леди Вайл прибегнет к отравлению, чем к удушению, дворецкий, напротив, выберет удушение с вероятностью в два раза большей, чем отравление, а Лидделл с равной вероятностью может выбрать любой из этих способов. Вопреки всем их усилиям, нет никакой гарантии, что лорд Вайл наверняка умрет в результате какой-нибудь из попыток убить его, однако в три раза более вероятно, что он станет жертвой удушения, чем отравления.

Сегодня лорд Вайл мертв. Какова вероятность того, что его убил дворецкий?

Решение. Запишем

$$\mathrm{P}(\text{м}\,|\,\text{у}) = \mathrm{P}(\text{мертв}\,|\,\text{удушен}) = 3r, \qquad \mathrm{P}(\text{м}\,|\,\text{о}) = \mathrm{P}(\text{мертв}\,|\,\text{отравлен}) = r$$

и

$$\mathrm{P}(\text{не пил виски}) = \mathrm{P}(\text{выпил один стакан}) = \frac{1}{e},$$

$$\mathrm{P}(\text{выпил два стакана или более}) = 1 - \frac{2}{e}.$$

Далее,

$$\mathrm{P}(\text{у}\,|\,\text{лВ}) = \mathrm{P}(\text{удушен}\,|\,\text{леди Вайл}) = \frac{1}{3},$$

$$\mathrm{P}(\text{о}\,|\,\text{лВ}) = \mathrm{P}(\text{отравлен}\,|\,\text{леди Вайл}) = \frac{2}{3},$$

$$\mathrm{P}(\text{у}\,|\,\text{д}) = \mathrm{P}(\text{удушен}\,|\,\text{дворецкий}) = \frac{2}{3},$$

$$\mathrm{P}(\text{о}\,|\,\text{д}) = \mathrm{P}(\text{отравлен}\,|\,\text{дворецкий}) = \frac{1}{3}$$

и

$$\mathrm{P}(\text{у}\,|\,\text{Лдл}) = \mathrm{P}(\text{удушен}\,|\,\text{Лидделл}) = \mathrm{P}(\text{о}\,|\,\text{Лдл}) = \mathrm{P}(\text{отравлен}\,|\,\text{Лидделл}) = \frac{1}{2}.$$

Тогда

$$\mathrm{P}(\text{д}\,|\,\text{м}) = \frac{\mathrm{P}(\text{м}\,|\,\text{д})\,\mathrm{P}(\text{д})}{\mathrm{P}(\text{м}\,|\,\text{д})\,\mathrm{P}(\text{д}) + \mathrm{P}(\text{м}\,|\,\text{лВ})\,\mathrm{P}(\text{лВ}) + \mathrm{P}(\text{м}\,|\,\text{Лдл})\,\mathrm{P}(\text{Лдл})} =$$

$$= \frac{\left(1-\frac{2}{e}\right)\left(\frac{3r\cdot 2}{3}+\frac{r}{3}\right)}{\left(1-\frac{2}{e}\right)\left(\frac{3r\cdot 2}{3}+\frac{r}{3}\right)+\frac{1}{e}\left(\frac{3r}{3}+\frac{r\cdot 2}{3}\right)+\frac{1}{e}\left(\frac{3r}{2}+\frac{r}{2}\right)},$$

что равно $\dfrac{e-2}{e-3/7} \approx 0{,}3137$. □

Задача 1.2.3. а) Докажите теорему Байеса.

б) На станции находятся три телефонных автомата, которые принимают монеты по 20 пенсов. Один из них никогда не работает, второй работает всегда, а третий работает с вероятностью 1/2. Отправляясь в столицу на целый день, я хотел бы определить, какой из телефонов надежный, чтобы по возвращении я мог им воспользоваться. На станции ни души, и у меня есть всего лишь три монеты в 20 пенсов. Я пробую один телефон, и он не работает, затем два раза подряд пробую звонить по другому телефону, и оба раза он работает. Какова вероятность того, что этот второй телефон является надежным?

Решение. б) Пусть A — это рассматриваемое событие: первый из опробованных телефонов не работал, а второй дважды работал. Очевидно, что

$$\mathrm{P}(A\,|\,\text{первый надежный}) = 0,$$

$$\mathrm{P}(A\,|\,\text{второй надежный}) =$$
$$\mathrm{P}(\text{первый никогда не работает}\,|\,\text{второй надежный}) +$$
$$+\frac{1}{2}\cdot\mathrm{P}(\text{первый работает наполовину}\,|\,\text{второй надежный}) = \frac{1}{2}+\frac{1}{2}\cdot\frac{1}{2}=\frac{3}{4}$$

и

$$\mathrm{P}(A\,|\,\text{третий надежный}) = \frac{1}{2}\cdot\frac{1}{2}\times$$
$$\times\,\mathrm{P}(\text{второй работает наполовину}\,|\,\text{третий надежный}) = \frac{1}{8}.$$

Тогда искомая вероятность равна

$$\frac{1/3\cdot 3/4}{1/3\cdot(0+3/4+1/8)} = \frac{6}{7}.$$ □

Задача 1.2.4. В парламенте имеется доля p членов партии лейбористов, которые неспособны менять свое мнение ни в одном вопросе, и доля $1-p$ членов партии консерваторов, которые случайным образом, с вероятностью r, могут менять свое мнение при последовательных голосованиях по одному и тому же вопросу. Выбранный наудачу парламентарий дважды проголосовал, не меняя своего мнения. Найдите вероятность того, что и при следующем голосовании он будет придерживаться того же мнения.

Решение. Введем следующие события:

$$A_1 = \{\text{выбран лейборист}\}, \quad A_2 = \{\text{выбран консерватор}\},$$
$$B = \{\text{выбранный парламентарий дважды проголосовал одинаково}\}.$$

Имеем $\mathsf{P}(A_1) = p$, $\mathsf{P}(A_2) = 1 - p$, $\mathsf{P}(B \mid A_1) = 1$, $\mathsf{P}(B \mid A_2) = 1 - r$. Мы хотим вычислить

$$\mathsf{P}(A_1 \mid B) = \frac{\mathsf{P}(A_1 \cap B)}{\mathsf{P}(B)} = \frac{\mathsf{P}(A_1)\,\mathsf{P}(B \mid A_1)}{\mathsf{P}(B)}$$

и $\mathsf{P}(A_2 \mid B) = 1 - \mathsf{P}(A_1 \mid B)$. Запишем

$$\mathsf{P}(B) = \mathsf{P}(A_1)\,\mathsf{P}(B \mid A_1) + \mathsf{P}(A_2)\,\mathsf{P}(B \mid A_2) = p \cdot 1 + (1-p)(1-r).$$

Тогда

$$\mathsf{P}(A_1 \mid B) = \frac{p}{p + (1-r)(1-p)}, \quad \mathsf{P}(A_2 \mid B) = \frac{(1-r)(1-p)}{p + (1-r)(1-p)},$$

и мы получаем ответ:

$$\mathsf{P}(\text{парламентарий проголосует, не изменяя своего мнения} \mid B) = \\ = \frac{p + (1-r)^2(1-p)}{p + (1-r)(1-p)}. \quad \square$$

Задача 1.2.5. Урновая схема Пойя задается следующим образом. В начальный момент времени мы имеем урну, содержащую один белый шар и один черный. Каждую секунду мы наугад извлекаем из урны один шар и возвращаем его обратно, добавляя при этом еще один шар того же цвета. Вычислите вероятность того, что, когда в урне будет находиться n шаров, i из них будут белого цвета.

Решение. Обозначим через P_n условную вероятность при условии, что в урне находится n шаров. При $n = 2$ и 3 имеем

$$\mathsf{P}_n(\text{один шар белый}) = \begin{cases} 1, & \text{если } n = 2, \\ \frac{1}{2}, & \text{если } n = 3, \end{cases}$$

и

$$\mathsf{P}_n(\text{два шара белые}) = \frac{1}{2}, \quad \text{если } n = 3.$$

Для доказательства по индукции введем следующее предположение:

$$\mathsf{P}_k(i \text{ белых шаров}) = \frac{1}{k-1}$$

для любых $k=2, \ldots, n-1$ и $i=1, \ldots, k-1$. Тогда после $n-1$ шага (когда число шаров в урне равно n) имеем

$$\begin{aligned}\mathrm{P}_n(i \text{ белых шаров}) = \mathrm{P}_{n-1}(i-1 \text{ белый шар}) \cdot \frac{i-1}{n-1} &+ \\ + \mathrm{P}_{n-1}(i \text{ белых шаров}) \cdot \frac{n-1-i}{n-1} &= \frac{1}{n-1}, \quad i=1, \ldots, n-1.\end{aligned}$$

Следовательно,

$$\mathrm{P}_n(i \text{ белых шаров}) = \frac{1}{n-1}, \quad i=1, 2, \ldots, n-1. \qquad \square$$

Задача 1.2.6. Имеется n урн, r-я урна содержит $r-1$ красных шаров и $n-r$ голубых шаров. Вы наугад выбираете урну (с равной вероятностью любую из урн) и извлекаете два шара без возвращения. Найдите вероятность того, что первый шар голубой, и условную вероятность того, что второй шар голубой, при условии, что первый шар красный.

Указание. Вам, возможно, понадобится тождество

$$\sum_{i=1}^{n-1} i(i-1) = \frac{1}{3}n(n-1)(n-2).$$

Решение. Каждая урна содержит $n-1$ шар, и с равной вероятностью первым может быть извлечен любой шар. Следовательно, мы получаем

$$\mathrm{P}(\text{первый шар голубой}) = \frac{\text{число голубых шаров}}{\text{общее число шаров}} = \frac{1}{2}$$

или, если применить формулу полной вероятности,

$$\begin{aligned}\mathrm{P}(\text{первый шар голубой}) = \sum_{1 \leqslant l \leqslant n} \mathrm{P}(\text{первый шар голубой} \mid \text{выбрана урна } l) &\times \\ \times \frac{1}{n} = \frac{1}{n} \sum_{1 \leqslant l \leqslant n} \frac{n-l}{n-1} = \frac{1}{n(n-1)} \sum_{l=1}^{n-1} l &= \frac{1}{2}.\end{aligned}$$

Итак, $\mathrm{P}(\text{первый шар красный}) = 1 - 1/2 = 1/2$. Теперь

$$\begin{aligned}&\mathrm{P}(\text{первый шар красный, второй шар голубой}) = \\ &= \sum_l \mathrm{P}(\text{первый шар красный, второй шар голубой} \mid \text{выбрана урна } l) \cdot \frac{1}{n} = \\ &= \frac{1}{n} \sum_l \frac{(l-1)(n-l)}{(n-1)(n-2)} = \frac{1}{n(n-1)(n-2)} \left[n \sum_{l=1}^{n} (l-1) - \sum_{l=1}^{n} l(l-1) \right] = \\ &= \frac{1}{n(n-1)(n-2)} \left[\frac{n(n-1)n}{2} - \frac{(n+1)n(n-1)}{3} \right] = \frac{n(n-1)}{n(n-1)(n-2)} \left(\frac{n}{2} - \frac{n+1}{3} \right) = \frac{1}{6}.\end{aligned}$$

Следовательно,

$$\mathrm{P}(\text{второй шар голубой} \mid \text{первый шар красный}) = \\ = \frac{\mathrm{P}(\text{первый шар красный, второй шар голубой})}{\mathrm{P}(\text{первый шар красный})} = \frac{1}{6} : \frac{1}{2} = \frac{1}{3}. \qquad \square$$

Задача 1.2.7. Власти Руритании приняли решение помиловать и освободить одного из трех заключенных X, Y и Z, которые содержатся в строгой изоляции в печально известной тюрьме Альказаф. Заключенные знают об этом, однако не могут догадаться, кто же из них этот счастливчик; ожидание мучительно. Сочувствующий, но продажный тюремщик подходит к заключенному X и предлагает, за определенную плату, назвать имя другого заключенного (не X), который приговорен остаться. Тюремщик говорит: «Это уменьшит твои шансы остаться здесь с 2/3 до 1/2: разве это не заставит тебя почувствовать себя лучше?» После некоторых колебаний X соглашается на предложение; тюремщик называет имя Y.

Предположим, что и в самом деле Y не будет освобожден. Определите условную вероятность

$$\mathrm{P}(\text{X остается} \mid \text{Y назван}) = \frac{\mathrm{P}(\text{X и Y остаются})}{\mathrm{P}(\text{Y назван})}$$

и таким образом проверьте утверждение тюремщика в следующих трех случаях:

а) когда тюремщик полностью беспристрастен (т. е. назовет любого из двух, Y или Z, с вероятностью 1/2, если пара Y, Z должна остаться в заключении),

б) когда он ненавидит Y и определенно назовет его, если Y должен остаться в заключении,

в) когда он ненавидит Z и определенно назовет его, если Z должен остаться в заключении.

Решение. В силу симметрии безусловная вероятность определяется так:

$$\mathrm{P}(\text{X освобожден}) = \frac{1}{3}, \quad \mathrm{P}(\text{X остается}) = \frac{2}{3},$$

и то же самое для Y, Z. Чтобы подсчитать условную вероятность, запишем

$$\mathrm{P}(\text{X остается} \mid \text{Y назван}) = \frac{\mathrm{P}(\text{X и Y остаются})}{\mathrm{P}(\text{Y назван})},$$

где $\mathrm{P}(\text{X и Y остаются}) = 1/3$.

Теперь для случая а) имеем

$$\mathrm{P}(\text{Y назван}) = \mathrm{P}(\text{X и Y остаются}) + \frac{1}{2}\mathrm{P}(\text{Y и Z остаются}) = \frac{1}{3} + \frac{1}{6} = \frac{1}{2}.$$

Следовательно, в случае а) получаем

$$\mathsf{P}(\text{X остается} \mid \text{Y назван}) = \frac{2}{3},$$

и тюремщик неправ: его информация не меняет шансы заключенного X.

Для случая б) имеем

$$\mathsf{P}(\text{Y назван}) = \mathsf{P}(\text{X и Y остаются}) + \mathsf{P}(\text{Y и Z остаются}) = \frac{2}{3}$$

и

$$\mathsf{P}(\text{X остается} \mid \text{Y назван}) = \frac{1}{2},$$

и тюремщик прав.

Наконец, для случая в) имеем

$$\mathsf{P}(\text{Y назван}) = \mathsf{P}(\text{X и Y остаются}) = \frac{1}{3}.$$

Таким образом, в этом случае

$$\mathsf{P}(\text{X остается} \mid \text{Y назван}) = 1,$$

и тюремщик неправ. □

Задача 1.2.8. Вы принимаете участие в игровом шоу и должны выбрать одну из трех дверей. За одной из них скрывается автомобиль; за двумя другими — по козе. Вы решили выбрать дверь 1, а ведущий шоу открывает дверь 3, за которой оказывается коза. Ведущий спрашивает, не желаете ли Вы изменить свое решение и выбрать дверь 2. Следует ли Вам менять свое решение?

Решение. Общепринято решать эту задачу в предположении, что ведущий шоу знает, за какой из дверей находится автомобиль, и поэтому открывает дверь не наугад, а именно одну из тех, за которой находится коза. На самом деле весьма поучительным будет рассмотреть оба случая: 1) когда ведущий знает, где находится автомобиль, и 2) когда он этого не знает. Во втором случае ваши шансы остаются прежними, в противном же случае вам следует менять свое решение. Действительно, рассмотрим следующие события:

$$Y_i = \{\text{вы выбираете дверь } i\},\ C_i = \{\text{автомобиль находится за дверью } i\},$$
$$H_i = \{\text{ведущий открывает дверь } i\},\ G_i = \{\text{коза находится за дверью } i\}.$$

Затем мы хотим подсчитать вероятность

$$\mathsf{P}(C_1 \mid Y_1 \cap H_3 \cap G_3) = \frac{\mathsf{P}(C_1 \cap H_3 \cap G_3 \mid Y_1)}{\mathsf{P}(H_3 \cap G_3 \mid Y_1)}.$$

Если ведущий знает, где находится автомобиль, то $H_3 \subseteq G_3$ (т. е. он открывает дверь 3, только если за ней коза) и $G_1 \cap H_3 \cap Y_1 = G_1 \cap G_3 \cap Y_1$

(т. е. если вы выбираете дверь 1, за которой находится коза, то он открывает дверь 3 всякий раз, когда за ней находится коза). Также имеем $\mathsf{P}(C_1 \mid Y_1) = 1/3$ и $\mathsf{P}(H_3 \mid Y_1 \cap C_1) = 1/2$ в силу симметрии.

В этом случае числитель $\mathsf{P}(C_1 \cap H_3 \cap G_3 \mid Y_1)$ равен

$$\mathsf{P}(C_1 \cap H_3 \mid Y_1) = \mathsf{P}(C_1 \mid Y_1)\,\mathsf{P}(H_3 \mid Y_1 \cap C_1) = \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6}.$$

С другой стороны, знаменатель $\mathsf{P}(H_3 \cap G_3 \mid Y_1)$ равен

$$\mathsf{P}(H_3 \mid Y_1) = \mathsf{P}(G_1 \cap H_3 \mid Y_1) + \mathsf{P}(C_1 \cap H_3 \mid Y_1) = \frac{\mathsf{P}(G_1 \cap H_3 \cap Y_1)}{\mathsf{P}(Y_1)} + \frac{1}{6}.$$

Первое слагаемое дает нам $\mathsf{P}(G_1 \cap G_3 \cap Y_1)/\mathsf{P}(Y_1) = \mathsf{P}(C_2 \mid Y_1) = 1/3$, опять в силу симметрии. Следовательно,

$$\mathsf{P}(C_1 \mid Y_1 \cap H_3 \cap G_3) = \frac{1}{6} : \left(\frac{1}{3} + \frac{1}{6}\right) = \frac{1}{3},$$

и $\mathsf{P}(C_2 \mid Y_1 \cap H_3 \cap G_3) = 1 - \mathsf{P}(C_1 \mid Y_1 \cap H_3 \cap G_3) = 1 - 1/3 = 2/3$.

Если ведущий не знает, где находится автомобиль, тогда события C_1, Y_1 и H_3 независимы. Более того, в силу симметрии

$$\mathsf{P}(C_1 \mid Y_1 \cap H_3 \cap G_3) = \frac{1}{2},$$

т. е. ваши шансы остаются прежними. □

Продолжая наше изложение, введем определение независимых событий. Понятие независимости было важным изобретением в теории вероятностей. Оно способствовало формированию теории на раннем этапе и считается одной из главных особенностей, определяющих место теории вероятностей в рамках более общей теории — теории меры.

Будем говорить, что события A и B являются *независимыми*, если

$$\mathsf{P}(A \cap B) = \mathsf{P}(A)\,\mathsf{P}(B). \tag{1.2.8}$$

Удобным критерием независимости является следующий: события A и B, для которых, скажем, $\mathsf{P}(B) > 0$, являются независимыми тогда и только тогда, когда $\mathsf{P}(A \mid B) = \mathsf{P}(A)$, т. е. знание того, что событие B произошло, не меняет вероятности события A.

Тривиальные примеры — пустое множество (невозможное событие) $\varnothing$ и все множество Ω (достоверное событие): они являются независимыми от любого события. Для следующего примера рассмотрим четыре элементарных события 00, 01, 10 и 11, которым приписаны вероятности 1/4; события $A = \{$первая цифра равна 1$\}$ и $B = \{$вторая цифра равна 0$\}$ являются независимыми: $\mathsf{P}(A) = p_{10} + p_{11} = 1/2$, $\mathsf{P}(B) = p_{10} + p_{00} = 1/2$ и $\mathsf{P}(A \cap B) = p_{10} = 1/4 = 1/2 \cdot 1/2$. Также и события $\{$первая цифра равна 0$\}$

и {обе цифры равны} являются независимыми, тогда как события {первая цифра равна 0} и {сумма цифр > 0} являются зависимыми.

Этот пример можно легко переформулировать в терминах задачи о подбрасывании двух симметричных монет. Важным свойством является то, что если A и B независимы, то $\bar{A}$ и B также независимы:

$$\mathsf{P}(\bar{A} \cap B) = \mathsf{P}(B \setminus (A \cap B)) = \mathsf{P}(B) - \mathsf{P}(A \cap B) = \quad \text{(в силу независимости)}$$
$$= \mathsf{P}(B) - \mathsf{P}(A)\,\mathsf{P}(B) = (1 - \mathsf{P}(A))\,\mathsf{P}(B) = \mathsf{P}(\bar{A})\,\mathsf{P}(B).$$

Далее, если а) события A_1 и B независимы, б) A_2 и B независимы и в) A_1 и A_2 несовместны, то события $A_1 \cup A_2$ и B независимы. Если выполняются условия а) и б) и $A_1 \subset A_2$, то события B и $A_2 \setminus A_1$ являются независимыми.

На интуитивном уровне независимость часто ассоциируется с «отсутствием всякой связи» между событиями. Существует известная шутка, связанная с именем А. Н. Колмогорова (1903—1987), знаменитого русского математика, которого считают отцом современной теории вероятностей. Его монография [Ко], которая первоначально появилась в Германии в 1933 г., сыграла революционную роль в понимании основ теории вероятностей и ее значения для математики и приложений.

Когда в 1930-х гг. эта монография была переведена на русский язык, советское правительство заинтересовалось вопросом, какова природа понятия независимости событий. Председатель Совнаркома желал знать, согласуется ли это понятие с материалистическим детерминизмом, ядром марксистско-ленинской философии, а также какие есть примеры таких событий. Колмогоров должен был найти выход из этого затруднительного положения и при этом проявить осторожность, как показывали последующие события (такие как постыдное осуждение властями генетики как «реакционной буржуазной лженауки»). Легенда гласит, что без единой секунды колебаний Колмогоров ответил: «Хорошо, давайте представим, что одна из отдаленных деревень пострадала от длительной засухи. И однажды доведенные до отчаяния местные крестьяне отправились в церковь, и священник предложил им помолиться о дожде. И на следующий день выпал дождь! Вот независимые события».

Попытаемся формализовать понятие независимости событий. Полезно посмотреть на понятие независимости как на геометрическое свойство. В вышеприведенном примере четыре вероятности p_{00}, p_{01}, p_{10} и p_{11} могут быть приписаны вершинам $\{(0; 0), (0; 1), (1; 0), (1; 1)\}$ единичного квадрата. Каждую из этих четырех точек можно спроектировать на горизонтальную и вертикальную оси. Проекциями являются точки 0 и 1 на каждой из осей, и любая вершина единственным образом может быть задана своими проекциями. Если спроектированные точки имеют вероятностную массу $1/2$ на каждой из осей, то каждая вершина будет иметь массу $p_{ij} = 1/2 \cdot 1/2 = 1/4$, $i, j = 0, 1$. В этом случае говорят, что четырехточечное вероятностное распределение $\{1/4, 1/4, 1/4, 1/4\}$ является произведением двух двухточечных распределений $\{1/2, 1/2\}$. Легко представить себе похожую ситуацию, когда существуют m точек по горизонтали и n по вертикали: в этом случае мы имеем mn пар (i, j) (узлы решетки), где

$i=0, \ldots, m-1$, $j=0, \ldots, n-1$, и каждая пара имеет вероятность $1/(mn)$. Более того, равномерное распределение может быть заменено более общим законом: $p_{ij} = p_i^{(1)} p_j^{(2)}$, где $p_i^{(1)}$, $i = 0, \ldots, m-1$, и $p_j^{(2)}$, $j = 0, \ldots, n-1$, являются вероятностными распределениями для двух проекций. Тогда любое событие, выраженное в терминах горизонтальной проекции (например, {первая цифра делится на 3}), не зависит от любого события, выраженного в терминах вертикальной проекции (например, {вторая цифра не превосходит $n/2$}). Этот пример является основным (а в некотором смысле и единственным) примером независимости.

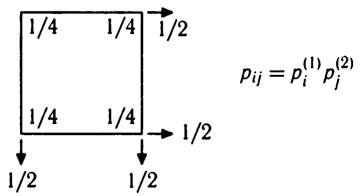


Рис. 1.1

Мы говорим, что события $A_1, \ldots, A_n$ взаимно независимы (или, коротко, независимы), если для любых подсистем $A_{i_1}, \ldots, A_{i_l}$ выполняется равенство

$$P(A_{i_1} \cap \ldots \cap A_{i_l}) = P(A_{i_1}) \ldots P(A_{i_l}). \qquad (1.2.9)$$

Это определение включает в себя всю совокупность $A_1, \ldots, A_n$. Важно отличать эту ситуацию от ситуации, когда формула (1.2.9) справедлива только для некоторых подсистем, например для пар A_i, A_j, $1 \leqslant i < j \leqslant n$, или для всей системы $A_1, \ldots, A_n$ (см. задачи 1.2.15 и 1.2.16).

Данное выше определение приводит к важной модели — последовательности независимых испытаний с двумя или более результатами. Существует много примеров таких моделей, и очень важно разобраться в них в начале нашего курса.

До сих пор мы предполагали, что общее число исходов ω конечно; в этой главе мы рассмотрим обобщение на случай, когда Ω — счетное множество, состоящее из точек $x_1, x_2, \ldots$, скажем, с данными вероятностями $p_i = P(\{x_i\})$, $i = 1, 2, \ldots$ Конечно, нумерация результатов может быть различной, например $i \in \mathbb{Z}$, где $\mathbb{Z}$ — множество целых чисел. Как и раньше, $p_i \geqslant 0$ для всех i, и $\sum\limits_i p_i = 1$.

Мы также можем работать с бесконечными последовательностями событий. Например, уравнения (1.2.6) и (1.2.7) не изменятся:

$$P(A) = \sum_{1 \leqslant j < \infty} P(A \mid B_j) P(B_j), \quad P(B_i \mid A) = \frac{P(A \mid B_i) P(B_i)}{\sum\limits_{1 \leqslant j < \infty} P(A \mid B_j) P(B_j)}. \qquad (1.2.10)$$

Задача 1.2.9. При каждом подбрасывании монетка падает вверх орлом с вероятностью p. Пусть π_n — вероятность того, что число орлов после n подбрасываний четно. Показав, что $\pi_{n+1} = (1-p)\pi_n + p(1-\pi_n)$, $n \geqslant 1$, или иным способом найдите π_n. (Число 0 будем считать четным.)

Решение 1. Как всегда в моделях с подбрасыванием монетки, мы будем считать, что результаты при различных бросках независимы. Положим $A_n = \{$при n-м подбрасывании выпадает орел$\}$, $\mathsf{P}(A_n) = p$, и $B_n = \{$число орлов после n бросков четно$\}$, $\pi_n = \mathsf{P}(B_n)$. Тогда при условиях A_{n+1} и A^c_{n+1} получим

$$\mathsf{P}(B_{n+1}) = \mathsf{P}(B_{n+1} \cap A_{n+1}) + \mathsf{P}(B_{n+1} \cap A^c_{n+1}) = \\ = \mathsf{P}(B_{n+1} \mid A_{n+1})\,\mathsf{P}(A_{n+1}) + \mathsf{P}(B_{n+1} \mid A^c_{n+1})\,\mathsf{P}(A^c_{n+1}).$$

Далее, $B_{n+1} \cap A_{n+1} = B^c_n \cap A_{n+1}$ и $B_{n+1} \cap A^c_{n+1} = B_n \cap A^c_{n+1}$. В силу независимости

$$\mathsf{P}(B^c_n \cap A_{n+1}) = \mathsf{P}(B^c_n)\,\mathsf{P}(A_{n+1}) \quad \text{и} \quad \mathsf{P}(B_n \cap A^c_{n+1}) = \mathsf{P}(B_n)\,\mathsf{P}(A^c_{n+1}),$$

откуда следует, что

$$\mathsf{P}(B_{n+1}) = \mathsf{P}(B^c_n)\,\mathsf{P}(A_{n+1}) + \mathsf{P}(B_n)\,\mathsf{P}(A^c_{n+1}) = (1 - \mathsf{P}(B_n))p + \mathsf{P}(B_n)(1-p),$$

$\mathsf{P}(B_0) = 1$, т. е.

$$\pi_{n+1} = (1-p)\pi_n + p(1-\pi_n) = (1-2p)\pi_n + p.$$

Подставляя $\pi_n = a(1-2p)^n + b$, получим $b = 1/2$, а при условии $\pi_0 = 1$ получим $a = 1/2$. Тогда $\pi_n = \frac{1}{2}(1-2p)^n + \frac{1}{2}$.

Эту рекуррентную формулу можно получить и более коротким путем, взяв условные вероятности относительно B_n и B^c_n:

$$\pi_{n+1} = \mathsf{P}(B_{n+1}) = \mathsf{P}(B_{n+1} \cap B_n) + \mathsf{P}(B_{n+1} \cap B^c_n) = \\ = \mathsf{P}(A^c_{n+1} \mid B_n)\,\mathsf{P}(B_n) + \mathsf{P}(A_{n+1} \mid B^c_n)\,\mathsf{P}(B^c_n) = (1-p)\pi_n + p(1-\pi_n).$$

Использование рекуррентных уравнений (подобно тому как это было сделано только что) является удобным методом при решении многих задач.

Решение 2. (См. это решение после чтения § 1.5.) Пусть $X_i = 0$ или 1 — результат i-й попытки, а $Y_n = X_1 + \ldots + X_n$ — общее число успехов при n попытках. Тогда производящая функция распределения Y_n равна

$$\psi(s) = (ps + (1-p))^n.$$

Следовательно, вероятность того, что n попыток дадут четное число успехов, равна

$$\frac{1}{2}(\psi(1) + \psi(-1)) = \frac{1}{2}[1 + (1-2p)^n]. \qquad \square$$

Задача 1.2.10. Моя тетя Агата подарила мне на день рождения заводной апельсин. Я поставил его на середину обеденного стола, длина которого ровно 2 метра. Через минуту после этого апельсин издал громкий жужжащий звук, выпустил облачко зеленого дыма и стал двигаться то на 10 см к левому краю с вероятностью 3/5, то на 10 см к правому краю с вероятностью 2/5. Это продолжается (направление каждого движения не зависит от предыдущего движения) с интервалом в одну минуту, до тех пор пока апельсин не достигнет края стола, с которого он немедленно упадет. Если он упадет с левого края, он разобьет мою вазу времен династии Минь, также подарок тети Агаты. Если он упадет с правого края, он благополучно приземлится в ведро с водой. Какова вероятность того, что ваза уцелеет?

Решение. Положим $p_\ell = \mathrm{P}$(упадет с правого края | был на расстоянии 10ℓ см от левого края), $1-p_\ell = \mathrm{P}$(упадет с левого края | был на расстоянии 10ℓ см от левого края). Тогда $p_0 = 0$, $p_{20} = 1$ и $p_\ell = \frac{3}{5}p_{\ell-1} + \frac{2}{5}p_{\ell+1}$, или

$$p_{\ell+1} = \frac{5}{2}p_\ell - \frac{3}{2}p_{\ell-1}.$$

Иными словами, $(p_\ell, p_{\ell+1}) = (p_{\ell-1}, p_\ell)A$, где матрица A имеет вид

$$A = \begin{pmatrix} 0 & -\frac{3}{2} \\ 1 & \frac{5}{2} \end{pmatrix}.$$

Получаем

$$(p_\ell, p_{\ell+1}) = (p_0, p_1)A^\ell = (0, p_1)A^\ell,$$

т. е. p_ℓ будет линейной комбинацией ℓ-х степеней собственных значений матрицы A. Эти собственные значения равны $\lambda_1 = 3/2$ и $\lambda_2 = 1$, и, таким образом,

$$p_\ell = b_1\left(\frac{3}{2}\right)^\ell + b_2.$$

Так как $b_1 + b_2 = 0$ и $1 = b_1\left(\frac{3}{2}\right)^{20} + b_2$, получаем $b_1 = -b_2 = \left(\left(\frac{3}{2}\right)^{20} - 1\right)^{-1}$ и

$$p_{10} = \left(\frac{3}{2}\right)^{10}\frac{1}{\left(\frac{3}{2}\right)^{20} - 1} - \frac{1}{\left(\frac{3}{2}\right)^{20} - 1} = \frac{1}{\left(\frac{3}{2}\right)^{10} + 1}. \qquad \square$$

Задача 1.2.11. Дубровский всю ночь играет в карты с друзьями офицерами. Каждый раз, когда Дубровский ставит u рублей, есть вероятность r, что он выиграет и получит обратно $2u$ рублей (включая свою ставку). В начале игры у него было 8 000 рублей. Если бы он имел 256 000 рублей, он бы женился на красавице Наташе, вышел бы в отставку и жил бы в своем загородном имении. В противном случае он покончит с собой. Дубровский решил следовать одной из двух стратегий:

а) ставить 1000 рублей каждый раз до конца игры;

б) каждый раз ставить все.

Посоветуйте ему, что делать, если $r=1/4$ и если $r=3/4$. Каковы шансы на счастливый конец в каждом случае, если Дубровский будет следовать вашему совету?

Решение. Пусть p_ℓ — вероятность того, что Дубровский выиграет 256 000, если его стартовый капитал был ℓ тысяч рублей, при условии что он следовал стратегии а). Рассуждая так же, как и в задаче 1.2.10, получим

$$p_\ell = b_1\lambda_1^\ell + b_2\lambda_2^\ell,$$

где $\lambda_1 = \dfrac{1-r}{r}$ и $\lambda_2 = 1$ — собственные значения матрицы $A = \begin{pmatrix} 0 & \dfrac{1-r}{r} \\ 1 & \dfrac{1}{r} \end{pmatrix}$.

Граничные условия $p_0 = 0$ и $p_{256} = 1$ дают $b_1 = -b_2 = \left(\left(\dfrac{1-r}{r}\right)^{256} - 1\right)^{-1}$.

В случае $r = 1/4$, $(1-r)/r = 3$ Дубровскому следует выбрать стратегию б), так как

$$p_8 = \frac{3^8 - 1}{3^{256} - 1},$$

что очень мало по сравнению с $(1/4)^5$ — вероятностью выиграть 256 000 за 5 последовательных раундов, каждый раз ставя на кон все, что он выиграет.

В случае $r=3/4$, $(1-r)/r = 1/3$ имеем $p_8 = \dfrac{1-(1/3)^8}{1-(1/3)^{256}}$, что значительно больше, чем $(3/4)^5$. Поэтому ему следует выбрать стратегию а). □

Замечание. В задачах 1.2.10 и 1.2.11 одно из собственных чисел матрицы рекурсии A равно единице. Это не случайно и происходит благодаря тому, что в уравнении (1.2.6) (которое приводит к рассматриваемому уравнению рекурсии) $\sum\limits_j \mathsf{P}(B_j) = 1$.

Задача 1.2.12. Я играю в кости против «удачливого» Петти Джея следующим образом. При каждом подбрасывании я бросаю две кости. Если при первом броске я получаю 7 или 11, то я выиграл, а если это 2, 3 или 12, то я проиграл. Если при первом броске я не получил ничего подобного, то я последовательно подбрасываю кости, пока не получу то же число, что и при первом броске; если мне это удастся, то я выиграл, если я в процессе получу 7, то я проиграл. Какова вероятность того, что я выиграю?

Решение. Запишем

$$\mathsf{P}(\text{я выиграю}) = \mathsf{P}(\text{я выиграю при первом броске}) + \\ + \mathsf{P}(\text{я выиграю, но не на первом броске}).$$

Вероятность P(я выиграю при первом броске) считается очевидным образом и равна

$$\sum_{i,j=1}^{6} \frac{1}{36} I(i+j=7) + \sum_{i,j=1}^{6} \frac{1}{36} I(i+j=11) = \frac{6}{36} + \frac{2}{36} = \frac{2}{9},$$

где $I(i+j=m)$ равно 1, если $i+j=m$, и 0 в других случаях. Для второй вероятности имеем

$$\mathsf{P}(\text{я выиграю, но не на первом броске}) = \sum_{i=4,5,6,8,9,10} p_i Q_i.$$

Здесь

$$p_i = \mathsf{P}(\text{при первом броске я получу } i)$$

и

$$Q_i = \mathsf{P}(\text{при последовательных бросках я получу } i \text{ прежде, чем } 7 \mid \text{при первом броске я получу } i) = \\ = \mathsf{P}(\text{при последовательных подбрасываниях я получу } i \text{ прежде, чем } 7).$$

Тогда для Q_i, взяв условные вероятности по результату первого броска, получим

$$Q_i = p_i + (1 - p_i - p_7) Q_i, \quad \text{т. е.} \quad Q_i = \frac{p_i}{p_i + p_7},$$

или, что то же самое,

$$Q_i = p_i + (1 - p_i - p_7) p_i + (1 - p_i - p_7)^2 p_i + \ldots$$

Таким образом,

$$Q_4 = \frac{3/36}{3/36 + 6/36} = \frac{3}{3+6} = \frac{1}{3}, \quad Q_5 = \frac{4/36}{4/36 + 6/36} = \frac{4}{4+6} = \frac{2}{5},$$

и аналогично

$$Q_6 = \frac{5/36}{5/36 + 6/36} = \frac{5}{5+6} = \frac{5}{11}, \quad Q_8 = \frac{5}{11}, \quad Q_9 = \frac{2}{5}, \quad Q_{10} = \frac{1}{3},$$

что дает для P(я выиграю, но не на первом броске) значение

$$\frac{1}{12} \cdot \frac{1}{3} + \frac{1}{9} \cdot \frac{2}{5} + \frac{5}{36} \cdot \frac{5}{11} + \frac{5}{36} \cdot \frac{5}{11} + \frac{1}{9} \cdot \frac{2}{5} + \frac{1}{12} \cdot \frac{1}{3} = \frac{134}{495}.$$

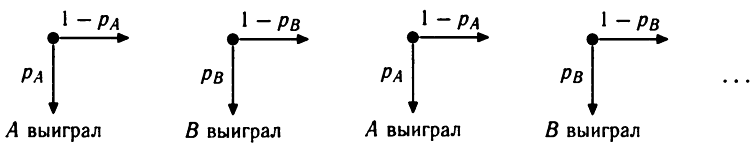
Тогда
$$\mathsf{P}(\text{я выиграю}) = \frac{2}{9} + \frac{134}{495} = \frac{244}{495}.$$

Интересно подсчитать среднюю продолжительность игры. Пусть X — результат первого броска. Тогда (см. определение математического ожидания в § 1.4)
$$\mathsf{E}[N] = \sum_{i=2}^{12} p_i \mathsf{E}[N \mid X = i].$$

Заметим, что $\mathsf{E}[N \mid X=i] = 1$ для $i=2, 3, 7, 11, 12$ и $\mathsf{E}[N \mid X=i] = 1 + \dfrac{1}{p_i + p_7}$ для $i = 4, 5, 6, 8, 9, 10$. Отсюда
$$\mathsf{E}[N] = \frac{557}{165} \approx 3{,}38. \qquad \square$$

Задача 1.2.13. Двое игроков бросают дротики в мишень, и выигрывает тот, кто попадет в центр первым. При каждом броске вероятность того, что игрок A попадет, равна p_A, и вероятность того, что игрок B попадет, равна p_B; результаты разных бросков независимы. Найдите вероятность того, что игрок A выиграет, при условии, что он начал первым.

Решение. Рассмотрим диаграмму



Если $Q = \mathsf{P}(A \text{ выиграл})$, то
$$Q = p_A + (1 - p_A)(1 - p_B)p_A + (1 - p_A)^2(1 - p_B)^2 p_A + \ldots =$$
$$= \frac{p_A}{1 - (1 - p_A)(1 - p_B)} = \frac{p_A}{p_A + p_B - p_A p_B}.$$

Или, при условии, что мы знаем результаты первого и второго бросков, имеем уравнение
$$\mathsf{P}(A \text{ выиграл}) = p_A + (1 - p_A)(1 - p_B)\,\mathsf{P}(A \text{ выиграл}),$$
решив которое, мы сразу получим требуемый результат. $\square$

Замечание. В задачах 1.2.12 и 1.2.13 мы использовали уравнение для вероятностей Q и Q_i, которые эквивалентны своим представлениям в виде рядов. Это еще одна полезная идея; например, она позволяет нам не использовать бесконечные пространства результатов (хотя скоро это станет неизбежным).

Задача 1.2.14. Симметричную монету подбрасывают до тех пор, пока не появится или последовательность OOO, что означает мой выигрыш, или последовательность POO, что означает ваш выигрыш. Какова вероятность того, что вы выиграете?

Решение. Вообще говоря, результатом игры может быть мой выигрыш или ваш выигрыш, либо игра будет продолжатся бесконечно. Заметим, что я выигрываю только в том случае, если OOO появится в начале: вероятность этого события $(1/2)^3 = 1/8$. Действительно, если OOO появляется не в начале, то должно появится POO и вы выигрываете. Но рано или поздно с вероятностью 1 появится OOO. Действительно, для любого N событие

$$A = \{OOO \text{ никогда не произойдет}\}$$

содержится в событии

$$A_N = \{OOO \text{ не выпадет среди первых } N \text{ последовательных троек}\}$$

(мы разбили первые $3N$ попыток на N последовательных троек). Таким образом, $\mathsf{P}(A) \leqslant \mathsf{P}(A_N)$. Но $\mathsf{P}(A_N) = (1 - 1/8)^N \to 0$ при $N \to \infty$. Следовательно, $\mathsf{P}(A) = 0$, и игра не может продолжаться до бесконечности, и вероятность того, что вы выиграете, равна $1 - 1/8 = 7/8$. □

Задача 1.2.15. 1. Приведите примеры

а) трех событий A, B, C, которые попарно независимы, но не независимы в совокупности;

б) трех событий, которые не независимы, но вероятность пересечения всех трех событий равна произведению вероятностей.

2. Три монетки при подбрасывании дают орел с вероятностью $3/5$ или решку. При подбрасывании первой монетки дается 10 очков за орла и 2 за решку, для второй монетки — по 4 за орла и решку, и для третьей монетки — 3 за орла и 20 за решку.

Вы и ваш соперник выбираете по монетке, причем вы не можете выбрать одну и ту же монетку. Каждый из вас бросает монетку один раз, и тот, у кого больше очков, выигрывает 10^{10} фунтов. Что вы предпочитаете — выбрать монетку первым или вторым?

Решение. 1. а) Подбросим две симметричные монетки, и пусть $A = \{\text{первый бросок дает } O\}$, $B = \{\text{второй бросок дает } O\}$, $C = \{\text{оба броска дают одинаковый результат}\}$. Тогда

$$\mathsf{P}(A \cap B) = p_{OO} = 1/4 = \mathsf{P}(A)\,\mathsf{P}(B),$$
$$\mathsf{P}(A \cap C) = p_{OO} = 1/4 = \mathsf{P}(A)\,\mathsf{P}(C),$$
$$\mathsf{P}(B \cap C) = p_{OO} = 1/4 = \mathsf{P}(B)\,\mathsf{P}(C)$$

и

$$\mathsf{P}(A \cap B \cap C) = p_{OO} = 1/4 \neq \mathsf{P}(A)\,\mathsf{P}(B)\,\mathsf{P}(C).$$

Другой способ: подбросим три кости, и пусть

$$A = \{\text{первая кость показывает нечетное число}\},$$
$$B = \{\text{вторая кость показывает нечетное число}\},$$
$$C = \{\text{сумма цифр на обоих костях нечетна}\}.$$

Тогда $\mathsf{P}(A) = \mathsf{P}(B) = \mathsf{P}(C) = 1/2$, $\mathsf{P}(A \cap B) = \mathsf{P}(A \cap C) = \mathsf{P}(B \cap C) = 1/4$, но $\mathsf{P}(A \cap B \cap C) = 0$.

б) Подбросим три монетки, и пусть

$$A = \{\text{первый бросок дает } O\}, \quad B = \{\text{третий бросок дает } O\},$$
$$C = \{OOO,\ OOP,\ OPP,\ PPP\} = \{\text{нет последовательных } PO\}.$$

Тогда $\mathsf{P}(A) = \mathsf{P}(B) = \mathsf{P}(C) = 1/2$,

$$A \cap B \cap C = \{OOO\} \quad \text{и} \quad \mathsf{P}(A \cap B \cap C) = 1/8 = (1/2)^3.$$

Но

$$A \cap C = \{OOO,\ OOP,\ OPP\}, \quad \mathsf{P}(A \cap C) = 3/8$$

и

$$B \cap C = \{OOO\} \quad \text{и} \quad \mathsf{P}(B \cap C) = 1/8.$$

Другой способ (как делает большинство студентов): пусть $A = \varnothing$, и возьмем любую зависимую пару событий B и C (скажем, $B = C$, $0 < \mathsf{P}(B) < 1$). Тогда $A \cap B \cap C = \varnothing$ и $\mathsf{P}(A \cap B \cap C) = 0 = \mathsf{P}(A)\,\mathsf{P}(B)\,\mathsf{P}(C)$, но $\mathsf{P}(B \cap C) \neq \mathsf{P}(B)\,\mathsf{P}(C)$.

2. Предположим, что я выбираю первую монетку, а вы — вторую; тогда $\mathsf{P}(\text{вы выиграли}) = 2/5$. Но если вы выбираете третью, то

$$\mathsf{P}(\text{вы выиграли}) = \frac{2}{5} + \frac{3}{5} \cdot \frac{2}{5} = \frac{16}{25} > \frac{1}{2}.$$

Аналогично если я выбираю вторую, а вы первую, то $\mathsf{P}(\text{вы выиграли}) = 3/5 > 1/2$. Наконец, если я выбираю третью монету, а вы вторую, то $\mathsf{P}(\text{вы выиграли}) = 3/5 > 1/2$. Значит, лучше всегда быть вторым. □

Задача 1.2.16. а) Пусть $A_1, \ldots, A_n$ — независимые события, и пусть $\mathsf{P}(A_i) < 1$. Докажите, что существует такое событие B, $\mathsf{P}(B) > 0$, что $B \cap A_i = \varnothing$ для $1 \leqslant i \leqslant n$.

б) Приведите пример трех событий A_1, A_2, A_3, зависимых, но таких, что для каждых $i \neq j$ события A_i и A_j независимы.

Решение. а) Обозначим через A^c дополнение события A; тогда

$$\mathsf{P}(A_1 \cup \ldots \cup A_n) = 1 - \mathsf{P}\left(\bigcap_{i=1}^{n} A_i^c\right) = 1 - \prod_{i=1}^{n} \mathsf{P}(A_i^c) < 1,$$

так как $\mathsf{P}(A_i^c) > 0\ \forall i$. Поэтому если $B = \left(\bigcup A_i\right)^c$, то $\mathsf{P}(B) > 0$ и $B \cap A_i = \varnothing\ \forall i$.

б) Положим

$$A_1 = \{1, 4\}, \quad A_2 = \{2, 4\}, \quad A_3 = \{3, 4\},$$

где $\mathsf{P}\{k\} = 1/4$, $k = 1, 2, 3, 4$. Тогда $\mathsf{P}(A_i) = 1/2$, $i = 1, 2, 3$, и

$$\mathsf{P}(A_i \cap A_j) = \mathsf{P}\{4\} = \left(\frac{1}{2}\right)^2 = \mathsf{P}(A_i)\,\mathsf{P}(A_j), \qquad 1 \leqslant i < j \leqslant 3,$$

но

$$\mathsf{P}(A_1 \cap A_2 \cap A_3) = \mathsf{P}\{4\} = \left(\frac{1}{2}\right)^2 \neq \left(\frac{1}{2}\right)^3 = \mathsf{P}(A_1)\,\mathsf{P}(A_2)\,\mathsf{P}(A_3). \qquad \square$$

Задача 1.2.17. Бросили две симметричные кости. Пусть A_s означает, что сумма чисел, которые выпали, равна s, и B_i означает, что на первой кости выпало i. Найдите, для каких значений s и i события A_s и B_i независимы.

Решение. Допустимыми значениями для s будут $2, \ldots, 12$, а для i — $1, \ldots, 6$. Имеем

$$\mathsf{P}(A_s) = \begin{cases} \dfrac{1}{36}(s-1), & 2 \leqslant s \leqslant 7, \\ \dfrac{1}{36}(6 - (s-6) + 1) = \dfrac{1}{36}(13 - s), & 8 \leqslant s \leqslant 12, \end{cases}$$

и

$$\mathsf{P}(B_i) = \frac{1}{6}, \qquad 1 \leqslant i \leqslant 6.$$

Очевидно, $\mathsf{P}(A_s \cap B_i) = 1/36$, если $1 \leqslant i \leqslant 6$, $i < s \leqslant i + 6$, и $\mathsf{P}(A_s \cap B_i) = 0$ в противном случае. Отсюда $\mathsf{P}(A_s \cap B_i) = \mathsf{P}(A_s)\,\mathsf{P}(B_i)$, когда или а) $s = 7$ и $1 \leqslant i \leqslant 6$, или б) по крайней мере одно из значений s и i недопустимо. $\square$

Замечание. Случай $s = 7$ особенный: тут i может принимать все допустимые значения $1, \ldots, 6$. Во всех остальных случаях значения s дают ограничения на допустимые значения i (например, если $s = 9$, то i не может быть равным 1 или 2). Это нарушает независимость.

Задача 1.2.18. Пусть n шаров случайно размещены в n ячейках. Найдите вероятность p_n того, что ровно 2 ячейки остались пустыми.

Решение. «Случайно» означает, что каждый шарик попадает в ячейку с вероятностью $1/n$, независимо от всех остальных шариков. Сначала рассмотрим случаи $n = 3$ и $n = 4$. Если $n = 3$, у нас будет одна ячейка с тремя шарами и 2 пустые. Таким образом,

$$p_3 = C_n^2 \cdot \frac{1}{n^n} = \frac{1}{9}.$$

Если $n = 4$, мы имеем две заполненные ячейки с двумя шарами в каждой (с вероятностью p_4') или одну заполненную ячейку с тремя шарами (с вероятностью p_4''). Поэтому

$$p_4 = p_4' + p_4'' = C_n^2 \cdot \frac{(n-2)}{n^n} \cdot \left(4 + \frac{n(n-1)}{4}\right) = \frac{21}{64}.$$

Тут 4 означает число способов выбрать три шара, которые попадут в одну ячейку, и $n(n-1)/4$ — число способов выбрать две пары шаров, которые попадут в данные ячейки.

В случае $n \geqslant 5$ для того чтобы иметь ровно две пустые ячейки, нужно или чтобы нашлись две ячейки, содержащие по два шара, и $n-4$ ячейки, содержащие по одному шару, или чтобы нашлась одна ячейка, содержащая 3 шара, и $n-3$ ячейки, содержащие по одному шару. Обозначим требуемые вероятности через p_n' и p_n'' соответственно. Тогда $p_n = p_n' + p_n''$. Далее,

$$p_n' = C_n^2 \cdot \frac{1}{2} C_n^2 C_{n-2}^2 \cdot \frac{n-2}{n}\frac{n-3}{n} \cdots \frac{1}{n} \cdot \left(\frac{1}{n}\right)^2 = \frac{1}{4}\frac{n!}{n^n} C_n^2 C_{n-2}^2.$$

Здесь первый множитель C_n^2 дает число способов, которыми можно выбрать две пустые ячейки из n возможных. Второй множитель $\frac{1}{2} C_n^2 C_{n-2}^2$ отвечает за выбор шара, который попадет в ячейку, уже содержащую два шара, а также за выбор ячейки. Наконец, третий множитель $\frac{n-2}{n}\frac{n-3}{n} \cdots \frac{1}{n}$ означает вероятность того, что $n-2$ шара попадут в $n-2$ ячейки по одному в каждую, и последний множитель $\left(\frac{1}{n}\right)^2$ означает вероятность того, что две пары шаров попадут в ячейки, предназначенные для двух шаров. Далее,

$$p_n'' = C_n^2 (n-2) \cdot C_n^3 \frac{(n-3)!}{n^n}.$$

Тут первый множитель C_n^2 отвечает за число способов выбрать две пустые ячейки из n возможных, второй множитель $(n-2)$ дает число способов, которыми можно выбрать ячейку с тремя шарами, третий множитель C_n^3 отвечает за число способов выбрать три шара, которые попадут в эту ячейку, и последний множитель описывает распределение всех шаров по соответствующим ячейкам. □

Задача 1.2.19. Вы играете в теннис с вашим противником, причем каждый раз подает он или вы. Если вы подаете, то вы выигрываете очко с вероятностью p_1, а если подает ваш противник — вы выигрываете с вероятностью p_2. Существует два возможных вида подачи:

а) поочередная;

б) игрок подает до тех пор, пока не теряет очко.

Вы подаете первым, и тот игрок, который выигрывает первым n очков, выигрывает игру. Покажите, что вероятность выигрыша не зависит от правил подачи а) и б).

Решение. Обе системы дают равные вероятности выигрыша. Действительно, предположим, что мы продолжим игру за пределы достижения результата, до тех пор пока у вас будет n подач, а у вашего противника $n-1$. (Согласно правилу а) вы просто продолжаете поочередные подачи или согласно правилу б) проигрывающий делает недостающее число подач.) Затем в любом случае, если вы выигрываете игру, вы выигрываете и продолженную игру, и наоборот (так как у вашего противника недостаточно очков, чтобы вас обогнать). Следовательно, достаточно рассмотреть продолженную игру.

Результатом этой игры будет последовательность $\omega = (\omega_1, \ldots, \omega_{2n-1})$ из $2n-1$ значений, скажем нулей (что означает, что вы проигрываете очко) и единиц (если вы выигрываете очко). Можно сказать, что $\omega_1, \ldots, \omega_n$ — результат ваших подач и $\omega_{n+1}, \ldots, \omega_{2n-1}$ — результат подач противника. Введем события

$$A_i = \{\text{вы выигрываете вашу } i\text{-ю подачу}\},$$
$$B_j = \{\text{вы выигрываете его } j\text{-ю подачу}\}.$$

Соответственно индикаторные функции этих событий равны

$$I_{A_i}(\omega) = \begin{cases} 1, & \omega \in A_i, \\ 0, & \omega \notin A_i, \ 1 \leqslant i \leqslant n, \end{cases}$$

и

$$I_{B_j}(\omega) = \begin{cases} 1, & \omega \in B_j, \\ 0, & \omega \notin B_j, \ 1 \leqslant j \leqslant n-1. \end{cases}$$

В обоих случаях а) и б) событие, состоящее в том, что вы выигрываете продолженную игру, равно

$$\left\{\omega = (\omega_1, \ldots, \omega_{2n-1}) \colon \sum_{1 \leqslant i \leqslant 2n-1} \omega_i \geqslant n\right\}$$

и вероятность результата ω равна

$$p_1^{\sum_i I_{A_i}(\omega)} (1-p_1)^{n-\sum_i I_{A_i}(\omega)} p_2^{\sum_j I_{B_j}(\omega)} (1-p_2)^{n-1-\sum_j I_{B_j}(\omega)}.$$

Поскольку это выражение не зависит от выбора правила, в обоих случаях вероятности совпадают. □

Замечание. Запись в терминах ω при таком решении была довольно удобной, и мы будем и далее ее использовать.

Задача 1.2.20. Ксерокс факультета состоит из трех частей A, B и C, которые могут поломаться. Вероятность того, что часть A сломается в процессе работы, равна 10^{-5}. Вероятность того, что поломается часть B, равна 10^{-1}, если сломается часть A, и 10^{-5} в противном случае. Вероятность того, что выйдет из строя часть C, равна 10^{-1} при условии, что сломается часть B, и 10^{-5}, если этого не произойдет. Сигнал «позвать мастера» загорается, если выйдут из строя две или три части. Если только две части выйдут из строя, я могу сам отремонтировать ксерокс, но если поломаны все три части, мои попытки принесут только вред. Если загорится сигнал «позвать мастера» и я готов рискнуть ухудшить ситуацию с вероятностью не более чем 1%, следует ли мне попытаться отремонтировать ксерокс, и почему?

Решение. Исходы имеют вид

Aс, Bс, Cс с вероятностью $10^{-5}\cdot 10^{-1}\cdot 10^{-1}=10^{-7}$,

Aс, Bр, Cс с вероятностью $10^{-5}\cdot 9\cdot 10^{-1}\cdot 10^{-5}=9\cdot 10^{-11}$,

Aс, Bс, Cр с вероятностью $10^{-5}\cdot 10^{-1}\cdot 9\cdot 10^{-1}=9\cdot 10^{-7}$,

Aр, Bс, Cс с вероятностью $(1-10^{-5})\cdot 10^{-5}\cdot 10^{-1}=(1-10^{-5})\cdot 10^{-6}$

(мы использовали сокращение «с» для «сломалась» и «р» для «работает»).

Значит,

$$\begin{aligned}
&\mathsf{P}(\text{три части сломались}\mid\text{две или три сломались})=\\
&=\frac{\mathsf{P}(\text{три сломались})}{\mathsf{P}(\text{две или три сломались})}=\\
&=\frac{\mathsf{P}(A,\,B,\,C\text{с})}{\mathsf{P}(A,\,B\text{с},\,C\text{р})+\mathsf{P}(A,\,C\text{с},\,B\text{р})+\mathsf{P}(B,\,C\text{с},\,A\text{р})+\mathsf{P}(A,\,B,\,C\text{с})}=\\
&=\frac{10^{-7}}{9\cdot 10^{-7}+9\cdot 10^{-11}+(1-10^{-5})\cdot 10^{-6}+10^{-7}}\approx\\
&\approx\frac{10^{-7}}{9\cdot 10^{-7}+10^{-6}+10^{-7}}=\frac{1}{20}>\frac{1}{100}.
\end{aligned}$$

Таким образом, вам не стоит пытаться отремонтировать ксерокс самостоятельно: вероятность ухудшить ситуацию составляет примерно 1/20. □

Задача 1.2.21. Я играю с моими родителями в теннис; вероятность того, что я выиграю, играя с мамой (М), равна p, а вероятность того, что я выиграю, играя с папой (П), равна q, где $0<q<p<1$. Мы договорились сыграть 3 игры, и их порядок может быть ПМП (здесь я играю с папой, потом с мамой, потом опять с папой) или МПМ. Результаты игр независимы.

При таких порядках игры вычислите вероятности следующих событий:

а) я выиграю по крайней мере одну игру,

б) я выиграю по крайней мере две игры,

в) я выиграю по крайней мере две последовательные игры (т. е. игры 1 и 2, или 2 и 3, или 1, 2 и 3),

г) я выиграю ровно две последовательные игры (т. е. игры 1 и 2 или 2 и 3, но не 1, 2 и 3),

д) я выиграю ровно две игры (т. е. 1 и 2, или 2 и 3, или 1 и 3, но не 1, 2 и 3).

В каждом из случаев а)—д) определите, какой порядок игры максимизирует вероятность события. В случае д) предположите, что $p+q>3pq$.

Решение. Имеем соотношения

а) $\mathrm{P}_{\text{ПМП}}(\text{выиграл} \geqslant 1) = 1-(1-q)^2\cdot(1-p) < 1-(1-p)^2(1-q) = \\ = \mathrm{P}_{\text{МПМ}}(\text{выиграл} \geqslant 1)$;

б) $\mathrm{P}_{\text{ПМП}}(\text{выиграл} \geqslant 2) = 2pq+q^2(1-2p) < 2pq+p^2(1-2q) = \\ = \mathrm{P}_{\text{МПМ}}(\text{выиграл} \geqslant 2)$;

в) $\mathrm{P}_{\text{ПМП}}(\text{выиграл} \geqslant 2 \text{ последовательно}) = 2pq-q^2p > 2pq-p^2q = \\ = \mathrm{P}_{\text{МПМ}}(\text{выиграл} \geqslant 2 \text{ последовательно})$

г) $\mathrm{P}_{\text{ПМП}}(\text{выиграл } 2 \text{ последовательно}) = 2pq-2q^2p > 2pq-2p^2q = \\ = \mathrm{P}_{\text{МПМ}}(\text{выиграл } 2 \text{ последовательно})$;

д) $\mathrm{P}_{\text{ПМП}}(\text{выиграл } 2) = 2pq+q^2(1-3p)$, $\mathrm{P}_{\text{МПМ}}(\text{выиграл } 2) = \\ = 2pq+p^2(1-3q)$.

При условии $p+q>3pq$ получаем

$$p^2(1-3q)-q^2(1-3p) = p^2-q^2+3pq(p-q) = (p-q)(p+q-3pq) > 0;$$

поэтому $q^2(1-3p) < p^2(1-3q)$ и $\mathrm{P}_{\text{ПМП}}(\text{выиграл } 2) < \mathrm{P}_{\text{МПМ}}(\text{выиграл } 2)$. □

Задача 1.2.22. В сумке лежат четыре монеты, каждая из которых при подбрасывании может приземлиться на любую из двух сторон. У одной из них орел на обеих сторонах, у остальных трех орел на одной стороне. Монетку выбирают случайным образом и три раза последовательно подбрасывают. Какова вероятность того, что при четвертом подбрасывании снова выпадет орел, если при предыдущих попытках тоже выпадал орел? Опишите пространство событий, которое вы используете, и аккуратно поясните вычисления.

Решение. Обозначим монетку с двумя орлами через 1, а остальные через 2, 3 и 4. Тогда пространство состояний Ω — это декартово произведение

$$\{1, 2, 3, 4\} \times \{O, P\}^4,$$

т. е. множество цепочек i; S_1, S_2, S_3, S_4, где i — это номер монеты, а $S_n = O$ или P — результат n-го подбрасывания, $n = 1, 2, 3, 4$. Имеем

$$\mathrm{P}(1;\, O, O, O, O) = \frac{1}{4} \quad \text{и} \quad \mathrm{P}(i;\, O, O, O, O) = \frac{1}{4} \cdot \frac{1}{2^4}, \qquad i = 2, 3, 4.$$

Это дает нам условную вероятность

$$\mathrm{P}(O, O, O, O \,|\, O, O, O) = \frac{\mathrm{P}(O, O, O, O)}{\mathrm{P}(O, O, O)} =$$

$$= \frac{\mathrm{P}(1;\, O, O, O, O) + \sum\limits_{i=2}^{4} \mathrm{P}(i;\, O, O, O, O)}{\mathrm{P}(1;\, O, O, O) + \sum\limits_{i=2}^{4} \mathrm{P}(i;\, O, O, O)} = \frac{\frac{1}{4} + 3 \cdot \frac{1}{4} \cdot \left(\frac{1}{2}\right)^4}{\frac{1}{4} + 3 \cdot \frac{1}{4} \cdot \left(\frac{1}{2}\right)^3} = \frac{2^4 + 3}{2^4 + 6} = \frac{19}{22}. \quad \square$$

§ 1.3. Формула включения-исключения. Задача о баллотировке

Естественный отбор — это механизм,
который генерирует в высшей степени невероятное.

Р. А. Фишер (1890—1962), британский статистик

Формула включения-исключения помогает вычислить вероятность $\mathrm{P}(A)$, где $A = A_1 \cup A_2 \ldots \cup A_n$ — объединение данного множества событий $A_1, \ldots, A_n$. Мы знаем (см. § 1.2), что если события попарно не пересекаются, то $\mathrm{P}(A) = \sum\limits_{1 \leqslant i \leqslant n} \mathrm{P}(A_i)$. В общем случае эта формула более сложна:

$$\mathrm{P}(A) = \mathrm{P}(A_1) + \ldots + \mathrm{P}(A_n) - \mathrm{P}(A_1 \cap A_2) - \mathrm{P}(A_1 \cap A_3) - \ldots - \mathrm{P}(A_{n-1} \cap A_n) +$$
$$+ \mathrm{P}(A_1 \cap A_2 \cap A_3) + \ldots + \mathrm{P}(A_{n-2} \cap A_{n-1} \cap A_n) + \ldots + (-1)^{n+1}\, \mathrm{P}(A_1 \cap \ldots \cap A_n) =$$
$$= \sum_{1 \leqslant k \leqslant n} (-1)^{k-1} \sum_{1 \leqslant i_1 < \ldots < i_k \leqslant n} \mathrm{P}\left(\bigcap_1^k A_{i_j}\right). \qquad (1.3.1)$$

Доказательство проводится прямым подсчетом: в правой части мы считаем каждое событие из $\bigcup\limits_{1 \leqslant i \leqslant n} A_i$ один раз. Но если мы возьмем сумму $\sum\limits_{1 \leqslant i \leqslant n} \mathrm{P}(A_i)$, то те события, которые входят больше чем в одно из $A_1, \ldots, A_n$, будут посчитаны более чем один раз. Если мы вычтем сумму $\sum\limits_{1 \leqslant i < j \leqslant n} \mathsf{P}(A_i \cap A_j)$, мы посчитаем точно один раз каждое событие, которое входит в два из событий $A_1, \ldots, A_n$, но возникнет проблема с теми событиями, которые входят в два и более событий. Поэтому мы должны прибавить $\sum\limits_{1 \leqslant i < j < k \leqslant n} \mathsf{P}(A_i \cap A_j \cap A_k)$, и т. д.

Формальное доказательство проводится индукцией по n. Удобно начать индукцию с $n = 2$ (для $n = 1$ формула тривиальна). Для двух событий A и B объединение $A \cup B$ совпадает с объединением непересекающихся событий $(A \setminus (A \cap B)) \cup (B \setminus (A \cap B)) \cup (A \cap B)$. Следовательно, $\mathsf{P}(A \cup B)$ может быть записано как

$$\begin{aligned}\mathsf{P}(A \setminus (A \cap B)) + \mathsf{P}(B \setminus (A \cap B)) + \mathsf{P}(A \cap B) = \mathsf{P}(A) - \mathsf{P}(A \cap B) + \\ + \mathsf{P}(B) - \mathsf{P}(A \cap B) + \mathsf{P}(A \cap B) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(A \cap B).\end{aligned}$$

По предположению индукции формула справедлива для всех множеств из n или менее событий. Для любого множества $A_1, \ldots, A_{n+1}$ из $n + 1$ событий вероятность $\mathsf{P}\left(\bigcup\limits_1^{n+1} A_i\right)$ равна

$$\begin{aligned}\mathsf{P}\left(\left(\bigcup_1^n A_i\right) \cup A_{n+1}\right) = \mathsf{P}\left(\bigcup_1^n A_i\right) + \mathsf{P}(A_{n+1}) - \mathsf{P}\left(\left(\bigcup_1^n A_i\right) \cap A_{n+1}\right) = \\ = \sum_{1 \leqslant k \leqslant n} (-1)^{k-1} \sum_{1 \leqslant i_1 < \ldots < i_k \leqslant n} \mathsf{P}\left(\bigcap_1^k A_{i_l}\right) + \mathsf{P}(A_{n+1}) - \mathsf{P}\left(\bigcup_1^n (A_i \cap A_{n+1})\right).\end{aligned}$$

Для последнего слагаемого мы получаем, снова применяя предположение индукции,

$$\begin{aligned}-\mathsf{P}\left(\bigcup_1^n (A_i \cap A_{n+1})\right) = \sum_{1 \leqslant k \leqslant n} (-1)^k \sum_{1 \leqslant i_1 < \ldots < i_k \leqslant n} \mathsf{P}\left(\bigcap_1^k (A_{i_j} \cap A_{n+1})\right) = \\ = \sum_{1 \leqslant k \leqslant n} (-1)^k \sum_{1 \leqslant i_1 < \ldots < i_k \leqslant n} \mathsf{P}\left(\left(\bigcap_1^k A_{i_j}\right) \cap A_{n+1}\right).\end{aligned}$$

Мы видим, что сумма в разложении $\mathsf{P}\left(\bigcup\limits_1^{n+1} A_i\right)$ включает в себя все возможные слагаемые из правой части формулы (1.3.1) для $n + 1$, причем с правильными знаками, что завершает доказательство.

Альтернативное доказательство (поучительное, так как оно показывает связь между разными понятиями теории вероятностей) будет дано в следующей главе, после того как мы введем определение случайной величины и математического ожидания.

Формула включения-исключения особенно эффективна при условии независимости и симметрии. Она также дает интересное асимптотическое понимание многих вероятностей.

Задача 1.3.1. Известно, что n студентов-физиологов не забыли о том, что надо пойти на семинар о рассеянности. После семинара ни один из них не мог узнать свою собственную шляпу и они взяли шляпы наугад. Далее, каждый из них способен, с вероятностью p и независимо от других, потерять шляпу по дороге домой. Сделав оптимистичное предположение, что они все вернулись домой, найдите вероятность того, что ни один студент в конце концов не имел при себе своей шляпы, и получите, что эта вероятность при больших значениях n приблизительно равна $e^{-(1-p)}$.

Решение. Пусть $A_j = \{$физиолог j пришел со своей шляпой$\}$, тогда событие $A = \{$никто не пришел со своей шляпой$\}$ является дополнением события $\bigcup\limits_{1\leqslant j\leqslant n} A_j$. Благодаря симметрии и формуле включения-исключения имеем

$$\mathrm{P}(A) = 1 - n\,\mathrm{P}(A_1) + \frac{n(n-1)}{2}\mathrm{P}(A_1 \cap A_2) - \ldots + (-1)^n\,\mathrm{P}(A_1 \cap \ldots \cap A_n).$$

Далее,

$$\mathrm{P}(A_1) = (1-p)\frac{(n-1)!}{n!},$$

$$\mathrm{P}(A_1 \cap A_2) = (1-p)^2\frac{(n-2)!}{n!},$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\mathrm{P}(A_1 \cap \ldots \cap A_n) = (1-p)^n\frac{1}{n!}.$$

Тогда

$$\mathrm{P}(A) = 1 - (1-p) + (1-p)^2/2! - \ldots + (-1)^n(1-p)^n/n!,$$

а это часть суммы в разложении $e^{-(1-p)}$. □

Задача 1.3.2. Выведите формулу включения-исключения для $\mathrm{P}(A_1 \cup \ldots \ldots \cup A_n)$.

Известно, что происходит по крайней мере одно, но не более двух событий $A_1, \ldots, A_n$. Зная, что $\mathrm{P}(A_i) = p_1$ для всех i и $\mathrm{P}(A_i \cap A_j) = p_2$ для всех i, j $(i \neq j)$, покажите, что

$$1 = np_1 - \frac{n(n-1)}{2}p_2.$$

Получите, что $p_1 \geqslant 1/n$ и $p_2 \leqslant 2/n$.

Решение. Так как $\mathrm{P}(A_{i_1} \cap \ldots \cap A_{i_k}) = 0$ для $k > 2$, формула включения-исключения дает $1 = np_1 - \frac{n(n-1)}{2} p_2$. Поэтому

$$p_1 = \frac{1}{n} + \frac{n-1}{2} p_2 \geqslant \frac{1}{n},$$

так как $p_2 \geqslant 0$. Наконец,

$$p_2 = \frac{2}{n-1}\left(p_1 - \frac{1}{n}\right) \leqslant \frac{2}{n-1}\left(1 - \frac{1}{n}\right) = \frac{2}{n}. \qquad \square$$

Задача 1.3.3. Собираясь провести долгую ночь в пабе «Заштопанный барабан», n удалых волшебников оставили свое снаряжение в прихожей. Возвратившись, каждый наобум взял свои вещи из общей кучи и, отправившись в свою комнату, попытался произнести заклинание от последствий похмелья. Если юный волшебник произнесет это заклинание в своем снаряжении, то он может превратиться в лягушку-быка с вероятностью p. Если он произнесет это заклинание в чужом снаряжении, то он точно превратится в лягушку-быка. Покажите, что вероятность того, что утром работник, перестилающий постели, найдет n недоумевающих лягушек-быков, при больших n приблизительно равна e^{p-1}.

Решение. Положим $A_i \, (= A_i(n)) = \{$волшебник i взял свои вещи$\}$. Тогда $\forall r = 1, \ldots, n$ и $1 \leqslant i_1 < \ldots < i_r \leqslant n$ мы имеем

$$\mathrm{P}(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_r}) = \frac{(n-r)!}{n!}$$

и существует $(n-r)!$ способов распределить вещи. По формуле включения-исключения вероятность $\mathrm{P}\left(\bigcup_{i=1}^{n} A_i\right)$ равна

$$\sum_{r=1}^{n} C_n^r (-1)^{r-1} \frac{(n-r)!}{n!} = \sum_{r=1}^{n} (-1)^{r-1} \frac{1}{r!} = 1 - \sum_{r=0}^{n} (-1)^r \frac{1}{r!},$$

что стремится к $1 - e^{-1}$ при $n \to \infty$. Аналогичные рассуждения справедливы для событий $A_i(k)$, определенных для данного множества из k волшебников, $1 \leqslant k \leqslant n$.

Если $P_r(k) = \mathrm{P}$(точно r из k волшебников взяли свои вещи), то

$$P_r(k) = C_k^r \frac{(k-r)!}{k!} P_0(k-r) = \frac{1}{r!} P_0(k-r), \quad 0 \leqslant r \leqslant k.$$

Далее,

$$P_0(k-r) = 1 - \mathrm{P}\left(\bigcup_{i=1}^{k-r} A_i(k-r)\right) = \sum_{i=1}^{k-r} (-1)^i \frac{1}{i!} \to e^{-1}$$

при $k \to \infty$. Итак, $P_r(k) \sim e^{-1}\frac{1}{r!}$ для достаточно больших k. Наконец,

$$\mathsf{P}(\text{все превратились в лягушек-быков}) =$$
$$= \sum_{r=0}^{n} P_r(n)p^r \approx \sum_{r=0}^{n} e^{-1}\frac{p^r}{r!} = e^{-1}e^p = e^{p-1}$$

для достаточно больших n. □

Замечание. Чтобы формально доказать сходимость $\sum\limits_{r=0}^{n} P_r(n)p^r \to e^{p-1}$, нужна теорема, которая гарантирует, что из почленной сходимости (в нашем случае $P_r(n)p^r \to e^{-1}\frac{p^r}{r!}$ для всех r) следует сходимость суммы. Например, подойдет следующая теорема: *если* $a_n(m) \to a_n$ $\forall n$ *при* $m \to \infty$, $|a_n(m)| \leqslant b_n$ *и* $\sum\limits_n b_n < \infty$, *то* $S(m) = \sum\limits_n a_n(m) \to \sum\limits_n a_n = S$. Тогда $P_r(n)p^r = = P_0(n-r)\frac{p^r}{r!} \leqslant \frac{p^r}{r!}$ $(=b_r)$ и ряд $\sum\limits_{r\geqslant 0}\frac{p^r}{r!}$ сходится к e^p для всех p (т. е. $\sum\limits_r b_r < \infty$). Это вычисление очень популярно, и мы приведем его в еще одной форме.

Задача 1.3.4. После вечеринки n гостей взяли куртки случайным образом из общей кучи из n курток. Вычислите вероятность того, что ни один из них не пошел домой в своей куртке.

Пусть $p(m, n)$ — вероятность того, что точно m гостей ушли домой в своих куртках. Связав $p(m, n)$ с $p(0, n-m)$ или как-нибудь иначе, определите $p(m, n)$ и покажите, что

$$\lim_{n\to\infty} p(m, n) = \frac{1}{e\cdot m!}.$$

Решение. Общее число возможных комбинаций распределения курток равно $n!$. Пусть $A_k = \{\text{гость } k \text{ взял свою куртку}\}$. Тогда

$$\mathsf{P}(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_r}) = \frac{(n-r)!}{n!}$$

и

$$\sum_{i_1<i_2<\ldots<i_r} \mathsf{P}(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_r}) = C_n^r \frac{(n-r)!}{n!} = \frac{1}{r!}.$$

Таким образом,

$$\mathsf{P}(\text{по крайней мере один гость ушел домой в своей куртке}) =$$
$$= \mathsf{P}\left(\bigcup_{i=1}^{n} A_i\right) = 1 - \frac{1}{2!} + \frac{1}{3!} - \ldots + (-1)^{n-1}\frac{1}{n!};$$

$$\mathbf{P}(\text{никто не ушел домой в своей куртке}) = 1 - \mathbf{P}\left(\bigcup_{i=1}^{n} A_i\right) =$$
$$= 1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \ldots + (-1)^n \frac{1}{n!} \to e^{-1}$$

при $n \to \infty$. Число случаев, когда никто не ушел в своей куртке, равно $n!\,p(0, n)$, и

$$p(m, n) = C_n^m \frac{(n-m)!\,p(0, n-m)}{n!}.$$

Таким образом,

$$p(m, n) = \frac{1}{m!} p(0, n-m) = \frac{1}{m!}\left(1 - \frac{1}{1!} + \ldots + (-1)^{n-m} \frac{1}{(n-m)!}\right) \to \frac{1}{m!} e^{-1}$$

при $n \to \infty$. □

Оставшаяся часть этого параграфа посвящена так называемой *задаче о баллотировке*. Ее первоначальная формулировка следующая: множество голосующих состоит из m консерваторов и n анархистов, которые голосуют за своих кандидатов, причем $m \geqslant n$. Какова вероятность того, что в процессе подсчета голосов представитель консерваторов никогда не будет позади анархиста? Этот вопрос возникал неоднократно (но, как ни странно, не появлялся до сих пор среди вопросов части IA «Математических треножников»). Тем не менее, в Университете Свонзи эта задача активно обсуждалась, правда, в несколько модифицированной форме, изложенной ниже. Пусть $m = n$. Предлагается $2n$ напитков, n из них джин и n тоник. В местной популярной игре участник с завязанными глазами выпивает все $2n$ стаканов поочередно и каждый стакан выбран случайно. Участник провозглашается победителем, если объем выпитого джина никогда не превосходит объем выпитого тоника. Мы проверим, что это может произойти с вероятностью $1/(n+1)$.

Рассмотрим случайное блуждание на множестве $\{-n, -n+1, \ldots, n\}$, причем частичка движется на один шаг вверх, если выбран стакан тоника, и на шаг вниз, если выбран стакан джина. Блуждание начинается из нуля (еще ничего не выпито) и после $2n$ шагов возвращается назад (число джинов равно числу тоников). На рис. 1.2 каждая траектория блуждания $X(t)$ начинается в точке $(0, 0)$ и заканчивается в точке $(2n, 0)$ и каждый раз прыгает или вверх и вправо, или вниз и вправо. Мы ищем вероятность того, что траектория все время находится выше линии $X = -1$.

Общее число траекторий из точки $(0, 0)$ в точку $(2n, 0)$ равно $\dfrac{(2n)!}{n!\,n!}$. Число траекторий, расположенных выше этой линии, равно числу траекторий из точки $(1, 1)$ в точку $(2n, 0)$ минус число траекторий из точки $(1, -3)$ в точку $(2n, 0)$. Действительно, первый шаг из $(0, 0)$ должен быть

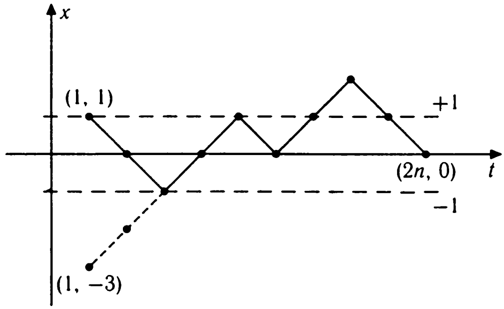


Рис. 1.2

вверх. Далее, если траектория из точки (0, 0) в точку ($2n$, 0) касается или пересекает линию $X = -1$, то мы можем зеркально отразить ее начальную часть и получить траекторию из точки (1, −3) в точку ($2n$, 0). Иногда это утверждение называют *принципом отражения*.

Таким образом, вероятность выигрыша равна

$$\left(\frac{(2n-1)!}{n!\,(n-1)!} - \frac{(2n-1)!}{(n+1)!\,(n-2)!}\right) \Big/ \frac{(2n)!}{n!\,n!} = \frac{1}{2n}\left[n - \frac{n(n-1)}{n+1}\right] = \frac{1}{n+1}.$$

Предположим теперь, что количество m тоников больше количества n джинов. Как и ранее, выиграть игру означает, что каждый раз число выпитых тоников не меньше, чем число выпитых джинов. Тогда общее число траекторий из точки (0, 0) в точку ($m+n$, $m-n$) равно $\frac{(m+n)!}{m!\,n!}$. Снова первый шаг выигрышной траектории всегда вверх. Общее число траекторий из точки (1, 1) в точку ($m+n$, $m-n$) равно $\frac{(m+n-1)!}{(m-1)!\,n!}$. Используя принцип отражения, мы видим, что число проигрышных траекторий равно общему числу траекторий из точки (1, −3) в точку ($m+n$, $m-n$), т. е. $\frac{(m+n-1)!}{(m+1)!\,(n-2)!}$. Наконец, вероятность выигрыша равна

$$\left(\frac{(m+n-1)!}{(m-1)!\,n!} - \frac{(m+n-1)!}{(m+1)!\,(n-2)!}\right) \Big/ \frac{(m+n)!}{m!\,n!} = \frac{m-n+1}{m+1}.$$

Применим эти результаты к следующей задаче.

Задача 1.3.5. Известно, что n женатых пар должны пересечь речку слева направо по узкому мосту, один за другим. Они решили, что в каждый момент времени на левом берегу не должно быть меньше мужчин, чем женщин; кроме этого требования, порядок пересечения может быть любым. Найдите вероятность того, что каждый мужчина пересечет речку после своей жены.

Решение. Положим

$$A = \{\text{каждый мужчина пересечет речку после своей жены}\},$$
$$B = \{\text{в любой момент число мужчин на левом берегу} \geqslant \text{числа женщин}\}.$$

Тогда

$$\mathsf{P}(A \mid B) = \mathsf{P}(A \cap B)/\mathsf{P}(B) = \mathsf{P}(A)/\mathsf{P}(B) = \frac{n+1}{2^n}. \qquad \square$$

Мы закончим этот параграф другой историей о Колмогорове. Он в юности всерьез занимался историей и в 19 лет завершил работу, посвященную принципам распределения и налогообложения пахотной земли Новгорода в XV и XVI веках. Новгород — древний русский город, княжество и республика в разные периоды существования, полностью или частично независимый, — был присоединен Москвой в 1478 г. В своей работе Колмогоров использовал математическую аргументацию, чтобы ответить на следующий вопрос. Верно ли, что а) вначале облагалась налогом деревня, а потом налог распределялся между домовладельцами, или б) наоборот, вначале облагались налогом домовладельцы, а потом платился общий налог с деревни? Источником информации был древние писцовые книги — кадастры и другие официальные рукописи того времени. Поскольку общая сумма, которую собирали с деревни, была всегда целым числом (меняющихся) денежных единиц, Колмогоров доказал, что использовалось правило б).

Он доложил о своем результате на семинаре по истории в Московском университете в 1922 г. Тем не менее, руководитель семинара, известный профессор (позже в честь него была названа одна московская улица), сказал, что результат молодого коллеги не может быть принят как окончательный, потому что «в истории каждое утверждение должно быть подкреплено несколькими доказательствами». После этого Колмогоров решил заняться дисциплиной, в которой «для правильного утверждения достаточно одного доказательства», т. е. математикой.

§ 1.4. Случайные величины. Математическое ожидание, условное математическое ожидание. Совместные распределения

> Celui qui a entendu dire la chose à douze mille témoins oculaires n'a que douze mille probabilités, égales à une forte probabilité, laquelle n'est pas égale à la certitude.
>
> F. M. A. Voltaire (Dictionnaire philosophique, chapitre sur la Vérité)

> 12000 показаний очевидцев — это всего 12000 весьма вероятных суждений, которые, будучи собраны вместе, дают одно очень вероятное суждение, но все еще ничего не доказывают.
>
> Ф. М. А. Вольтер (1694—1778), французский философ

Этот параграф значительно больше, чем предыдущие: многообразие представленных в нем задач такое, что, прорешав их, студент может считать, что он разобрался в одной трети кембриджского курса «Вероятность IA».

Определение *случайной величины* (с. в.) следующее: это функция X на множестве всех событий Ω, обычно действительная, иногда комплексная, $X(\omega)$, $\omega \in \Omega$ (во втором случае мы можем рассматривать пару, состоящую из действительной и мнимой частей функции X). Простой (и важный) пример с. в. — это индикаторная функция события:

$$\omega \in \Omega \mapsto I_A(\omega) = \begin{cases} 1, & \text{если } \omega \in A, \\ 0, & \text{если } \omega \notin A. \end{cases} \tag{1.4.1}$$

Очевидно, что произведение $I_{A_1}(\omega) I_{A_2}(\omega)$ равно 1 тогда и только тогда, когда $\omega \in A_1 \cap A_2$, т. е. $I_{A_1} I_{A_2} = I_{A_1 \cap A_2}$. С другой стороны, $I_{A_1 \cup A_2} = I_{A_1} + I_{A_1} - I_{A_1 \cap A_2}$. В случае конечного числа событий любая с. в. является линейной комбинацией индикаторных функций; в случае бесконечного числа событий нужно взять бесконечную линейную комбинацию.

Математическое ожидание (или *среднее значение*, или просто *среднее*) случайной величины X, принимающей значения $x_1, \ldots, x_m$ с вероятностями $p_1, \ldots, p_m$, определяется как сумма

$$\mathsf{E}X = \sum_{1 \leqslant i \leqslant m} p_i x_i = \sum_{1 \leqslant i \leqslant m} x_i \, \mathsf{P}(X = x_i), \tag{1.4.2}$$

или

$$\mathsf{E}X = \sum_{\omega \in \Omega} p(\omega) X(\omega). \tag{1.4.3}$$

Если $X \equiv b$ есть постоянная с. в., то $\mathsf{E}X = b$.

Это определение также имеет смысл для с. в., принимающих счетное число значений $x_1, x_2, \ldots$ с вероятностями $p_1, p_2, \ldots$:

$$\mathsf{E}X = \sum_i p_i x_i$$

при условии, что ряд сходится абсолютно: $\sum\limits_i p_i |x_i| < \infty$. (Если $\sum\limits_i p_i |x_i| = \infty$, говорят, что случайная величина X имеет бесконечное математическое ожидание.)

Во многих задачах полезно рассматривать математическое ожидание $\mathsf{E}X$ как положение центра масс системы точек $x_1, x_2, \ldots$ с массами $p_1, p_2, \ldots$

Первое (и очень полезное) наблюдение показывает, что математическое ожидание индикатора события I_A равно вероятности:

$$\mathsf{E}I_A = \sum_{\omega \in \Omega} p(\omega) I_A(\omega) = \sum_{\omega \in A} p(\omega) = \mathsf{P}(A). \tag{1.4.4}$$

Далее, если для с. в. X и Y справедливо неравенство $X \leqslant Y$, то $\mathsf{E}X \leqslant \mathsf{E}Y$.

Математическое ожидание линейной комбинации с. в. равно линейной комбинации их математических ожиданий. Кратчайшее доказательство в ω-обозначениях таково:

$$\mathsf{E}(c_1X_1 + c_2X_2) = \sum_{\omega} p(\omega)(c_1X_1(\omega) + c_2X_2(\omega)) =$$
$$= c_1 \sum_{\omega} p(\omega)X_1(\omega) + c_2 \sum_{\omega} p(\omega)X_2(\omega) = c_1\,\mathsf{E}X_1 + c_2\,\mathsf{E}X_2.$$

Этот факт (называемый линейностью математического ожидания) может быть легко распространен на n слагаемых:

$$\mathsf{E}\left(\sum_{1\leqslant k\leqslant n} c_kX_k\right) = \sum_{1\leqslant k\leqslant n} c_k\,\mathsf{E}X_k;$$

в частности, если $\mathsf{E}X_k = \mu$ для любого k, то $\mathsf{E}\sum\limits_{1\leqslant k\leqslant n} X_k = \mu n$. Похожее свойство справедливо для бесконечной последовательности случайных величин $X_1, X_2, \ldots$:

$$\mathsf{E}\left(\sum_{k} c_kX_k\right) = \sum_{k} c_k\,\mathsf{E}X_k \qquad (1.4.5)$$

при условии, что ряд в правой части сходится абсолютно: $\sum\limits_k |c_k\,\mathsf{E}X_k| < \infty$. (Точная формулировка: если $\sum\limits_k |c_k\,\mathsf{E}X_k| < \infty$, то ряд $\sum\limits_{k\geqslant 1} c_kX_k$ определяет с. в. с конечным средним, равным сумме $\sum\limits_{k\geqslant 1} c_k\,\mathsf{E}X_k$.)

Для заданной функции $g\colon \mathbb{R} \to \mathbb{R}$ справедливо равенство

$$\mathsf{E}g(X) = \sum_{\omega\in\Omega} p(\omega)g(X(\omega)) = \sum_{i} p_ig(x_i) \qquad (1.4.6)$$

при условии, что $\sum\limits_i p_i|g(x_i)| < \infty$. Действительно, записывая $Y = g(X)$ и обозначив значения Y через $y_1, y_2, \ldots$, имеем

$$\mathsf{E}Y = \sum_{j} y_j\,\mathsf{P}(g(X) = y_j) = \sum_{j} y_j \sum_{i} I_{g(x_i)=y_j}\,\mathsf{P}(X = x_i),$$

что равно $\sum\limits_i p_ig(x_i)$ при условии, что $\sum\limits_i p_i|g(x_i)| < \infty$, так как тогда мы можем суммировать, группируя слагаемые.

Замечание. Формула (1.4.6) является дискретной версией формулы, известной под названием *правило незадачливого статистика*. См. формулу (2.2.8) ниже.

Пусть даны две с. в. X и Y, принимающие значения $X(\omega)$ и $Y(\omega)$; рассмотрим события $\{\omega\colon X(\omega)=x_i,\ Y(\omega)=y_j\}$ (или, коротко, $\{X=x_i,\ Y=y_j\}$) для любой пары их значений x_i и y_j. Вероятность $\mathsf{P}(X=x_i,\ Y=y_j)$ означает *совместное распределение* пары X, Y. «Одномерные» вероятности $\mathsf{P}(X=x_i)$ и $\mathsf{P}(Y=y_j)$ получаются при суммировании по всем возможным значениям второй с. в.:

$$\begin{aligned}\mathsf{P}(X=x_i)&=\sum_{y_j}\mathsf{P}(X=x_i,\ Y=y_j),\\ \mathsf{P}(Y=y_j)&=\sum_{x_i}\mathsf{P}(X=x_i,\ Y=y_j).\end{aligned}\tag{1.4.7}$$

Мы также можем рассмотреть условные вероятности

$$\mathsf{P}(X=x_i\mid Y=y_j)=\frac{\mathsf{P}(X=x_i,\ Y=y_j)}{\mathsf{P}(Y=y_j)}.\tag{1.4.8}$$

Аналогичные правила применяются в случае нескольких с. в. $X_1, \ldots, X_n$.

Например, для суммы $X+Y$ случайных величин X и Y с совместным распределением $\mathsf{P}(X=x_i,\ Y=y_j)$ имеем

$$\begin{aligned}\mathsf{P}(X+Y=u)&=\sum_{x_i,y_j\colon x_i+y_j=u}\mathsf{P}(X=x_i,\ Y=y_j)=\\ &=\sum_x\mathsf{P}(X=x,\ Y=u-x)=\sum_y\mathsf{P}(X=u-y,\ Y=y).\end{aligned}\tag{1.4.9}$$

Аналогично для произведения XY получаем

$$\begin{aligned}\mathsf{P}(XY=u)&=\sum_{x_i,y_j\colon x_iy_j=u}\mathsf{P}(X=x_i,\ Y=y_j)=\\ &=\sum_x\mathsf{P}(X=x,\ Y=u/x)=\sum_y\mathsf{P}(X=u/y,\ Y=y).\end{aligned}\tag{1.4.10}$$

Мощным инструментом является *формула условного математического ожидания*: если X и N — две с. в., то

$$\mathsf{E}X=\mathsf{E}[\mathsf{E}(X\mid N)].\tag{1.4.11}$$

Тут в правой части внешнее математическое ожидание берется по вероятностям $\mathsf{P}(N=n_j)$, с которыми с. в. N принимает значение n_j:

$$\mathsf{E}[\mathsf{E}(X\mid N)]=\sum_j\mathsf{P}(N=n_j)\,\mathsf{E}(X\mid N=n_j).$$

Внутреннее математическое ожидание берется по отношению к условным вероятностям $\mathsf{P}(X=x_i\mid N=n_j)$ того, что с. в. X принимает значение x_i при условии, что $N = n_j$:

$$\mathsf{E}(X \mid N = n_j) = \sum_i x_i \mathsf{P}(X = x_i \mid N = n_j).$$

Чтобы доказать формулу (1.4.11), мы просто используем определение условной вероятности:

$$\mathsf{E}[\mathsf{E}(X \mid N)] = \sum_j \mathsf{P}(N = n_j) \sum_i x_i \mathsf{P}(X = x_i \mid N = n_j) =$$
$$= \sum_{x_i, n_j} x_i \mathsf{P}(X = x_i,\ N = n_j) = \sum_{x_i} x_i \mathsf{P}(X = x_i) = \mathsf{E}X.$$

Мы видим, что равенства (1.4.11) — просто перефразировка соотношения (1.2.6). Опять-таки заметим, что это есть следствие перехода к условным вероятностям относительно с. в. N.

Следующая формула удобна, когда N принимает неотрицательные целые значения:

$$\mathsf{E}N = \sum_{n \geqslant 1} \mathsf{P}(N \geqslant n). \tag{1.4.12}$$

Действительно,

$$\sum_{n \geqslant 1} \mathsf{P}(N \geqslant n) = \sum_{n \geqslant 1} \sum_{k \geqslant n} \mathsf{P}(N = k) =$$
$$= \sum_{k \geqslant 1} \mathsf{P}(N = k) \sum_{1 \leqslant n \leqslant k} 1 = \sum_{k \geqslant 1} \mathsf{P}(N = k)k = \mathsf{E}N. \tag{1.4.13}$$

Мы говорим, что с. в. X и Y, принимающие значения x_i и y_j, *независимы*, если для любых пар значений выполняется равенство

$$\mathsf{P}(X = x_i,\ Y = y_j) = \mathsf{P}(X = x_i)\, \mathsf{P}(Y = y_j). \tag{1.4.14}$$

В случае трех случайных величин X, Y и Z нужно, чтобы для любой тройки значений было выполнено равенство $\mathsf{P}(X = x_i,\ Y = y_j,\ Z = z_k) = \mathsf{P}(X = x_i) \times$ $\times\ \mathsf{P}(Y = y_j)\, \mathsf{P}(Z = z_k)$. Это определение может быть обобщено на случай любого числа с. в. $X_1, \ldots, X_n$: $\forall x_1, \ldots, x_n$ имеет место равенство

$$\mathsf{P}(X_1 = x_1,\ X_2 = x_2,\ \ldots,\ X_n = x_n) = \prod_{i=1}^{n} \mathsf{P}(X_i = x_i). \tag{1.4.15}$$

Далее, с. в., образующие бесконечную последовательность, $X_1, X_2, \ldots$, называются независимыми, если для любого n величины $X_1, \ldots, X_n$ независимы.

Возникает вопрос, почему в формуле (1.4.14) достаточно рассматривать «полное» произведение $\prod_{i=1}^{n}$, в то время как в определении независимых событий требовалось, чтобы выполнялось равенство $\mathrm{P}\left(\bigcap A_{i_k}\right) = \prod \mathrm{P}(A_{i_k})$ для любой подсистемы (см. формулу (1.2.9)). Ответ следующий: потому что мы требуем, чтобы формула выполнялась для любой системы значений $x_1, \ldots, x_n$ случайных величин $X_1, \ldots, X_n$. Действительно, если не учитывать некоторые с. в., то нужно взять сумму по всем их значениям:

$$\mathrm{P}(X_2 = x_2, \ldots, X_n = x_n) = \sum_{x_1} \mathrm{P}(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n),$$

так как события $\{X_1 = x_1, \ldots, X_{n-1} = x_{n-1}, X_n = x_n\}$ попарно несовместны для различных значений x_1 и их объединение дает $\{X_2 = x_2, \ldots, X_n = x_n\}$. Значит, если $\mathrm{P}(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} \mathrm{P}(X_i = x_i)$, то

$$\mathrm{P}(X_2 = x_2, \ldots, X_n = x_n) = \sum_{x_1} \mathrm{P}(X_1 = x_1)\, \mathrm{P}(X_2 = x_2) \ldots \mathrm{P}(X_n = x_n) =$$
$$= \mathrm{P}(X_2 = x_2) \ldots \mathrm{P}(X_n = x_n),$$

т. е. подсистема $X_2, \ldots, X_n$ автоматически является независимой.

Тем не менее, обратное не всегда верно. Например, если в каждой паре (X, Y), (X, Z) и (Y, Z) с. в. независимы, то необязательно, что и тройка X, Y, Z независима. Пример дан в задаче 1.2.16: для решения нужно просто взять I_{A_1}, I_{A_2} и I_{A_3}.

Важное понятие, которое возникло сейчас и будет использоваться в оставшейся части книги, — это понятие *независимых одинаково распределенных* (н. о. р.) с. в. $X_1, X_2, \ldots$ Тут вдобавок к независимости предполагается, что вероятности $\mathrm{P}(X_i = x)$ одинаковы для всех $i = 1, 2, \ldots$ Хорошей моделью является подбрасывание монеты: X_n может быть функцией результата n-го подбрасывания (т. е. $X_n = I(n$-е подбрасывание дает орел$)$). Вообще, мы можем разбить испытания на непересекающиеся «блоки» данной длины l и рассматривать в качестве X_n функцию событий в n-м блоке.

Немедленным следствием этого определения будет то, что для независимых с. в. X, Y математическое ожидание произведения равно произведению математических ожиданий:

$$\mathrm{E}(XY) = \sum_{x_i, y_j} x_i y_j\, \mathrm{P}(X = x_i, Y = y_j) = \sum_{x_i, y_j} x_i y_j\, \mathrm{P}(X = x_i)\, \mathrm{P}(Y = y_j) =$$
$$= \sum_{x_i} x_i\, \mathrm{P}(X = x_i) \sum_{y_j} y_j\, \mathrm{P}(Y = y_j) = \mathrm{E}X\, \mathrm{E}Y. \quad (1.4.16)$$

Тем не менее, обратное неверно: существуют такие зависимые с. в., что $\mathsf{E}(XY) = \mathsf{E}X\,\mathsf{E}Y$. Простой пример:

$$X = \begin{cases} -1 & \text{с вероятностью } 1/3, \\ 0 & \text{с вероятностью } 1/3, \\ 1 & \text{с вероятностью } 1/3, \end{cases} \qquad Y = -\frac{2}{3} + X^2.$$

Тут

$$\mathsf{E}X = 0, \quad \mathsf{E}Y = 0, \quad \mathsf{E}(XY) = -\frac{2}{3}\,\mathsf{E}X + \mathsf{E}X^3 = 0.$$

Но очевидно, что X и Y зависимы.

С другой стороны, невозможно построить пример зависимых случайных величин X и Y, принимающих каждая по два значения и таких, что $\mathsf{E}(XY) = \mathsf{E}X\,\mathsf{E}Y$ (см. задачу 2.2.20).

Дисперсия с. в. X, принимающей действительные значения, со средним $\mathsf{E}X = \mu$ равна

$$\operatorname{Var} X = \mathsf{E}(X - \mu)^2; \tag{1.4.17}$$

раскрывая скобки, используя линейность математического ожидания и то, что математическое ожидание постоянной равно этой постоянной, получаем

$$\begin{aligned} \operatorname{Var} X = \mathsf{E}(X^2 - 2X\mu + \mu^2) = \mathsf{E}X^2 - 2\mu\,\mathsf{E}X + \mathsf{E}(\mu)^2 = \\ = \mathsf{E}X^2 - 2(\mu)^2 + (\mu)^2 = \mathsf{E}X^2 - (\mathsf{E}X)^2. \end{aligned} \tag{1.4.18}$$

В частности, если X — постоянная с. в.: $X \equiv b$, то $\operatorname{Var} X = b^2 - b^2 = 0$.

Дисперсию считают мерой «отклонения» с. в. X от среднего значения. (Величина $\sqrt{\operatorname{Var} X}$ называется *стандартным отклонением* или средним квадратическим отклонением.)

Из первого определения мы видим, что $\operatorname{Var} X \geqslant 0$, т. е. $(\mathsf{E}X)^2 \leqslant \mathsf{E}X^2$. Это частный случай так называемого *неравенства Коши—Шварца*[1]

$$|\,\mathsf{E}(X\overline{Y})|^2 \leqslant \mathsf{E}|X|^2\,\mathsf{E}|Y|^2, \tag{1.4.19}$$

справедливого для любой пары действительных или комплексных с. в. X и Y; тут $\overline{Y}$ означает комплексное сопряжение. (Это неравенство предполагает, что если математическое ожидание величин $|X|^2$ и $|Y|^2$ конечно, то математическое ожидание величины XY также конечно.) Неравенство Коши—Шварца названо в честь двух знаменитых математиков: француза О. Л. Коши (1789—1857), который доказал его в дискретном случае, и немца Г. А. Шварца (1843—1921), распространившего неравенство на непрерывный случай.

[1] В отечественной литературе Коши—Буняковского. — *Прим. ред.*

Доказательство неравенства Коши—Шварца может служить элегантным алгебраическим упражнением; для простоты мы приведем его для действительных с. в. Заметим, что $\forall\lambda \in \mathbb{R}$ случайная величина $(X+\lambda Y)^2$ неотрицательна и, таким образом, $\mathsf{E}(X+\lambda Y)^2 \geqslant 0$. Представив это математическое ожидание как функцию от λ, получим квадратичный полином

$$\mathsf{E}(X+\lambda Y)^2 = \mathsf{E}X^2 + 2\lambda\,\mathsf{E}XY + \lambda^2\,\mathsf{E}Y^2.$$

Для того чтобы этот полином был неотрицательным для всех $\lambda \in \mathbb{R}$, он должен иметь отрицательный или равный нулю дискриминант, т. е.

$$4(\mathsf{E}XY)^2 \leqslant 4\,\mathsf{E}X^2\,\mathsf{E}Y^2.$$

Величина, тесно связанная с понятием дисперсии и называемая *ковариацией* двух с. в. X и Y, определяется как

$$\mathrm{Cov}(X, Y) = \mathsf{E}(X-\mathsf{E}X)(Y-\mathsf{E}Y) = \mathsf{E}(XY) - \mathsf{E}X\,\mathsf{E}Y. \tag{1.4.20}$$

Для независимых случайных величин $\mathrm{Cov}(X, Y) = 0$. Опять-таки обратное неверно: существуют зависимые случайные величины X и Y, для которых $\mathrm{Cov}(X, Y) = 0$. Из неравенства Коши—Шварца следует, что $|\mathrm{Cov}(X, Y)|^2 \leqslant (\mathrm{Var}\,X)(\mathrm{Var}\,Y)$.

Для дисперсии $\mathrm{Var}(X+Y)$ суммы $X+Y$ справедливо следующее представление:

$$\mathrm{Var}(X+Y) = \mathrm{Var}\,X + \mathrm{Var}\,Y + 2\,\mathrm{Cov}(X, Y). \tag{1.4.21}$$

Действительно,

$$\begin{aligned}\mathrm{Var}(X+Y) &= \mathsf{E}((X-\mathsf{E}X) + (Y-\mathsf{E}Y))^2 = \\ &= \mathsf{E}(X-\mathsf{E}X)^2 + 2\,\mathsf{E}(X-\mathsf{E}X)(Y-\mathsf{E}Y) + \mathsf{E}(Y-\mathsf{E}Y)^2,\end{aligned}$$

что равно правой части равенства (1.4.20).

Важным следствием является то, что дисперсия суммы независимых с. в. равна сумме дисперсий:

$$\mathrm{Var}(X+Y) = \mathrm{Var}\,X + \mathrm{Var}\,Y.$$

Этот факт легко обобщается на любое конечное число слагаемых:

$$\mathrm{Var}(X_1 + \ldots + X_n) = \mathrm{Var}\,X_1 + \ldots + \mathrm{Var}\,X_n. \tag{1.4.22}$$

В случае н. о. р. с. в. $\mathrm{Var}\,X_i$ не зависит от i и если $\mathrm{Var}\,X_i = \sigma^2$ (стандартное вероятностное обозначение), то $\mathrm{Var}\Big(\sum\limits_{1\leqslant j\leqslant n} X_j\Big) = n\sigma^2$.

С другой стороны, если c — действительная постоянная, то $\mathrm{Var}(cX) = \mathsf{E}(cX)^2 - (\mathsf{E}(cX))^2 = c^2(\mathsf{E}X^2 - (\mathsf{E}X)^2) = c^2\,\mathrm{Var}\,X$. Таким образом, $\mathrm{Var}\,nX = n^2\,\mathrm{Var}\,X$. Это означает, что, суммируя идентичные с. в., мы получим, что дисперсия возрастает квадратично, в то время как суммируя н. о. р. с. в. мы получим только линейный рост. (Средние значения в обоих случаях возрастают линейно.)

Постоянная с. в., принимающая одно значение ($X \equiv b$), не зависит от любой другой с. в. (или множества с. в.). Следовательно, $\operatorname{Var}(X + b) = \operatorname{Var} X + \operatorname{Var} b = \operatorname{Var} X$.

Итак, для независимых случайных величин $X_1, X_2, \ldots$ и действительных чисел $c_1, c_2, \ldots$ имеем

$$\operatorname{Var} \sum_i c_i X_i = \sum_i c_i^2 \operatorname{Var} X_i$$

при условии, что ряд в правой части равенства сходится абсолютно. (Более точно, если $\sum_i c_i^2 \operatorname{Var} X_i < \infty$, то ряд $\sum_i c_i X_i$ определяет с. в. с конечной дисперсией $\sum_i c_i^2 \operatorname{Var} X_i$.)

Наконец, в случае независимых случайных величин формулы (1.4.9), (1.4.10) выглядят так:

$$\mathsf{P}(X+Y=u)=\sum_y \mathsf{P}(X=u-y)\,\mathsf{P}(Y=y)=\sum_x \mathsf{P}(X=x)\,\mathsf{P}(Y=u-x) \quad (1.4.23)$$

и

$$\mathsf{P}(XY=u)=\sum_y \mathsf{P}(X=u/y)\,\mathsf{P}(Y=y)=\sum_x \mathsf{P}(X=x)\,\mathsf{P}(Y=u/x). \quad (1.4.24)$$

Уравнение (1.4.23) часто называют *формулой свертки.*

Замечание. В обозначениях бывают некоторые отклонения: многие авторы пишут $\mathsf{E}[X]$, $\operatorname{Var}[X]$ и/или $\operatorname{Var}(X)$. Так как мы придерживаемся стиля «Математических треножников», такие обозначения также будут иногда появляться в нашей книге.

Теперь мы приведем альтернативное доказательство формулы включения-исключения (1.3.1) для вероятности $\mathsf{P}(A)$ объединения $A = \bigcup\limits_{1\leqslant i\leqslant n} A_i$. Используя индикаторные функции, формулу можно записать следующим образом:

$$\mathsf{E} I_A = \sum_{1\leqslant k\leqslant n} (-1)^{k+1} \sum_{1\leqslant i_1<\ldots<i_k\leqslant n} \mathsf{E}(I_{A_{i_1}} \ldots I_{A_{i_k}}).$$

Пусть $\omega \in \bigcup\limits_{1\leqslant i\leqslant n} A_i$ и $A_{j_1}, \ldots, A_{j_s}$ — множество событий, содержащих ω, т. е. $\omega \in A_{j_1} \cap \ldots \cap A_{j_s}$ и $\omega \notin A_j$ для всех A_j не из этого списка. Не ограничивая общности, предположим, что $\{j_1, \ldots, j_s\} = \{1, \ldots, s\}$; в противном случае мы соответственно переобозначим события. Тогда для этого ω правая часть должна содержать $\mathsf{E}(I_{A_{i_1}} \dots I_{A_{i_k}})$ только при $1 \leqslant k \leqslant s$ и $1 \leqslant i_1 < \dots < i_k \leqslant s$. Точнее, мы хотим проверить, что

$$1 = I_{\bigcup_{1 \leqslant i \leqslant n} A_i}(\omega) = \sum_{1 \leqslant k \leqslant s} (-1)^{k+1} \sum_{1 \leqslant i_1 < \dots < i_k \leqslant s} I_{A_{i_1}}(\omega) \dots I_{A_{i_k}}(\omega).$$

Но правая часть равна

$$\sum_{1 \leqslant k \leqslant s} (-1)^{k+1} C_s^k = 1 - \left(1 + \sum_{1 \leqslant k \leqslant s} (-1)^k C_s^k\right) = 1 - (1-1)^s = 1.$$

Взяв математическое ожидание, мы получим требуемый результат.

Задача 1.4.1. Я пришел домой с банкета и пытаюсь открыть мою входную дверь одним из трех ключей, находящихся в моем кармане. (Мы предполагаем, что в точности один ключ откроет дверь, при условии, что он выбран правильно.) Найдите среднее число a попыток, которые я должен сделать, чтобы открыть дверь, при условии, что я выбираю ключ случайно и бросаю его на пол, если он не подходит. Найдите среднее число b необходимых попыток при условии, что я кладу ключ, который не подошел, обратно в карман. Найдите a и b и покажите, что $b - a = 1$.

Решение. Обозначим ключи через 1, 2 и 3 и предположим, что ключи 2 и 3 не подходят. Рассмотрим следующие случаи: а) первый выбранный ключ подошел, б) второй выбранный ключ подошел и в) третий выбранный ключ подошел. Тогда для a получим

$$a = \left(\frac{1}{3} \cdot 1\right) + \left(2 \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot 2\right) + \left(2 \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot 3\right) = 2.$$

Для нахождения b используем разложение по условным вероятностям относительно результата первой попытки:

$$b = \frac{1}{3} \cdot 1 + \frac{2}{3} \cdot (1 + b), \quad \text{откуда } b = 3.$$

Множитель $(1 + b)$ отображает тот факт, что если первая попытка неудачна, то ситуация повторяется в силу независимости. □

Замечание. Уравнение для b аналогично выведенным ранее уравнениям для вероятностей Q и Q_i; см. задачи 1.2.12 и 1.2.13.

Задача 1.4.2. Пусть N — неотрицательная целочисленная случайная величина со средним μ_1 и дисперсией σ_1^2 и $X_1, X_2, \dots$ — одинаково распределенные случайные величины со средним μ_2 и дисперсией σ_2^2; предположим, что величины $N, X_1, X_2, \dots$ независимы. Вычислите среднее и дисперсию с. в. $S_N = X_1 + X_2 + \dots + X_N$.

Решение 1. Переходя к условным математическим ожиданиям, получим

$$\mathsf{E}S_N = \mathsf{E}[\mathsf{E}(S_N \mid N)] = \sum_{n=0}^{N} \mathsf{P}(N=n)\,\mathsf{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{n=0}^{N} \mathsf{P}(N=n) n\mu_2 = \mu_1\mu_2.$$

Далее,

$$\mathsf{E}S_N^2 = \mathsf{E}[\mathsf{E}(S_N^2 \mid N)] = \sum_{n=0}^{N} \mathsf{P}(N=n)\,\mathsf{E}\left(\sum_{i=1}^{n} X_i\right)^2 =$$

$$= \sum_{n=0}^{N} \mathsf{P}(N=n)\left[\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) + \left(\mathsf{E}\sum_{i=1}^{n} X_i\right)^2\right] =$$

$$= \sum_{n=0}^{N} \mathsf{P}(N=n)[n\sigma_2^2 + (n\mu_2)^2] = \mu_1\sigma_2^2 + \mu_2^2\,\mathsf{E}N^2 = \mu_1\sigma_2^2 + \mu_2^2(\mu_1^2 + \sigma_1^2).$$

Поэтому

$$\mathrm{Var}(S_N) = \mathsf{E}S_N^2 - (\mathsf{E}S_N)^2 = \mu_1\sigma_2^2 + \mu_2^2\sigma_1^2.$$

Решение 2. (Читайте решение после § 1.5, посвященного производящим функциям.) Запишем

$$p(z) = \mathsf{E}(z^{S_N}) = g(h(z)),$$

где $h(z) = \mathsf{E}(z^{X_1})$, $g(z) = \mathsf{E}(z^N)$. Дифференцирование дает $p'(z) = g'(h(z)) \times$ $\times\, h'(z)$, и тогда

$$\mathsf{E}(S_N) = p'(1) = g'(h(1))h'(1) = g'(1)h'(1) = \mu_1\mu_2.$$

Далее, $p''(z) = g''(h(z))h'(z)^2 + g'(h(z))h''(1)$. Получим

$$\mathsf{E}(S_N(S_N - 1)) = p''(1) = g''(1)h'(1)^2 + g'(1)h''(1)$$

и

$$\mathrm{Var}(S_N) = p''(1) + p'(1) - (p'(1))^2 = \mu_1\sigma_2^2 + \mu_2^2\sigma_1^2. \qquad \square$$

Задача 1.4.3. Пусть X и Y — неотрицательные целочисленные независимые случайные величины с конечными средними. Положим $U = \min[X, Y]$ и $V = \max[X, Y]$.

а) Докажите, что

$$\mathsf{E}U = \sum_{n\geqslant 1} \mathsf{P}(X \geqslant n)\,\mathsf{P}(Y \geqslant n),$$

$$\mathsf{E}V = \sum_{n\geqslant 1} (\mathsf{P}(X \geqslant n) + \mathsf{P}(Y \geqslant n) - \mathsf{P}(X \geqslant n)\,\mathsf{P}(Y \geqslant n))$$

и

$$\mathsf{E}UV = \sum_{n,n'\geqslant 1} \mathsf{P}(X \geqslant n)\,\mathsf{P}(Y \geqslant n').$$

б) Предположим вдобавок, что случайные величины X и Y не являются постоянными и одинаково распределены. Рассмотрев вероятность $\mathsf{P}(V \leqslant x < U)$ для некоторого x (или как-нибудь иначе), докажите, что U и V — зависимые с. в.

Решение. а) Запишем

$$\mathsf{E}U = \mathsf{E}\min[X,\ Y] = \sum_{n\geqslant 1}\mathsf{P}(\min[X,\ Y] \geqslant n) = \sum_{n\geqslant 1}\mathsf{P}(X \geqslant n)\,\mathsf{P}(Y \geqslant n).$$

Так как $U + V = X + Y$, мы получаем

$$\mathsf{E}V = \mathsf{E}X + \mathsf{E}Y - \mathsf{E}U = \sum_{n\geqslant 1}(\mathsf{P}(X \geqslant n) + \mathsf{P}(Y \geqslant n) - \mathsf{P}(X \geqslant n)\,\mathsf{P}(Y \geqslant n)).$$

Аналогично $UV = XY$, и тогда

$$\mathsf{E}UV = \mathsf{E}XY = \mathsf{E}X\,\mathsf{E}Y = \sum_{n,n'\geqslant 1} \mathsf{P}(X \geqslant n)\,\mathsf{P}(Y \geqslant n').$$

б) Для всех x имеем $\mathsf{P}(V \leqslant x < U) = 0$. Но $\mathsf{P}(V \leqslant x) = \mathsf{P}(X \leqslant x)^2$ и $\mathsf{P}(U > x) = \mathsf{P}(X > x)^2$, и тогда если случайные величины X, Y не константы, то существует такое x, что $\mathsf{P}(X \leqslant x)$, $\mathsf{P}(X > x) > 0$. Поэтому $\mathsf{P}(V \leqslant x)\,\mathsf{P}(U > x) \neq 0$, и, таким образом, случайные величины U и V зависимы. □

Задача 1.4.4. а) Пусть $X_1, X_2, \ldots, X_n$ — случайные величины. Покажите, что

$$\mathrm{Var}(X_1 + X_2 + \ldots + X_n) = \sum_{i,j=1}^{n} \mathrm{Cov}(X_i,\ X_j).$$

б) Десять человек сидят в кругу, и каждый подбрасывает симметричную монету. Пусть N — число людей, монетка которых упала той же стороной, что и монетка соседей справа и слева. Найдите $\mathsf{P}(N=9)$ и $\mathsf{P}(N=10)$. Представив N как сумму случайных величин (или другим способом), найдите ее среднее и дисперсию.

Решение. а) Пусть $Y_i = X_i - \mathsf{E}X_i$. Тогда

$$\mathrm{Var}(X_1 + X_2 + \ldots + X_n) = \mathsf{E}(Y_1 + \ldots + Y_n)^2 = \mathsf{E}\Big(\sum_{i,j=1}^{n} Y_iY_j\Big) = \sum_{i,j=1}^{n}\mathsf{E}(Y_iY_j).$$

Но $\mathsf{E}(Y_iY_j) = \mathrm{Cov}(X_i,\ X_j)$, что и дает требуемый результат.

б) Заметим, что $\mathsf{P}(N = 9) = 0$. Действительно, не может быть такой ситуации, чтобы результаты бросков девяти человек совпадали с результатами обоих соседей, а результат десятого не совпадал. Далее, $\mathsf{P}(N = 10) = 2 \cdot \left(\frac{1}{2}\right)^{10}$: существует 2 способа зафиксировать сторону монетки, и $\left(\frac{1}{2}\right)^{10}$ — вероятность того, что все 10 монеток выпадут данной стороной.

Наконец, запишем $N = I_{A_1} + \ldots + I_{A_{10}}$, где A_i — событие, состоящее в том, что у двух соседей i-го человека монетки выпали той же стороной, что и у него. В силу симметрии

$$\mathsf{E}N = 10\,\mathsf{P}(A_1) = 10 \cdot \frac{1}{4} = 2{,}5.$$

Величины I_{A_i} и I_{A_j} попарно независимы, если i не сосед j. Действительно, когда i и j не имеют общих соседей, очевидно, что каждая с. в. зависит от непересекающегося множества результатов. Предположим, что у i и j есть (один) общий сосед. Тогда $\mathsf{P}(I_{A_i} = 1,\ I_{A_j} = 1) = 1/32 + 1/32 = 1/16$. В то же время $\mathsf{P}(I_{A_i} = 1) = \mathsf{P}(I_{A_j} = 1) = 1/4$.

Таким образом,

$$\operatorname{Var}(N) = 10\operatorname{Var}(I_{A_1}) + 20\operatorname{Cov}(I_{A_1}, I_{A_2}) = 10\left(\frac{1}{4} - \left(\frac{1}{4}\right)^2\right) + 20\left(\mathsf{E}(I_{A_1}I_{A_2}) - \frac{1}{16}\right).$$

Далее,

$$\mathsf{E}(I_{A_1}I_{A_2}) = \mathsf{P}(I_{A_1} = 1 \mid I_{A_2} = 1)\,\mathsf{P}(I_{A_2} = 1) = \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}.$$

Таким образом, $\operatorname{Var} N = 10 \cdot \frac{3}{16} + 20\left(\frac{1}{8} - \frac{1}{16}\right) = 3{,}125.$ □

Задача 1.4.5. Пусть $X_1, \ldots, X_n$ — независимые одинаково распределенные случайные величины со средним μ и дисперсией σ^2. Найдите среднее случайной величины

$$S = \sum_{i=1}^{n}(X_i - \overline{X})^2, \quad \text{где} \quad \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Решение. Рассмотрим сначала среднее значение и дисперсию случайной величины $\overline{X}$:

$$\mathsf{E}\overline{X} = \frac{1}{n}\,\mathsf{E}\sum_{i=1}^{n} X_i = \frac{1}{n}\sum_{i=1}^{n}\mathsf{E}X_i = \frac{1}{n} \cdot n\mu = \mu,$$

$$\operatorname{Var}\overline{X} = \frac{1}{n^2}\sum_{i=1}^{n}\operatorname{Var} X_i = \frac{\sigma^2}{n}.$$

Далее,

$$\mathsf{E}S = \mathsf{E}\left(\sum_{i=1}^{n} X_i^2 - 2\bar{X}\sum_{i=1}^{n} X_i + n\bar{X}^2\right) = \mathsf{E}\left(\sum_{i=1}^{n} X_i^2 - 2n\bar{X}^2 + n\bar{X}^2\right) =$$

$$= \mathsf{E}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right) = \sum_{i=1}^{n} \mathsf{E}X_i^2 - n\,\mathsf{E}\bar{X}^2 = n(\mu^2 + \sigma^2) - n((\mathsf{E}\bar{X})^2 + \operatorname{Var}\bar{X}) =$$

$$= n(\mu^2 + \sigma^2) - n\left(\mu^2 + \frac{\sigma^2}{n}\right) = \sigma^2(n-1). \quad \square$$

Задача 1.4.6. Пусть (X_k) — последовательность независимых одинаково распределенных положительных случайных величин, причем существуют $\mathsf{E}(X_k) = a$ и $\mathsf{E}X_k^{-1} = b$. Пусть $S_n = \sum\limits_{i=1}^{n} X_i$. Покажите, что $\mathsf{E}(S_m/S_n) = m/n$, если $m \leqslant n$, и $\mathsf{E}(S_m/S_n) = 1 + (m-n)a\,\mathsf{E}(S_n^{-1})$, если $m \geqslant n$.

Решение. 1. При $m \leqslant n$ запишем

$$\mathsf{E}\frac{S_m}{S_n} = \mathsf{E}\frac{X_1 + \ldots + X_m}{S_n} = \sum_{i=1}^{m} \mathsf{E}\frac{X_i}{S_n} = m\,\mathsf{E}\frac{X_1}{S_n}.$$

Но при $m = n$ мы имеем

$$1 = n\,\mathsf{E}\frac{X_1}{S_n}.$$

Поэтому $\mathsf{E}\dfrac{S_m}{S_n} = \dfrac{m}{n}$.

2. При $m > n$ имеем

$$\mathsf{E}\frac{S_m}{S_n} = \mathsf{E}\frac{S_n}{S_n} + \mathsf{E}\sum_{j=n+1}^{m}\frac{X_j}{S_n} = 1 + \sum_{j=n+1}^{m} \mathsf{E}X_j\,\mathsf{E}\frac{1}{S_n} = 1 + (m-n)a\,\mathsf{E}\frac{1}{S_n}.$$

Тут важно отметить, что случайные величины S_n и S_m зависимы. Среднее $\mathsf{E}\dfrac{1}{S_n}$ не превосходит b, поскольку $\dfrac{1}{S_n} \leqslant \dfrac{1}{X_1}$. $\square$

Задача 1.4.7. Предположим, что X и Y — независимые случайные величины, причем $|X|$, $|Y| \leqslant K$ при некотором K. Пусть $Z = XY$; покажите, что

$$\operatorname{Var} Z = \operatorname{Var} X \operatorname{Var} Y + \operatorname{Var} Y(\mathsf{E}X)^2 + \operatorname{Var} X(\mathsf{E}Y)^2,$$

указав свойства математического ожидания, которые вы используете.

Решение. Запишем

$$\operatorname{Var}(XY) = \mathsf{E}(XY)^2 - (\mathsf{E}(XY))^2 = \mathsf{E}X^2\,\mathsf{E}Y^2 - (\mathsf{E}X)^2(\mathsf{E}Y)^2 =$$

$$= (\operatorname{Var} X + (\mathsf{E}X)^2)(\operatorname{Var} Y + (\mathsf{E}Y)^2) - (\mathsf{E}X)^2(\mathsf{E}Y)^2 =$$

$$= \operatorname{Var} X \operatorname{Var} Y + (\mathsf{E}X)^2 \operatorname{Var} Y + (\mathsf{E}Y)^2 \operatorname{Var} X.$$

Мы использовали линейность математического ожидания, формулы $\mathsf{E}Z = \mathsf{E}X\,\mathsf{E}Y$ и $\mathsf{E}Z^2 = \mathsf{E}X^2\,\mathsf{E}Y^2$, которые справедливы благодаря независимости, и конечность всех средних, которая следует из ограниченности случайных величин X и Y. □

Задача 1.4.8. Пусть $a_1, a_2, \ldots, a_n$ — последовательность ежегодного количества дождей в Кембридже в последующие n лет; предположим, что $a_1, a_2, \ldots, a_n$ — независимые одинаково распределенные случайные величины. Будем говорить, что k — рекордный год, если $a_k > a_i$ для всех $i < k$ (таким образом, первый год всегда рекордный). Пусть $Y_i = 1$, если i-й год рекордный, и $Y_i = 0$, если нет. Найдите распределение случайной величины Y_i и покажите, что $Y_1, Y_2, \ldots, Y_n$ независимы. Вычислите среднее и дисперсию числа рекордных лет за последующие n лет.

Решение. Упорядочивая $a_1, \ldots, a_n$ в невозрастающем порядке, получаем случайную перестановку множества $1, 2, \ldots, n$. В силу симметрии все такие перестановки равновероятны. Соответственно, упорядочивая первые i количества дождей, получаем случайную перестановку множества $1, 2, \ldots, i$. Поэтому

$$\mathsf{P}(Y_i = 1) = \frac{1}{i}, \quad \mathsf{P}(Y_i = 0) = \frac{i-1}{i}, \quad \mathsf{E}Y_i = \frac{1}{i}, \quad \operatorname{Var} Y_i = \frac{1}{i} - \frac{1}{i^2}.$$

Заметим, что с. в. $Y_1, \ldots, Y_n$ независимы. Действительно, положим

$$X_i = \text{место года } i \text{ среди } 1, \ldots, i \text{ после упорядочения}$$

и

$$\mathsf{P}(X_i = l) = \frac{1}{i}, \quad 1 \leqslant l \leqslant i.$$

Тогда Y_i зависит только от X_i. Поэтому достаточно проверить, что все X_i независимы. Действительно, для всех множеств таких попарно различных номеров l_i, $i = 1, \ldots, n$, что $l_i \leqslant i$, событие $\{X_i = l_i,\ 1 \leqslant i \leqslant n\}$ соответствует единственной перестановке из $n!$ возможных. Таким образом,

$$\frac{1}{n!} = \mathsf{P}(X_i = x_i,\ 1 \leqslant i \leqslant n) = \prod_{1 \leqslant i \leqslant n} \mathsf{P}(X_i = x_i).$$

Наконец,

$$\mathsf{E}\sum_1^n Y_i = \sum_1^n \frac{1}{i},$$

и

$$\operatorname{Var}\Big(\sum_1^n Y_i\Big) = \sum_{i=1}^n \operatorname{Var} Y_i = \sum_1^n \Big(\frac{1}{i} - \frac{1}{i^2}\Big). \qquad \square$$

Задача 1.4.9. Чтобы поймать двух диких яков для размножения, в Гималаи послали экспедицию. Предположим, что яки бродят по горам в одиночестве в случайных направлениях. Вероятность $p \in (0, 1)$ того, что пойманный як — самец, не зависит от предыдущих результатов. Пусть N — число яков, которых необходимо поймать, для того чтобы получить пару.

а) Покажите, что математическое ожидание случайной величины N равно $1 + p/q + q/p$, где $q = 1 - p$.

б) Найдите дисперсию случайной величины N.

Решение. Ясно, что $\mathsf{P}(N = n) = p^{n-1}q + q^{n-1}p$ при $n \geqslant 2$. Тогда $\mathsf{E}N$ равно

$$q\sum_{n=2}^{\infty} np^{n-1} + p\sum_{n=2}^{\infty} nq^{n-1} = pq\left[\frac{1}{1-p} + \frac{1}{1-q} + \frac{1}{(1-p)^2} + \frac{1}{(1-q)^2}\right],$$

что дает $1 + p/q + q/p$. Далее,

$$\operatorname{Var} N = \mathsf{E}N(N-1) + \mathsf{E}N - (\mathsf{E}N)^2,$$

и

$$\mathsf{E}N(N-1) = pq\sum_{n=2}^{\infty} n(n-1)p^{n-2} + pq\sum_{n=2}^{\infty} n(n-1)q^{n-2} =$$

$$= \frac{2pq}{(1-p)^3} + \frac{2pq}{(1-q)^3} = \frac{2p}{q^2} + \frac{2q}{p^2}.$$

Поэтому

$$\operatorname{Var} N = \frac{2p}{q^2} + \frac{2q}{p^2} + \frac{p}{q} + \frac{q}{p} + 1 - \left(\frac{p}{q} + \frac{q}{p} + 1\right)^2 = \frac{1}{q^2} + \frac{1}{p^2} - \frac{p}{q} - \frac{q}{p} - 4. \quad \square$$

Задача 1.4.10. Тарелка спагетти Лиама содержит n штук спагетти. Он выбирает случайным образом два конца и соединяет их вместе. Эту процедуру он проделывает до тех пор, пока не останется свободных концов. Каково математическое ожидание числа колец спагетти в тарелке?

Решение. Полагая

$$X_i = \begin{cases} 1, & \text{если } i\text{-е склеивание дало кольцо,} \\ 0 & \text{иначе,} \end{cases}$$

найдем

$$\mathsf{P}(X_i = 1) = \frac{1}{2n - 2i + 1}.$$

(При i-м склеивании мы имеем $2n - 2(i-1)$ несвязанных концов; для выбранного кольца существует $2n - 2i + 1$ возможностей выбрать другой конец, и только один такой выбор приведет к кольцу.) Таким образом,

$$\mathsf{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathsf{P}(X_i = 1) = \sum_{i=1}^{n} \frac{1}{2n - 2i + 1} = \sum_{i=0}^{n-1} \frac{1}{2i + 1}. \qquad \square$$

Задача 1.4.11. Сара собирает фигурки из пакетов кукурузных хлопьев. В каждом пакете лежит одна фигурка, и n различных фигурок составляют полный набор. Покажите, что среднее число пакетов, которые должна купить Сара, чтобы собрать полный набор фигурок, равно

$$n \sum_{i=1}^{n} \frac{1}{i}.$$

Решение. Число необходимых пакетов равно

$$N = 1 + Y_1 + \ldots + Y_{n-1},$$

где Y_1 — число пакетов, необходимых для того, чтобы найти вторую фигурку, и т. д. Каждое Y_j имеет геометрическое распределение: $\mathsf{P}(Y_j = s) = \left(\frac{j}{n}\right)^{s-1} \frac{n-j}{n}$ и $\mathsf{E} Y_j = n/(n-j)$. Поэтому

$$\mathsf{E} N = n \sum_{j=0}^{n-1} \frac{1}{n-j} = n \sum_{1}^{n} \frac{1}{j} \approx n \ln n. \qquad \square$$

Задача 1.4.12. а) Случайная величина N принимает неотрицательные целые значения. Покажите, что ее среднее значение равно

$$\mathsf{E}(N) = \sum_{k=0}^{\infty} \mathsf{P}(N > k)$$

при условии, что ряд сходится.

б) Каждый пакет завтрака из пшеничных хлопьев содержит точно одну эмблему, и эти эмблемы могут быть трех цветов — голубые, белые и красные. Предположим, что каждая найденная эмблема может быть с одинаковой вероятностью любого цвета и (случайные) цвета разных эмблем независимы.

Найдите вероятность того, что, ища среди k пакетов, вы не найдете эмблем всех цветов.

Пусть N — число пакетов, необходимых для того, чтобы найти эмблемы всех цветов. Покажите, что $\mathsf{E}(N) = 11/2$.

Решение. а) Запишем

$$\sum_{n=1}^{\infty} \left(\mathsf{P}(N = n) \sum_{r=1}^{n} 1 \right) = \sum_{n=1}^{\infty} \sum_{r=n}^{\infty} \mathsf{P}(N = r) = \sum_{n=1}^{\infty} \mathsf{P}(N \geqslant n) = \sum_{k=0}^{\infty} \mathsf{P}(N > k).$$

б) Далее, вероятность $\mathsf{P}(N > k)$ равна

$$\begin{aligned}
&\mathsf{P}(\text{нет всех цветов после } k \text{ испытаний}) = \\
&\quad = \mathsf{P}(\text{после } k \text{ испытаний все эмблемы красные или голубые}) + \\
&\quad + \mathsf{P}(\text{после } k \text{ испытаний все эмблемы красные или белые}) + \\
&\quad + \mathsf{P}(\text{после } k \text{ испытаний все эмблемы белые или голубые}) - \\
&- \mathsf{P}(\text{все красные}) - \mathsf{P}(\text{все голубые}) - \mathsf{P}(\text{все белые}) = 3\left(\left(\frac{2}{3}\right)^k - \left(\frac{1}{3}\right)^k\right).
\end{aligned}$$

Тогда

$$\mathsf{E}N = 3 + 3\left(\sum_{k=3}^{\infty}\left(\frac{2}{3}\right)^k - \sum_{k=3}^{\infty}\left(\frac{1}{3}\right)^k\right) = 3 + 3\left(3\left(\frac{2}{3}\right)^3 - \frac{3}{2}\left(\frac{1}{3}\right)^3\right) = \frac{11}{2}. \qquad \square$$

Задача 1.4.13. Искатель приключений Дипковский играет в рулетку в казино и использует для игры свой метод. Он ставит 5 долларов на красное, и вероятность выигрыша равна 1/2. Если он выигрывает, он уходит, если нет, он ставит 10 долларов снова на красное. Если он выигрывает, он уходит, если нет, он ставит 20 долларов на красное и потом в любом случае прекращает игру. Результаты всех игр независимы. Какова вероятность того, что, когда он закончит игру, он проиграет? Найдите $\mathsf{E}X$ и $\operatorname{Var}X$, если X обозначает величину проигрыша или выигрыша.

В общем случае, если Дипковский использует тот же метод, но ставит S_1, S_2 и S_3 в трех раундах, каковы $\mathsf{P}(X < 0)$, $\mathsf{E}X$ и $\operatorname{Var}X$? Покажите, что если он ставит те же суммы, но в другом порядке, то $\mathsf{P}(X < 0)$ и $\operatorname{Var}X$ минимальны, когда меньшая сумма ставится вначале, а большая — в конце.

Решение. Составим таблицу:

результаты	вероятность	X
выиграла первая ставка	1/2	5
первая ставка проиграла, вторая выиграла	1/4	5
первая и вторая ставки проиграли, третья выиграла	1/8	5
все 3 проиграли	1/8	−35

Тогда $\mathsf{P}(\text{проигрыш}) = 1/8$ и

$$\mathsf{E}X = \frac{5}{2} + \frac{5}{4} + \frac{5}{8} - \frac{35}{8} = 0,$$

$$\operatorname{Var}X = \mathsf{E}X^2 = \frac{25}{2} + \frac{25}{4} + \frac{25}{8} + \frac{1225}{8} = 175.$$

В общем случае

$$\mathsf{E}X = \frac{S_1}{2} + \frac{S_2 - S_1}{4} + \frac{S_3 - S_2 - S_1}{8} - \frac{S_1 + S_2 + S_3}{8} = 0,$$

в то время как

$$\mathrm{P}(\text{проигрыш}) = \frac{1}{8} + \frac{1}{4}I(S_2 < S_1) + \frac{1}{8}I(S_3 < S_1 + S_2)$$

и

$$\operatorname{Var} X = \frac{S_1^2}{2} + \frac{(S_2 - S_1)^2}{4} + \frac{(S_3 - S_2 - S_1)^2}{8} + \frac{(S_1 + S_2 + S_3)^2}{8} = S_1^2 + \frac{S_2^2}{2} + \frac{S_3^2}{4}$$

достигает минимума, когда $S_1 < S_2 < S_3$. □

Задача 1.4.14. Дипковский, сорвиголова с Дикого Запада, окружен бандой противника и сражается не на жизнь, а на смерть за свое спасение. У него $m > 1$ ружей, стреляющих в разных направлениях, и он пытается их использовать последовательно, чтобы создать впечатление, что у него есть сообщники. Когда он оборачивается к следующему ружью и обнаруживает, что оно заряжено, он стреляет с вероятностью $1/2$ и переходит к следующему. Если ружье не заряжено, он с вероятностью $3/4$ заряжает его или просто переходит к следующему с вероятностью $1/4$. Если он решает зарядить его, он стреляет (или не стреляет) с вероятностью $1/2$ и после этого в любом случае переходит к следующему ружью.

Изначально каждое ружье было заряжено независимо от других с вероятностью p. Покажите, что если после каждого перехода это распределение сохраняется, то $p = 3/7$. Вычислите $\mathrm{E}N$ и $\operatorname{Var} N$ числа N заряженных ружей при таком распределении.

Решение. Уравнение инвариантности для p имеет вид

$$\frac{p}{2} + \frac{(1-p)\cdot 3\cdot 1}{4\cdot 2} = p,$$

откуда $p = 3/7$. Действительно, $N = \sum\limits_{1\leqslant j\leqslant m} X_j$, где

$$X_j = \begin{cases} 1, & \text{если ружье } j \text{ заряжено; вероятность } \dfrac{3}{7}, \\ 0, & \text{если ружье } j \text{ не заряжено; вероятность } \dfrac{4}{7}, \end{cases}$$

и эти величины независимы. Тогда

$$\mathrm{E}X = \frac{3}{7}, \quad \operatorname{Var} X = \mathrm{E}X^2 - (\mathrm{E}X)^2 = \frac{12}{49}.$$

Таким образом, $\mathrm{E}N = m\,\mathrm{E}X = \dfrac{3m}{7}$, $\operatorname{Var} N = m \operatorname{Var} X = \dfrac{12m}{49}$. □

Задача 1.4.15. Гамлет, Розенкранц и Гильденстерн подбрасывают монеты. «Отличный от других» выигрывает монеты остальных; если все монеты одинаковы, ничего не происходит. Найдите среднее число бросков, необходимых для того, чтобы один человек вышел из игры, при условии, что первоначально у Гамлета 14 монет, а у Розенкранца и Гильденстерна по 6 монет.

Указание. Рассмотрите среднее в виде $Klmn$, где l, m и n — начальные количества монет и K — неизвестная функция от общего числа монет Гамлета, Розенкранца и Гильденстерна, т. е. она зависит только от общего числа монет $l+m+n$.

Решение. При таких условиях игры общее число монет всегда постоянно. Поэтому если H — число монет Гамлета, R — Розенкранца, G — Гильденстерна, то

$$H+R+G=26. \tag{1.4.25}$$

Такой игре отвечает случайное блуждание на трехмерной решетке, удовлетворяющей равенству (1.4.25), при $H \geqslant 0$, $R \geqslant 0$ и $G \geqslant 0$. Игра заканчивается тогда, когда по крайней мере одно из неравенств превращается в равенство. Возможны переходы

$$\begin{aligned}
&(H,\ R,\ G) \to (H+2,\ R-1,\ G-1) \quad &&\text{с вероятностью } \frac{1}{4},\\
&(H,\ R,\ G) \to (H-1,\ R+2,\ G-1) \quad &&\text{с вероятностью } \frac{1}{4},\\
&(H,\ R,\ G) \to (H-1,\ R-1,\ G+2) \quad &&\text{с вероятностью } \frac{1}{4},\\
&(H,\ R,\ G) \to (H,\ R,\ G) \quad &&\text{с вероятностью } \frac{1}{4}.
\end{aligned}$$

Блуждание стартует при $H=14$, $R=G=6$.

Пусть $E_{H,R,G}$ — среднее число бросков, необходимое для того, чтобы закончить игру, если начальные количества равны H, R, G. Тогда по формуле условного математического ожидания

$$E_{H,R,G} = \frac{1}{4}(1+E_{H,R,G}) + \frac{1}{4}(1+E_{H+2,R-1,G-1}) + \\ + \frac{1}{4}(1+E_{H-1,R-1,G+2}) + \frac{1}{4}(1+E_{H-1,R+2,G-1})$$

при условии, что нам известен результат первого броска. Имеем

$$\frac{3}{4}E_{H,R,G} - 1 = \frac{1}{4}(E_{H+2,R-1,G-1} + E_{H-1,R+2,G-1} + E_{H-1,R-1,G+2})$$

при граничных условиях $E_{0,R,G} = E_{H,0,G} = E_{H,R,0} = 0$.

Предположительный вид искомой функции $E_{H,R,G} = K(H+G+R)HGR$, откуда получаем, что $K = \dfrac{4}{3(H+G+R-2)}$ и

$$E_{14,6,6} = \frac{4 \cdot 14 \cdot 6 \cdot 6}{3 \cdot (26-2)} = \frac{336}{12} = 28.$$

Замечание. Плодотворная догадка такова: $E_{H,R,G}$ является симметричной функцией от H, R, G. Поэтому она является функцией симметрических полиномов от H, R, G, т. е. $H+R+G$, $HG+HR+GR$, HRG и т. д. Из граничного условия следует, что HRG должно появиться в качестве множителя. □

Задача 1.4.16. Мы с вами играем в следующую игру с подбрасыванием монетки. Каждый из нас подбрасывает одну (симметричную) монетку; если монетки упадут одинаковыми сторонами, то я получаю обе монетки; если разными, то вы получаете обе монетки. Первоначально у меня m монет и n монет у вас. Пусть $\mathsf{E}(m, n)$ — средняя продолжительность игры, до тех пор пока у одного из нас не останется монет. Опишите линейную зависимость между $\mathsf{E}(m, n)$, $\mathsf{E}(m-1, n+1)$ и $\mathsf{E}(m+1, n-1)$. Представьте ее как функцию от $m+n$ и $m-n$, получите, что $\mathsf{E}(m, n)$ — квадратичная функция от m и n, и, таким образом, используя подходящие начальные условия, покажите, что

$$\mathsf{E}(m, n) = mn.$$

Решение. Как и ранее, получим

$$\mathsf{E}(m, n) = \frac{1}{2}\mathsf{E}(m-1, n+1) + \frac{1}{2}\mathsf{E}(m+1, n-1) + 1.$$

Ответ: $\mathsf{E}(m, n) = mn$. □

Задача 1.4.17. Каково среднее значение времени, которое пройдет до тех пор, пока заводной апельсин не упадет со стола в задаче про тетю Агату (см. задачу 1.2.10)?

Решение. Продолжая решение задачи 1.2.10, обозначим через E_ℓ (условное) математическое ожидание времени падения, если апельсин начинает двигаться с расстояния 10ℓ см от левого края стола. Тогда $E_0 = E_{20} = 0$ и

$$E_\ell = \frac{3}{5}E_{\ell-1} + \frac{2}{5}E_{\ell+1} + 1,$$

или $E_{\ell+1} = \frac{5}{2}E_\ell - \frac{3}{2}E_{\ell-1} - \frac{5}{2}$. Векторы $u_\ell = (E_\ell, E_{\ell+1})$ и $v = (0, -5/2)$ удовлетворяют соотношению $u_\ell = u_{\ell-1}A + v$. Здесь, как и ранее, матрица

$$A = \begin{pmatrix} 0 & -\frac{3}{2} \\ 1 & \frac{5}{2} \end{pmatrix}$$

имеет собственные значения $\lambda_1 = 3/2$ и $\lambda_2 = 1$ и собственные векторы $e_1 = (1, 3/2)$ и $e_2 = (1, 1)$. Заметим, что $v = 5(e_2 - e_1)$.

Итерирование дает

$$u_\ell = u_0 A^l + v \sum_{1 \leqslant j \leqslant l-1} A^j,$$

и если $u_0 = a_1 e_1 + a_2 e_2$, то

$$u_l = \left(a_1 \lambda_1^l + a_2 - 5\left(\frac{\lambda_1^\ell - 1}{\lambda_1 - 1} - \ell\right), \frac{3}{2} a_1 \lambda_1^\ell + a_2 - 5\left(\frac{3}{2}\frac{\lambda_1^\ell - 1}{\lambda_1 - 1} - \ell\right)\right).$$

Подставляя в первую компоненту $E_0 = 0$ (для $\ell = 0$), получим

$$a_1 + a_2 = 0,$$

а подставляя во вторую компоненту $E_{20} = 0$ (для $\ell = 19$), получим

$$a_1 \left(\frac{3}{2}\right)^{20} + a_2 = 5\left(\frac{(3/2)^{20} - 3/2}{1/2} - 19\right) = 5\left(\frac{(3/2)^{20} - 1}{1/2} - 20\right).$$

Таким образом,

$$a_1 = -a_2 = \frac{10((3/2)^{20} - 1) - 100}{(3/2)^{20} - 1} = 10 - \frac{100}{(3/2)^{20} - 1},$$

и

$$E_{10} = \left(10 - \frac{100}{(3/2)^{20} - 1}\right)\left(\left(\frac{3}{2}\right)^{10} - 1\right) + 50 - 10\left(\left(\frac{3}{2}\right)^{10} - 1\right) =$$
$$= 50 + 10\left(\left(\frac{3}{2}\right)^{10} - 1\right) - 10\left(\left(\frac{3}{2}\right)^{10} - 1\right) - \frac{100}{(3/2)^{10} + 1} = 50 - \frac{100}{(3/2)^{10} + 1}.$$

Грубая оценка дает $E_{10} \geqslant 48$.

Замечание. Гораздо более короткое решение (студенты иногда предлагают его, чувствуя, что оно правильно, но не могут строго обосновать) следующее. Пусть τ обозначает время падения. Как было показано в решении задачи 1.2.10, положение S_τ в момент τ равно

$$S_\tau = \begin{cases} 20 & \text{с вероятностью } \dfrac{1}{(3/2)^{10} + 1}, \\ 0 & \text{с вероятностью } 1 - \dfrac{1}{(3/2)^{10} + 1}. \end{cases}$$

После n продвижений положение равно сумме независимых с. в.:

$$S_n = S_0 + X_1 + \ldots + X_n,$$

где S_0 — начальное положение (в нашем случае равное 10) и X_i — приращение при j-м продвижении:

$$X_i = \begin{cases} +1 & \text{с вероятностью } \dfrac{2}{5}, \\ -1 & \text{с вероятностью } \dfrac{3}{5}. \end{cases}$$

Записывая

$$U_n = S_0 + (X_1 - \mathsf{E}X_1) + \ldots + (X_n - \mathsf{E}X_n) = S_n - n\,\mathsf{E}X_1,$$

получим $\mathsf{E}U_n = \mathsf{E}S_0$. То же самое справедливо и для случайного момента времени τ: $\mathsf{E}U_\tau = \mathsf{E}S_0$, но этот факт основан на теории мартингалов. Поэтому

$$10 = 20\frac{1}{(3/2)^{10}+1} - \mathsf{E}\tau\left(\frac{2}{5} - \frac{3}{5}\right)$$

и

$$\mathsf{E}\tau = 50 - \frac{100}{(3/2)^{10}+1} = E_{10},$$

как и ранее. □

Аналогичные соображения позволяют дать и короткое решение задачи 1.4.15. Обозначим через H_n, R_n и G_n число монет у Гамлета, Розенкранца и Гильденстерна после n бросаний. Пусть

$$Y_n = H_n R_n G_n + \frac{3}{4}(H_n + R_n + G_n - 2)n.$$

Ясно, что $\frac{3}{4}(H_n + R_n + G_n - 2) = 18$, поскольку $H_n + G_n + R_n = 26$. Мы оставляем в качестве упражнения проверку того, что условное ожидание равно

$$\mathsf{E}(H_{n+1}R_{n+1}G_{n+1} \mid H_n, R_n, G_n) = H_n G_n R_n - 18.$$

Тогда ясно, что

$$\mathsf{E}Y_{n+1} = \mathsf{E}H_{n+1}R_{n+1}G_{n+1} + 18(n+1) = \mathsf{E}[\mathsf{E}(H_{n+1}R_{n+1}G_{n+1} \mid H_n, R_n, G_n)] + \\ + 18(n+1) = \mathsf{E}H_nR_nG_n + 18n = \mathsf{E}Y_n = \mathsf{E}Y_0 = 14\cdot 6\cdot 6 = 504.$$

Ключевой факт (требующий для своего обоснования теории мартингалов) состоит в том, что это равенство справедливо и в случайный момент τ, когда игра заканчивается (т. е. один из игроков остается без монет): $\mathsf{E}Y_\tau = 504$. С другой стороны $\mathsf{E}Y_\tau = 18\,\mathsf{E}\tau$, поскольку $H_\tau R_\tau G_\tau = 0$. Таким образом, $\mathsf{E}\tau = 504/18 = 28$.

Задача 1.4.18. а) Определите ковариацию $\mathrm{Cov}(X, Y)$ случайных величин X и Y. Покажите, что

$$\mathrm{Var}(X+Y) = \mathrm{Var}\,X + \mathrm{Var}\,Y + 2\,\mathrm{Cov}(X, Y).$$

б) Симметричная кость имеет две зеленые стороны, две красные, и две белые; ее подбрасывают один раз.

Пусть

$$X = \begin{cases} 1, & \text{если зеленая сторона сверху,} \\ 0 & \text{иначе,} \end{cases}$$

$$Y = \begin{cases} 1, & \text{если синяя сторона сверху,} \\ 0 & \text{иначе.} \end{cases}$$

Найдите $\mathrm{Cov}(X, Y)$.

Решение. а) Положим $\mu_X = \mathrm{E}X$, $\mu_Y = \mathrm{E}Y$ и $\mathrm{E}(X+Y) = \mathrm{E}X + \mathrm{E}Y$. Тогда

$$\mathrm{Cov}(X, Y) = \mathrm{E}(X - \mu_X)(Y - \mu_Y),$$

$$\mathrm{Var}(X+Y) = \mathrm{E}(X + Y - \mu_X - \mu_Y)^2 = \mathrm{E}(X - \mu_X)^2 + \mathrm{E}(Y - \mu_Y)^2 + \\ + 2\,\mathrm{E}(X - \mu_X)(Y - \mu_Y) = \mathrm{Var}\,X + \mathrm{Var}\,Y + 2\,\mathrm{Cov}(X, Y).$$

б) Следовательно,

$$\mathrm{P}(\text{зеленая сторона сверху}) = \mathrm{P}(\text{синяя сторона сверху}) = \frac{1}{3},$$

поэтому

$$\mu_X = \mu_Y = \frac{1}{3}.$$

Наконец,

$$\mathrm{Cov}(X, Y) = \mathrm{E}XY - \mu_X\mu_Y.$$

Но $XY = 0$ с вероятностью 1, поэтому $\mathrm{E}XY = 0$. Таким образом, $\mathrm{Cov}(X, Y) = -1/9$. □

Задача 1.4.19. Такси ездит между четырьмя поселками W, X, Y, Z, расположенными в углах прямоугольника. Четыре дороги, соединяющие поселки, расположены по сторонам прямоугольника; расстояние от W до X и от Y до Z — 5 миль, от W до Z и от Y до X — 10 миль. После доставки пассажира такси ждет следующего вызова, потом едет за новым пассажиром и привозит его на место назначения. Звонки могут поступать из любого поселка с вероятностью $1/4$, и каждый пассажир может ехать в любой из поселков с вероятностью $1/3$. Естественно, проезжая между парой соседних поселков, такси выбирает прямую дорогу, иначе (когда такси едет от W к Y, или от X к Z, или наоборот) это неважно. Расстояния внутри поселков малы. Пусть D — расстояние, которое нужно проехать, чтобы забрать и отвезти одного пассажира. Найдите вероятности, с которыми D принимает свои возможные значения. Найдите $\mathrm{E}D$ и $\mathrm{Var}\,D$.

Решение. Пусть стороны прямоугольника равны $a = 5$, $b = 10$, $c = 5$ и $d = 10$ миль. Без потери общности предположим, что такси расположено в вершине, смежной с a и b, тогда

$$D = \begin{cases} 0 + a = 5 & \text{с вероятностью } 1/12, \\ 0 + b = 10 & \text{с вероятностью } 1/12, \\ a + a = 10 & \text{с вероятностью } 1/12, \\ a + b = 15, & \text{с вероятностью } 1/4, \\ b + b = 20 & \text{с вероятностью } 1/12, \\ 2a + b = 20 & \text{с вероятностью } 1/6, \\ a + 2b = 25 & \text{с вероятностью } 1/6, \\ 2a + 2b = 30 & \text{с вероятностью } 1/12. \end{cases}$$

Поэтому (в милях)

$$\mathsf{E}D = \frac{210}{12} = \frac{35}{2}, \ \mathsf{E}D^2 = \frac{4250}{12} = \frac{2125}{6}, \quad \operatorname{Var} D = \mathsf{E}D^2 - (\mathsf{E}D)^2 = \frac{575}{12}. \ \square$$

Задача 1.4.20. Пусть X — целочисленная случайная величина с распределением

$$\mathsf{P}(X = n) = n^{-s}/\zeta(s),$$

где $s > 1$ и

$$\zeta(s) = \sum_{n \geqslant 1} n^{-s}.$$

Пусть $1 < p_1 < p_2 < p_3 < \ldots$ — простые числа, и пусть A_k — событие $\{X$ делится на $p_k\}$. Найдите $\mathsf{P}(A_k)$ и покажите, что события $A_1, A_2, \ldots$ независимы. Получите, что

$$\prod_{k=1}^{\infty} (1 - p_k^{-s}) = 1/\zeta(s).$$

Решение. Запишем

$$\mathsf{P}(A_k) = \sum_{n \text{ делится на } p_k} \frac{n^{-s}}{\zeta(s)} = \sum_{n:\, n = p_k l} \frac{n^{-s}}{\zeta(s)} = p_k^{-s} \sum_{l \geqslant 1} \frac{l^{-s}}{\zeta(s)} = p_k^{-s}.$$

Аналогично $\forall A_{k_1}, \ldots, A_{k_i}$ $(1 \leqslant k_1 < \ldots < k_i)$ мы имеем

$$\mathsf{P}(A_{k_1} \cap \ldots \cap A_{k_i}) = p_{k_1}^{-s} \ldots p_{k_i}^{-s} = \prod_{l=1}^{i} \mathsf{P}(A_{k_l}),$$

что означает, что события $A_1, A_2, \ldots$ независимы.

Наконец, вероятность того, что X не делится на p_k, равна $1-p_k^{-s}$. Тогда $\prod\limits_{k=1}^{\infty}(1-p_k^{-s})$ — это вероятность того, что X не делится ни на одно простое число, т. е. $X=1$. Эта вероятность равна $1/\zeta(s)$. □

Замечание. Функция $\zeta(s)$ — это знаменитая *дзета-функция Римана*, которая является предметом гипотезы Римана, одной из немногих задач, поставленных в XIX в. и не решенных до сих пор. Задача, сформулированная выше, дает представление дзета-функции в виде бесконечного произведения простых чисел.

Г. Ф. Б. Риман (1826—1866), замечательный немецкий математик, внес также большой вклад в геометрию (римановы многообразия, риманова метрика). В нашей книге мы будем говорить об *интегрировании в смысле Римана*, отличающемся от более общего интегрирования в смысле Лебега; см. ниже.

§ 1.5. Биномиальное, пуассоновское и геометрическое распределения. Производящие функции распределений, производящие функции моментов и характеристические функции

La vie n'est bonne qu'à deux choses:
dècouvrir les mathèmatiques et enseigner les mathèmatiques.

Только две вещи придают жизни смысл:
математические открытия и преподавание математики.

С.-Д. Пуассон (1781—1840), французский математик

Биномиальное распределение естественным образом появляется в задачах с подбрасыванием монетки. Рассмотрим случайную величину X, равную числу орлов, выпадающих при n испытаниях с монетками одинакового типа. Ее удобно представить в виде

$$X = Y_1 + \ldots + Y_n,$$

где с. в. $Y_1, \ldots, Y_n$ независимы и одинаково распределены,

$$Y_j = \begin{cases} 1, & \text{если } j\text{-е испытание дает орел,} \\ 0, & \text{если } j\text{-е испытание дает решку.} \end{cases}$$

Предполагая, что вероятность того, что выпадет орел, $\mathsf{P}(Y_j=1)=p$ и вероятность того, что выпадет решка, $\mathsf{P}(Y_j=0)=q=1-p$, получим

$$\mathsf{P}(X=k) = C_n^k p^k q^{n-k}, \quad 0 \leqslant k \leqslant n. \tag{1.5.1}$$

Такое вероятностное распределение (и сами с. в.) называется *биномиальным* (или (n, p)-биномиальным), потому что оно связано с биномиальным разложением $\sum\limits_{0\leqslant k\leqslant n} C_n^k p^k q^{n-k} = (p+q)^n = 1$.

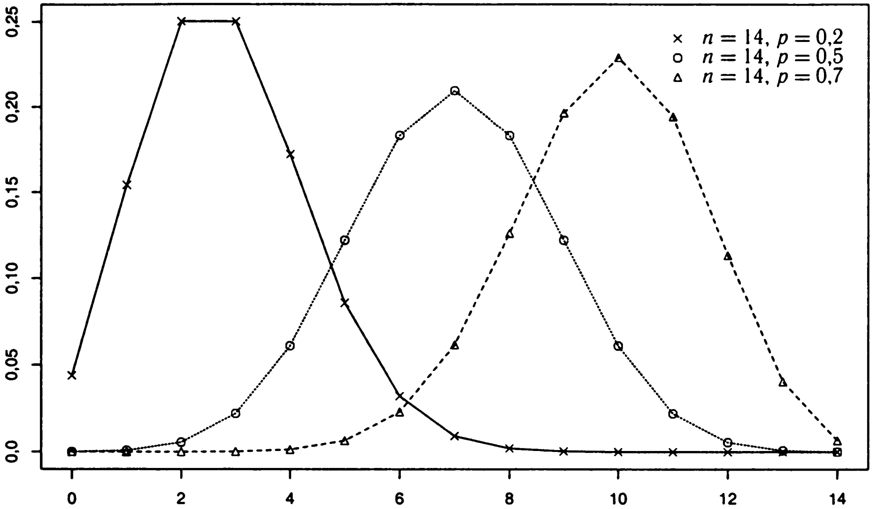


Рис. 1.3. Биномиальное распределение

Значения биномиальных вероятностей показаны на рис. 1.3.

Благодаря представлению $X = Y_1 + \ldots + Y_n$ сумма $X + X'$ независимых (n, p)- и (n', p)-биномиальных с. в. X и X' есть $((n+n'), p)$-биномиальная с. в. Из этого разложения также следует, что

$$\mathsf{E}X = n\,\mathsf{E}Y_1 = np, \qquad \operatorname{Var} X = n \operatorname{Var} Y_1 = npq. \tag{1.5.2}$$

Мы будем обозначать биномиальную с. в. X следующим образом: $X \sim \sim \operatorname{Bin}(n, p)$.

Другие хорошо известные разложения также дают полезные вероятностные распределения. Например, при подбрасывании монетки до первого появления орла результатами будут числа 0, 1, ... (означающие число решек до появления первого орла). Пусть X обозначает число решек до появления первого орла. Тогда

$$\mathsf{P}(X = k) = pq^k, \qquad k = 0, 1, 2, \ldots, \tag{1.5.3}$$

что равно вероятности появления последовательности $PP\ldots PO$, в которой на первых k местах решки P и на $(k+1)$-м месте — орел O. Сумма всех вероятностей (сумма геометрической прогрессии) равна $p/(1-q) = 1$. Вероятность «решек» равна $\mathsf{P}(X \geqslant k) = q^k$, $k \geqslant 0$.

Говорят, что с. в. X, дающая число испытаний до появления первого орла, имеет *геометрическое распределение* (или является геометри-

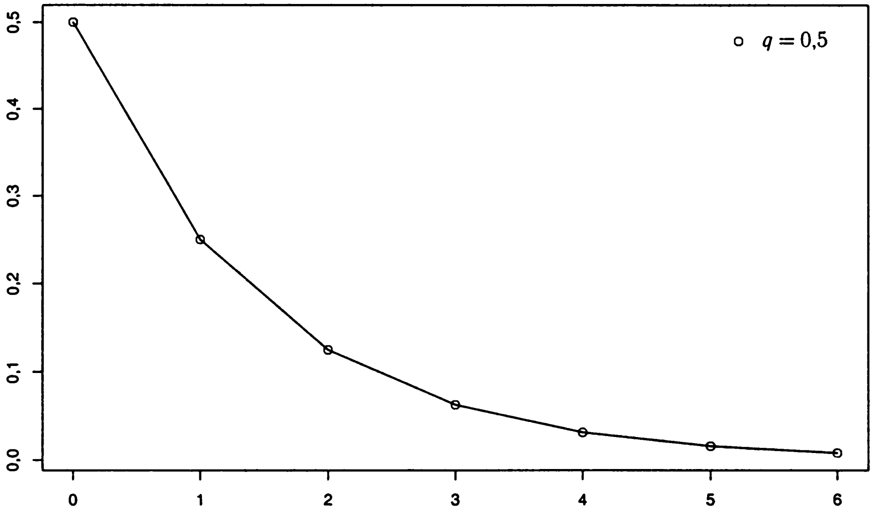


Рис. 1.4. Геометрическое распределение

ческой) с параметром q. На рис. 1.4 приведен график геометрического распределения.

Среднее значение и дисперсия геометрического распределения вычисляются по формулам

$$\mathsf{E}X = \sum_{k\geqslant 1}(1-q)kq^k = \sum_{k\geqslant 1} q^k = \frac{q}{p}, \tag{1.5.4}$$

$$\begin{aligned}\operatorname{Var} X = \mathsf{E}X^2 - (\mathsf{E}X)^2 = \sum_{k\geqslant 1} k^2 pq^k - \frac{q^2}{p^2} =\\ = pq^2\sum_{k\geqslant 2} k(k-1)q^{k-2} + pq\sum_{k\geqslant 1} kq^{k-1} - \frac{q^2}{p^2} =\\ = pq^2\frac{d^2}{dq^2}\cdot\frac{1}{p} + \frac{q}{p} - \frac{q^2}{p^2} = pq^2\frac{2}{p^3} + \frac{q}{p} - \frac{q^2}{p^2} = \frac{q^2}{p^2} + \frac{q}{p} = \frac{q}{p^2}.\end{aligned} \tag{1.5.5}$$

Отметим особенность геометрической с. в.: для любого $m < n$ выполняется равенство

$$\mathsf{P}(X \geqslant n \mid X \geqslant m) = \mathsf{P}(X \geqslant n - m), \tag{1.5.6}$$

называемое *свойством отсутствия памяти*. Другой особенностью геометрического распределения является то, что $\min\,[X, X']$ двух независимых геометрических с. в. X и X' с параметрами q и q' также геометрическая с. в., с параметром qq': $\mathsf{P}(\min[X, X'] \geqslant k) = (qq')^k$, $k \geqslant 0$.

Иногда геометрическое распределение определяют при помощи вероятностей $p_k = pq^{k-1}$, $k = 1, 2, \ldots$, соответствующих количеств испытаний до (включительно) первого орла. Это приводит к другому значению среднего $1/p$; дисперсия при этом не меняется. Мы будем писать $X \sim \mathrm{Geom}(q)$ для геометрической с. в. X (в обоих определениях).

Геометрическое распределение появляется во многих ситуациях. Например, число взмахов крыльями, которые делает птица, перед тем как взлететь, имеет геометрическое распределение.

Обобщением геометрического распределения является *отрицательное биномиальное распределение*. В этом случае соответствующая с. в. X обозначает число решек до появления r-го орла. Иными словами, $X = X_1 + \ldots + X_r$, где $X_i \sim \mathrm{Geom}(q)$ и X_i независимы. Прямые вычисления показывают, что

$$\mathsf{P}(X = k) = C^k_{k+r-1} p^r q^k, \qquad k = 0, 1, 2, \ldots$$

Действительно, равенство $X = k$ означает, что было $k + r$ испытаний, из которых k раз выпала решка и r раз орел, причем при последнем броске выпал орел.

Мы будем обозначать отрицательное биномиальное распределение $X \sim \mathrm{NegBin}(q, r)$.

Еще один пример — это пуассоновское распределение с параметром $\lambda \geqslant 0$; оно появляется в разложении $e^\lambda = \sum\limits_{k \geqslant 0} \dfrac{\lambda^k}{k!}$ и названо в честь С.-Д. Пуассона (1781—1840), выдающегося французского математика и физика. Мы снова приписываем неотрицательным числам вероятности, и вероятность, приписанная числу k, равна $\dfrac{\lambda^k}{k!} e^{-\lambda}$. Случайная величина X, для которой

$$\mathsf{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \qquad k = 0, 1, 2, \ldots, \tag{1.5.7}$$

называется *пуассоновской*. Эти вероятности появляются из биномиальных при $n \to \infty$ и $p = \lambda/n \to 0$:

$$\frac{n!}{(n-k)!\,k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{n(n-1)\ldots(n-k+1)}{n^k} \frac{\lambda^k}{k!} \frac{(1-\lambda/n)^n}{(1-\lambda/n)^k},$$

а это выражение стремится к $\dfrac{\lambda^k}{k!} e^{-\lambda}$.

Приведенные выше рассуждения объясняют тот факт, что сумма $X + X'$ независимых пуассоновских с. в. X и X' с параметрами λ и λ' также пуассоновская, с параметром $\lambda + \lambda'$. Этот факт также может быть доказан непосредственными вычислениями.

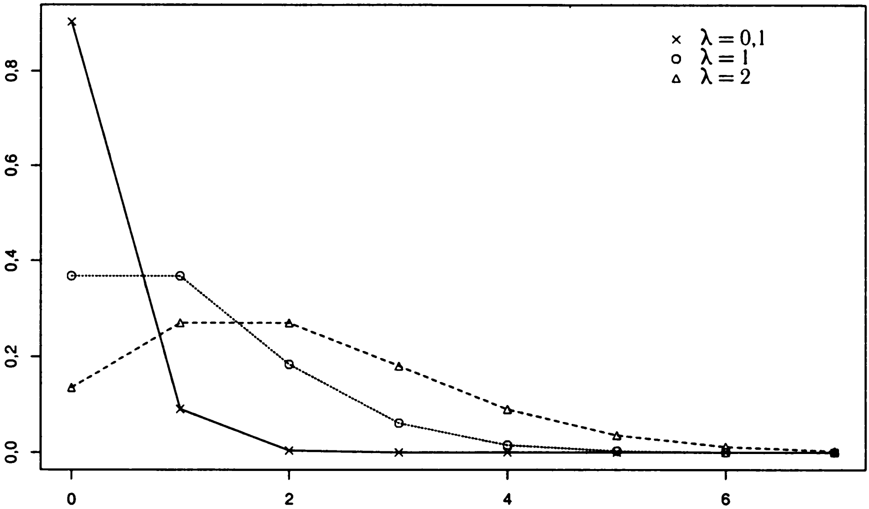


Рис. 1.5. Распределение Пуассона

График значений пуассоновских вероятностей представлен на рис. 1.5. Среднее значение $\mathsf{E}X$ и дисперсия $\mathrm{Var}\,X$ пуассоновских с. в. равны λ:

$$\sum_{k\geqslant 0}\frac{\lambda^k}{k!}e^{-\lambda}k=\lambda,\quad \sum_{k\geqslant 0}\frac{\lambda^k}{k!}e^{-\lambda}k^2-\lambda^2=\lambda. \tag{1.5.8}$$

Мы будем писать $X \sim \mathrm{Po}(\lambda)$ для пуассоновской с. в. X.

Пуассоновское распределение широко используется во многих ситуациях. Знаменитый (хотя и устрашающий) пример (с которого оба автора начали свое изучение теории вероятностей): число прусских кавалеристов в каждом из шестнадцати корпусов, убитых ударом копыта лошади в каждом из 1875—1894 гг. Этот пример идеально подходит под пуассоновское распределение! Удивительно, что он проникал в большинство учебников до конца XX в. и энтузиазм нескольких поколений студентов от этого не угасал.

Производящая функция $\varphi(s)$ $(=\varphi_X(s))$ распределения случайной величины X, принимающей конечное или счетное число действительных значений x_j с вероятностями p_j, определяется следующим образом:

$$\varphi_X(s)=\mathsf{E}s^X=\sum_j p_j s^{x_j}=\sum_{\omega\in\Omega}p(\omega)s^{X(\omega)}; \tag{1.5.9}$$

обычно рассматривают только случай $s>0$. Функция

$$M_X(t)=\varphi_X(e^t)=\mathsf{E}e^{tX} \tag{1.5.10}$$

называется *производящей функцией моментов*. Производящие функции распределений и моментов могут не существовать для некоторых значений аргумента (название «производящая функция моментов» происходит от представления $M_X(t)=\sum\limits_{n=0}^{\infty}(\mathsf{E}X^n)t^n/n!$, так как математическое ожидание $\mathsf{E}X^n$ называется n-м *моментом* случайной величины X).

С другой стороны, *характеристическая функция* $\psi(t)$ $(=\psi_X(t))$ определена для всех действительных t:

$$\psi(t)=\mathsf{E}e^{itX}=\sum_j p_j e^{itx_j}=\sum_{\omega\in\Omega}p(\omega)e^{itX(\omega)}, \tag{1.5.11}$$

при этом она принимает комплексные значения, по модулю меньшие 1. Полезность этих двух функций видна из следующих свойств.

1. Если X и Y — независимые с. в., то

$$\varphi_{X+Y}(s)=\varphi_X(s)\varphi_Y(s),\qquad \psi_{X+Y}(t)=\psi_X(t)\psi_Y(t). \tag{1.5.12}$$

Действительно,

$$\varphi_{X+Y}(s)=\mathsf{E}s^{X+Y}=\mathsf{E}s^Xs^Y=\mathsf{E}s^X\,\mathsf{E}s^Y=\varphi_X(s)\varphi_Y(s),$$

и аналогично в случае характеристических функций.

2. Математическое ожидание $\mathsf{E}X$ и дисперсия $\operatorname{Var}X$ (а также и другие моменты $\mathsf{E}X^j$) выражаются в терминах производных функций φ и ψ при $s=1$ и $t=0$ соответственно:

$$\mathsf{E}X=\frac{d}{ds}\varphi_X(s)\Big|_{s=1}=\frac{1}{i}\cdot\frac{d}{dt}\psi_X(t)\Big|_{t=0}, \tag{1.5.13}$$

$$\operatorname{Var}X=\left(\frac{d^2}{ds^2}\varphi_X(s)+\frac{d}{ds}\varphi_X(s)-\left(\frac{d}{ds}\varphi_X(s)\right)^2\right)\Big|_{s=1}=$$
$$=\left(-\frac{d^2}{dt^2}\psi_X(t)+\left(\frac{d}{dt}\psi_X(t)\right)^2\right)\Big|_{t=0}. \tag{1.5.14}$$

3. Функции $\varphi_X(s)$ и $\psi_X(t)$ однозначно определяют распределение с. в.: если $\varphi_X(s)=\varphi_Y(s)$ или $\psi_X(t)=\psi_Y(t)$ во всей области (т. е. при $s>0$ и $t\in\mathbb{R}$ соответственно), то с. в. X и Y принимают одинаковые значения с одинаковыми вероятностями. В этом случае мы будем писать $X\sim Y$.

Последнее свойство очень полезно для определения многих распределений, но доказательство этого факта выходит за рамки нашей книги. Тем не менее, если случайная величина X принимает конечное число целых значений $n_1,\ldots,n_m$ с вероятностями $p_1,\ldots,p_m$, то $\varphi_X(s)=\sum s^{n_j}p_j=$ $=s^{\underline{n}}\sum s^{n_j-\underline{n}}p_j=s^{\underline{n}}\underline{\varphi}(s)$, где $\underline{n}=\min n_j$ и $\underline{\varphi}(s)$ — многочлен. Далее, мы знаем, что если $\varphi_X(s)\equiv\varphi_Y(s)$, то φ_X и φ_Y имеют одинаковые коэффициенты при одинаковых степенях s; это и означает, что $X\sim Y$.

Однозначность также справедлива для характеристических функций: если $\psi_X(t) \equiv \psi_Y(t)$, то $X \sim Y$. Если X и Y принимают целые значения, то этот факт имеет отношение к единственности функции с данным разложением Фурье. Опять-таки мы не будем обсуждать это в настоящем томе.

Задача 1.5.1. Пусть N — дискретная случайная величина,

$$\mathrm{P}(N = k) = (1 - p)^{k-1} p, \quad k = 1, 2, 3, \ldots,$$

где $0 < p < 1$. Покажите, что $\mathrm{E}N = 1/p$.

Решение. Справедливо равенство

$$\mathrm{E}[N] = \sum_{k=1}^{\infty} k(1 - p)^{k-1} p = -p \frac{d}{dp}\left(\frac{1}{1 - (1 - p)}\right) = \frac{1}{p}.$$

Другое решение:

$$\mathrm{E}[N] = p \cdot 1 + (1 - p)(1 + \mathrm{E}[N]) \Rightarrow \mathrm{E}[N] = \frac{1}{p}. \qquad \square$$

Задача 1.5.2. а) Предположим, что случайная величина X имеет геометрическое распределение, $\mathrm{P}(X = n) = pq^n$, $n = 0, 1, \ldots$, где $q = 1 - p$, $0 < p < 1$. Покажите, что

$$\mathrm{E}\frac{1}{(X + 1)} = -\frac{p}{q} \ln p.$$

б) Две кости подбрасывают последовательно, до тех пор пока ни на одной из них не выпадает 6. Найдите вероятность того, что при последнем броске по крайней мере на одной из костей выпадет 5.

Предположим, что ваш выигрыш — это сумма Y значений, выпавших на двух костях при последнем броске, а время, которое понадобилось, чтобы получить этот выигрыш, равно числу N бросков.

Найдите средний выигрыш $\mathrm{E}Y$ и средний доход $\mathrm{E}\frac{Y}{N}$, используя аппроксимацию $\ln \frac{36}{25} \approx \frac{11}{30}$.

Решение. а) Заметим, что $N = X + 1$, а вероятность успеха равна $p = 25/36$. Тогда

$$\mathrm{E}\frac{1}{N} = \sum_{n \geqslant 1} \frac{1}{n} q^{n-1} p = \frac{p}{q} \sum_{n \geqslant 1} \frac{1}{n} q^n = \frac{p}{q} \sum_{n \geqslant 1} \int_0^q x^{n-1}\, dx = \frac{p}{q} \int_0^q \frac{dx}{1 - x} = \frac{p}{q} \int_p^1 \frac{du}{u} = \frac{p}{q} \ln \frac{1}{p}.$$

б) В силу симметрии

$$\mathrm{P}(\alpha) = \frac{\mathrm{P}(\beta)}{\mathrm{P}(\gamma)} = \frac{9}{25},$$

где $\alpha =$ (по крайней мере одна пятерка при последнем броске), $\beta =$ (по крайней мере одна пятерка и ни одной шестерки при последнем броске), $\gamma =$ (ни одной шестерки при последнем броске).

Мы имеем $Y = Y_1 + Y_2$, где Y_i — число, выпавшее на i-й кости. Таким образом, $\mathsf{E}Y = 2\,\mathsf{E}Y_1$, $\mathsf{E}Y_1 = \frac{1}{5}(1+2+3+4+5) = 3$, $\mathsf{E}Y = 6$. Далее, $N \sim \mathrm{Geom}(q)$ с параметром $q = 11/36$; с. в. Y и N независимы:

$$\mathsf{P}(Y = y,\ N = n) = q^{n-1}p\,\mathsf{P}(\text{сумма } y \text{ при } n\text{-м броске} \mid \text{нет шестерки}) = \\ = \mathsf{P}(Y = y)\,\mathsf{P}(N = n).$$

Таким образом, $\mathsf{E}\dfrac{Y}{N} = \mathsf{E}Y\,\mathsf{E}\dfrac{1}{N} = 6 \cdot \dfrac{25}{11}\ln\left(\dfrac{36}{25}\right) \approx 5.$ □

Задача 1.5.3. 1. Предположим, что X и Y — дискретные с. в. с конечными средними и дисперсиями. Докажите следующие утверждения:

а) $\mathsf{E}(X+Y) = \mathsf{E}X + \mathsf{E}Y$;

б) $\mathrm{Var}(X+Y) = \mathrm{Var}\,X + \mathrm{Var}\,Y + 2\,\mathrm{Cov}(X,\ Y)$;

в) из того, что X и Y независимы, следует, что $\mathrm{Cov}(X,\ Y) = 0$.

2. Монетка падает вверх орлом с вероятностью $p > 0$ и решкой с вероятностью $q = 1 - p$. Пусть X_n — количество подбрасываний, необходимых для получения n орлов. Найдите производящую функцию распределения X_1 и вычислите ее среднее значение и дисперсию. Каково среднее и дисперсия для X_n?

Решение. 1. а) Справедливы равенства $\mathsf{E}X = \sum\limits_a a\,\mathsf{P}(X = a)$, $\mathsf{E}Y = \sum\limits_b b\,\mathsf{P}(Y = b)$, $\{\omega\colon X = a\} = \bigcup\limits_b \{\omega\colon X = a,\ Y = b\}$;

$$\mathsf{E}X + \mathsf{E}Y = \sum_{a,b} a\,\mathsf{P}(X = a,\ Y = b) + b\,\mathsf{P}(X = a,\ Y = b) = \\ = \sum_c c \sum_{a+b=c} \mathsf{P}(X = a,\ Y = b) = \sum_c c\,\mathsf{P}(X + Y = c) = \mathsf{E}(X+Y).$$

б) По определению $\mathrm{Var}(X) = \mathsf{E}((X - \mathsf{E}X)^2) = \mathsf{E}X^2 - (\mathsf{E}X)^2$. Тогда

$$\mathrm{Var}(X+Y) = \mathsf{E}(X + Y - (\mathsf{E}X + \mathsf{E}Y))^2 = \mathsf{E}(X - \mathsf{E}X)^2 + \mathsf{E}(Y - \mathsf{E}Y))^2 + \\ + 2\,\mathsf{E}(X - \mathsf{E}X)(Y - \mathsf{E}Y) = \mathrm{Var}\,X + \mathrm{Var}\,Y + 2\,\mathrm{Cov}(X,\ Y).$$

в) Если случайные величины X, Y независимы, то

$$\mathsf{E}XY = \sum_c c\,\mathsf{P}(XY = c) = \sum_c c \sum_{ab=c} \mathsf{P}(X = a,\ Y = b) = \\ = \sum_{a,b} ab\,\mathsf{P}(X = a)\,\mathsf{P}(Y = b) = \left(\sum_a a\,\mathsf{P}(X = a)\right)\left(\sum_b b\,\mathsf{P}(Y = b)\right) = \mathsf{E}X\,\mathsf{E}Y.$$

Если X, Y независимы, то таковыми же являются $X - \mathsf{E}X$, $Y - \mathsf{E}Y$ и, таким образом

$$\mathrm{Cov}(X, Y) = \mathsf{E}(X - \mathsf{E}X)(Y - \mathsf{E}Y) = 0.$$

2. Мы имеем $\mathsf{P}(X_1 = k) = pq^{k-1}$, тогда $\varphi(s) = \varphi_{X_1}(s) = \dfrac{ps}{1-qs}$ и $\varphi'(s) = \dfrac{p}{(1-qs)^2}$. Поэтому

$$\mathsf{E}X_1 = \varphi'(1) = \frac{1}{p}, \quad \mathsf{E}X_1^2 = \varphi''(1) + \varphi'(1) = \frac{1+q}{p^2} \quad \text{и} \quad \mathrm{Var}\,X_1 = \frac{q}{p^2}.$$

Значит, $\mathsf{E}X_n = n/p$, $\mathrm{Var}\,X_n = nq/p^2$. □

Задача 1.5.4. Каждая из случайных величин U и V принимает значение ± 1. Их совместное распределение имеет вид

$$\mathsf{P}(U = 1) = \mathsf{P}(U = -1) = \frac{1}{2},$$
$$\mathsf{P}(V = 1 \mid U = 1) = \mathsf{P}(V = -1 \mid U = -1) = \frac{1}{3},$$
$$\mathsf{P}(V = -1 \mid U = 1) = \mathsf{P}(V = 1 \mid U = -1) = \frac{2}{3}.$$

а) Найдите вероятность того, что уравнение $x^2 + Ux + V = 0$ имеет по крайней мере один действительный корень.

б) Найдите среднее значение большего корня уравнения $x^2 + Ux + V = 0$ при условии, что это уравнение имеет по крайней мере один действительный корень.

в) Найдите вероятность того, что уравнение $x^2 + (U + V)x + U + V = 0$ имеет по крайней мере один действительный корень.

Решение. Запишем

$$\mathsf{P}(U = 1,\ V = 1) = \frac{1}{6}, \quad \mathsf{P}(U = -1,\ V = 1) = \frac{1}{3},$$
$$\mathsf{P}(U = 1,\ V = -1) = \frac{1}{3}, \quad \mathsf{P}(U = -1,\ V = -1) = \frac{1}{6}.$$

а) Уравнение $x^2 + Ux + V$ имеет действительный корень тогда и только тогда, когда $U^2 - 4V \geqslant 0$, что означает, что $V = -1$. Ясно, что если $V = -1$, то $U^2 - 4V = 5$. Поэтому вероятность того, что существует действительный корень, равна $1/2$.

б) Среднее значение большего корня равно

$$\frac{(-1+\sqrt{5})}{2}\mathsf{P}(U = 1 \mid V = -1) + \frac{(1+\sqrt{5})}{2}\mathsf{P}(U = -1 \mid V = -1) =$$
$$= \frac{1}{\mathsf{P}(V = -1)}\left(\frac{(-1+\sqrt{5})}{2}\frac{1}{3} + \frac{(1+\sqrt{5})}{2}\frac{1}{6}\right) = \frac{\sqrt{5}}{2} - \frac{1}{6}.$$

в) Уравнение $x^2 + Wx + W$ имеет действительный корень, если $W^2 - 4W \geqslant 0$. Если $W = U + V$, то W принимает значения 2, 0, −2, и уравнение имеет действительный корень, если $W = 0$ или -2. Тогда $\mathrm{P}(W = 0) = \frac{2}{3}$ и $\mathrm{P}(W = 0 \text{ или } -2) = \frac{5}{6}$. □

Задача 1.5.5. Читатель может брать из публичной библиотеки не более m книг за один раз, и вероятность того, что он взял k книг, равна $C_m^k \frac{1}{2^m}$, $k = 0, 1, \ldots, m$. Когда читатель приходит в библиотеку, он возвращает (или не возвращает) каждую книгу независимо с вероятностью 1/2. При заданном количестве невозвращенных книг r читатель берет n новых книг с вероятностью

$$C_{m-r}^n \left(\frac{1}{3}\right)^n \left(\frac{2}{3}\right)^{m-r-n}, \quad n = 0, 1, \ldots, m - r.$$

Докажите, что количество книг у читателя после посещения библиотеки имеет то же самое распределение, что и до посещения.

Число читателей, посещающих библиотеку в течение дня, имеет пуассоновское распределение со средним λ, и читатели ведут себя независимо друг от друга. Найдите среднее значение и дисперсию общего числа возвращенных и взятых книг в течение дня.

Решение. 1. Количество книг, которые взял один читатель до момента визита, равно $\mathrm{Bin}(m, 1/2)$ и может быть представлено как

$$H = \sum_{1 \leqslant i \leqslant m} X_i, \quad \text{где } X_i = \begin{cases} 0 & \text{с вероятностью } \frac{1}{2}, \\ 1 & \text{с вероятностью } \frac{1}{2}. \end{cases}$$

Тогда количество возвращенных книг равно

$$T = \sum_{1 \leqslant i \leqslant m} X_i R_i, \quad \text{где } R_i = \begin{cases} 0 & \text{с вероятностью } \frac{1}{2}, \\ 1 & \text{с вероятностью } \frac{1}{2}. \end{cases}$$

Значит, $T \sim \mathrm{Bin}(m, 1/4)$. Количество оставшихся книг равно

$$Q = \sum_{1 \leqslant i \leqslant m} X_i(1 - R_i) = H - T;$$

поэтому ясно, что $Q \sim T$. Наконец, количество книг, взятых во время данного посещения, равно

$$B = \sum_{1 \leqslant i \leqslant m} (1 - X_i(1 - R_i))Z_i,$$

и общее количество книг у читателя после визита равно

$$S = \sum_{1 \leqslant i \leqslant m} (X_i(1-R_i) + (1 - X_i(1-R_i))Z_i),$$

где случайные величины $Z_1, \ldots, Z_m$ независимы и

$$Z_i = \begin{cases} 0 & \text{с вероятностью } \frac{2}{3}, \\ 1 & \text{с вероятностью } \frac{1}{3} \end{cases}$$

независимо друг от друга. Обозначив $Y_i = X_i(1-R_i) + (1 - X_i(1-R_i))Z_i$, легко увидеть, что

$$Y_i = \begin{cases} 0 & \text{с вероятностью } \frac{1}{2}, \\ 1 & \text{с вероятностью } \frac{1}{2} \end{cases}$$

также независимо друг от друга. Таким образом, $S \sim H$.

2. Мы имеем $M_T(t) = \mathsf{E}e^{tT} = \left(\frac{1}{4}e^t + \frac{3}{4}\right)^m$. Общее количество возвращений имеет производящую функцию моментов $\boldsymbol{M}(t)$, равную

$$\sum_{n \geqslant 0} \frac{\lambda^n}{n!} e^{-\lambda} (\mathsf{E}e^{tT})^n = \exp(\lambda(\mathsf{E}e^{tT} - 1)) = \exp\left(\lambda\left(\left(\frac{1}{4}e^t + \frac{3}{4}\right)^m - 1\right)\right).$$

Таким образом,

$$\mathsf{E}(\text{общего числа возвращенных книг}) = \frac{d}{dt}\boldsymbol{M}(t)\Big|_{t=0} = \lambda\,\mathsf{E}T = \frac{\lambda m}{4},$$

$$\mathrm{Var}(\text{общего числа возвращенных книг}) = \frac{d^2}{dt^2}\boldsymbol{M}(t)\Big|_{t=0} - (\lambda\,\mathsf{E}T)^2 =$$

$$= \lambda\,\mathsf{E}T^2 + \lambda^2(\mathsf{E}T)^2 - (\lambda\,\mathsf{E}T)^2 = \lambda\,\mathsf{E}T^2 =$$

$$= \lambda\frac{d^2}{dt^2}\left(\frac{1}{4}e^t + \frac{3}{4}\right)^m\Big|_{t=0} = \lambda\frac{m}{4} + \lambda\frac{m(m-1)}{16} = \lambda\frac{m}{4}\left(1 + \frac{m-1}{4}\right).$$

3. Для общего числа взятых книг ответ такой же. □

Задача 1.5.6. В последовательности испытаний Бернулли X означает число испытаний до (включительно) a-го успеха. Покажите, что

$$\mathsf{P}(X = r) = C_{r-1}^{a-1} p^a q^{r-a}, \qquad r = a, a+1, \ldots$$

Проверьте, что производящая функция этого распределения равна $p^a s^a (1 - qs)^{-a}$. Покажите, что $\mathsf{E}X = a/p$ и $\mathrm{Var}\,X = aq/p^2$. Покажите, как случайная величина X может быть представлена в виде суммы a независимых одинаково распределенных случайных величин. Используйте это представление для нахождения среднего и дисперсии случайной величины X.

Решение. Мы имеем

$$\mathsf{P}(X=r)=\mathsf{P}(\{a-1 \text{ успех при } r-1 \text{ испытаниях}\}\cap$$
$$\cap\{\text{успех при } r\text{-м испытании}\})=C_{r-1}^{a-1}p^{a-1}q^{r-1-a+1}p=$$
$$=C_{r-1}^{a-1}p^a q^{r-a}, \quad r\geqslant a.$$

Производящая функция этого распределения равна

$$\varphi(s)=p^a s^a\sum_{r\geqslant a}C_{r-1}^{a-1}(qs)^{r-a}=p^a s^a\sum_{k\geqslant 0}C_{k+a-1}^{a-1}(qs)^k.$$

Заметим, что C_{k+a-1}^{a-1} совпадает с числом способов, которыми можно представить k в виде суммы a неотрицательных целых чисел:

$$k=k_1+\ldots+k_a.$$

Этот факт может быть подан геометрически следующим образом. Возьмем k «звездочек» и разделим их $a-1$ «черточками», как на диаграмме, приведенной ниже:

$$|*\ldots*|*|*|||*\ldots*|*.$$

Число звездочек между $(j-1)$-й и j-й черточками равно k_j; k_1 равно числу звездочек до первой черточки, и k_a — число звездочек после $(a-1)$-й черточки. Но число различных диаграмм равно C_{k+a-1}^{a-1}.

Используя разложение

$$\left(\sum_k b_k\right)^a=\sum_k\sum_{\substack{k_1,\ldots,k_a:\\ k_1+\ldots+k_a=k}}\prod_j b_{k_j},$$

мы получим

$$\sum_{k\geqslant 0}C_{k+a-1}^{a-1}(qs)^k=\left(\sum_{k\geqslant 0}(qs)^k\right)^a=(1-qs)^{-a}$$

и $\varphi(s)=p^a s^a(1-qs)^{-a}$. Далее,

$$\mathsf{E}X=(p^a s^a(1-qs)^{-a})'\big|_{s=1}=\frac{a}{p};$$

$$\mathsf{E}(X(X-1))=(p^a s^a(1-qs)^{-a})''\big|_{s=1}=a(a-1)+\frac{2a^2(1-p)}{p}+\frac{a(a+1)(1-p)^2}{p^2}.$$

Таким образом,

$$\operatorname{Var}X=\mathsf{E}X(X-1)+\mathsf{E}X-(\mathsf{E}X)^2=\varphi''(1)+\varphi'(1)-(\varphi'(1))^2=$$
$$=\frac{a(a-1)p^2+2a^2p(1-p)+a(a+1)(1-p)^2}{p^2}+\frac{a}{p}-\frac{a^2}{p^2}=\frac{aq}{p^2}.$$

Так как $\varphi(s) = (\psi(s))^a$, где $\psi(s) = ps(1 - qs)^{-1}$, мы приходим к заключению, что X может быть представлено суммой $\sum\limits_{j=1}^{a} Y_j$, где $Y_1, \ldots, Y_a$ — н. о. р. с. в. с производящей функцией $\psi(s)$. Действительно, Y_j равно числу испытаний между j-м и $(j + 1)$-м успехами, включая испытание при $(j + 1)$-м успехе. Таким образом,

$$\mathsf{E}X = a\,\mathsf{E}Y_j, \qquad \operatorname{Var} X = a \operatorname{Var} Y_j.$$

Так как $Y_j \sim \mathrm{Geom}(q)$, мы получаем $\mathsf{E}Y_j = 1/p$ и $\operatorname{Var}(Y_j) = q/p^2$. □

Задача 1.5.7. В моем супе случайное число N инородных частиц со средним μ и конечной дисперсией. С вероятностью p выбранная частица является мухой, иначе это паук; типы разных частиц независимы. Пусть F — количество мух и S — количество пауков.

а) Покажите, что производящая функция распределения случайной величины F равна $\varphi_F(s) = \varphi_N(ps + 1 - p)$.

б) Предположим, что случайная величина N имеет пуассоновское распределение с параметром μ. Покажите, что F имеет пуассоновское распределение с параметром $p\mu$, а случайные величины F и S независимы.

в) Пусть $p = 1/2$, и предположим, что F и S независимы. Покажите, что $\varphi_N(s) = \varphi_N\left(\frac{1}{2}(1 + s)\right)^2$. Используя уравнение на функцию $H(s) = \varphi_N(1 - s)$ (или другим путем), убедитесь, что случайная величина N имеет пуассоновское распределение.

Указание. Можно предположить, что $(1 + x/n + o(n^{-1}))^n \to e^x$ при $n \to \infty$. См. также доказательство уравнения (2.3.5).

Решение. а) По определению $\varphi_F(s) = \mathsf{E}s^F = \sum\limits_{n\geqslant 0} s^n P_F(n)$, $P_F(n) = \mathsf{P}(F = n)$, и аналогично для $\varphi_N(s)$. Тогда

$$\begin{aligned}\varphi_F(s) &= \sum_n s^n \sum_{m\geqslant n} P_N(m) C_m^n p^n (1-p)^{m-n} = \\ &= \sum_m P_N(m) \sum_{n=0}^{m} C_m^n (sp)^n (1-p)^{m-n} = \\ &= \sum_m P_N(m)(ps + (1-p))^m = \varphi_N(ps + (1-p)),\end{aligned}$$

или, короче,

$$\begin{aligned}\mathsf{E}s^F = \mathsf{E}[\mathsf{E}(s^F \mid N)] &= \sum_m P_N(m)\,\mathsf{E}(s^F \mid N = m) = \\ &= \sum_m P_N(m)(ps + (1 - p))^m = \varphi_N(ps + (1 - p)).\end{aligned}$$

б) Если $N \sim \mathrm{Po}(\mu)$, то $\varphi_N(s) = e^{\mu(s-1)}$. Тогда

$$\varphi_F(s) = e^{\mu(ps+1-p-1)} = e^{\mu p(s-1)},$$

т. е. $F \sim \mathrm{Po}(\mu p)$. Аналогично $\varphi_S(s) = e^{\mu(1-p)(s-1)}$, т. е. $F \sim \mathrm{Po}(\mu(1-p))$. Кроме того, $\varphi_N(s) = \varphi_F(s)\varphi_S(s)$. Далее,

$$\mathrm{P}(F = k,\ S = l) = \mathrm{P}(F = k \mid N = k + l)\,\mathrm{P}(N = k + l) =$$
$$= C_{k+l}^k p^k (1-p)^l \frac{\mu^{k+l} e^{-\mu}}{(k+l)!} = \frac{(p\mu)^k e^{-p\mu}}{k!} \frac{((1-p)\mu)^l e^{-\mu(1-p)}}{l!} = \mathrm{P}(X = k)\,\mathrm{P}(Y = l),$$

т. е. случайные величины F и S независимы.

Другой путь доказательства того, что $F \sim \mathrm{Po}(p\mu)$, заключается в том, чтобы прямо записать представление для $\mathrm{P}(F = k)$:

$$\sum_{n \geqslant k} \mathrm{P}(F = k \mid N = n)\,\mathrm{P}(N = n) = \sum_{n \geqslant k} \frac{n!}{k!\,(n-k)!} p^k (1-p)^{n-k} \frac{\mu^n e^{-\mu}}{n!} =$$
$$= \frac{(p\mu)^k e^{-p\mu}}{k!} \sum_{n-k \geqslant 0} \frac{e^{-\mu(1-p)}(\mu(1-p))^{n-k}}{(n-k)!} = \frac{(p\mu)^k e^{-p\mu}}{k!}.$$

в) Благодаря независимости, симметрии и функциональному уравнению $\varphi_F(s) = \varphi_N(ps + (1-p))$ получим

$$\varphi_N(s) = \varphi_F(s)\varphi_S(s) = \varphi_N\left(\frac{1}{2}s + 1 - \frac{1}{2}\right)^2.$$

Тогда функция $H(s) = \varphi_N(1-s)$ удовлетворяет равенствам

$$H(s) = H\left(\frac{s}{2}\right)^2 = \ldots = H\left(\frac{s}{2^n}\right)^{2^n}.$$

Таким образом, если z мало, разложение Тейлора дает

$$H(z) = \varphi_N(1-z) = \varphi_N(1) - \varphi_N'(1)z + o(z^2) = 1 - \mu z + O(z^2).$$

Поэтому при $n \to \infty$ мы получаем

$$H(s) = \left(1 - \frac{\mu s}{2^n} + o\left(\frac{s^2}{4^n}\right)\right)^{2^n} \to e^{-\mu s}. \qquad \square$$

Задача 1.5.8. а) Что такое пуассоновское распределение?

б) Предположим, что X и Y — независимые пуассоновские с. в. с параметрами λ и μ соответственно. Покажите, что $X + Y$ есть пуассоновская с. в. с параметром $\lambda + \mu$. Являются ли X и $X + Y$ независимыми? Какова условная вероятность $\mathrm{P}(X = r \mid X + Y = n)$?

Решение. а) Пуассоновское распределение приписывает вероятности

$$p_r = e^{-\lambda} \frac{\lambda^r}{r!}$$

неотрицательным целым числам $r = 0, 1, \ldots$ Здесь $\lambda > 0$ — параметр распределения.

б) Производящая функция пуассоновского распределения равна

$$\varphi_X(s) = \mathsf{E}s^X = \sum_{r=0}^{\infty} e^{-\lambda} s^r \frac{\lambda^r}{r!} = e^{\lambda(s-1)}.$$

Благодаря независимости

$$\varphi_{X+Y}(s) = \varphi_X(s)\varphi_Y(s) = e^{(\lambda+\mu)(s-1)},$$

и по теореме единственности $X + Y \sim \mathrm{Po}(\lambda + \mu)$. Аналогичный вывод следует из прямых вычислений:

$$\mathsf{P}(X + Y = r) = \sum_{i,j:\, i+j=r} \mathsf{P}(X = i,\, Y = j) = \sum_{i,j:\, i+j=r} \mathsf{P}(X = i)\,\mathsf{P}(Y = j) =$$

$$= \sum_{i,j:\, i+j=r} \frac{e^{-\lambda}\lambda^i}{i!} \frac{e^{-\mu}\mu^j}{j!} = \frac{e^{-(\lambda+\mu)}}{r!} \sum_{i,j:\, i+j=r} C_r^i \lambda^i \mu^j = \frac{e^{-(\lambda+\mu)}}{r!} (\lambda + \mu)^r.$$

Тем не менее, случайные величины X и $X + Y$ зависимы, так как $\mathsf{P}(X + Y = 2 \,|\, X = 4) = 0$. Действительно, $\forall r = 0, 1, \ldots, n$ выполняется равенство

$$\mathsf{P}(X = r \,|\, X + Y = n) = \frac{\mathsf{P}(X = r,\, Y = n - r)}{\mathsf{P}(X + Y = n)} = \frac{\mathsf{P}(X = r)\,\mathsf{P}(Y = n - r)}{\mathsf{P}(X + Y = n)} =$$

$$= \frac{e^{-\lambda}\lambda^r e^{-\mu}\mu^{n-r} n!}{e^{-(\lambda+\mu)}(\lambda + \mu)^n r!\,(n - r)!} = C_n^r \left(\frac{\lambda}{\lambda + \mu}\right)^r \left(\frac{\mu}{\lambda + \mu}\right)^{n-r},$$

т. е. $(X \,|\, X + Y = n) \sim \mathrm{Bin}\left(n, \dfrac{\lambda}{\lambda + \mu}\right)$. □

Задача 1.5.9. а) Пусть X — целочисленная положительная с. в. Обозначим через φ_X ее производящую функцию. Покажите, что если V_1 и V_2 — независимые положительные целочисленные с. в., то

$$\varphi_{V_1+V_2} = \varphi_{V_1}\varphi_{V_2}.$$

б) Нестандартная пара костей — это пара шестигранных симметричных костей, грани которых перенумерованы положительными целыми числами в произвольном порядке (например, (2, 2, 2, 3, 5, 7) и (1, 1, 5, 6, 7, 8)). Покажите, что существует такая нестандартная пара костей A и B, что при их подбрасывании

$$\mathsf{P}(\text{общая сумма на костях } A \text{ и } B \text{ равна } n) =$$
$$= \mathsf{P}(\text{общая сумма на обыкновенных костях равна } n)$$

для всех n, $2 \leqslant n \leqslant 12$.

Указание. Воспользуйтесь равенством $x + x^2 + x^3 + x^4 + x^5 + x^6 = x(x+1)(1+x^2+x^4) = x(1+x+x^2)(1+x^3)$.

Решение. а) Используем производящие функции: для с. в. V с конечным или счетным числом значений v имеем $\varphi_V(x) = \mathsf{E}x^V = \sum\limits_s x^v \mathsf{P}(V=v)$. Производящая функция однозначно определяет распределение: если $\varphi_U(x) \equiv \varphi_V(x)$, то $U \sim V$ в том смысле, что $\mathsf{P}(U=v) \equiv \mathsf{P}(V=v)$. Если $V = V_1 + V_2$, где случайные величины V_1 и V_2 независимы, то $\varphi_V(x) = \varphi_{V_1}(x)\varphi_{V_2}(x)$.

б) Пусть S_2 — общая сумма, выпавшая при подбрасывании пары стандартных костей, тогда

$$\varphi_{S_2}(x) = \varphi_S^2(x) = \frac{1}{36}\left(\sum_{k=1}^{6} x^k\right)^2 = \frac{1}{36}x(1+x)(1+x^2+x^4)x(1+x+x^2)(1+x^3),$$

где S — число, которое выпало при подбрасывании одной кости.

Если мы возьмем такие кости A и B, что производящая функция числа T_A, выпавшего на кости A, равна

$$\varphi_A(x) = \frac{1}{6}x(1+x^2+x^4)(1+x^3) = \frac{1}{6}(x+x^3+x^4+x^5+x^6+x^8)$$

и производящая функция числа T_B, выпавшего на кости B, равна

$$\varphi_B(x) = \frac{1}{6}x(1+x+x^2)(1+x) = \frac{1}{6}(x+2x^2+2x^3+x^4),$$

то производящая функция суммы T_{A+B} равна производящей функции суммы S_2. Таким образом, нужными костями будут кость A с гранями 1, 3, 4, 5, 6 и 8 и кость B с гранями 1, 2, 2, 3, 3 и 4. □

Задача 1.5.10. а) Вероятность выпадения орла при подбрасывании несимметричной монетки равна p, $0 < p < 1$. Покажите, что производящая функция числа орлов при n подбрасываниях равна

$$G(s) = (ps + 1 - p)^n.$$

б) Предположим, что монетку подбросили N раз, где N — случайная величина со средним μ_N и дисперсией σ_N^2, и пусть Y — число выпавших орлов. Покажите, что производящая функция $G_Y(s)$ с. в. Y равна

$$G_Y(s) = G_N(ps + 1 - p),$$

где $G_N(s)$ — производящая функция случайной величины N. Таким способом (или иначе) найдите $\mathsf{E}Y$ и $\operatorname{Var} Y$.

Предположим, что $\mathsf{P}(N=k) = \dfrac{e^{-\lambda}\lambda^k}{k!}$, $k = 0, 1, 2, \ldots$ Покажите, что случайная величина Y имеет пуассоновское распределение с параметром λp.

Решение. а) Обозначим через X число орлов после n подбрасываний. Тогда $X \sim \mathrm{Bin}(n, p)$:

$$\mathsf{P}(X = k) = C_n^k p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n,$$

и

$$G_X(s) = \mathsf{E}s^X = \sum_{k=0}^{n} C_n^k (sp)^k (1-p)^{n-k} = (ps + 1 - p)^n.$$

б) Далее,

$$G_Y(s) = \mathsf{E}s^Y = \mathsf{E}[\mathsf{E}(s^Y \mid N)] = \mathsf{E}(ps + 1 - p)^N = G_N(ps + 1 - p),$$

где $G_N(s) = \mathsf{E}s^N$. Тогда

$$G'_Y(s) = pG'_N(ps + 1 - p), \quad \text{поэтому} \quad G'_Y(1) = pG'_N(1), \quad \text{т. е.} \quad \mathsf{E}Y = p\mu_N.$$

Далее,

$$G''_Y(s) = p^2 G''_N(ps + 1 - p),$$

и

$$\mathrm{Var}\, Y = p^2 G''_N(1) + pG'_N(1) - p^2\mu_N^2 = p^2\sigma_N^2 + \mu_N p(1-p).$$

Таким образом, если $N \sim \mathrm{Po}(\lambda)$ и

$$G_N(s) = \sum_{k=0}^{\infty} s^k e^{-\lambda} \frac{\lambda^k}{k!} = \exp(\lambda(s-1)),$$

то

$$G_Y(s) = \exp(\lambda(ps + 1 - p - 1)) = \exp(\lambda p(s-1))$$

и $Y \sim \mathrm{Po}(\lambda p)$. □

Задача 1.5.11. а) Покажите, что если X и Y — независимые пуассоновские с. в. с параметрами λ и μ, то $X + Y$ — пуассоновская случайная величина с параметром $\lambda + \mu$.

б) Я получил от издателей верстку моего трактата о биномиальной теореме, который, на мой взгляд, будет иметь европейскую известность. К своему ужасу, я обнаружил, что количество опечаток, сделанное издателями на каждой странице, имеет пуассоновское распределение с параметром λ. При вычитывании книги я обнаруживаю каждую опечатку независимо от другой с вероятностью p. Покажите, что если я вычитаю книгу один раз, то количество оставшихся ошибок на каждой странице будет иметь пуассоновское распределение, и найдите параметр этого распределения.

В моей книге 256 страниц, количество опечаток на каждой странице не зависит от других страниц, среднее число опечаток на странице равно 2 и $p = 3/4$. Сколько раз мне следует вычитать книжку, чтобы вероятность того, что в книге не будет опечаток, была больше $1/2$?

Решение. б) Мы можем использовать производящие функции

$$\varphi_X(t) = \mathsf{E}t^X = \sum_{k\geqslant 0} \frac{e^{-\lambda}(t\lambda)^k}{k!} = e^{\lambda(t-1)}$$

и, аналогично, $\varphi_Y(t) = e^{\mu(t-1)}$, где $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t) = e^{(\lambda+\mu)(t-1)}$ благодаря независимости. Тогда в силу единственности $X + Y \sim \mathrm{Po}(\lambda + \mu)$.

Другой путь заключается в прямых вычислениях: по формуле свертки

$$\mathsf{P}(X+Y=r) = \sum_{k=0}^{r} \mathsf{P}(X=k)\,\mathsf{P}(Y=r-k) = \sum_{k=0}^{r} \frac{e^{-\lambda}\lambda^k}{k!}\,\frac{e^{-\mu}\mu^{r-k}}{(r-k)!} =$$

$$= \frac{e^{-\lambda-\mu}}{r!}\sum_{k=0}^{r}\frac{r!}{k!\,(r-k)!}\lambda^k\mu^{r-k} = \frac{e^{-\lambda-\mu}(\lambda+\mu)^r}{r!}.$$

Если X — число первоначальных и Z — число оставшихся опечаток, то $Z \sim \mathrm{Po}(p\lambda)$. Действительно, производящую функцию распределения $p_Z(t) = \mathsf{E}t^Z$ можно представить в виде $\mathsf{E}[\mathsf{E}(t^Z\,|\,X)]$, и она равна

$$\sum_{k\geqslant 0}\frac{e^{-\lambda}\lambda^k}{k!}\,\mathsf{E}(t^Z\,|\,X=k) = \sum_{k\geqslant 0}\frac{e^{-\lambda}\lambda^k}{k!}\sum_{r=0}^{k} t^r\frac{k!}{r!\,(k-r)!}p^r(1-p)^{k-r} =$$

$$= \sum_{k\geqslant 0}\frac{e^{-\lambda}\lambda^k}{k!}(tp+1-p)^k = e^{\lambda(-1+tp-p+1)} = e^{\lambda p(t-1)}.$$

Отсюда $Z \sim \mathrm{Po}(p\lambda)$. Снова вычисляем непосредственно:

$$\mathsf{P}(Z=r) = \sum_{k\geqslant r}\mathsf{P}(Z=r\,|\,X=k)\,\mathsf{P}(X=k) = \sum_{k\geqslant r}\frac{e^{-\lambda}\lambda^k}{k!}\,\frac{k!}{r!\,(k-r)!}p^r(1-p)^{k-r} =$$

$$= \frac{e^{-p\lambda}(p\lambda)^r}{r!}\sum_{k-r\geqslant 0}\frac{e^{-(1-p)\lambda}(\lambda(1-p))^{k-r}}{(k-r)!} = \frac{e^{-p\lambda}(p\lambda)^r}{r!}.$$

Наконец, после n прочтений получаем $Z_i \sim \mathrm{Po}(p^n\lambda)$, $1 \leqslant i \leqslant 256$, $p = 3/4$ и $\lambda = 2$. Тогда с. в. $R = \sum\limits_{i=1}^{256} Z_i$ распределена как $R \sim \mathrm{Po}\left(256\cdot 2\cdot\left(\frac{3}{4}\right)^n\right)$ и

$$\mathsf{P}(R=0) = e^{-512\cdot(3/4)^n} \geqslant \frac{1}{2}.$$

Поэтому должно выполняться неравенство $512\cdot\left(\frac{3}{4}\right)^n \leqslant \ln 2$, т. е. $\left(\frac{3}{4}\right)^n \leqslant \frac{\ln 2}{2^8}$, или

$$n \geqslant \frac{8\ln 2 - \ln\ln 2}{\ln(4/3)} \approx 20{,}54939 \quad \text{т. е.} \quad n \geqslant 21. \qquad \square$$

Задача 1.5.12. а) Дайте определение производящей функции распределения случайной величины. Найдите производящую функцию биномиальной случайной величины с параметрами n и p и используйте ее для вычисления среднего и дисперсии этой случайной величины.

б) Пусть X — биномиальная случайная величина с параметрами n и p, Y — биномиальная случайная величина с параметрами m и p, причем X и Y независимы. Найдите распределение случайной величины $X+Y$, т. е. определите вероятности $\mathsf{P}(X+Y=k)$ для всех возможных значений k.

Решение. а) Для $X \sim \mathrm{Bin}(n, p)$ производящая функция распределения имеет вид

$$G(z) = \sum_{x=0}^{n} C_n^x p^x (1-p)^{n-x} z^x = (pz + 1 - p)^n,$$

где

$$\mathsf{E}X = G'(1) = np, \qquad \operatorname{Var} X = G''(1) + np - (np)^2 = np(1-p).$$

б) Производящая функция распределения имеет вид $\mathsf{E}[z^{X+Y}] = \mathsf{E}[z^X] \times \times \mathsf{E}[z^Y] = (pz+1-p)^m (pz+1-p)^n = (pz+1-p)^{m+n}$. Поэтому $X+Y \sim \sim \mathrm{Bin}(m+n, p)$. □

Задача 1.5.13. а) Определите биномиальное и пуассоновское распределения и соотношение между ними.

б) За один розыгрыш лотереи продается в среднем 10^7 билетов. Чтобы выиграть первый приз, нужно угадать 7 номеров из 50. Предположим, что отдельные ставки независимы и случайны (это весьма нереалистичное предположение). Для заданного $n \geqslant 0$ запишите пуассоновскую аппроксимацию вероятности того, что в данном розыгрыше будет по меньшей мере n выигравших первый приз. Дайте приблизительную оценку этой вероятности для $n = 5$; 10. (Формулу Стирлинга можно использовать без доказательства.)

Решение. а) Равенство

$$\lim_{n\to\infty} C_n^k \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = e^{-\lambda}\frac{\lambda^k}{k!}$$

означает, что если $X_n \sim \mathrm{Bin}\left(n, \dfrac{\lambda}{n}\right)$, то $X_n \to Y \sim \mathrm{Po}(\lambda)$ при $n \to \infty$. Этот факт используется для аппроксимации различных биномиальных вероятностей.

б) В данном примере

$$n = 10^7, \qquad p = \frac{1}{C_{50}^7}, \qquad \lambda = \frac{10^7}{C_{50}^7}.$$

Далее,

$$C_{50}^7 \approx 10^8, \qquad \lambda \approx 0{,}1, \qquad e^{-\lambda} \approx 0{,}9.$$

Поэтому из равенства

$$\mathsf{P} = \mathsf{P}(\text{число выигравших} \geqslant n) = e^{-\lambda} \sum_{k=n}^{\infty} \frac{\lambda^k}{k!} \approx e^{-\lambda} \frac{\lambda^n}{n!}$$

следует, что

$$\text{если } n = 5, \text{ то } \mathsf{P} \approx 0{,}9 \cdot \frac{10^{-5}}{120} \approx 7{,}5 \cdot 10^{-8},$$
$$\text{если } n = 10, \text{ то } \mathsf{P} \approx 0{,}9 \cdot \frac{10^{-10}}{3 \cdot 10^6} \approx 3 \cdot 10^{-17}. \qquad \square$$

Дж. Стирлинг (1692—1770) был шотландским математиком и инженером. Помимо многочисленных математических достижений (формула Стирлинга — одно из них), он известен своим интересом к таким прикладным задачам, как форма Земли. В решении этих задач он достиг успехов и высоко ценился современниками. Его жизнь была нелегкой, поскольку он был активным якобитом, сторонником Джеймса, претендента на английский престол. Ему пришлось бежать за границу и провести несколько лет в Венеции, где он продолжал свою академическую работу.

Задача 1.5.14. а) Пусть $S_n = X_1 + \ldots + X_n$, $S_0 = 0$ и $X_1, \ldots, X_n$ — н. о. р. с. в.,

$$X_i = \begin{cases} 1 & \text{с вероятностью } p, \\ -1 & \text{с вероятностью } q, \end{cases}$$

$p \geqslant q$. Пусть $\tau_b = \min[n\colon S_n = b]$, $b > 0$, и $\tau_1 = \tau$. Докажите, что

$$\varphi(\lambda) = \mathsf{E}[e^{-\lambda\tau}] = \frac{1}{2q}(e^{\lambda} - \sqrt{e^{2\lambda} - 4pq})$$

и

$$\mathsf{E}\tau_b = b\frac{1}{2p-1}, \qquad \mathrm{Var}(\tau_b) = b\frac{4pq}{(2p-1)^3}.$$

б) Пусть $p < q$. Докажите, что

$$\mathsf{P}(\tau < \infty) = \frac{p}{q},$$
$$\mathsf{E}[e^{-\lambda\tau} I_{\tau<\infty}] = \frac{1}{2q}(e^{\lambda} - \sqrt{e^{2\lambda} - 4pq})$$

и

$$\mathsf{E}[\tau I_{\tau<\infty}] = \frac{p}{(1-p)(1-2p)}.$$

в) Вычислите $\mathrm{Var}(\tau I_{\tau<\infty})$.

Решение. а) Запишем $\tau_i = \sum_{j=1}^{i} T_j$, где $T_j \sim \tau_1$ и T_j независимы. Тогда

$$\mathsf{E}\tau_b = b\,\mathsf{E}\tau, \qquad \mathrm{Var}(\tau_b) = b\,\mathrm{Var}(\tau).$$

Далее, пусть $x = \varphi(\lambda)$. Тогда

$$x = pe^{-\lambda} + q\,\mathsf{E}[e^{-\lambda(1+\tau'+\tau'')}],$$

где τ' и τ'' независимы и распределены так же, как τ. Поэтому $x = pe^{-\lambda} + qe^{-\lambda}x^2$, или

$$x = \frac{1}{2q}(e^{\lambda} \pm \sqrt{e^{2\lambda} - 4pq}).$$

Знак $+$ нужно отбросить, так как $x \to 0$ при $\lambda \to \infty$. Более того,

$$\varphi'(\lambda) = \frac{1}{2(1-p)}\left(e^{\lambda} - \frac{e^{2\lambda}}{\sqrt{e^{2\lambda} - 4pq}}\right),$$

откуда следует, что

$$\varphi'(0) = -\frac{1}{2p-1}.$$

В качестве альтернативного способа решения можно получить уравнение $x = p + q(1 + 2x)$, из которого следует тот же результат.

Используя формулу $\mathrm{Var}(\tau) = \varphi''(0) - (\varphi'(0))^2$, получим

$$\mathrm{Var}(\tau) = \frac{1}{2(1-p)}\left(1 - \frac{2}{(2p-1)} + \frac{1}{2(2p-1)^3}\right) - \frac{1}{(2p-1)^2} = \\ = \frac{8p^3 - 20p^2 + 14p - 2 + 4p^2 - 6p + 2}{2(1-p)(2p-1)^3} = \frac{4pq}{(2p-1)^3}.$$

Рассуждая иным способом, получим для $y = \mathrm{Var}(\tau) = \mathsf{E}(\tau - \mathsf{E}\tau)^2$ следующее уравнение:

$$y = p(1 - \mathsf{E}\tau)^2 + (1-p)\,\mathsf{E}(1 + \tau_2 - \mathsf{E}\tau)^2,$$

где $\tau_2 = \tau' + \tau''$, τ' и τ'' независимы и распределены так же, как и τ. Отсюда следует, что

$$y = p(1 - \mathsf{E}\tau)^2 + (1-p)\,\mathsf{E}[1 + \mathsf{E}\tau + (\tau_2 - 2\,\mathsf{E}\tau)]^2 = \\ = p(1 - \mathsf{E}\tau)^2 + (1-p)[\mathrm{Var}(\tau_2) + (1 + \mathsf{E}\tau)^2],$$

поскольку $\mathsf{E}(\tau_2 - 2\,\mathsf{E}\tau) = 0$. Наконец, заметим, что $\mathrm{Var}(\tau_2) = 2\,\mathrm{Var}(\tau)$, и получим

$$y = p\left(1 - \frac{1}{2p-1}\right)^2 + (1-p)\left[2y + \left(\frac{1}{2p-1}\right)^2\right] \Rightarrow y = \frac{4pq}{(2p-1)^3}.$$

б) Вероятность $\mathsf{P}(\tau < \infty)$ является наименьшим решением уравнения $\pi = p + q\pi^2$. Уравнение для $\mathsf{E}[e^{-\lambda\tau}I_{\tau<\infty}]$ имеет тот же вид

$$x = pe^{-\lambda} + qe^{-\lambda}x^2,$$

откуда следует, что

$$\mathsf{E}[\tau I_{\tau<\infty}] = -\left[\frac{1}{2(1-p)}\left(1 - \frac{1}{\sqrt{(p-q)^2}}\right)\right] = \frac{p}{(1-p)(1-2p)}.$$

в) Дифференцируя дважды $\mathsf{E}[e^{-\lambda\tau} I_{\tau<\infty}]$ по λ, получим

$$\mathrm{Var}(\tau I_{\tau<\infty}) = \frac{1}{2(1-p)(1-2p)^3}[(1-2p)^3 - 2(1-2p)^2 + 1] - \\ - \left(\frac{p}{(1-p)(1-2p)}\right)^2 = \frac{p(1-4p^2+4p^3)}{(1-p)^2(1-2p)^3}. \qquad \square$$

§ 1.6. Неравенства Чебышёва и Маркова. Неравенство Йенсена. Закон больших чисел и теорема Муавра—Лапласа

Вероятностники это делают с помощью больших чисел.

(Из серии «Как они делают это».)

Неравенство Чебышёва — возможно, самый знаменитый результат из всей теории вероятностей и, наверное, самое выдающееся достижение великого русского математика П. Л. Чебышёва (1821—1894). Согласно этому неравенству для любой случайной величины X с конечными математическим ожиданием и дисперсией и для любого $\varepsilon > 0$ мы имеем

$$\mathsf{P}(|X - \mathsf{E}X| \geqslant \varepsilon) \leqslant \frac{1}{\varepsilon^2}\,\mathrm{Var}\,X. \qquad (1.6.1)$$

Неравенство Чебышёва допускает ряд обобщений. Одно из них — *неравенство Маркова*, названное по имени автора — ученика Чебышёва А. А. Маркова (1856—1922), другого выдающегося русского математика начала XX столетия. Неравенство Маркова утверждает, что для любой неотрицательной случайной величины Y с конечным математическим ожиданием и любого $\varepsilon > 0$ мы имеем

$$\mathsf{P}(Y \geqslant \varepsilon) \leqslant \frac{1}{\varepsilon}\,\mathsf{E}Y. \qquad (1.6.2)$$

Неравенство Чебышёва можно получить из неравенства Маркова, полагая $Y = |X - \mathsf{E}X|^2$ и учитывая, что события $\{|X - \mathsf{E}X| \geqslant \varepsilon\}$ и $\{|X - \mathsf{E}X|^2 \geqslant \varepsilon^2\}$ тождественные.

Имена Чебышёва и Маркова ассоциируются с подъемом русской (точнее, Санкт-Петербургской) школы теории вероятностей. Они оба были неординарными людьми. Чебышёв интересовался широким кругом проблем современной ему науки, а также политикой, экономикой и социальными вопросами. Кроме того, он изучал баллистику, помогая своему брату, также незаурядному человеку, генералу от артиллерии, служившему в русской царской армии.

Марков был известным либералом, оппозиционером к царскому режиму: в 1913 г., когда Россия праздновала 300-ю годовщину царствования семьи Романовых, он совместно со своими коллегами организовал празднование 200-й годовщины закона больших чисел, и это было весьма вызывающе.

Докажем теперь неравенство Маркова. Это несложно: если значениям $x_1, x_2, \ldots$ соответствуют вероятности $p_1, p_2, \ldots$, то

$$\mathsf{E}X = \sum_j x_j p_j \geqslant \sum_{j:\, x_j \geqslant \varepsilon} x_j p_j \geqslant \varepsilon \sum_{j:\, x_j \geqslant \varepsilon} p_j = \varepsilon\, \mathsf{P}(X \geqslant \varepsilon). \tag{1.6.3}$$

Это можно записать короче, используя индикаторы:

$$\mathsf{E}X \geqslant \mathsf{E}(X I_{X \geqslant \varepsilon}) \geqslant \varepsilon\, \mathsf{E} I_{X \geqslant \varepsilon} = \varepsilon\, \mathsf{P}(X \geqslant \varepsilon). \tag{1.6.4}$$

Здесь мы рассуждаем таким образом: сперва заметим, что имеет место неравенство $X \geqslant X I_{X \geqslant \varepsilon}$, так как $X \geqslant 0$ (и, конечно, $1 \geqslant I_{X \geqslant \varepsilon}$). Значит, $\mathsf{E}X \geqslant \mathsf{E}(X I_{X \geqslant \varepsilon})$. Аналогично из неравенства $X I_{X \geqslant \varepsilon} \geqslant \varepsilon I_{X \geqslant \varepsilon}$ следует, что $\mathsf{E}(X I_{X \geqslant \varepsilon}) \geqslant \mathsf{E}\varepsilon I_{X \geqslant \varepsilon}$. Последнее выражение равно $\varepsilon\, \mathsf{E} I_{X \geqslant \varepsilon}$, что, в свою очередь, равно $\varepsilon\, \mathsf{P}(X \geqslant \varepsilon)$.

Заметим, что число $\varepsilon > 0$ в неравенствах Чебышёва и Маркова не обязательно должно быть большим или малым, неравенство верно для любого положительного числа.

Пусть теперь $g\colon \mathbb{R} \to \mathbb{R}$ — монотонно неубывающая функция, X — действительная случайная величина, причем $\mathsf{E}g(X)$ конечно. Тогда для любого такого $x \in \mathbb{R}$, что $g(x) > 0$, мы имеем

$$\mathsf{P}(X \geqslant x) \leqslant \frac{1}{g(x)} \mathsf{E}g(X); \tag{1.6.5}$$

часто рассматривается случай, когда $g(x) = e^{ax}$, $a > 0$:

$$\mathsf{P}(X \geqslant x) \leqslant \frac{1}{e^{ax}} \mathsf{E}e^{aX}$$

(это *неравенство Чернова*).

Область применения этих неравенств очень обширна и не ограничивается теорией вероятностей. Мы обсудим одно из применений, а именно закон больших чисел (ЗБЧ).

Другим примером выдающегося неравенства, используемого во многих областях математики, является *неравенство Йенсена*. Оно названо в честь Дж. Л. Йенсена (1859—1925), датского математика, который первым применил это неравенство в статье, опубликованной в 1906 г. На самом деле это неравенство было получено в 1889 г. немецким математиком О. Гёльдером, но по некоторым причинам не было названо его именем (возможно, потому что уже существовало неравенство Гёльдера, чрезвычайно важное в математическом анализе и теории дифференциальных

4 — 1104

уравнений). Пусть X — случайная величина со значениями в интервале, полуинтервале или отрезке $J \subseteq \mathbb{R}$ (возможно, неограниченном, т. е. совпадающем со всей осью или с какой-либо полуосью). Пусть математическое ожидание $\mathsf{E}X$ конечно, а $g\colon J \to \mathbb{R}$ — выпуклая вниз (вверх) действительная функция с конечным математическим ожиданием $\mathsf{E}g(X)$. В неравенстве Йенсена утверждается, что

$$\mathsf{E}g(X) \geqslant g(\mathsf{E}X) \quad (\text{соответственно } \mathsf{E}g(X) \leqslant g(\mathsf{E}X)). \tag{1.6.6}$$

При этом функция g, заданная на J, называется *выпуклой вниз* (*вверх*), если для любых точек $x_1, \ldots, x_n \in (a, b)$ и вероятностей $p_1, \ldots, p_n$ ($p_1, \ldots$ $\ldots, p_n \geqslant 0$ и $p_1 + \ldots + p_n = 1$) выполняется неравенство

$$g\Big(\sum_j p_j x_j\Big) \leqslant \sum_j p_j g(x_j) \quad (\text{соответственно } g\Big(\sum_j p_j x_j\Big) \geqslant \sum_j p_j g(x_j)) \tag{1.6.7}$$

(см. рис. 1.6).

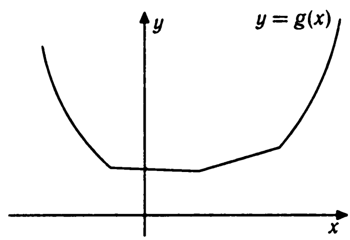


Рис. 1.6

При наличии этого определения неравенство Йенсена для случайной величины с конечным числом значений тривиально, поскольку неравенство (1.6.6) тогда не более чем краткая форма неравенства (1.6.7). Если X принимает бесконечное число значений, то для доказательства неравенства Йенсена нужны дополнительные аргументы, которые здесь не приводятся (например, нужно использовать тот факт, что выпуклая вниз (вверх) функция g непрерывна на том интервале (полуинтервале, отрезке) J, где она определена; она может не быть дифференцируемой, но не более чем в счетном числе точек $x \in J$, и в каждой точке $x \in J$, где она дважды дифференцируема, $g''(x) \leqslant 0$ ($\geqslant 0$)).

Если использовать более слабое определение выпуклости (вогнутости):

$$g\Big(\frac{1}{2}x_1 + \frac{1}{2}x_2\Big) \leqslant(\geqslant)\ \frac{1}{2}g(x_1) + \frac{1}{2}g(x_2) \quad \forall x_1, x_2 \in J,$$

то еще нужно проверить, что отсюда следует неравенство (1.6.6); это упражнение для студентов I курса, изучающих математический анализ (см. задачу 1.6.10).

Неравенство Йенсена можно усилить, охарактеризовав те случаи, когда имеет место равенство. Назовем выпуклую вниз (вверх) функцию g строго выпуклой вниз (вверх), если равенство в формуле (1.6.7) достигается тогда и только тогда, когда либо $x_1 = \ldots = x_n$, либо все вероятности p_j, кроме одной, равны нулю (а эта оставшаяся равна 1). Для такой функции g равенство в формуле (1.6.6) достигается тогда и только тогда, когда X — постоянная случайная величина.

Немедленным следствием неравенства Йенсена при $g(x) = x^s$, $x \in \in [0, \infty)$, является то, что для любого $s \geqslant 1$ выполняется неравенство $(\mathsf{E}X)^s \leqslant \mathsf{E}X^s$ для любой случайной величины $X \geqslant 0$ с конечным математическим ожиданием $\mathsf{E}X^s$. Для $s < 1$ в неравенстве меняется знак: $(\mathsf{E}X)^s \geqslant \mathsf{E}X^s$. Другим следствием при $g(x) = \ln x$, $x \in (0, \infty)$, является неравенство $\sum\limits_j x_j p_j \geqslant \prod\limits_j x_j^{p_j}$ для любых положительных $x_1, \ldots, x_n$ и вероятностей $p_1, \ldots, p_n$. При $p_1 = \ldots = p_n$, получаем знаменитое неравенство между средним арифметическим и средним геометрическим

$$\frac{1}{n}(x_1 + \ldots + x_n) \geqslant (x_1 \cdot \ldots \cdot x_n)^{1/n}. \tag{1.6.8}$$

Обратимся теперь к *закону больших чисел* (ЗБЧ). Его *слабая* форма звучит и доказывается очень просто. Пусть $X_1, \ldots, X_n$ — н. о. р. с. в. с конечным средним значением $\mathsf{E}X_j = \mu$ и дисперсией $\operatorname{Var} X = \sigma^2$. Положим

$$S_n = X_1 + \ldots + X_n. \tag{1.6.9}$$

Тогда для любого $\varepsilon > 0$ мы имеем

$$\mathsf{P}\left(\left|\frac{1}{n}S_n - \mu\right| \geqslant \varepsilon\right) \to 0 \quad \text{при} \quad n \to \infty. \tag{1.6.10}$$

Иными словами, усредненная сумма S_n/n н. о. р. с. в. $X_1, \ldots, X_n$ со средним $\mathsf{E}X_j = \mu$ и $\operatorname{Var} X_j = \sigma^2$ сходится по вероятности к μ.

При доказательстве используется неравенство Чебышёва

$$\mathsf{P}\left(\left|\frac{1}{n}S_n - \mu\right| \geqslant \varepsilon\right) = \mathsf{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)\right| \geqslant \varepsilon\right) \leqslant$$
$$\leqslant \frac{1}{\varepsilon^2}\frac{1}{n^2}\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{\varepsilon^2 n^2}\sum_{i=1}^{n}\operatorname{Var} X_i = \frac{1}{\varepsilon^2 n^2} n\sigma^2 = \frac{\sigma^2}{n\varepsilon^2},$$

а это выражение стремится к нулю при $n \to \infty$.

Нелишне отметить, что доказательство проходит, если просто предположить, что $\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) = o(n^2)$. Для н. о. р. с. в.

$$X_i = \begin{cases} 1 & \text{с вероятностью } p, \\ 0 & \text{с вероятностью } 1-p \end{cases}$$

(орлы и решки, возникающие в результате бросания несимметричной монеты) $\mathsf{E}X_i = p$ и ЗБЧ говорит, что после большого числа испытаний доля орлов будет близка к p, т. е. к вероятности выпадения орла при одном бросании. В таком контексте ЗБЧ был известен математикам XVII и XVIII столетий, например Якобу Бернулли (1654—1705). Он был одним из представителей выдающейся династии Бернулли фламандского происхождения, давшей миру математиков и естествоиспытателей. Несколько поколений этой династии проживало в Пруссии, России, Швейцарии и других странах. Они во многом определили развитие науки на европейском континенте.

На самом деле предположение о том, что дисперсия σ^2 конечна, не является необходимым для ЗБЧ. В связи с ЗБЧ возникает несколько видов сходимости; некоторые из них будут обсуждаться в последующих параграфах. Сейчас мы упомянем только *усиленную* форму ЗБЧ: для н. о. р. с. в. $X_1, X_2, \ldots$ с конечным средним μ выполняется равенство

$$\mathsf{P}\left(\lim_{n\to\infty} \frac{S_n}{n} = \mu\right) = 1.$$

Следующий шаг в изучении сумм вида (1.6.9) — *центральная предельная теорема* (ЦПТ). Рассмотрим н. о. р. с. в. $X_1, X_2, \ldots$ с конечным средним $\mathsf{E}X_j = a$, дисперсией $\operatorname{Var} X_j = \sigma^2$ и конечным моментом высшего порядка $\mathsf{E}|X_j - \mathsf{E}X_j|^{2+\delta} = \rho$ для некоторого фиксированного $\delta > 0$. ЦПТ в *интегральной* форме утверждает, что имеет место следующая сходимость: $\forall x \in \mathbb{R}$

$$\lim_{n\to\infty} \mathsf{P}\left(\frac{S_n - na}{\sigma\sqrt{n}} < x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2}\, dy. \tag{1.6.11}$$

Отображение

$$\Phi\colon x \in \mathbb{R} \mapsto \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2}\, dy \tag{1.6.12}$$

определяет так называемую стандартную нормальную, или N(0, 1), кумулятивную функцию распределения $\Phi(x)$, объект первостепенной важности в теории вероятностей и статистике. Ее называют также гауссовской функцией распределения, по имени К.-Ф. Гаусса (1777—1855), великого немецкого математика, астронома и физика, существенно повлиявшего на ряд областей математики. Он получил это распределение, изучая теорию ошибок астрономических наблюдений. Гауссовское распределение гораздо лучше описывает ошибки наблюдений, чем двойное экспоненциальное распределение, ранее использовавшееся Лапласом.

Значения $\Phi(x)$ (или $\overline{\Phi}(x) = 1 - \Phi(x)$) вычислены с большой степенью точности для значений x с очень узким интервалом между ними. Они содержатся в вероятностных и статистических таблицах в любом справочнике.

x	Значения $1 - \Phi(x)$	
	+0,00	+0,05
0,0	0,5000	0,4801
0,1	0,4602	0,4404
0,2	0,4207	0,4013
0,3	0,3821	0,3632
0,4	0,3446	0,3264
0,5	0,3085	0,2912
0,6	0,2743	0,2578
0,7	0,2420	0,2266
0,8	0,2119	0,1977
0,9	0,1841	0,1711
1,0	0,1587	0,1469
1,1	0,1357	0,1251
1,2	0,1151	0,1056
1,3	0,0968	0,0885
1,4	0,0808	0,0735

x	Значения $1 - \Phi(x)$	
	+0,00	+0,05
1,5	0,0668	0,0606
1,6	0,0548	0,0495
1,7	0,0446	0,0401
1,8	0,0359	0,0322
1,9	0,0288	0,0256
2,0	0,0228	0,0202
2,1	0,0179	0,0158
2,2	0,0129	0,0122
2,3	0,0107	0,0094
2,4	0,0082	0,0071
2,5	0,0062	0,0054
2,6	0,0047	0,0040
2,7	0,0035	0,0030
2,8	0,0026	0,0022
2,9	0,0019	0,0016

Данная таблица содержит значения $1 - \Phi(x)$ для $0 \leqslant x < 3$ с шагом 0,05. На данном этапе полезно отметить четыре факта.

1. Справедливы равенства

$$\lim_{x \to -\infty} \Phi(x) = 0, \qquad \lim_{x \to \infty} \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2}\, dy = 1$$

(это свойства всех функций распределения).

2. Справедливо равенство $\Phi(x) = 1 - \Phi(-x)$ $\forall x \in \mathbb{R}$, откуда следует, что

$$\Phi(0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{-y^2/2}\, dy = \frac{1}{2}$$

(это означает, что медиана стандартного гауссовского распределения равна 0). Это свойство следует из свойства 1 и того факта, что подынтегральное выражение $e^{-y^2/2}$ — четная функция.

3. Справедливо равенство

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y e^{-y^2/2}\, dy = 0$$

(это означает, что среднее значение стандартного гауссовского распределения равно нулю). Это свойство следует из того, что $ye^{-y^2/2}$ — нечетная функция.

4. Справедливо равенство

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 e^{-y^2/2}\, dy = 1$$

(это означает, что дисперсия стандартного гауссовского распределения равна 1). Это свойство можно получить интегрированием по частям; все понятия и факты, упомянутые, но не объясненные ранее, будут прокомментированы далее.

Доказательство утверждения 1, состоящего в том, что $\Phi(\infty) = 1$, т. е.

$$\int_{\mathbb{R}} e^{-x^2/2}\, dx = \sqrt{2\pi},$$

достаточно элегантно. Если $\int_{\mathbb{R}} e^{-x^2/2}\, dx = G$, то

$$G^2 = \int_{\mathbb{R}} e^{-x^2/2}\, dx \int_{\mathbb{R}} e^{-y^2/2}\, dy = \int_{\mathbb{R}^2} e^{-(x^2+y^2)/2}\, dy\, dx,$$

что в полярных координатах равно

$$\int_0^{\infty} re^{-r^2/2} \int_0^{2\pi} d\theta\, dr = 2\pi \int_0^{\infty} re^{-r^2/2}\, dr = 2\pi \int_0^{\infty} e^{-u}\, du = 2\pi.$$

Таким образом, $G = \sqrt{2\pi}$. Более того, свойство 4 можно получить так:

$$\int_{\mathbb{R}} y^2 e^{-y^2/2}\, dy = (-ye^{-y^2/2})\Big|_{-\infty}^{\infty} + \int_{\mathbb{R}} e^{-y^2/2}\, dy = \sqrt{2\pi}.$$

Имеет место также ЦПТ в *локальной* форме, и эта теорема содержит подробный анализ вероятностей $\mathsf{P}\left(\frac{S_n - na}{\sigma\sqrt{n}} = x_n\right) = \mathsf{P}(S_n = na + x_n\sigma\sqrt{n})$ для определенных значений $x_n \in \mathbb{R}$. Цель такого анализа — получить асимптотику этих вероятностей в виде $\frac{1}{\sqrt{2\pi n}\sigma} \exp\left(-\frac{1}{2}x_n^2\right)$, т. е. выразить искомые вероятности только через среднее значение a и дисперсию σ^2.

Современный метод доказательства ЦПТ основан на том факте, что сходимость в формуле (1.6.11) эквивалентна сходимости характеристических функций $\psi_{(S_n-na)/(\sigma\sqrt{n})}(t)$ к характеристической функции гауссовского распределения. Мы обсудим этот метод позднее.

В оставшейся части данного параграфа сосредоточимся на ЦПТ для упомянутого ранее примера с бросанием монеты, где

$$X_i = \begin{cases} 1 & \text{с вероятностью } p, \\ 0 & \text{с вероятностью } 1-p. \end{cases}$$

ЦПТ в рамках этого примера была получена еще в XVIII столетии, и ее часто называют теоремой Муавра—Лапласа (тМЛ) в честь двух французских математиков.

Один из них, А. де Муавр (1667—1754), бежал в Англию после отмены Нантского эдикта и сделал блестящую исследовательскую и преподавательскую карьеру. Другой, П.-С. Лаплас (1749—1827), внес выдающийся вклад во многие области математики. Он служил также (недолго и не слишком успешно) министром внутренних дел при Наполеоне (Лаплас экзаменовал юного Наполеона по математике во французском королевском артиллерийском корпусе и дал перспективному офицеру наивысшую оценку). Несмотря на принадлежность к наполеоновскому окружению, Лаплас первым голосовал во французском сенате за отстранение Наполеона от власти в 1814 г. После реставрации Бурбонов его титул возвысился от графа до маркиза.

Первоначальный вариант теоремы был сформулирован в 1733 г. де Муавром. Он был первооткрывателем нормального распределения (оно было названо именем Гаусса сто лет спустя).

Говоря неформальным языком, теорема Муавра—Лапласа утверждает, что для последовательности таких н. о. р. с. в. $X_1, X_2, \ldots$, что $\mathrm{P}(X_j=1)=p$, $\mathrm{P}(X_j=0)=1-p$ и $\mathrm{E}X_j=p$, $\mathrm{Var}\,X_j=p(1-p)$, случайная величина $\dfrac{S_n-np}{\sqrt{np(1-p)}}$ приближенно имеет стандартное $\mathrm{N}(0,1)$-распределение. На уровне формального изложения вновь нужно различать локальную и интегральную теоремы Муавра—Лапласа. Для заданного положительного целого числа $m \leqslant n$ запишем

$$\mathrm{P}(S=m)=\mathrm{P}\left(\frac{S_n-np}{\sqrt{np(1-p)}}=z_n(m)\right), \quad \text{где } z_n(m)=\frac{m-np}{\sqrt{np(1-p)}}.$$

Как было сказано ранее, локальная теорема Муавра—Лапласа утверждает, что если $n, m \to \infty$, то

$$\mathrm{P}(S_n=m)\Big/\left(\frac{1}{\sqrt{2\pi np(1-p)}}\exp\left(-\frac{1}{2}z_n(m)^2\right)\right)\to 1, \tag{1.6.13}$$

если только $m-np=o(n^{2/3})$. Говоря точнее, формула имеет место равномерно по n, m, для которых выражение $(m-np)n^{-2/3}$ содержится в ограниченном интервале и стремится к нулю. Напомним, что $\mathsf{P}(S_n = m)$ — это вероятность того, что событие, вероятность которого равна p, появится m раз в n независимых испытаний.

Как и ранее, интегральная форма теоремы Муавра—Лапласа касается функции распределения и утверждает, что $\forall x \in \mathbb{R}$

$$\lim_{n\to\infty} \mathsf{P}\left(\frac{S_n - np}{\sqrt{np(1-p)}} < x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2}\, dy, \tag{1.6.14}$$

или, что эквивалентно, $\forall a,\ b,\ -\infty \leqslant a < b \leqslant \infty$, выполняется равенство

$$\lim_{n\to\infty} \mathsf{P}\left(a < \frac{S_n - np}{\sqrt{np(1-p)}} < b\right) = \frac{1}{\sqrt{2\pi}} \int_{a}^{b} e^{-x^2/2}\, dx. \tag{1.6.15}$$

Хотя интегральная форма теоремы Муавра—Лапласа выглядит более простой, ее доказательство длиннее доказательства локальной формы (и использует последнюю как свою часть). Ни одна из этих теорем не доказывается формально в кембриджском курсе «Вероятность IA», хотя они широко используются в последующих курсах (в частности, в курсе второго года «Статистика» и в следующем за ним курсе третьего года «Принципы статистики», а также в ряде магистерских курсов статистики). Доказательство, приведенное ниже, дано для полноты картины. Оно не используется в задачах из этого тома, но проливает свет на то, как нормальное распределение появляется в виде асимптотического распределения нормированных должным образом сумм независимых случайных величин.

Локальный вариант теоремы Муавра—Лапласа можно доказать прямыми рассуждениями. Основной шаг опирается на тот факт, что для любой такой последовательности $\{m_n\}$ натуральных чисел, что $m_n \leqslant n$, $m_n \to \infty$ и $n - m_n \to \infty$ при $n \to \infty$, выполняется соотношение

$$\mathsf{P}(S_n = m_n) \Big/ \left(\left(2\pi n \frac{m_n}{n}\left(1 - \frac{m_n}{n}\right)\right)^{-1/2} \exp\left(-n h_{m_n/n}(p)\right)\right) \to 1. \tag{1.6.16}$$

Здесь

$$h_{m_n/n}(p) = \frac{m_n}{n} \ln \frac{m_n}{np} + \frac{n - m_n}{n} \ln \frac{n - m_n}{n(1-p)}, \tag{1.6.17}$$

что является частным случаем более общего определения: для $p^* \in (0, 1)$ выполняется соотношение

$$h_{p^*}(p) = p^* \ln \frac{p^*}{p} + (1 - p^*) \ln \frac{1 - p^*}{1 - p}. \tag{1.6.18}$$

Сходимость (1.6.13) получаем теперь для $m_n - np = o(n^{2/3})$, так как для p^*, близких к p, из формулы Тейлора следует равенство

$$h_{p^*}(p) = \frac{1}{2}\left(\frac{1}{p} + \frac{1}{1-p}\right)(p^* - p)^2 + O(|p^* - p|^3) \tag{1.6.19}$$

(поскольку $h_{p^*}(p)\big|_{p^*=p} = \left(\frac{d}{dp^*}h_{p^*}(p)\right)\Big|_{p^*=p} = 0$).

Замечание. Соотношение (1.6.16), а также показатель $-nh_{p^*}(p)$ экспоненты при $p^* = \frac{m_n}{n}$ и $h_{p^*}(p)$, заданном равенством (1.6.19), важны для асимптотического анализа различных вероятностей, касающихся сумм независимых случайных величин. В частности, формулы (1.6.17)—(1.6.19) играют важную роль в теории информации и теории больших уклонений. Функция $h_{p^*}(p)$ называется относительной *энтропией* вероятностного распределения $\{p^*, 1 - p^*\}$ по отношению к распределению $\{p, 1 - p\}$.

Доказательство равенства (1.6.19) непосредственно вытекает из *формулы Стирлинга*

$$n! \approx \sqrt{2\pi n}\, n^n e^{-n} \tag{1.6.20}$$

(хорошо известно, что у этой формулы есть более точный вариант: $n! = \sqrt{2\pi n}\, n^n e^{-n+\theta(n)}$, где $\frac{1}{12n+1} < \theta(n) < \frac{1}{12n}$, но для наших целей достаточно и соотношения (1.6.20)). Тогда, опуская индекс n в символе m_n, получим, что вероятность $\mathsf{P}(S_n = m)$ равна

$$C_n^m p^m (1-p)^{n-m} \approx \left(\frac{n}{2\pi m(n-m)}\right)^{1/2} \frac{n^n}{m^m (n-m)^{n-m}} p^m (1-p)^{n-m} =$$
$$= \left(2\pi n \frac{m}{n}\left(1 - \frac{m}{n}\right)\right)^{-1/2} \times$$
$$\times \exp\left(-m \ln \frac{m}{n} - (n-m)\ln\frac{n-m}{n} + m \ln p + (n-m)\ln(1-p)\right).$$

При этом правая часть последнего равенства совпадает со знаменателем в формуле (1.6.16).

Чтобы получить интегральную форму теоремы Муавра—Лапласа, рассмотрим сумму

$$\sum_m \mathsf{P}(S_n = m) I(a < z_n(m) < b)$$

и интерпретируем ее как площадь фигуры, образованной осью Ox и графиком кусочно постоянной функции

$$\Pi_n : x \in \mathbb{R} \mapsto \sqrt{np(1-p)}\, \mathsf{P}(S_n = m) \quad \text{для } z_n(m) \leqslant x < z_n(m+1)$$

на отрезке $z_n(\underline{m}) \leqslant x \leqslant z_n(\overline{m})$, где $\underline{m}$ и $\overline{m}$ однозначно определяются условием

$$a \leqslant z_n(\underline{m}) < a + \frac{1}{\sqrt{np(1-p)}}, \quad b \leqslant z_n(\overline{m}) < b + \frac{1}{\sqrt{np(1-p)}}.$$

Очевидно, эта площадь равна интегралу

$$\int\limits_{z_n(\underline{m})}^{z_n(\overline{m})} \Pi_n(x)\,dx.$$

Два обстоятельства затрудняют вычисления: подынтегральное выражение Π_n и переменные пределы $z_n(\underline{m})$ и $z_n(\overline{m})$ зависят от n. Для начала рассмотрим случай, когда a и b — действительные числа. Тогда $z_n(\underline{m}) \to a$ и $z_n(\overline{m}) \to b$ и предыдущий интеграл отличается от

$$\int\limits_a^b \Pi_n(x)\,dx$$

на величину, стремящуюся к нулю. Далее, на интервале (a, b) можно применить локальную теорему Муавра—Лапласа, согласно которой

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

С помощью некоторых дополнительных соображений отсюда можно вывести, что

$$\int\limits_a^b \Pi_n(x)\,dx \to \int\limits_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx.$$

Для завершения доказательства рассмотрим случай, когда a и/или b бесконечны. При этом используется тот факт, что сходимость в локальной теореме Муавра—Лапласа равномерна, если $z_n(m_n)$ остается в некотором конечном интервале, а предельная функция $x \mapsto \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ монотонно убывает, когда $|x|$ растет, и является интегрируемой. Детали мы опускаем.

Один из важных аспектов центральной предельной теоремы состоит в том, что предельное распределение для всевозможных допредельных — нормальное.

Задача 1.6.1. Игральный кубик подбрасывают n раз, Y_n — общее число выпавших очков. Покажите, что $\mathsf{E} Y_n = \frac{7n}{2}$ и $\operatorname{Var} Y_n = \frac{35n}{12}$. Используя неравенство Чебышёва, найдите такое n, для которого

$$\mathsf{P}\left(\left|\frac{Y_n}{n} - 3{,}5\right| > 0{,}1\right) \leqslant 0{,}1.$$

Решение. Запишем $Y_n = \sum_{i=1}^{n} X_i$, где X_i — число очков, выпавших при i-м бросании. Случайные величины $X_1, \ldots, X_n$ независимы и одинаково распределены, причем $\mathsf{P}(X_i = r) = 1/6$, $r = 1, 2, \ldots, 6$. Поэтому

$$\mathsf{E}X_i = 3\frac{1}{2}, \quad \mathrm{Var}(X_i) = 2\frac{11}{12}, \quad \mathsf{E}Y_n = \frac{7n}{2}, \quad \mathrm{Var}(Y_n) = \frac{35n}{12}.$$

Неравенство Чебышёва можно записать в виде $\mathsf{P}(|X - \mathsf{E}X| \geqslant \varepsilon) \leqslant \leqslant \frac{1}{\varepsilon^2}\mathrm{Var}\,X$, и оно верно для любой случайной величины X с конечным средним и дисперсией.

В силу неравенства Чебышёва и независимости случайных величин X_i мы получаем

$$\mathsf{P}\left(\left|\frac{Y_n}{n} - 3{,}5\right| > 0{,}1\right) = \mathsf{P}(|Y_n - \mathsf{E}Y_n| > 0{,}1n) = \frac{\mathrm{Var}\,Y_n}{0{,}1^2 \cdot n^2} = \frac{35n/12}{0{,}01n^2} = \frac{1750}{6n}.$$

Нам нужно неравенство $1750/(6n) \leqslant 0{,}1$, откуда получаем $n \geqslant 2920$. Отметим, что если применить нормальное приближение (теорема Муавра—Лапласа), то оценка серьезно понижается: $n \geqslant \frac{35}{12}(1{,}96 \cdot 10)^2 \approx 794$. □

Задача 1.6.2. При подбрасывании монеты орел выпадает с вероятностью $p > 0$, а решка — с вероятностью $q = 1 - p$. Обозначим через X_n число бросаний, необходимое для выпадения n орлов. Найдите производящую функцию распределения случайной величины X_1, ее среднее и дисперсию. Чему равно среднее и дисперсия случайной величины X_n?

Пусть теперь $p = 1/2$. Как оценивается $\mathsf{P}(X_{100} > 400)$ с помощью неравенства Чебышёва?

Решение. Поскольку $\mathsf{P}(X_1 = k) = pq^{k-1}$, производящая функция моментов имеет вид $g(s) = \frac{ps}{1 - qs}$. Поэтому $g'(s) = \frac{p}{(1 - qs)^2}$ и $\mathsf{E}X_1 = g'(1) = 1/p$, $\mathrm{Var}(X_1) = q/p^2$. Значит, $\mathsf{E}X_n = n/p$, $\mathrm{Var}\,X = nq/p^2$. Теперь положим $p = q = 1/2$, $\mathsf{E}X_{100} = 200$, $\mathrm{Var}(X_{100}) = 200$ и запишем

$$\mathsf{P}(X_{100} \geqslant 400) \leqslant \mathsf{P}(|X_{100} - 200| > 200) \leqslant \frac{200}{200^2} = \frac{1}{200}.$$ □

Замечание. Несомненно, X_n имеет отрицательное биномиальное распределение с параметрами (n, q).

Задача 1.6.3. В этой задаче и ряде последующих используется термин «выборка» вместо «исход». Эта терминология чаще всего используется в статистике, см. часть B.

а) Случайную выборку какого размера с заданным распределением нужно образовать для того, чтобы с вероятностью не менее 0,99 выборочное среднее отличалось от среднего значения распределения не более чем на два стандартных отклонения? Используя неравенство Чебышёва, определите указанный размер выборки независимо от распределения.

б) Случайную выборку какого размера с нормальным распределением нужно образовать для того, чтобы с вероятностью не менее 0,99 выборочное среднее отличалось от среднего значения распределения не более чем на стандартное отклонение?

Указание. Используйте равенство $\Phi(2{,}58) = 0{,}995$.

Решение. а) Выборочное среднее $\overline{X}$ имеет среднее значение μ и дисперсию σ^2/n. Поэтому в силу неравенства Чебышёва

$$\mathsf{P}(|\overline{X} - \mu| \geqslant 2\sigma) \leqslant \frac{\sigma^2}{n(2\sigma)^2} = \frac{1}{4n}.$$

Таким образом, достаточна выборка объема $n = 25$. Чем больше известно о распределении X_i, тем меньший размер выборки достаточен; случай нормального распределения рассмотрен в п. б).

б) Если $X_i \sim \mathrm{N}(\mu, \sigma^2)$, то $\overline{X} - \mu \sim \mathrm{N}\left(0, \frac{\sigma^2}{n}\right)$ и

$$\mathsf{P}(|\overline{X} - \mu| \geqslant \sigma) = \mathsf{P}\left(|\overline{X} - \mu| \Big/ \left(\frac{\sigma^2}{n}\right)^{1/2} \geqslant \sigma \Big/ \left(\frac{\sigma^2}{n}\right)^{1/2}\right) = \mathsf{P}(|Z| \geqslant n^{1/2}),$$

где $Z \sim \mathrm{N}(0, 1)$. Но $\mathsf{P}(|Z| \geqslant 2{,}58) = 0{,}99$, поэтому нужно потребовать, чтобы выполнялось неравенство $n^{1/2} \geqslant 2{,}58$, т. е. $n \geqslant 7$. Как видим, зная, что распределение нормально, можно взять выборку гораздо меньшего размера даже для более узкого интервала. □

Задача 1.6.4. Что означают слова «функция $g\colon (0, \infty) \to \mathbb{R}$ выпукла вниз»? Пусть функция $f\colon (0, \infty) \to \mathbb{R}$ такова, что функция $-f$ выпукла вниз. Покажите, что

$$\int\limits_n^{n+1} f(x)\,dx \geqslant \frac{f(n) + f(n+1)}{2},$$

и выведите отсюда, что

$$\int\limits_1^N f(x)\,dx \geqslant \frac{1}{2}f(1) + f(2) + \ldots + f(N-1) + \frac{1}{2}f(N)$$

для всех целых чисел $N \geqslant 2$. Выбирая подходящую функцию f, докажите, что

$$N^{N+\frac{1}{2}} e^{-(N-1)} \geqslant N!$$

для всех целых чисел $N \geqslant 2$.

Решение. Функция $g\colon (0, \infty) \to \mathbb{R}$ выпукла вниз, если $\forall x, y \in (0, \infty)$ и $p \in (0, 1)$ выполняется неравенство $g(px + (1-p)y) \leqslant pg(x) + (1-p)g(y)$.

Если функция g дважды дифференцируема, ее выпуклость вниз следует из неравенства $g''(x) > 0$, $x > 0$.

Если функция $-f$ выпукла вниз, то

$$-f(pn + (1-p)(n+1)) \leqslant -pf(n) - (1-p)f(n+1).$$

Отсюда для $x = pn + (1-p)(n+1) = n + 1 - p$ получаем

$$f(x) \geqslant pf(n) + (1-p)f(n+1),$$

а тогда

$$\int\limits_n^{n+1} f(x)\,dx \geqslant \int\limits_n^{n+1} (pf(n) + (1-p)f(n+1))\,dx.$$

Поскольку $dx = -dp = d(1-p)$, мы получаем, что

$$\int\limits_n^{n+1} f(x)\,dx \geqslant \int\limits_0^1 ((1-p)f(n) + pf(n+1))\,dp = \frac{1}{2}f(n) + \frac{1}{2}f(n+1).$$

Суммируя эти неравенства от $n = 1$ до $n = N-1$, получим первое неравенство.

Выберем теперь $f(x) = \ln x$; тогда $f'(x) = 1/x$, $f''(x) = -1/x^2$. Значит, функция $-f$ выпукла вниз. Согласно предыдущему неравенству

$$\int\limits_1^N \ln x\,dx \geqslant \ln 2 + \ldots + \ln(N-1) + \ln(N^{1/2}),$$

т. е.

$$\frac{1}{2}\ln N + \int\limits_1^N \ln x\,dx = \frac{1}{2}\ln N + (x\ln x - x)\Big|_1^N = \left(N + \frac{1}{2}\right)\ln N - (N-1) \geqslant \ln(N!),$$

а это то же, что неравенство $N! \leqslant N^{N+1/2}e^{-(N-1)}$. □

Задача 1.6.5. Рассматривается случайная выборка с целью найти долю тех избирателей, кто голосовал за лейбористов. Найдите такой размер выборки, чтобы вероятность того, что ошибка измерения будет меньше 0,04, была не менее 0,99.

Ответ. Справедлива оценка $0{,}04\sqrt{n} \geqslant 2{,}58\sqrt{pq}$, где $pq \leqslant 1/4$. Таким образом, $n \approx 1040$. □

Задача 1.6.6. Рассмотрим индикаторную функцию I_A события A.

Пусть I_i — индикатор события A_i, $1 \leqslant i \leqslant n$, и пусть $N = \sum\limits_{i=1}^{n} I_i$ — число тех i, при которых произошло событие A_i. Покажите, что $\mathsf{E}(N) = \sum\limits_i p_i$, где

$p_i = \mathsf{P}(A_i)$, и выразите $\operatorname{Var} N$ через вероятности $p_{ij} = \mathsf{P}(A_i \cap A_j)$. Используя неравенство Чебышёва или иные соображения, покажите, что

$$\mathsf{P}(N=0) \leqslant \frac{\operatorname{Var} N}{(\mathsf{E}(N))^2}.$$

Решение. Индикаторная функция события A — это

$$I_A(\omega) = \begin{cases} 1, & \omega \in A, \\ 0, & \omega \in A^c, \end{cases}$$

причем $\mathsf{E} I_A = \mathsf{P}(A)$.

Поскольку $\mathsf{E} I_i = p_i$, можно записать, что $\mathsf{E}N = \sum\limits_{i=1}^{n} \mathsf{E} I_i = \sum\limits_{i=1}^{n} p_i$ и $(\mathsf{E}N)^2 = \left(\sum\limits_{i=1}^{n} p_i\right)^2$. Далее, $I_i^2 = I_i$, откуда следует, что $\mathsf{E} I_i^2 = p_i$. Кроме того, $\mathsf{E} I_i I_j = p_{ij}$. Значит,

$$\mathsf{E}(N^2) = \mathsf{E}\left(\sum_{i=1}^{n} I_i\right)^2 = \mathsf{E}\left(\sum_{1 \leqslant i \leqslant n} I_i + 2\,\mathsf{E} \sum_{1 \leqslant i < j \leqslant n} I_i I_j\right) = \sum_i p_i + \sum_{i \neq j} p_{ij}.$$

Тогда

$$\operatorname{Var} N = \mathsf{E}N^2 - (\mathsf{E}N)^2 = \sum_i p_i - \left(\sum_i p_i\right)^2 + \sum_{i \neq j} p_{ij}.$$

Наконец,

$$\mathsf{P}(N=0) \leqslant \mathsf{P}(|N - \mathsf{E}N| \geqslant \mathsf{E}N) \leqslant \frac{\operatorname{Var} N}{(\mathsf{E}(N))^2}$$

в силу неравенства Чебышёва. □

Замечание. Событие $\{N=0\}$ совпадает с $\bigcap\limits_i I_i^c = \Omega \setminus \left(\bigcup\limits_i I_i\right)$. По формуле включения-исключения получаем

$$\mathsf{P}(N=0) = 1 - \mathsf{P}\left(\bigcup_i I_i\right) = 1 - \sum_{1 \leqslant i \leq n} p_i + \sum_{1 \leqslant i_1 < i_2 \leq n} p_{i_1 i_2} - \\ - \sum_{1 \leqslant i_1 < i_2 < i_3 \leq n} p_{i_1 i_2 i_3} + \ldots + (-1)^n p_{1 \ldots n}.$$

Здесь для любого l, $1 \leq l \leqslant n$, и любых индексов $i_1, \ldots, i_l$, $1 \leqslant i_1 < i_2 < \ldots < i_l \leqslant n$, выполняется равенство $p_{i_1 \ldots i_l} = \mathsf{P}(A_{i_1} \cap \ldots \cap A_{i_l})$. Поэтому полученная выше оценка для $\mathsf{P}(N=0)$ дает также оценку для знакопеременной суммы, содержащей все $p_{i_1 \ldots i_l}$ (независимо от их значений), через p_i и p_{ij}:

$$1 - \sum_{1 \leqslant i \leq n} p_i + \sum_{1 \leqslant i_1 < i_2 \leq n} p_{i_1 i_2} - \sum_{1 \leqslant i_1 < i_2 < i_3 \leq n} p_{i_1 i_2 i_3} + \ldots + (-1)^n p_{1 \ldots n} \leqslant$$

$$\leqslant \left(\sum_i p_i\right)^{-2} \left(\sum_i p_i - \left(\sum_i p_i\right)^2 + \sum_{i \neq j} p_{ij}\right).$$

Задача 1.6.7. а) Сформулируйте и докажите неравенство Чебышёва.

Покажите, что для н. о. р. с. в. $X_1, X_2, \ldots$ с конечными средним μ и дисперсией σ^2 справедливо соотношение

$$\mathsf{P}\left(\left|n^{-1} \sum_{i=1}^{n} X_i - \mu\right| \geqslant \varepsilon\right) \to 0$$

при $n \to \infty$ для всех $\varepsilon > 0$.

б) Предположим, что $Y_1, Y_2, \ldots$ — н. о. р. с. в., причем $\mathsf{P}(Y_j = 4^r) = 2^{-r}$ для всех целых $r \geqslant 1$. Покажите, что P(хотя бы одна случайная величина $Y_1, Y_2, \ldots, Y_{2^n}$ приняла значение 4^n) $\to 1 - e^{-1}$ при $n \to \infty$, и выведите отсюда, что для всех K выполняется условие

$$\mathsf{P}\left(2^{-n} \sum_{i=1}^{2^n} Y_i > K\right) \not\to 0.$$

Решение. а) Неравенство Чебышёва имеет вид

$$\mathsf{P}(|X - \mathsf{E}X| \geqslant b) \leqslant \frac{1}{b^2} \operatorname{Var} X \quad \forall b > 0,$$

а доказательство его вытекает из следующих соображений:

$$\operatorname{Var} X = \mathsf{E}(X - \mathsf{E}X)^2 \geqslant \mathsf{E}((X - \mathsf{E}X)^2 I((X - \mathsf{E}X)^2 \geqslant b^2)) \geqslant$$

$$\geqslant b^2\, \mathsf{E} I((X - \mathsf{E}X)^2 \geqslant b^2) = b^2\, \mathsf{P}((X - \mathsf{E}X)^2 \geqslant b^2) = b^2\, \mathsf{P}(|X - \mathsf{E}X| \geqslant b).$$

Применим неравенство Чебышёва к $n^{-1} \sum\limits_{i=1}^{n} X_i$ и получим

$$\mathsf{P}\left(n^{-1} \left|\sum_{i=1}^{n} (X_i - \mu)\right| > \varepsilon\right) \leqslant \frac{1}{n^2 \varepsilon^2} \operatorname{Var}\left(\sum_{i=1}^{n} (X_i - \mu)\right) = \frac{1}{n^2 \varepsilon^2} n \operatorname{Var}(X_1) \to 0$$

при $n \to \infty$.

б) Для случайных величин $Y_1, Y_2, \ldots$, о которых идет речь в задаче,

$$
\begin{aligned}
q_n := \mathsf{P}(&\text{хотя бы одна случайная величина}\\
&Y_1, Y_2, \ldots, Y_{2^n} \text{ приняла значение } 4^n) =\\
&= 1 - \prod_{i=1}^{2^n} \mathsf{P}(Y_i \neq 4^n) = 1 - (\mathsf{P}(Y_1 \neq 4^n))^{2^n} =\\
&\qquad = 1 - (1 - \mathsf{P}(Y_1 = 4^n))^{2^n} = 1 - (1 - 2^{-n})^{2^n} \to 1 - e^{-1}.
\end{aligned}
$$

Поэтому для $2^n > K$ получаем

$$\mathsf{P}\left(2^{-n} \sum_{i=1}^{2^n} Y_i > K\right) \geqslant q_n \to 1 - e^{-1} > 0.$$

Таким образом, если случайные величины Y_i не имеют конечного среднего, то их среднее арифметическое может не сходиться к конечному значению. □

Задача 1.6.8. а) Предположим, что X и Y — дискретные случайные величины с конечным числом значений. Покажите, что $\mathsf{E}(X+Y) = \mathsf{E}X + \mathsf{E}Y$.

б) По сухой дороге студент едет на велосипеде на лекцию со скоростью 20 миль в час, по мокрой — со скоростью 10 миль в час. Расстояние между его домом и зданием, где проходят лекции, равно 3 милям, а лекционный курс состоит из 24 лекций, и начинаются они в 9 утра. Вероятность того, что утром дорога сухая, равна 1/2, но нет причин полагать, что сухая и влажная утренняя погода чередуются независимо. Найдите ожидаемое время, необходимое для того, чтобы доехать на велосипеде на одну из лекций, и среднее время, затраченное на дорогу, для всего лекционного курса.

Приятель этого студента (не математик) предлагает простой ответ: среднее время, затраченное на весь лекционный курс, равно

$$\frac{3 \cdot 24}{\frac{1}{2} \cdot 10 + \frac{1}{2} \cdot 20} = 4 \text{ часа } 48 \text{ минут.}$$

Объясните, почему это время меньше истинного.

Решение. а) Для случайных величин X и Y, принимающих конечное число значений вида x и y соответственно, выполняется равенство

$$
\begin{aligned}
\mathsf{E}(X+Y) &= \sum_{x,y} (x+y)\,\mathsf{P}(X=x,\ Y=y) =\\
&= \sum_x x \sum_y \mathsf{P}(X=x,\ Y=y) + \sum_y y \sum_x \mathsf{P}(X=x,\ Y=y).
\end{aligned}
$$

Внутренние суммы $\sum\limits_y \mathsf{P}(X=x,\ Y=y)$ и $\sum\limits_x \mathsf{P}(X=x,\ Y=y)$ равны соответственно $\mathsf{P}(X=x)$ и $\mathsf{P}(Y=y)$. Поэтому

$$\mathsf{E}(X+Y)=\sum_x x\,\mathsf{P}(X=x)+\sum_y y\,\mathsf{P}(Y=y)=\mathsf{E}X+\mathsf{E}Y.$$

б) Если T — время, затраченное на дорогу при поездке на одну из лекций, то

$$\mathsf{E}(\text{время на дорогу})=3\,\mathsf{E}\left(\frac{1}{\text{скорость}}\right)=3\left(\frac{1}{2}\cdot\frac{1}{10}+\frac{1}{2}\cdot\frac{1}{20}\right)=13{,}5\ \text{мин.}$$

Общее среднее время равно $24\cdot 13{,}5 = 5$ часов 24 мин. Предположение о зависимости здесь не нужно, так как равенство $\mathsf{E}\sum\limits_i T_i=\sum\limits_i \mathsf{E}T_i$ верно в любом случае. «Простой» ответ дает меньшее время

$$\frac{3\cdot 24}{\frac{1}{2}\cdot 10+\frac{1}{2}\cdot 20}<\frac{1}{2}\cdot\frac{3\cdot 24}{10}+\frac{1}{2}\cdot\frac{3\cdot 24}{20}.$$

Однако средняя скорость не равна $\frac{1}{2}\cdot 10+\frac{1}{2}\cdot 20=15$. Это частный случай неравенства Йенсена для строго выпуклой вниз функции $g(x)=\dfrac{3\cdot 24}{x}$, $x\in(0,\infty)$. □

Задача 1.6.9. Пусть $X_1, X_2, \ldots$ — независимые пуассоновские случайные величины с параметром $\lambda>0$. Найдите среднее и дисперсию суммы $\frac{1}{n}\sum\limits_{j=1}^{n} X_j$ и распределение суммы $\sum\limits_{j=1}^{n} X_j$. Исходя из этого или из других соображений найдите

$$\lim_{n\to\infty}\sum_{0\leqslant k\leqslant \mu n}\frac{(\lambda n)^k}{k!}e^{-\lambda n}$$

для всех положительных λ и μ, включая $\lambda=\mu$.

Решение. Сумма $\sum\limits_j X_j$ независимых пуассоновских величин с параметрами (λ_j) есть пуассоновская величина с параметром $\left(\sum\limits_j \lambda_j\right)$: $S_n=\sum\limits_{j=1}^{n} X_j$ — пуассоновская величина с параметром $(n\lambda)$;

$$\mathsf{E}\left(\frac{1}{n}\sum_{j=1}^{n}X_j\right)=\frac{1}{n}n\,\mathsf{E}X_1=\mathsf{E}X_1=\lambda,$$

$$\mathrm{Var}\left(\frac{1}{n}\sum_{j=1}^{n}X_j\right)=\frac{1}{n^2}n\,\mathrm{Var}(X_1)=\frac{\lambda}{n}$$

в силу независимости рассматриваемых случайных величин. Согласно закону больших чисел $\forall \varepsilon > 0$

$$\mathsf{P}\left(\left|\frac{1}{n}S_n - \lambda\right| \leqslant \varepsilon\right) \to 1, \quad \mathsf{P}\left(\left|\frac{1}{n}S_n - \lambda\right| \geqslant \varepsilon\right) \to 0$$

при $n \to \infty$. Иными словами,

$$\mathsf{P}((-\varepsilon + \lambda)n \leqslant S_n \leqslant (\varepsilon + \lambda)n) \to 1,$$
$$\mathsf{P}(S_n \leqslant (-\varepsilon + \lambda)n \ \text{ или } \ S_n \geqslant (\varepsilon + \lambda)n) \to 0$$

при $n \to \infty$. Сумма $\sum\limits_{0 \leqslant k \leqslant \mu n} \frac{(\lambda n)^k}{k!} e^{-\lambda n}$ — это вероятность $\mathsf{P}\left(\frac{1}{n}S_n \leqslant \mu\right)$, где $S_n = X_1 + \ldots + X_n$ и $X_1, X_2, \ldots$ — н. о. р. с. в., пуассоновские, с параметром λ. При этом

$$\mathsf{E}S_n = \mathrm{Var}\, S_n = n\lambda.$$

Тогда по закону больших чисел

$$\mathsf{P}(S_n \leqslant n\mu) \to \begin{cases} 1, & \text{если } \mu > \lambda, \\ 0, & \text{если } \mu < \lambda. \end{cases}$$

Наконец, в силу центральной предельной теоремы для всех x, y, $-\infty \leqslant x < y \leqslant \infty$ выполняется условие

$$\mathsf{P}\left(x < \frac{S_n - n\lambda}{\sqrt{n\lambda}} < y\right) \to \Phi(y) - \Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_x^y e^{-u^2/2}\, du$$

при $n \to \infty$, откуда для $x = -\infty$, $y = 0$ получаем

$$\mathsf{P}(S_n \leqslant n\lambda) = \mathsf{P}\left(\frac{S_n - n\lambda}{\sqrt{n\lambda}} \leqslant 0\right) \to \Phi(0) = \frac{1}{2}.$$

Таким образом, для $\mu = \lambda$ при $n \to \infty$ получаем

$$\lim_{n \to \infty} \sum_{0 \leqslant k \leqslant \lambda n} \frac{(\lambda n)^k}{k!} e^{-\lambda n} \to \frac{1}{2}. \qquad \square$$

Задача 1.6.10. а) Что такое «функция, выпуклая вниз»? Сформулируйте и докажите неравенство Йенсена для выпуклой вниз функции от случайной величины, принимающей конечное число значений.

б) В частности, докажите, что для неотрицательной случайной величины X с конечным числом значений выполняются неравенства

$$\mathsf{E}[X] \leqslant (\mathsf{E}[X^2])^{1/2} \leqslant (\mathsf{E}[X^3])^{1/3} \leqslant \ldots$$

Решение. а) Напомним, что функция g на $[a, b]$ называется выпуклой вниз, если $\forall x, y \in [a, b]$, $\forall\lambda \in (0, 1)$ выполняется неравенство

$$g(\lambda x + (1 - \lambda)y) \leqslant \lambda g(x) + (1 - \lambda)g(y),$$

или, другими словами, для выпуклой вниз функции $g\colon J \to \mathbb{R}$, определенной на интервале, полуинтервале или отрезке $J \subseteq \mathbb{R}$, для любых $x_1, x_2 \in J$ и $p_1, p_2 \in [0, 1]$, $p_1 + p_2 = 1$, справедливо неравенство

$$g(p_1x_1 + p_2x_2) \leqslant p_1g(x_1) + p_2g(x_2).$$

Далее, в неравенстве Йенсена утверждается, что для любых $x_1, \ldots, x_K \in J$ и $p_1, \ldots, p_K \in [0, 1]$, $\sum\limits_{i=1}^{K} p_i = 1$, мы имеем $g(p_1x_1 + \ldots + p_Kx_K) \leqslant p_1g(x_1) + \ldots$ $\ldots + p_Kg(x_K)$. Это легко проверить индукцией по K. Действительно, запишем

$$g\left(\sum_{i=1}^{K} p_ix_i\right) \leqslant p_Kg(x_K) + (1 - p_K)g\left(\sum_{i=1}^{K-1} \frac{p_i}{1 - p_K}x_i\right)$$

и используем предположение индукции для вероятности вида $p_i/(1 - p_K)$, $1 \leqslant i \leqslant K - 1$. Отсюда получаем оценку

$$p_Kg(x_K) + (1 - p_K)g\left(\sum_{i=1}^{K-1} \frac{p_i}{1 - p_K}x_i\right) \leqslant p_Kg(x_K) + \sum_{i=1}^{K-1} p_ig(x_i) = \sum_{i=1}^{K} p_ig(x_i).$$

б) Теперь рассмотрим функцию $g(x) = x^{(n+1)/n}$, строго выпуклую на $J = [0, \infty)$. В силу неравенства Йенсена

$$(\mathsf{E}X)^{(n+1)/n} \leqslant \mathsf{E}X^{(n+1)/n}.$$

Наконец, положим $Y^n = X$ и получим

$$(\mathsf{E}Y^n)^{1/n} \leqslant (\mathsf{E}Y^{n+1})^{1/(n+1)}. \qquad \square$$

Задача 1.6.11. а) Покажите, что для ограниченной случайной величины X выполняется неравенство

$$\mathsf{P}(X \geqslant \lambda) \leqslant e^{-\lambda}\,\mathsf{E}(e^X).$$

б) С помощью разложения в ряд Тейлора или иным способом докажите, что

$$\operatorname{ch} t \leqslant e^{t^2/2}.$$

Покажите, что если Y — случайная величина, для которой $\mathsf{P}(Y = a) = \mathsf{P}(Y = -a) = 1/2$, то

$$\mathsf{E}(e^{\theta Y}) \leqslant e^{a^2\theta^2/2}.$$

Пусть $Y_1, Y_2, \ldots, Y_n$ — независимые случайные величины, для которых $\mathsf{P}(Y_k=a_k)=\mathsf{P}(Y_k=-a_k)=1/2$ и $Z=Y_1+Y_2+\ldots+Y_n$. Аккуратно докажите, что

$$\mathsf{E}(e^{\theta Z}) \leqslant e^{A^2\theta^2/2},$$

где A^2 можно выразить через a_k.

Используя п. а) или иным способом, покажите, что для $\lambda > 0$ выполняется неравенство

$$\mathsf{P}(Z \geqslant \lambda) \leqslant e^{(A^2\theta^2-2\lambda\theta)/2}$$

для всех $\theta > 0$. Найдите значение θ, минимизирующее $e^{(A^2\theta^2-2\lambda\theta)/2}$, и покажите, что

$$\mathsf{P}(Z \geqslant \lambda) \leqslant e^{-\lambda^2/(2A^2)}.$$

Объясните, почему

$$\mathsf{P}(|Z| \geqslant \lambda) \leqslant 2e^{-\lambda^2/(2A^2)}.$$

в) Покажите, что если $a_1 = a_2 = \ldots = a_n = 1$ в п. б), то

$$\mathsf{P}(|Y_1 + Y_2 + \ldots + Y_n| \geqslant (2n \ln \varepsilon^{-1})^{1/2}) \leqslant 2\varepsilon$$

для любых $\varepsilon > 0$.

Решение. а) Это неравенство Чернова, и оно доказывается так: $\mathsf{P}(X \geqslant \lambda) \leqslant e^{-\lambda}\,\mathsf{E}[e^X I(X \geqslant \lambda)] \leqslant e^{-\lambda}\,\mathsf{E}e^X$.

б) Имеет место разложение в ряд $\operatorname{ch} t = \sum\limits_{n=0}^{\infty} \dfrac{t^{2n}}{(2n)!}$ и $e^{t^2/2} = \sum\limits_{n=0}^{\infty} \dfrac{(t^2/2)^n}{(n)!} = \sum\limits_{n=0}^{\infty} \dfrac{t^{2n}}{2^n n!}$. Далее, $(2n)(2n-1)\ldots(n+1) > 2^n$ для $n \geqslant 2$. Поэтому $\operatorname{ch} t \leqslant e^{t^2/2}$ и $\mathsf{E}(e^{\theta Y}) = \operatorname{ch}(\theta a) \leqslant e^{a^2\theta^2/2}$.

Аналогично

$$\mathsf{E}(e^{\theta Y}) = \mathsf{E}(e^{\theta X_1+\ldots+\theta X_n}) = \prod_{k=1}^{n} \operatorname{ch}(\theta a_k) \leqslant \prod_{k=1}^{n} e^{a_k^2\theta^2/2} = e^{A^2\theta^2/2},$$

где $A^2 = \sum\limits_{k=1}^{n} a_k^2$. Отсюда следует, что

$$\mathsf{P}(Z \geqslant \lambda) \leqslant e^{-\lambda\theta}\,\mathsf{E}(e^{\theta Z}) \leqslant e^{(A^2\theta^2-2\lambda\theta)/2}$$

для всех $\theta > 0$.

Функция $f = e^{(A^2\theta^2-2\lambda\theta)/2}$ достигает своего минимума при $\theta = \lambda/A^2$. Поэтому

$$\mathsf{P}(Z \geqslant \lambda) \leqslant e^{-\lambda^2/2A^2}.$$

Теперь рассмотрим $-Z = -(Y_1 + \ldots + Y_n) = Y'_1 + \ldots + Y'_n$, где случайные величины $Y'_k = -Y_k$, $k = 1, \ldots, n$, имеют то же распределение, что и Y_k, $k = 1, \ldots, n$. Тогда

$$\mathsf{P}(-Z \geqslant \lambda) \leqslant e^{-\lambda^2/2A^2}$$

и

$$\mathsf{P}(|Z| \geqslant \lambda) \leqslant 2e^{-\lambda^2/2A^2}.$$

в) Если $a_1 = a_2 = \ldots = a_n = 1$, то $A^2 = n$ и

$$\mathsf{P}(|Y_1 + Y_2 + \ldots + Y_n| \geqslant (2n \ln \varepsilon^{-1})^{1/2}) \leqslant 2e^{\left(-\frac{2n \ln \varepsilon^{-1}}{2n}\right)} = 2\varepsilon. \qquad \square$$

Задача 1.6.12. Пусть $\{X_k\}$ — последовательность н. о. р. с. в. со средним μ и дисперсией σ^2. Покажите, что

$$\sum_{k=1}^{n}(X_k - \bar{X})^2 = \sum_{k=1}^{n}(X_k - \mu)^2 - n(\bar{X} - \mu)^2,$$

где $\bar{X} = \frac{1}{n}\sum\limits_{k=1}^{n} X_k$. Докажите, что при условии $\mathsf{E}(X_1 - \mu)^4 < \infty$ для любого $\varepsilon > 0$ выполняется соотношение

$$\mathsf{P}\left(\left|\frac{1}{n}\sum_{k=1}^{n}(X_k - \bar{X})^2 - \sigma^2\right| > \varepsilon\right) \to 0$$

при $n \to \infty$.

Указание. Согласно неравенству Чебышёва

$$\mathsf{P}\left(\left|\frac{1}{n}\sum_{k=1}^{n}(X_k - \mu)^2 - \sigma^2\right| > \varepsilon/2\right) \to 0. \qquad \square$$

Задача 1.6.13. Пусть $x_1, x_2, \ldots, x_n$ — положительные действительные числа. Тогда их среднее геометрическое находится между средним гармоническим и средним арифметическим:

$$\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}\right)^{-1} \leqslant \left(\prod_{i=1}^{n} x_i\right)^{1/n} \leqslant \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Докажите эти неравенства.

Решение. Для доказательства неравенства между средним арифметическим и средним геометрическим применим индукцию по n. Для $n = 2$ неравенство имеет вид $4x_1x_2 \leqslant (x_1 + x_2)^2$.

Теперь заметим, что это неравенство эквивалентно следующему:

$$\frac{1}{n}\sum_{i=1}^{n}\ln x_i \leqslant \ln\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right).$$

Функция $\ln y$ строго выпукла вверх на $[0, \infty)$ (т. е. $(\ln y)'' < 0$). Поэтому для любого $\alpha \in [0, 1]$ и любых $y_1, y_2 > 0$ выполняется неравенство

$$\ln(\alpha y_1 + (1-\alpha)y_2) \geqslant \alpha \ln y_1 + (1-\alpha)\ln y_2.$$

Положим $\alpha = 1/n$, $y_1 = x_1$, $y_2 = \dfrac{1}{n-1}\sum\limits_{j=2}^{n} x_j$ и получим

$$\ln\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) \geqslant \frac{1}{n}\ln x_1 + \frac{n-1}{n}\ln\left(\frac{1}{n-1}\sum_{i=2}^{n} x_i\right).$$

А теперь применим предположение индукции, согласно которому

$$\ln\left(\frac{1}{n-1}\sum_{i=2}^{n} x_i\right) \geqslant \frac{1}{n-1}\sum_{i=2}^{n}\ln x_i.$$

Чтобы доказать неравенство между средним гармоническим и средним геометрическим, применим предыдущее неравенство к $1/x_1, \ldots, 1/x_n$:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i} \geqslant \prod_{i=1}^{n}\left(\frac{1}{x_i}\right)^{1/n}.$$

Поэтому

$$\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}\right)^{-1} \leqslant \left(\prod_{i=1}^{n}\left(\frac{1}{x_i}\right)^{1/n}\right)^{-1} = \left(\prod_{i=1}^{n} x_i\right)^{1/n}. \qquad \square$$

Задача 1.6.14. Пусть X — положительная случайная величина с конечным числом значений. Покажите, что

$$\mathsf{E}\frac{1}{X} \geqslant \frac{1}{\mathsf{E}X}$$

и что неравенство строгое, за исключением случая, когда $\mathsf{P}(X = \mathsf{E}X) = 1$.

Решение. Пусть величина X принимает значения $x_1, x_2, \ldots, x_n > 0$ с вероятностями p_i. Тогда искомое неравенство запишем в виде $\mathsf{E}X \geqslant \left(\mathsf{E}\dfrac{1}{X}\right)^{-1}$, т. е.

$$\sum_{i=1}^{n} p_i x_i \geqslant \left(\sum_{i=1}^{n} p_i \frac{1}{x_i}\right)^{-1}. \tag{1.6.21}$$

Докажем следующее обобщение неравенства о среднем арифметическом и среднем геометрическом:

$$\prod_{i=1}^{n} x_i^{p_i} \leqslant \sum_{i=1}^{n} p_i x_i \tag{1.6.22}$$

и, применяя его к числам $1/x_i$, получим

$$\left(\sum_{i=1}^{n} p_i \frac{1}{x_i}\right)^{-1} \leqslant \prod_{i=1}^{n} x_i^{p_i},$$

откуда следует и неравенство (1.6.21).

Чтобы доказать неравенство (1.6.22), предположим, что все x_j различны, и применим индукцию по n. Если $n = 1$, то (1.6.22) является очевидным равенством. Предположим, что неравенство (1.6.22) имеет место для $n - 1$, и докажем, что

$$\sum_{i=1}^{n} p_i \ln x_i \leqslant \ln\left(\sum_{i=1}^{n} p_i x_i\right).$$

Для этого вновь используем строгую выпуклость вверх натурального логарифма:

$$\ln(\alpha y_1 + (1-\alpha) y_2) \geqslant \alpha \ln y_1 + (1-\alpha)\ln y_2.$$

Полагая $\alpha = p_1$, $y_1 = x_1$, $y_2 = \frac{1}{1-p_1}\sum\limits_{j=2}^{n} p_j x_j$, получим

$$\ln\sum_{i=1}^{n} p_i x_i = \ln\left(p_1 x_1 + (1-p_1)\sum_{i=2}^{n} p_i' x_i\right) \geqslant p_1 \ln x_1 + (1-p_1)\ln\left(\sum_{i=2}^{n} p_i' x_i\right),$$

где $p_i' = \frac{p_i}{1-p_1}$, $i = 2, \ldots, n$. Воспользуемся теперь предположением индукции, согласно которому

$$\ln\left(\sum_{i=2}^{n} p_i' x_i\right) \geqslant \sum_{i=2}^{n} p_i' \ln x_i,$$

и получим требуемый результат. Равенство имеет место тогда и только тогда, когда либо $p_1(1-p_1) = 0$, либо $x_1 = \sum\limits_{i=2}^{n} p_i' x_i$. Перенося рассуждения на $x_2, \ldots, x_n$ и т. д., убеждаемся, что равенство имеет место тогда и только тогда, когда либо $p_i(1-p_i) = 0$ для некоторого (а значит, для всех) i, либо $x_1 = \ldots = x_n$. Согласно условию это означает, что $n = 1$, т. е.

$$\mathsf{P}(X = \mathsf{E}X) = 1. \qquad \square$$

Задача 1.6.15. Пусть $b_1, b_2, \ldots, b_n$ — произвольная перестановка положительных действительных чисел $a_1, a_2, \ldots, a_n$. Докажите, что

$$\sum_{i=1}^{n} \frac{a_i}{b_i} \geqslant n.$$

Указание. Используйте равенство $\prod\limits_{i=1}^{n} \frac{a_i}{b_i} = 1$. $\square$

Задача 1.6.16. Пусть X — случайная величина, для которой $\mathsf{E}X = \mu$ и $\mathsf{E}(X-\mu)^4 = \beta_4$. Докажите, что

$$\mathsf{P}(|X-\mu| \geqslant t) \leqslant \frac{\beta_4}{t^4}.$$

Указание. Используйте неравенство Маркова для $|X-\mu|^4$. □

Задача 1.6.17. С помощью неравенства Йенсена докажите неравенство между арифметическим средним и геометрическим средним в виде

$$\frac{a_1 + a_2 + \ldots + a_n}{n} \geqslant (a_1 a_2 \ldots a_n)^{1/n},$$

где $a_1, a_2, \ldots, a_n > 0$.

(Можно воспользоваться тем, что функция с положительной второй производной выпукла вниз.)

Решение. Функция $f\colon (a, b) \to \mathbb{R}$ выпукла вниз, если $\forall x, x' \in (a, b)$ и $p \in (0, 1)$ выполняется неравенство

$$f(px + (1-p)x') \leqslant pf(x) + (1-p)f(x').$$

Неравенство Йенсена для случайных величин с конечным числом значений $x_1, \ldots, x_n$ утверждает, что для любых таких функций f, как определено выше,

$$f\left(\sum_{i=1}^{n} p_i x_i\right) \leqslant \sum_{i=1}^{n} p_i f(x_i)$$

$\forall x_1, \ldots, x_n \in (a, b)$ и $p_1, \ldots, p_n \in (0, 1)$, $p_1 + \ldots + p_n = 1$.

Для доказательства используем индукцию по n. Для $n = 1$ неравенство тривиально. (Для $n = 2$ оно эквивалентно определению функции, выпуклой вниз.) Предположим, что неравенство справедливо для некоторого n. Тогда для $n+1$ пусть $x_1, \ldots, x_{n+1} \in (a, b)$ и $p_1, \ldots, p_{n+1} \in (0, 1)$, $p_1 + \ldots + p_{n+1} = 1$. Из равенства $p'_i = p_i/(1-p_{n+1})$ получаем $p'_1, \ldots, p'_n \in (0, 1)$ и $p'_1 + \ldots + p'_n = 1$. Тогда, используя определение функции, выпуклой вниз, а также предположение индукции, получаем

$$f\left(\sum_{i=1}^{n+1} p_i x_i\right) = f\left((1-p_{n+1})\sum_{i=1}^{n} p'_i x_i + p_{n+1}x_{n+1}\right) \leqslant$$
$$\leqslant (1-p_{n+1}) f\left(\sum_{i=1}^{n} p'_i x_i\right) + p_{n+1} f(x_{n+1}) \leqslant$$
$$\leqslant (1-p_{n+1}) \sum_{i=1}^{n} p'_i f(x_i) + p_{n+1} f(x_{n+1}) = \sum_{i=1}^{n+1} p_i f(x_i).$$

Следовательно, неравенство справедливо для $n+1$, а тогда и для любого n.

Для функции $f\colon (0, \infty) \to \mathbb{R}$, $f(x) = -\ln x$, имеем $f'(x) = -1/x$ и $f''(x) = 1/x^2 > 0$. Поэтому функция f выпукла вниз. Согласно неравенству Йенсена при $p_i = 1/n$, $i = 1, \ldots, n$, получаем

$$f\Big(\frac{1}{n}\sum a_i\Big) \leqslant \frac{1}{n}\sum f(a_i),$$

т. е.

$$-\ln\Big(\frac{1}{n}\sum a_i\Big) \leqslant -\frac{1}{n}\sum \ln a_i,$$

или

$$\ln\Big(\frac{1}{n}\sum a_i\Big) \geqslant \ln\Big(\prod a_i\Big)^{1/n}.$$

Отсюда следует, что

$$\frac{1}{n}\sum a_i \geqslant \Big(\prod a_i\Big)^{1/n}. \qquad \square$$

Задача 1.6.18. В ящике находятся N фишек, помеченных номерами $1, \ldots, N$. Эксперимент состоит в вытаскивании n из этих фишек из ящика, $n \leqslant N$. Предположим, что каждая из фишек может быть вынута равновероятно с другими и что вынутые фишки не возвращаются в ящик. Пусть $X_1, \ldots, X_n$ — случайные величины, причем X_i — число, которым помечена i-я вынутая фишка, $i = 1, \ldots, n$. Положим $Y = X_1 + X_2 + \ldots + X_n$.

а) Проверьте, что $\mathrm{Var}(X_i) = (N+1)(N-1)/12$.

Указание. Используйте равенство $\sum\limits_{i=1}^{N} i^2 = \dfrac{N(N+1)(2N+1)}{6}$.

б) Проверьте, что $\mathrm{Cov}(X_i, X_j) = -(N+1)/12$, $i \neq j$.

в) Используя формулу

$$\mathrm{Var}(Y) = \sum_{i=1}^{N} \mathrm{Var}(X_i) + \sum_{i\neq j} \mathrm{Cov}(X_i, X_j)$$

(или иным путем), докажите, что

$$\mathrm{Var}(Y) = \frac{n(N+1)(N-n)}{12}.$$

Решение. а) Очевидно, $\mathsf{E}X_i = (N+1)/2$. Тогда

$$\mathrm{Var}(X_i) = \sum_{k=1}^{N}\frac{k^2}{N} - \Big(\frac{N+1}{2}\Big)^2 = \frac{1}{N}\,\frac{N(N+1)(2N+1)}{6} - \Big(\frac{N+1}{2}\Big)^2 = \frac{(N+1)(N-1)}{12}.$$

б) Поскольку X_i и X_j не равны, мы получаем

$$\frac{1}{N(N-1)}\sum_{k\neq s}\left(k-\frac{N+1}{2}\right)\left(s-\frac{N+1}{2}\right)=$$

$$=\frac{1}{N(N-1)}\sum_{k=1}^{N}\left(k-\frac{N+1}{2}\right)\sum_{s=1}^{N}\left(s-\frac{N+1}{2}\right)-\frac{1}{N-1}\sum_{k=1}^{N}\left(k-\frac{N+1}{2}\right)^2\frac{1}{N}.$$

Первая сумма равна нулю, а вторая равна $-(\mathrm{Var}\,X_i)/(N-1)$. Поэтому $(X_i, X_j) = -(N+1)/12$.

в) Мы имеем

$$\mathrm{Var}(Y) = n\frac{(N+1)(N-1)}{12} - n(n-1)\frac{(N+1)}{12} = \frac{n(N+1)(N-n)}{12}. \qquad \square$$

§ 1.7. Ветвящиеся процессы

Life is a school of probability.

Жизнь — это школа вероятности.

У. Бэйджхот (1826—1877), британский экономист

Ветвящиеся процессы — интереснейший раздел теории вероятностей. Исторически понятие *ветвящегося процесса* возникло при вычислении вероятности вырождения знатных семейств. Упомянем здесь имена У. Р. Гамильтона (1805—1865), известного ирландского математика, Ф. Гальтона (1822—1911), английского ученого и путешественника, и Г. У. Ватсона (1827—1903), английского математика. Начиная примерно с 1940 г. ветвящиеся процессы интенсивно применялись в различных естественных науках, в частности для расчетов, связанных с продуктами деления атомного ядра при цепной реакции (физика), а также для расчетов размера популяции (биология). Позже они нашли широкое применение в компьютерных науках (алгоритмы на логических деревьях) и в других дисциплинах.

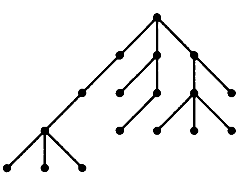


Рис. 1.7

Модель, приводящая к ветвящемуся процессу, проста и элегантна. Вначале имеется некая особь (частица либо биологический организм), производящая случайное число потомков, каждый из которых производит случайное число потомков, и т. д. Этот процесс порождает структуру, сходную с «деревом», где имеется связь потомка со своим родителем и несколько его связей со своими потомками (см. рис. 1.7).

Каждый узел возникающего (случайного) дерева имеет путь, соединяющий его с основным прародителем (называемым началом или корнем дерева). Длина пути, равная числу отрезков этого пути, измеряет число поколений, предшествующих данному узлу (и особи, которую он представляет). Каждый узел порождает поддерево, растущее из него (некоторые узлы продолжения не имеют, т. е. число потомков равно нулю).

Основное предположение состоит в максимальной независимости и однородности: число потомков, произведенное данным родителем, не зависит от числа потомков, произведенных другими особями. Говоря точнее, рассмотрим случайные величины X_0, X_1, X_2, ..., где X_n — это размер популяции в n-м поколении. Тогда

$$X_0 = 1,$$

$$X_1 = \text{число потомков после первого деления},$$

$$X_2 = \text{число потомков после второго деления},$$

и т. д. Случайные величины X_n и X_{n+1} связаны рекуррентно следующим образом:

$$X_{n+1} = \sum_{i=1}^{X_n} Y_i^{(n)}, \tag{1.7.1}$$

где $Y_i^{(n)}$ — число потомков, произведенных i-м представителем n-го поколения. Случайные величины $Y_i^{(n)}$ считаются независимыми и одинаково распределенными, и их распределение, собственно, и определяет ветвящийся процесс.

Первое важное упражнение — вычислить $\mathrm{E}X_n$, т. е. ожидаемую численность n-го поколения. Используя условные математические ожидания, получим

$$\begin{aligned}\mathrm{E}X_n = \mathrm{E}[\mathrm{E}(X_n \,|\, X_{n-1})] = \sum_m \mathrm{P}(X_{n-1} = m)\,\mathrm{E}(X_n \,|\, X_{n-1} = m) = \\ = \sum_m \mathrm{P}(X_{n-1} = m)\,\mathrm{E}\sum_{i=1}^{m} Y_i = \sum_m \mathrm{P}(X_{n-1} = m) m\,\mathrm{E}Y_i = \\ = \mathrm{E}Y \sum_m \mathrm{P}(X_{n-1} = m) m = \mathrm{E}Y\,\mathrm{E}X_{n-1}.\end{aligned} \tag{1.7.2}$$

Здесь Y — случайное число потомков заданного родителя, участвующего в ветвящемся процессе, и мы опустим верхний индекс $(n-1)$ в обозначении $Y_i^{(n-1)}$. Рекуррентный подсчет дает

$$\mathrm{E}X_1 = \mathrm{E}Y, \quad \mathrm{E}X_2 = (\mathrm{E}Y)^2, \quad \ldots, \quad \mathrm{E}X_n = (\mathrm{E}Y)^n, \quad \ldots \tag{1.7.3}$$

Мы видим, что если $\mathrm{E}Y < 1$, то $\mathrm{E}X_n \to 0$ при $n \to \infty$, т. е. процесс в конце концов вырождается. Этот случай часто называют *докритическим*.

Наоборот, если $\mathsf{E}Y > 1$ (*надкритический* процесс), то $\mathsf{E}X_n \to \infty$. Пограничный случай $\mathsf{E}Y = 1$ называется *критическим*.

Замечание. При выводе равенств (1.7.3) не использовалось предположение о независимости.

Удобной характеристикой ветвящегося процесса является производящая функция распределения $\varphi(s) = \mathsf{E}s^Y$, общая для случайных величин $Y_i^{(n)}$. Здесь важен тот факт, что если $\Phi_n(s) = \mathsf{E}s^{X_n}$ — производящая функция распределения численности n-го поколения, то $\Phi_1(s) = \varphi(s)$ и, рекуррентно,

$$\Phi_{n+1}(s) = \Phi_n(\varphi(s)), \quad n \geqslant 1. \tag{1.7.4}$$

См. по этому поводу задачи 1.7.2 и 1.7.9. Иными словами,

$$\Phi_n(s) = \varphi(\varphi(\ldots\varphi(s)\ldots)) = \underbrace{\varphi \circ \ldots \circ \varphi}_{n\ \text{раз}}(s), \tag{1.7.5}$$

где $\varphi \circ \ldots \circ \varphi$ означает итерацию отображения $s \mapsto \varphi(s)$. В частности, $\Phi_{n+1}(s) = \varphi(\Phi_n(s))$.

Эта конструкция приводит к интересному анализу *вероятностей вырождения*

$$\pi_n := \mathsf{P}(X_n = 0) = \Phi_n(0) = \varphi^{\circ n}(0). \tag{1.7.6}$$

В силу того что $\pi_{n+1} = \varphi(\pi_n)$, интуитивно ожидается, что предел $\pi = \lim\limits_{n\to\infty} \pi_n$ будет неподвижной точкой отображения $s \mapsto \varphi(s)$, т. е. решением уравнения $z = \varphi(z)$. Одна такая точка равна 1 (поскольку $\varphi(1) = 1$), но может быть и другое решение, лежащее между 0 и 1. Здесь важен тот факт, что функция φ выпукла вниз и значение $\varphi(0)$ находится между 0 и 1. Если $\varphi'(1) > 1$, то на интервале $(0, 1)$ уравнение $z = \varphi(z)$ будет иметь корень. В противном случае $z = 1$ — наименьший положительный корень (см. рис. 1.8).

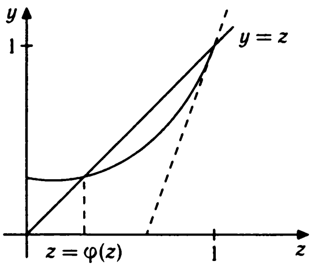


Рис. 1.8

Нетрудно проверить, что предельная вероятность вырождения π существует и

$$\pi = \text{наименьшее неотрицательное решение уравнения } z = \varphi(z). \tag{1.7.7}$$

Действительно, если существует значение $z \in (0, 1)$, при котором $z = \varphi(z)$, то оно единственно и, кроме того, $0 < \varphi'(z) < 1$ и $0 < \varphi(0) = \mathsf{P}(Y = 0) < z$ (поскольку функция φ выпукла вниз и $\varphi(1) = 1$). Значит, $\varphi(0) < \varphi(\varphi(0)) < \ldots < z$, поскольку функция φ монотонна и $\varphi(z) = z$. Поэтому последовательность $\varphi^{\circ n}(0)$ должна иметь предел, который в силу непрерывности совпадает с z.

Если наименьшая неотрицательная неподвижная точка — это $z = 1$, то все предыдущие рассуждения остаются в силе и приводят к равенству $\pi = 1$. Значит, если $\mathsf{P}(Y=0) > 0$, то $\pi > 0$ (на самом деле $\pi > \mathsf{P}(Y=0)$). С другой стороны, если $\varphi(0) = 0$ (т. е. $\mathsf{P}(Y=0) = 0$), то вполне тривиально получаем, что $\pi = 0$. Таким образом, верно равенство (1.7.7). Мы видим, что даже в надкритическом случае (если $\varphi'(1) > 1$) предельная вероятность вырождения π может быть сколь угодно близка к 1.

Небольшая модификация предыдущих рассуждений необходима, если первоначальное число особей больше одной (возможно, случайное число X_0).

Задача 1.7.1. Каждый индивидуум, участвующий в ветвящемся процессе, имеет вероятность p_k породить ровно k потомков, $k = 0, 1, \ldots$, и индивидуумы в каждом поколении производят потомство независимо друг от друга и от индивидуумов в других поколениях. Пусть X_n — численность n-го поколения. Предположим, что $X_0 = 1$ и $p_0 > 0$, и пусть $F_n(s)$ — производящая функция распределения X_n, т. е.

$$F_n(s) = \mathsf{E}s^{X_n} = \sum_{k=0}^{n} p_k s^k.$$

а) Докажите, что

$$F_{n+1}(s) = F_n(F_1(s)).$$

б) Докажите, что для $n < m$ выполняется равенство

$$\mathsf{E}[s^{X_n} \mid X_m = 0] = \frac{F_n(sF_{m-n}(0))}{F_m(0)}. \qquad \square$$

Решение. а) По определению

$$F_{n+1}(s) = \mathsf{E}s^{X_{n+1}} = \sum_{k=0}^{\infty} \mathsf{P}(X_{n+1} = k)s^k,$$

где X_n означает численность n-го поколения. Запишем

$$\mathsf{E}s^{X_{n+1}} = \sum_{l} p_l^{(n)} \mathsf{E}(s^{X_{n+1}} \mid X_n = l),$$

где $p_l^{(n)} = \mathsf{P}(X_n = l)$, а условное математическое ожидание равно

$$\mathsf{E}(s^{X_{n+1}} \mid X_n = l) = \sum_{k=l}^{\infty} \mathsf{P}(X_{n+1} = k \mid X_n = l)s^k.$$

Заметим, что

$$\mathsf{E}(s^{X_{n+1}} \mid X_n = l) = (\mathsf{E}s^{X_1})^l = (F_1(s))^l$$

по двум причинам: 1) при условии $X_n = l$ выполняется равенство

$$X_{n+1} = \sum_{j=1}^{l} \tilde{X}_j,$$

где $\tilde{X}_j$ — число потомков, порожденных j-м индивидуумом в n-м поколении; 2) все $\tilde{X}_j$ — н. о. р. с. в., и $\mathsf{E}(s^{\tilde{X}_j} \mid X_n = l) = \mathsf{E}s^{X_1} = F_1(s)$. Из этого равенства следует, что

$$\mathsf{E}s^{X_{n+1}} = \sum p_l^{(n)}(F_1(s))^l = F_n(F_1(s)).$$

б) Обозначим через $I_0^{(m)}$ случайную величину, заданную как

$$I_0^{(m)} = \begin{cases} 1, & \text{если } X_m = 0, \\ 0 & \text{в противном случае.} \end{cases}$$

Тогда $\mathsf{P}(I_0^{(m)} = 1) = \mathsf{E}I_0^{(m)} = \mathsf{P}(X_m = 0) = F_m(0)$. Кроме того,

$$\mathsf{E}[s^{X_n} \mid X_m = 0] = \mathsf{E}(s^{X_n} I_0^{(m)})/F_m(0).$$

Значит, достаточно проверить, что

$$\mathsf{E}(s^{X_n} I_0^{(m)}) = F_n(sF_{n-m}(0)).$$

Но

$$\mathsf{E}(s^{X_n} I_0^{(m)}) = \sum_k s^k \,\mathsf{P}(X_n = k,\, X_m = 0) = \sum_k s^k \,\mathsf{P}(X_n = k)\,\mathsf{P}(X_m = 0 \mid X_n = k),$$

а теперь, поскольку $\mathsf{P}(X_m = 0 \mid X_n = k) = F_{m-n}(0)^k$, мы получаем

$$\mathsf{E}(s^{X_n} I_0^{(m)}) = F_n(sF_{n-m}(0)). \qquad \square$$

Задача 1.7.2. В лаборатории размещена популяция тли. Вероятность того, что день для конкретной особи пройдет без событий, равна $q < 1$. При условии, что за день состоятся какие-либо события, с вероятностью r тля производит одного потомка, с вероятностью s — двух потомков и с вероятностью t она умрет от истощения, причем $r + s + t = 1$. Потомки готовы к воспроизводству на следующий день. Судьбы различных особей независимы друг от друга, так же как и события разных дней. Лаборатория начинает работу, имея одну особь.

Пусть X_n — число особей в конце n-го дня. Получите производящую функцию $\varphi_n(z)$ распределения величины X_n. Каково ожидаемое значение величины X_n?

Покажите, что вероятность вырождения не зависит от q и что в случае, когда $2r + 3s \leqslant 1$, тли наверняка вымрут. Найдите вероятность вырождения, если $r = 1/5$, $s = 2/5$ и $t = 2/5$.

Решение. Обозначим через $\varphi_{X_1}(z)$ производящую функцию распределения величины X_1, т. е. числа тлей, порожденных к концу первого дня одной особью (включая эту начальную особь). Тогда

$$\varphi_{X_1}(z) = (1-q)t + qz + (1-q)rz^2 + (1-q)sz^3, \quad z > 0.$$

Запишем $\mathsf{E}X_n = (\mathsf{E}X_1)^n$, где $\mathsf{E}X_1 = \varphi'_{X_1}(1) = q + 2(1-q)r + 3(1-q)s$. В самом деле, из равенства $\varphi_{X_n}(z) = \varphi_{X_{n-1}}(\varphi_{X_1}(z))$ следует, что $\mathsf{E}X_n = \mathsf{E}X_{n-1}\,\mathsf{E}X_1$, или $\mathsf{E}X_n = (\mathsf{E}X_1)^n$. Вероятность вырождения в конце n-го дня равна $\varphi_{X_n}(0)$. Она не убывает по n и при $n \to \infty$ стремится к пределу, который и равен вероятности вырождения π. Как уже известно, $\pi < 1$ тогда и только тогда, когда $\mathsf{E}X_1 > 1$. Вероятность вырождения π — это наименьший положительный корень уравнения

$$(1-q)t + qz + (1-q)rz^2 + (1-q)sz^3 = z,$$

или, после деления на $(1-q)$ (поскольку $q < 1$),

$$t - z + rz^2 + sz^3 = 0.$$

Последнее уравнение не зависит от q, значит, и π не зависит от q. Условие $\mathsf{E}X_1 \leqslant 1$ эквивалентно неравенству $2r + 3s \leqslant 1$. В случае $r = 1/5$, $s = 2/5$, $t = 2/5$ уравнение принимает вид $2z^3 + z^2 - 5z + 2 = 0$. Разделив на $(z-1)$ (поскольку $z = 1$ — корень), получим квадратное уравнение $2z^2 + 3z - 2 = 0$, корни которого равны $z_\pm = \dfrac{(-3 \pm 5)}{4}$. Положительный корень равен $1/2$, и это есть вероятность вырождения. □

Задача 1.7.3. а) Объясните, что такое «ветвящийся процесс».

б) Пусть X_n — численность n-го поколения ветвящегося процесса, в котором численность каждой семьи имеет производящую функцию распределения G, и предположим, что $X_0 = 1$. Покажите, что производящая функция распределения G_n величины X_n удовлетворяет соотношению $G_{n+1}(s) = G_n(G(s))$ для $n \geqslant 1$.

в) Покажите, что $G(s) = 1 - \alpha(1-s)^\beta$ — производящая функция распределения неотрицательной целочисленной случайной величины, если $\alpha, \beta \in (0, 1)$, и найдите явно G_n при такой функции G.

г) Найдите вероятность того, что $X_n = 0$, и покажите, что при $n \to \infty$ она сходится к $1 - \alpha^{1/(1-\beta)}$. Объясните, почему отсюда следует, что вероятность вырождения равна $1 - \alpha^{1/(1-\beta)}$.

Решение. а) Ветвящийся процесс — это такая последовательность случайных величин $X_1, X_2, \ldots$, что

$$X_{n+1} = \sum_{i=1}^{X_n} F_i,$$

где F_i — это независимые копии случайной величины F с неотрицательными целыми значениями, причем F — случайное число потомков одного родителя, F_i не зависят от X_n.

б) Из п. а) следует, что

$$G_{n+1}(s) = \mathsf{E}s^{X_{n+1}} = \mathsf{E}s^{\sum_1^{X_n} F_i} = \mathsf{E}[\mathsf{E}(s^{X_{n+1}} \mid X_n)] = \sum_m \mathsf{P}(X_n = m)\,\mathsf{E}s^{\sum_1^{m} F_i} =$$
$$= \sum_m \mathsf{P}(X_n = m)(\mathsf{E}s^F)^m = \sum_m \mathsf{P}(X_n = m)(G(s))^m = G_n(G(s)).$$

в), г) Запишем

$$G(s) = 1 - \alpha\left(1 + \sum_{i=1}^{\infty} \frac{\beta(\beta-1)\dots(\beta-i+1)}{i!}(-s)^i\right) =$$
$$= (1-\alpha) + \alpha \sum_{1}^{\infty} \frac{\beta(1-\beta)\dots(i-1-\beta)}{i!} s^i.$$

Это степенной ряд с положительными коэффициентами, сумма которых равна $G(1) = 1$. Поэтому G — производящая функция распределения. По индукции $G_n(s) = 1 - \alpha^{1+\beta+\dots+\beta^{n-1}}(1-s)^{\beta^n}$. Поэтому при $n \to \infty$ получаем

$$\mathsf{P}(X_n = 0) = G_n(0) = 1 - \alpha^{\frac{1-\beta^n}{1-\beta}} \to 1 - \alpha^{\frac{1}{(1-\beta)}}.$$

Наконец,

$$\mathsf{P}(\text{окончательное вырождение}) = \mathsf{P}(X_n = 0 \text{ для некоторого } n) =$$
$$= \mathsf{P}\left(\bigcup_n \{X_n = 0\}\right) = \lim_{n\to\infty} \mathsf{P}(X_n = 0),$$

поскольку события $\{X_n = 0\}$ возрастают по n. □

Задача 1.7.4. В момент времени $t = 0$ эксперимент начинается с одной красной кровяной клетки. В конце первой минуты красная клетка погибает и заменяется одной из следующих комбинаций с указанными вероятностями:

две красные клетки с вероятностью $\frac{1}{4}$;

одна красная, одна белая с вероятностью $\frac{2}{3}$;

две белые клетки с вероятностью $\frac{1}{12}$.

Каждая красная кровяная клетка живет одну минуту и порождает потомков таким же образом, как и родительская клетка. Каждая белая кровяная клетка живет одну минуту и погибает без воспроизведения. Предположим, что индивидуальные клетки ведут себя независимо друг от друга.

а) Найдите вероятность того, что ни одной белой клетки еще не появилось в момент времени $t = n + 1/2$ минут после начала эксперимента.

б) Найдите вероятность того, что вся колония клеток в конце концов выродится.

Решение. а) Событие «до момента времени $n + 1/2$ не появилось ни одной белой клетки» означает, что общее число клеток равно общему числу красных клеток и равно 2^n. Тогда, обозначая вероятность этого события p_n, получим

$$p_0 = 1, \quad p_1 = \frac{1}{4} \quad \text{и} \quad p_{n+1} = p_n\left(\frac{1}{4}\right)^{2^n}, n \geqslant 1,$$

откуда следует, что

$$p_n = \left(\frac{1}{4}\right)^{2^{n-1}} \cdot \left(\frac{1}{4}\right)^{2^{n-2}} \cdot \ldots \cdot \left(\frac{1}{4}\right)^{2^0} = \left(\frac{1}{4}\right)^{2^n - 1}, \quad n \geqslant 1.$$

б) Вероятность вырождения π удовлетворяет уравнению

$$\pi = \frac{1}{4}\pi^2 + \frac{2}{3}\pi + \frac{1}{12}, \quad \text{или} \quad \frac{1}{4}\pi^2 - \frac{1}{3}\pi + \frac{1}{12} = 0,$$

откуда следует, что

$$\pi = \frac{2}{3} \pm \frac{1}{3}.$$

Нужно взять наименьший корень $\pi = 1/3$. □

Задача 1.7.5. Пусть $\{X_n\}$ — такой ветвящийся процесс, что $X_0 = 1$, $\mathsf{E}X_1 = \mu$. Докажите, что если $Y_n = X_0 + \ldots + X_n$ и для $0 \leqslant s \leqslant 1$ справедливо тождество

$$\Psi_n(s) \equiv \mathsf{E}s^{Y_n},$$

то

$$\Psi_{n+1}(s) = s\varphi(\Psi_n(s)),$$

где $\varphi(s) \equiv \mathsf{E}s^{X_1}$. Проверьте, что для $Y = \sum\limits_{n \geqslant 0} X_n$ функция $\Psi(s) \equiv \mathsf{E}s^Y$ удовлетворяет уравнению

$$\Psi(s) = s\varphi(\Psi(s)), \quad 0 \leqslant s \leqslant 1.$$

Если $\mu < 1$, докажите, что $\mathsf{E}Y = (1 - \mu)^{-1}$.

Решение. Запишем $Y_{n+1} = 1 + Y_n^1 + \ldots + Y_n^{X_1}$, где Y_n^j — общее число потомков, порожденное индивидуумом j из первого поколения. Тогда Y_n^j — это н. о. р. с. в. и их производящая функция распределения равна $\Psi_n(s)$.

Поэтому

$$\Psi_{n+1}(s) = \mathsf{E}s^{Y_{n+1}} = s\sum_j \mathsf{P}(X_1 = j)\prod_{l=1}^{j}\Psi_n(s) =$$
$$= s\sum_j \mathsf{P}(X_1 = j)(\Psi_n(s))^j = s\varphi(\Psi_n(s)).$$

Бесконечный ряд $Y = \sum\limits_{n\geqslant 0} X_n$ имеет производящую функцию распределения $\Psi(s) = \lim\limits_{n\to\infty} \Psi_n(s)$. Поэтому она удовлетворяет уравнению

$$\Psi(s) = s\varphi(\Psi(s)).$$

По индукции $\mathsf{E}Y_n = 1+\mu+\ldots+\mu^n$. В самом деле, $Y_1 = 1+X_1$ и $\mathsf{E}Y_1 = 1+\mu$. Предположим что данная формула верна для $n \leqslant k-1$. Тогда

$$\mathsf{E}Y_k = (\Psi_k(1))' = \varphi(\Psi_{k-1}(1)) + \varphi'(\Psi_{k-1}(1))\Psi'_{k-1}(1) =$$
$$= 1+\mu(1+\mu+\ldots+\mu^{k-1}) = 1+\mu+\mu^2+\ldots+\mu^k,$$

что и завершает индукцию. Итак, $\mathsf{E}Y = (1-\mu)^{-1}$. □

Задача 1.7.6. При болезни Грина волосы становятся розовыми, а ногти — голубыми, но других симптомов она не имеет. Заболевший может заразить n следующих здоровых людей ($n = 0, 1, 2, 3$) с вероятностью p_n, но не может заразить более трех людей. [Обычные предположения по поводу простых ветвящихся процессов.] Запишите выражение для ожидаемого числа e индивидуумов, инфицированных одним заболевшим.

Найдите вероятность π того, что эпидемия заглохнет, если вначале был один заболевший.

Пусть e_A и π_A — значения e и π при $p_0 = 2/5$, $p_1 = p_2 = 0$ и $p_3 = 3/5$. Пусть e_B и π_B — значения e и π при $p_0 = p_1 = 1/10$, $p_2 = 4/5$ и $p_3 = 0$. Покажите, что $e_A > e_B$, но $\pi_A > \pi_B$.

Решение. Выражение для e имеет вид $e = p_1 + 2p_2 + 3p_3$. Пусть X_j — число заболевших индивидуумов в j-м поколении и $X_0 = 1$. Предположим, что каждый заболевший умирает, заразив $n \leqslant 3$ других. Обозначим через π вероятность того, что эпидемия прекратится:

$$\pi = \sum_k \mathsf{P}(X_1 = k)\pi^k = p_0 + p_1\pi + p_2\pi^2 + p_3\pi^3.$$

Простые вычисления показывают, что $e_A = 9/5$, $e_B = 17/10$. Найдем π_A как наименьший положительный корень уравнения

$$0 = p_0 + (p_1 - 1)\pi + p_2\pi^2 + p_3\pi^3 = (\pi - 1)\left(\frac{3}{5}\pi^2 + \frac{3}{5}\pi - \frac{2}{5}\right).$$

Тогда

$$\pi_A = \frac{-3+\sqrt{33}}{6} \approx 0{,}46.$$

Аналогично π_B — наименьший положительный корень уравнения

$$0 = (\pi - 1)\left(\frac{4}{5}\pi - \frac{1}{10}\right),$$

и $\pi_B = 1/8$. Таким образом, $e_A > e_B$, но $\pi_A > \pi_B$. □

Задача 1.7.7. Предположим, что $(X_r,\ r \geqslant 0)$ — ветвящийся процесс, $X_0 = 1$ и функция распределения вероятностей для X_1 равна $G(s)$.

а) Найдите итерационную формулу для производящей функции распределения $G_r(s)$ случайной величины X_r. Сформулируйте результат относительно вероятности вырождения в терминах $G(s)$.

б) Предположим, что вероятность того, что особь оставит после себя k потомков в следующем поколении, равна $p_k = 1/2^{k+1}$, $k \geqslant 0$. Выведите из п. а), что вырождение произойдет с вероятностью 1. Затем докажите, что

$$G_r(s) = \frac{r-(r-1)s}{(r+1)-rs}, \qquad r \geqslant 1,$$

и получите вероятность того, что r-е поколение вымрет.

в) Предположим, что каждая особь порождает не более трех потомков в следующем поколении и что вероятность рождения k потомков в следующем поколении равна $p_k = C_3^k \cdot \frac{1}{2^3}$, $0 \leqslant k \leqslant 3$. Какова вероятность вырождения?

Решение. а) Пусть Y_i^n — число потомков индивидуума i в n-м поколении. Таким образом,

$$X_{n+1} = Y_1^n + \ldots + Y_{X_n}^n,$$

$$\mathsf{E}(s^{X_{n+1}}) = \mathsf{E}(\mathsf{E}(s^{X_{n+1}} \mid X_n)) = \sum_{k=0}^{\infty} \mathsf{P}(X_n = k)\,\mathsf{E}(s^{Y_1^n+\ldots+Y_k^n}) =$$

$$= \sum_{k=0}^{\infty} \mathsf{P}(X_n = k)\,\mathsf{E}(s^{Y_1^n})^k = \sum_{k=0}^{\infty} \mathsf{P}(X_n = k) G(s)^k = G_n(G(s)),$$

и $G_{n+1}(s) = G_n(G(s))$. Вероятность вырождения — это наименьшее такое $s \in (0,\ 1]$, что $s = G(s)$.

б) Вычислим производящую функцию распределения

$$G(s) = \frac{1}{2} + \frac{1}{2^2}s + \ldots = \frac{1}{2-s}.$$

Решая уравнение $s = \frac{1}{2-s}$, или $(s-1)^2 = 0$, видим, что вырождение произойдет с вероятностью 1.

Докажем формулу для $G_r(s)$ по индукции:

$$G_{r+1}(s) = \frac{r - (r-1)\dfrac{1}{2-s}}{r+1-r\dfrac{1}{2-s}} = \frac{(r+1) - rs}{(r+2) - (r+1)s}.$$

Поэтому вероятность того, что r-е поколение вымрет, равна $G_r(0) = \dfrac{r}{r+1}$.

в) В этом случае $G(s) = \frac{1}{2^3}(1+s)^3$, и, преобразуя это уравнение, получим $2^3 s = (1+s)^3$, или $(s-1)(s^2+4s-1) = 0$, значит, решения есть $s = 1$, $s = \pm\sqrt{5} - 2$, а вероятность вырождения равна $\sqrt{5} - 2 \approx 0{,}24$. □

Задача 1.7.8. Рассмотрим такой *процесс Гальтона—Ватсона* (т. е. ветвящийся процесс, в котором число потомков случайно и независимо на любом этапе деления), что $\mathrm{P}(X_1 = 0) = 2/5$ и $\mathrm{P}(X_1 = 2) = 3/5$. Нужно вычислить распределение случайной величины X_2. Для третьего поколения найдите все вероятности $\mathrm{P}(X_3 = 2k)$, $k = 0, 1, 2, 3, 4$. Найдите вероятность вырождения в этой модели.

Решение. Вероятность вырождения $\pi = 2/3$, и производящая функция распределения случайной величины X_2 равна

$$\varphi_{X_2}(s) = \frac{2}{5} + \frac{12}{125} + \frac{36}{125}s^2 + \left(\frac{3}{5}\right)^3 s^4.$$

Производящая функция распределения имеет вид $\varphi_{X_3}(s) = \varphi_{X_2}(2/5 + 3s^2/5)$. Тогда

$$\mathrm{P}(X_2 = 0) = \frac{2}{5} + \frac{12}{125},\ \mathrm{P}(X_2 = 2) = \frac{36}{125},\ \mathrm{P}(X_2 = 4) = \left(\frac{3}{5}\right)^3;$$

$$\mathrm{P}(X_3 = 0) = \frac{2}{5} + \frac{12}{125} + \frac{36}{125}\left(\frac{2}{5}\right)^2 + \left(\frac{3}{5}\right)^3\left(\frac{2}{5}\right)^4 = 0{,}54761;$$

$$\mathrm{P}(X_3 = 2) = \frac{2^4 \cdot 3^3}{5^5} + \frac{2^5 \cdot 3^4}{5^7} = 0{,}17142;$$

$$\mathrm{P}(X_3 = 4) = \frac{4 \cdot 3^4}{5^5} + \frac{4^2 \cdot 3^5}{5^7} + \frac{8 \cdot 3^5}{5^7} = 0{,}17833;$$

$$\mathrm{P}(X_3 = 6) = \frac{3^6 \cdot 2^3}{5^7} = 0{,}07465;$$

$$\mathrm{P}(X_3 = 8) = \left(\frac{3}{5}\right)^7 = 0{,}02799.$$

□

Задача 1.7.9. Рассматривая вероятность вырождения (или другим путем), решите следующую задачу.

Никто, будучи в здравом уме, не захочет остановиться в гостинице «Виртуальная реальность». Там комнаты пронумерованы от 0 до $(3^N - 3)/2$, где N — очень большое целое число. Если $0 \leqslant i \leqslant (3^{N-1} - 3)/2$, а $j = 1, 2, 3$, то имеется дверь между комнатами i и $3i + j$, через которую, если она не заперта, гости могут проходить в обоих направлениях. Кроме того, каждая комната с номером, большим чем $(3^{N-1} - 3)/2$, имеет открытое окно, через которое гости могут удрать на улицу (и будут так поступать). Других дверей и окон для гостей нет. Рис. 1.9 демонстрирует часть плана этажа гостиницы.

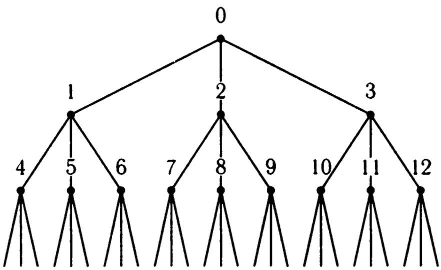


Рис. 1.9

Любая дверь гостиницы заперта с вероятностью 1/3 независимо от других. Прибывшего гостя размещают в комнате 0, и он может свободно перемещаться, насколько позволяют запертые двери. Покажите, что шансы гостя выбраться близки к $(9 - \sqrt{27})/4$.

Решение. Обозначим через X_r число доступных комнат на уровне r удаления от 0. Запишем $\varphi_r(t) = \mathsf{E}t^{X_r}$, где $\varphi_1 = \varphi$,

$$\varphi_{r+1}(t) = \mathsf{E}t^{X_{r+1}} = \sum_i \mathsf{P}(X_r = i)\,\mathsf{E}(t^{X_{r+1}} \mid X_r = i) =$$
$$= \sum \mathsf{P}(X_r = i)(\mathsf{E}t^{X_1})^i = \sum \mathsf{P}(X_r = i)(\varphi(t))^i = \varphi_r(\varphi(t)).$$

Тогда $\varphi_{r+1}(t) = \varphi(\varphi(\ldots\varphi(t)\ldots)) = \varphi(\varphi_r(t))$ и P(уровня r нельзя достичь) $=$ $= \varphi(\varphi_r(0))$.

Теперь

$$\varphi(t) = \left(\frac{1}{3}\right)^3 + 3\frac{2}{3}\left(\frac{1}{3}\right)^2 t + 3\left(\frac{2}{3}\right)^2 \frac{1}{3}t^2 + \left(\frac{2}{3}\right)^3 t^3 = \frac{1}{27}(1 + 6t + 12t^2 + 8t^3).$$

Значит, уравнение $\varphi(t) = t$ принимает вид

$$27t = 1 + 6t + 12t^2 + 8t^3, \quad \text{т. е.} \quad 1 - 21t + 12t^2 + 8t^3 = 0.$$

Разлагая на множители многочлен $1 - 21t + 12t^2 + 8t^3 = (t-1)(8t^2 + 20t - 1)$, находим корни

$$t = 1, \ \frac{-5 \pm \sqrt{27}}{4}.$$

Между 0 и 1 находится корень $(\sqrt{27}-5)/4$. Последовательность $\varphi_n(0)$ монотонно возрастает и ограничена сверху. Поэтому она сходится:

$$\mathsf{P}(\text{не выбраться}) \to \frac{\sqrt{27}-5}{4},$$

и

$$\mathsf{P}(\text{выбраться}) \to 1 - \frac{\sqrt{27}-5}{4} = \frac{9-\sqrt{27}}{4} \approx 0{,}950962. \qquad \square$$

Задача 1.7.10. а) Зрелый индивидуум производит потомков согласно производящей функции распределения $F(s)$. Предположим, что все начинается с популяции, состоящей из k индивидуумов, не достигших зрелости, каждый из которых достигает зрелости с вероятностью p и затем размножается независимо от других индивидуумов. Найдите производящую функцию распределения числа индивидуумов (не достигших зрелости) в следующем поколении.

б) Найдите производящую функцию распределения числа зрелых индивидуумов в следующем поколении, если известно, что в родительском поколении имеется k зрелых индивидуумов.

Решение. а) Пусть R — общее число (не достигших зрелости) потомков от одного (зрелого) родителя. Положим $F(s) = \mathsf{E}s^R$. Пусть Y — общее число (не достигших зрелости) потомков от (не достигшего зрелости) предка. Положим $G(s) = \mathsf{E}s^Y$. Пусть $M = 1$ или 0 в зависимости от того, достиг зрелости данный потомок или нет. Тогда $g(s) = \mathsf{E}s^M = sp + 1 - p$. Далее,

$$G(s) = \mathsf{E}[\mathsf{E}(s^Y \mid M)] = p\,\mathsf{E}s^R + (1-p)s^0 = (1-p) + pF(s) = g(F(s)).$$

Наконец, если $X^{(k)}$ — общее число (не достигших зрелости) потомков от k не достигших зрелости предков, то $X^{(k)} = Y_1 + \ldots + Y_k$, где случайные величины $Y_1, \ldots, Y_k$ независимы и распределены так же, как Y. Поэтому

$$G_k(s) = \mathsf{E}s^{X^{(k)}} = \mathsf{E}s^{Y_1+\ldots+Y_k} = (\mathsf{E}s^Y)^k = [g(F(s))]^k.$$

б) Пусть U — общее число зрелых потомков одного зрелого родителя. Положим

$$H(s) = \mathsf{E}s^U = \mathsf{E}[\mathsf{E}(s^U \mid R)].$$

Но

$$\mathsf{E}(s^U \mid R = r) = \sum_{k=0}^{r} C_r^k s^k p^k (1-p)^{r-k} = (ps + 1 - p)^r,$$

поэтому

$$H(s) = \mathsf{E}(ps + 1 - p)^R = F(ps + 1 - p) = F(g(s)).$$

Наконец, если $Z^{(k)}$ — общее число зрелых потомков от k зрелых родителей, то $Z^{(k)} = U_1 + \ldots + U_k$, где случайные величины $U_1, \ldots, U_k$ независимы и распределены так же, как U. Поэтому

$$H_k(s) = \mathsf{E}s^{Z^{(k)}} = \mathsf{E}s^{U_1+\ldots+U_k} = (\mathsf{E}s^U)^k = [F(g(s))]^k. \qquad \square$$

Задача 1.7.11. Покажите, что распределения в пп. а) и б) из задачи 1.7.10 имеют одинаковое среднее, но не обязательно одинаковую дисперсию.

Решение. Поскольку $G'(s) = g'(F(s))F'(s)$,

$$\mu_G = G'(1) = F'(1)g'(F(1)) = \mu_F\mu_g$$

и

$$\mu_H = H'(1) = g'(1)F'(g(1)) = \mu_F\mu_g = \mu_G,$$

средние одинаковы. Теперь подсчитаем вторые производные:

$$G''(s) = F''(s)g'(F(s)) + (F'(s))^2g''(F(s)),$$

$$\begin{aligned}
\sigma_G^2 &= G''(1) + G'(1) - (G'(1))^2 = \\
&= F''(1)g'(F(1)) + (F'(1))^2g''(F(1)) + \mu_F\mu_g - \mu_F^2\mu_g^2 = \\
&= F''(1)\mu_g + \mu_F^2g''(1) + \mu_F\mu_g - \mu_F^2\mu_g^2 = \\
&= (\sigma_F^2 + \mu_F^2 - \mu_F)\mu_g + \mu_F^2(\sigma_g^2 + \mu_g^2 - \mu_g) + \mu_F\mu_g - \mu_F^2\mu_g^2 = \mu_g\sigma_F^2 + \mu_F^2\sigma_g^2,
\end{aligned}$$

$$H''(s) = g''(s)F'(g(s)) + (g'(s))^2F''(g(s)),$$

$$\sigma_H^2 = \mu_F\sigma_g^2 + \mu_g^2\sigma_F^2,$$

а это выражение не обязательно равно $\sigma_G^2 = \mu_g\sigma_F^2 + \mu_F^2\sigma_g^2$. В данном примере $\sigma_g^2 = p(1-p)$. $\square$

Глава 2

Непрерывные пространства элементарных исходов

§ 2.1. Равномерное распределение. Плотность распределения вероятностей. Случайные величины. Независимость

Probabilists do it continuously but discretely.

Вероятностники делают это непрерывно, но прерывисто.[1]

(Из серии «Как они делают это».)

Bye, Bi, Variate[2]

(Из серии «Фильмы, которые не вышли на большой экран».)

После изучения основных вероятностных моделей с дискретными исходами мы в состоянии продвинуться далее и поработать с моделями, в которых участвует несчетное число исходов. Теперь события ассоциируются с подмножествами непрерывного пространства (действительная прямая $\mathbb{R}$, интервал (a, b), плоскость $\mathbb{R}^2$, квадрат и т. д.). Простейший случай — когда пространство исходов Ω представимо «хорошим» ограниченным множеством и вероятностное распределение соответствует единичной массе, равномерно по нему распределенной. Тогда событие (т. е. подмножество) $A \subseteq \Omega$ имеет вероятность

$$\mathsf{P}(A) = \frac{v(A)}{v(\Omega)}, \tag{2.1.1}$$

где $v(A)$ — стандартный евклидов объем (либо площадь, либо длина) множества A и $v(\Omega)$ — то же для Ω. Термин «равномерно распределенный» здесь ключевой; как показывает приведенная ниже задача 2.1.2, нужно тщательно разобраться, что это в точности значит в контексте данного примера. Полезное наблюдение состоит в том, что можно работать с плотностью *равномерного распределения*, которая приписывает точке $x \in \Omega$

[1] Игра слов: «discretely» означает в англ. языке «прерывисто», а «discreetly» — осторожно, благоразумно.

[2] Непереводимая игра слов.

значение

$$f_{\Omega}^{\mathrm{uni}}(x) = \frac{1}{v(\Omega)} I_{\Omega}(x), \tag{2.1.2}$$

а тогда вероятность события $A \subseteq \Omega$ вычисляется как интеграл

$$\mathsf{P}(A) = \int\limits_A f_{\Omega}^{\mathrm{uni}}(x)\,dx = \frac{1}{v(\Omega)} \int\limits_A dx \tag{2.1.3}$$

и мы приходим к тому же ответу, что и в формуле (2.1.1). Поскольку $f_{\Omega}^{\mathrm{uni}} \geqslant 0$ и $\int\limits_{\Omega} f_{\Omega}^{\mathrm{uni}}(x)\,dx = 1$, вероятность события $A \subseteq \Omega$ всегда находится между 0 и 1. Заметим, что масса, приписываемая единичному исходу ω, представимому в виде точки из Ω, равна нулю. Поэтому масса, приписываемая любому конечному или счетному множеству исходов, равна нулю (поскольку это сумма масс, приписываемых каждому исходу); чтобы иметь положительную массу (а значит, и положительную вероятность), событие A должно быть несчетным.

Задача 2.1.1. Алиса и Боб условились встретиться в «Медном чайнике» после своих субботних лекций. Каждый из них появляется там в некий момент времени независимо друг от друга, и это время равномерно распределено между 12 и 13 часами. Каждый ожидает другого в течение s минут, а затем уходит. Найдите наименьшее значение s, для которого вероятность встречи не менее 1/2.

Решение. Множество Ω — это единичный квадрат $\mathcal{S}$ с координатами $0 \leqslant x,\, y \leqslant 1$ (время измеряется долями одного часа между 12:00 и 13:00). Каждая точка $\omega = (x, y) \in \Omega$ означает время x прибытия Алисы и y — Боба. Тогда событие «разница между моментами прибытия не превосходит s минут» — это полоска вокруг диагонали $x = y$:

$$A = \left\{(x, y) \in \mathcal{S}\colon |x - y| \leqslant \frac{s}{60}\right\}.$$

Дополнение этого множества состоит из двух треугольников площадью $\frac{1}{2}\left(1 - \frac{s}{60}\right)^2$ каждый (см. рис. 2.1). Поэтому площадь $v(A)$ равна

$$1 - \left(1 - \frac{s}{60}\right)^2,$$

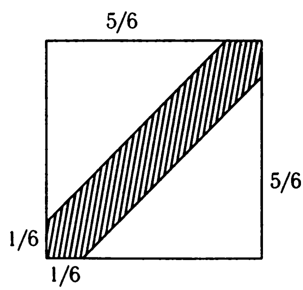


Рис. 2.1

и мы хотим, чтобы эта вероятность была не меньше 1/2. Отсюда получаем $s \geqslant 60\left(1 - \frac{\sqrt{2}}{2}\right) \approx 18$ минут. □

Задача 2.1.2. Следующая задача известна как *парадокс Бертрана*. На круге радиуса r случайным образом выбирается хорда. Найдите вероятность того, что она длиннее, чем сторона равностороннего треугольника, вписанного в круг. Рассмотрим три различных случая:

а) середина хорды равномерно распределена внутри круга,

б) один конец хорды фиксирован, а второй равномерно распределен по окружности,

в) расстояние между серединой хорды и центром круга равномерно распределено на отрезке $[0, r]$.

Решение. В случае а) середина хорды должна находиться внутри круга, вписанного в треугольник. Поэтому

$$\mathrm{P}(\text{хорда длиннее}) = \frac{\text{площадь вписанного круга}}{\text{площадь исходного круга}} = \frac{\pi r^2/4}{\pi r^2} = \frac{1}{4}.$$

В случае б) второй конец окружности должен находиться на противоположной трети окружности. Поэтому

$$\mathrm{P}(\text{хорда длиннее}) = \frac{1}{3}.$$

Наконец, в случае в) середина хорды должна находиться на расстоянии не больше $r/2$ от центра. Поэтому

$$\mathrm{P}(\text{хорда длиннее}) = \frac{1}{2}. \qquad \square$$

Задача 2.1.3. Палочку разламывают на две части случайным образом; затем ту часть, что длиннее, вновь разламывают надвое случайным образом. Какова вероятность того, что из трех частей можно образовать треугольник?

Решение. Пусть длина палочки ℓ. Если x — точка первого разлома, то $0 \leqslant x \leqslant \ell$ и точка x равномерно распределена на интервале $(0, \ell)$. Если $x \geqslant \ell/2$, то вторая точка разлома y равномерно распределена на интервале $(0, x)$. См. рис. 2.2. В противном случае точка y равномерно распределена на (x, ℓ).

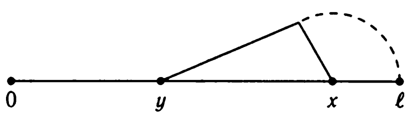


Рис. 2.2

Поэтому

$$\Omega = \{(x, y)\colon 0 \leqslant x,\ y \leqslant \ell;\ y \leqslant x \text{ для } x \geqslant \ell/2 \text{ и } x \leqslant y \leqslant \ell \text{ для } x \leqslant \ell/2\}.$$

Для того чтобы можно было образовать треугольник, точка (x, y) должна лежать в области A, где

$$A = \left\{\max[x, y] > \frac{\ell}{2},\ \min[x, y] < \frac{\ell}{2},\ |x - y| < \frac{\ell}{2}\right\}.$$

В силу симметрии решение дается следующим вычислением:

$$\mathrm{P}(\text{треугольник}) = \frac{2}{\ell}\int\limits_0^{\ell/2} \frac{1}{l-x} \int\limits_{\ell/2}^{\ell/2+x} dy\,dx = \frac{2}{\ell}\int\limits_0^{\ell/2} \frac{x}{l-x}\,dx,$$

что приводит к ответу $2\ln 2 - 1 \approx 0{,}3862944$. □

Задача 2.1.4. Палочку разламывают в двух заранее намеченных совершенно случайным образом по всей ее длине местах. Какова вероятность того, что три полученные части образуют треугольник?

Ответ. $1/4$. Интересно, что это значение меньше, чем в задаче 2.1.3. Одно из отличий состоит в том, что в последней задаче пространство исходов Ω есть квадрат $(0, \ell) \times (0, \ell)$ (см. рис. 2.3), т. е. больше, чем в задаче 2.1.3 (поскольку теперь, возможно, будет разломана на две части более короткая часть, что исключалось в предыдущем случае). С другой стороны, интересующее нас событие отвечает одному и тому же множеству A. Интуитивно это должно привести к меньшей вероятности. И это действительно так, несмотря на то что вероятностная масса распределена в двух задачах по-разному. □

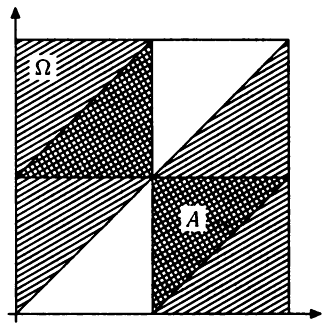


Рис. 2.3

Задача 2.1.5. Монета в один экю — это диск диаметром $4/5$. В традиционной игре с бросанием монеты монету бросают случайным образом на квадратную решетку, образованную линиями, отстоящими друг от друга на расстоянии 1. Если монета покрывает часть линии, вы теряете свой экю. Если нет — вы все равно теряете свой экю, но автомат исполняет государственный гимн по вашему выбору. Какова вероятность того, что вы получите возможность выбрать для себя государственный гимн?

Решение. Не ограничивая общности, предположим, что центр монеты оказался внутри квадрата с вершинами $(0; 0)$, $(0; 1)$, $(1; 0)$, $(1; 1)$. Вы услышите гимн, если центр попал внутрь квадрата S вида

$$\frac{2}{5} \leqslant x \leqslant \frac{3}{5}, \quad \frac{2}{5} \leqslant y \leqslant \frac{3}{5}.$$

Поэтому

$$\mathrm{P}(\text{автомат исполнит гимн}) = \text{площадь квадрата } S = \frac{1}{25}. \qquad \square$$

Есть ряд серьезных вопросов, возникающих сейчас, а обсудим мы их позже. Один из них — так называемая измеримость: есть странные множества $A \subset \Omega$ (даже если Ω — единичный интервал $(0, 1)$), которые не имеют разумно определенного объема (или длины). Вообще, как измерить объем множества A в непрерывном пространстве? Это не обязательно множество, которое достаточно трудно описать (например, канторовский континуум $\mathcal{K}$ имеет корректно определенную длину), но вычисление объема, площади или длины лежит за границами возможностей стандартного интегрирования по Риману. (На самом деле корректно определенная длина множества $\mathcal{K}$ равна нулю.) Чтобы изложить целостную теорию, необходимо прибегнуть к так называемому интегрированию по Лебегу, названному в честь А. А. Лебега (1875—1941), известного французского математика. (Будучи человеком очень простого происхождения, Лебег стал математиком высокого класса, что было необычно для того времени. Он славился безупречными и элегантными докладами и статьями.) В свою очередь, для интегрирования по Лебегу необходимы понятия сигма-алгебры и сигма-аддитивной меры, которые глубоко обобщают понятия длины, площади и объема — частные случаи общей концепции меры. Эти вопросы будут обсуждаться в следующих томах.

Теперь обсудим такой вопрос: а что будет, если распределение массы неравномерно? Этот вопрос имеет не только теоретический интерес. Во многих практических моделях Ω — это неограниченное подмножество евклидова пространства $\mathbb{R}^d$ и объем этого множества бесконечен (например, это все пространство $\mathbb{R}^d$ или $\mathbb{R}_+ = [0, \infty)$ при $d = 1$). Тогда знаменатель $v(\Omega)$ в формуле (2.1.1) бесконечен. Выход из этой ситуации следующий: рассмотрим такую функцию $f \geqslant 0$, что $\int\limits_{\Omega} f(x)\,dx = 1$, и положим

$$\mathrm{P}(A) = \int\limits_A f(x)\,dx, \qquad A \subseteq \Omega$$

(ср. с (2.1.3)). Такая функция f интерпретируется как плотность распределения вероятностей (п. р. в.). Далее в задачах возникают следующие естественные и важные примеры.

Равномерное распределение на интервале (a, b), $a < b$: здесь $\Omega = (a, b)$ и

$$f(x) = \frac{1}{b-a} I(a < x < b). \tag{2.1.4а}$$

Гауссовское, или *нормальное*, *распределение* с $\Omega = \mathbb{R}$ и

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right), \qquad x \in \mathbb{R}. \tag{2.1.4б}$$

Здесь $\mu \in \mathbb{R}$, $\sigma > 0$ — параметры, определяющие распределение.

Графики нормальных п. р. в. на интервале около начала координат и на удалении от него приведены на рис. 2.4, 2.5.

Это известная кривая, о которой великий французский математик А. Пуанкаре (1854—1912) сказал: «Экспериментаторы думают, что это математическая теорема, а математики верят, что это экспериментальный факт».

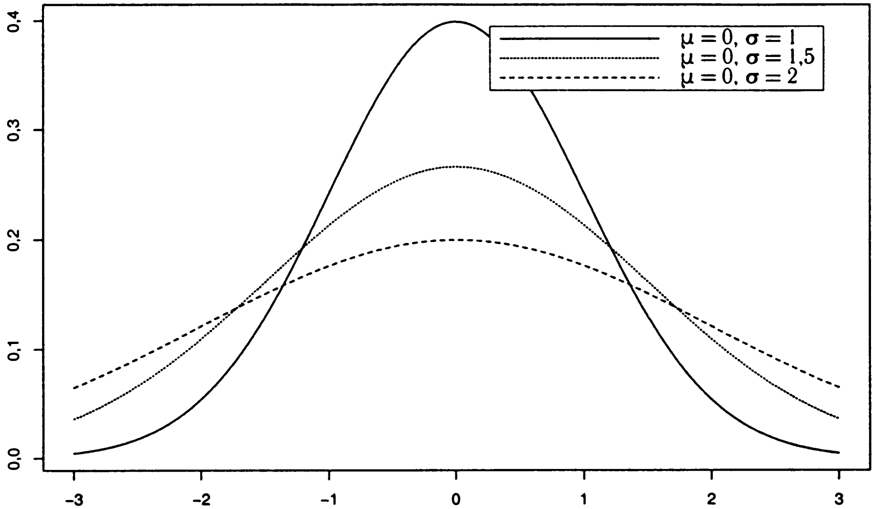


Рис. 2.4. Плотности нормального распределения, I

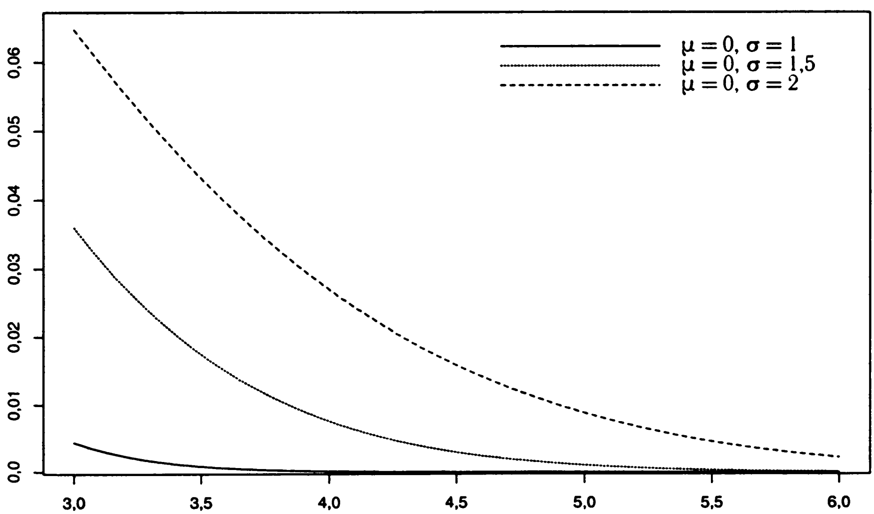


Рис. 2.5. Плотности нормального распределения, II

Экспоненциальное распределение: здесь $\Omega = \mathbb{R}_+$ и

$$f(x) = \lambda e^{-\lambda x} I(x \geqslant 0), \quad x \in \mathbb{R}. \tag{2.1.4в}$$

Параметр $\lambda > 0$ определяет распределение.

Графики экспоненциальных п. р. в. показаны на рис. 2.6.

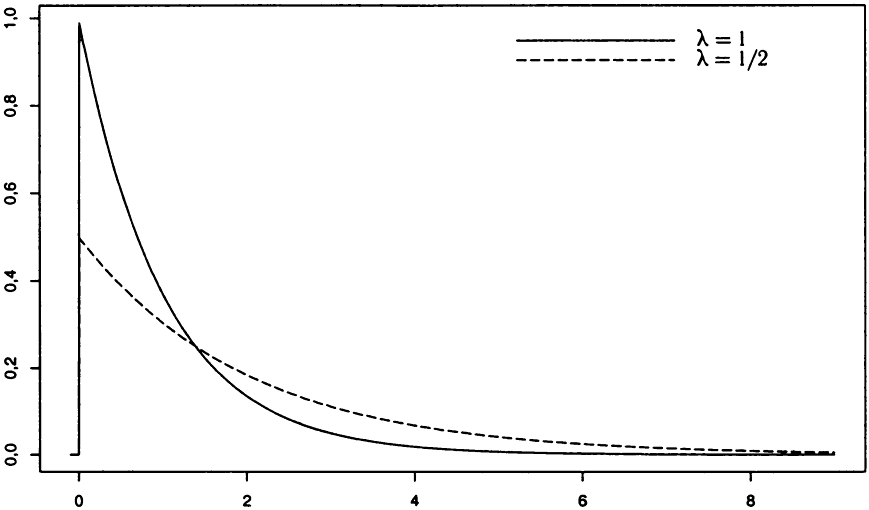


Рис. 2.6. Плотности экспоненциального распределения

Обобщение функции (2.1.4в) — это *гамма-распределение*. Здесь снова $\Omega = \mathbb{R}_+$ и п. р. в. равна

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} I(x \geqslant 0) \tag{2.1.4г}$$

с параметрами α, $\lambda > 0$. Напомним, что $\Gamma(\alpha) = \int\limits_0^\infty x^{\alpha-1} e^{-x}\,dx$ (гамма-функция), здесь это нормирующая постоянная. (Напомним также, что для положительного аргумента $\Gamma(n) = (n-1)!$; вообще $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$ для $\alpha > 1$.) Гамма-распределение играет выдающуюся роль в статистике и будет часто возникать в последующих главах. Графики гамма-п. р. в. показаны на рис. 2.7.

Другой пример — это *распределение Коши*, для которого $\Omega = \mathbb{R}$ и

$$f(x) = \frac{1}{\pi} \cdot \frac{\tau}{\tau^2 + (x - \alpha)^2}, \quad x \in \mathbb{R}, \tag{2.1.4д}$$

с параметрами $\alpha \in \mathbb{R}$ и $\tau > 0$. Интересно, что распределение Коши открыл Пуассон в 1824 г., когда построил контрпример к центральной предельной теореме. Рассказ об этом будет далее. Графики п. р. в. Коши приведены на рис. 2.8.

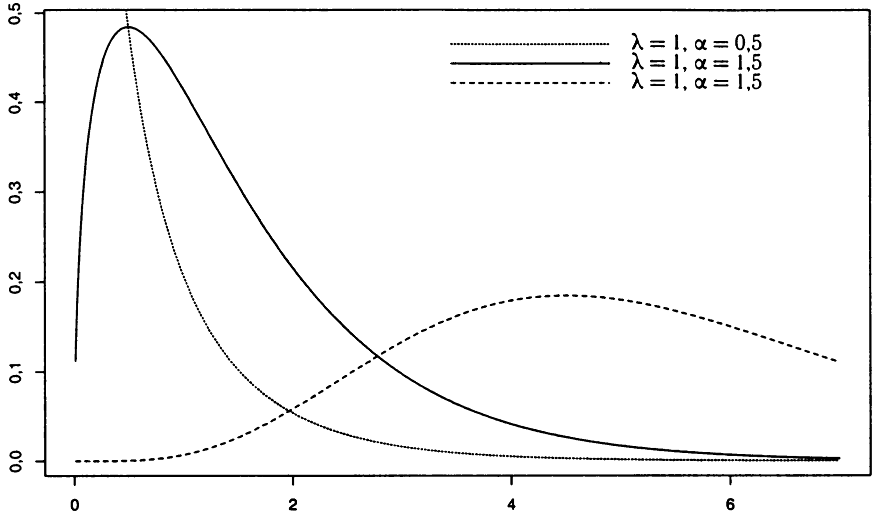


Рис. 2.7. Плотности гамма-распределения

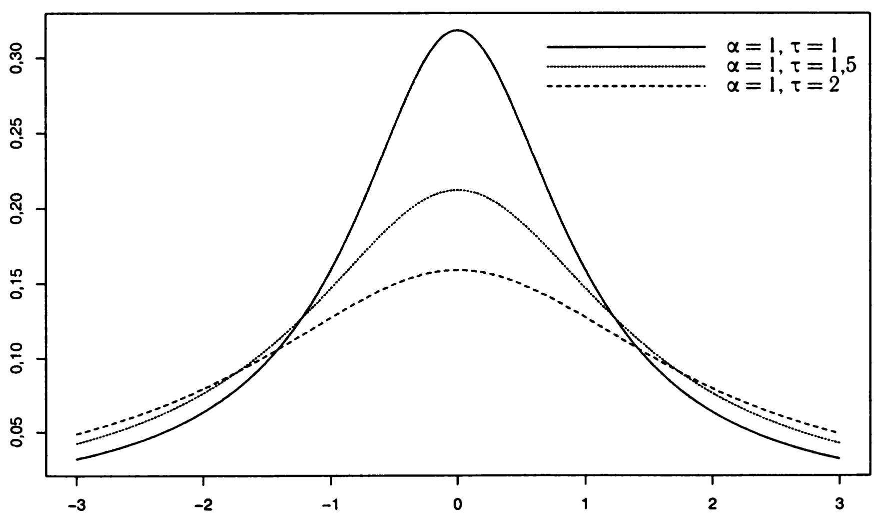


Рис. 2.8. Плотности распределения Коши

Огюстен Коши был стойким роялистом и ревностным католиком и, в отличие от многих выдающихся французских ученых того периода, имел сложные отношения с республиканским режимом. В 1830 г., во время одной из французских революций XIX столетия, он добровольно отправился в ссылку в Турин и Прагу, где давал частные уроки математики наследникам королевской семьи Бурбонов. Его избрание во Французскую академию состоялось лишь в 1838 г., когда он вернулся в Париж.

Гауссовское распределение будет подробно изучаться далее. На данном этапе мы только приведем его обобщение на многомерный случай, когда $\Omega = \mathbb{R}^d$ и

$$f(\boldsymbol{x}) = \frac{1}{(\sqrt{2\pi})^d (\det\Sigma)^{1/2}} \exp\Big(-\frac{1}{2}\langle \boldsymbol{x} - \boldsymbol{\mu},\ \Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\rangle\Big). \tag{2.1.5}$$

Здесь $\boldsymbol{x}$ и $\boldsymbol{\mu}$ — действительные d-мерные векторы:

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_d \end{pmatrix} \in \mathbb{R}^d,$$

а Σ — обратимая положительно определенная $(d \times d)$-действительная матрица с определителем $\Sigma > 0$ и обратной матрицей $\Sigma^{-1} = (\Sigma_{ij}^{-1})$. (Матрица Σ называется *положительно определенной*, если ее можно представить в виде произведения $\Sigma = AA^*$, и строго положительно определенной, если в этом представлении матрица A *обратима*, т. е. существует *обратная матрица* A^{-1} (в этом случае обратная матрица $\Sigma^{-1} = A^{*-1}A^{-1}$ также существует). Легко видеть, что положительно определенная матрица Σ всегда симметрична (эрмитова), т. е. $\Sigma^* = \Sigma$. Значит, положительно определенная матрица имеет ортонормированный базис из собственных векторов, а ее собственные значения неотрицательны (положительны, если она строго положительно определена).) Далее, $\langle\ ,\ \rangle$ означает евклидово скалярное произведение в $\mathbb{R}^d$:

$$\langle \boldsymbol{x} - \boldsymbol{\mu},\ \Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\rangle = \sum_{i,j=1}^{d} (x_i - \mu_i)\Sigma_{ij}^{-1}(x_j - \mu_j).$$

Плотность распределения вероятностей указанного вида называется *многомерным нормальным* или *гауссовским* распределением.

Как и прежде, возникает вопрос: функции f какого вида могут быть п. р. в.? (Типичный пример — функция $f(x) = I(x \in (0,\ 1) \setminus \mathcal{K})$, где $\mathcal{K} \subset (0,\ 1)$ — канторово множество. Здесь $f \geqslant 0$ по определению, но как определить $\int_0^1 f(x)\,dx$?) И снова ответ можно найти в теории интегрирования по Лебегу. К счастью, в «реалистичных» моделях такие проблемы возникают редко и остаются в тени других более практических вопросов.

Итак, с этого момента и до конца главы наша базовая модель такова, что исходы ω пробегают «дозволенные» подмножества $A \subseteq \Omega$ (такие подмножества называются измеримыми и будут введены позже). Очень часто в качестве Ω будет выступать $\mathbb{R}^d$. Вероятность $\mathrm{P}(A)$ можно вычислить для любого такого множества A (называемого *событием* из Ω) как

$$\mathrm{P}(A) = \int\limits_A f(\boldsymbol{x})\, d\boldsymbol{x}. \qquad (2.1.6)$$

Здесь f — заданная п. р. в., $f \geqslant 0$ и $\int\limits_\Omega f(\boldsymbol{x})\, d\boldsymbol{x} = 1$.

Как и в дискретном случае, имеется интуитивно приемлемое свойство аддитивности: если $A_1, A_2, \ldots$ — конечная или счетная последовательность попарно не пересекающихся событий, то

$$\mathrm{P}\Big(\bigcup_j A_j\Big) = \sum_j \mathrm{P}(A_j), \qquad (2.1.7)$$

тогда как в общем случае $\mathrm{P}\Big(\bigcup\limits_j A_j\Big) \leqslant \sum\limits_j \mathrm{P}(A_j)$. Поскольку $\mathrm{P}(\Omega) = 1$, получаем, что для дополнения $A^c = \Omega \setminus A$ выполняется равенство $\mathrm{P}(A^c) = 1 - \mathrm{P}(A)$, а для теоретико-множественной разности $A \setminus B$ мы имеем $\mathrm{P}(A \setminus B) = \mathrm{P}(A) - \mathrm{P}(A \cap B)$. Конечно, имеют место и более глубокие факты, упомянутые в дискретной теории, например формула включения-исключения.

Неудивительно, что понятие *случайной величины* возникает в новой постановке из своего дискретного аналога: случайная величина — это функция

$$X\colon \omega \in \Omega \mapsto X(\omega)$$

с действительными или комплексными значениями $X(\omega)$ (в комплексном случае вновь рассматривается пара действительных случайных величин, представляющих действительную и мнимую части). Действительная случайная величина должна иметь то свойство, что $\forall x \in \mathbb{R}$ множество $\{\omega \in \Omega\colon X(\omega) < x\}$ — это событие из Ω, которому нужно приписать вероятность $\mathrm{P}(X < x)$. Тогда с каждой действительнозначной с. в. ассоциируется ее (*кумулятивная*) *функция распределения* (ф. р.)

$$x \in \mathbb{R} \mapsto F_X(x) = \mathrm{P}(X < x), \qquad (2.1.8)$$

которая монотонно меняется от 0 до 1, когда x возрастает от $-\infty$ до ∞ (см. рис. 2.9). Величина

$$\bar{F}_X(x) = 1 - F_X(x) = \mathrm{P}(X \geqslant x), \qquad (2.1.9)$$

которая описывает *хвосты распределения*, также используется довольно часто.

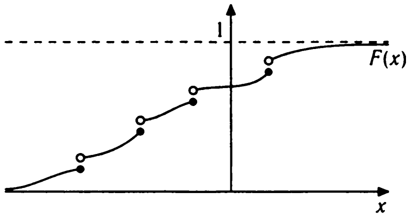


Рис. 2.9

Заметим, что определение (2.1.8) ведет к ф. р. $F_X(x)$, непрерывной слева (на графике она представлена с помощью черных точек). Это означает, что $F_X(x_n) \nearrow F_X(x)$ при $x_n \nearrow x$. С другой стороны, для любого x предел справа $\lim\limits_{x_n \searrow x} F_X(x_n)$ также существует, но равенства этого предела значению $F_X(x)$ не гарантируется (это отмечено кружочками на графике). Конечно, и хвост распределения $\bar{F}_X(x)$ непрерывен слева.

Однако если принять определение $F_X(x) = \mathbf{P}(X \leqslant x)$ (в некоторых учебниках так и сделано), то функция F_X становится непрерывной справа и то же верно для $\bar{F}_X(x)$.

Пример. Если $X \equiv b$ — это действительная постоянная, то ф. р. F_X — это функция Хевисайда

$$F_X(y) = I(y > b). \tag{2.1.10а}$$

Если $X = I_A$ (т. е. индикатор события A), то $\mathbf{P}(X < y)$ равно 0 для $y \leqslant 0$, равно $1 - \mathbf{P}(A)$ для $0 < y \leqslant 1$ и равно 1 для $y > 1$. Обобщим: если X принимает дискретное множество действительных значений, т. е. конечное или счетное множество, без точек накопления на $\mathbb{R}$, пусть это $y_j \in \mathbb{R}$, $y_j < y_{j+1}$, то функция $F_X(y)$ постоянна на каждом полуинтервале $y_j < x \leqslant y_{j+1}$ и имеет скачки в точках y_j размера $\mathbf{P}(X = y_j)$. □

Заметим, что все обсуждавшиеся ранее примеры с. в. на дискретных пространствах можно рассмотреть в рамках нынешнего подхода. Например, если $X \sim \mathrm{Bin}(n, p)$, то

$$F_X(y) = \sum_{0 \leqslant m < y, m \leqslant n} C_n^m p^m (1-p)^{n-m} I(y > 0). \tag{2.1.10б}$$

Если $X \sim \mathrm{Po}(\lambda)$, то

$$F_X(y) = e^{-\lambda} \sum_{0 \leqslant n < y} \frac{\lambda^n}{n!} I(y > 0). \tag{2.1.10в}$$

На рис. 2.10 приводятся графики функций $F_X \sim \mathrm{Po}(\lambda)$.

Если $X \sim \mathrm{Geom}(q)$, то

$$F_X(y) = I(y > 0)(1-q) \sum_{0 \leqslant n < y} q^n. \tag{2.1.10г}$$

График ф. р. для с. в. $X \sim \mathrm{Geom}(q)$ приведен на рис. 2.11 вместе с графиком ф. р. пуассоновской с. в. с $\lambda = 1$ (обе с. в. имеют одно и то же среднее 1).

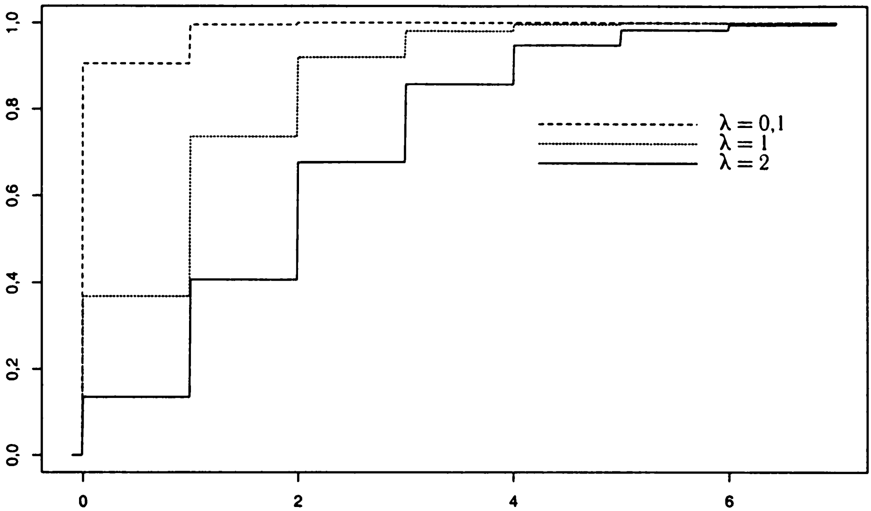


Рис. 2.10. Функции распределения распределения Пуассона

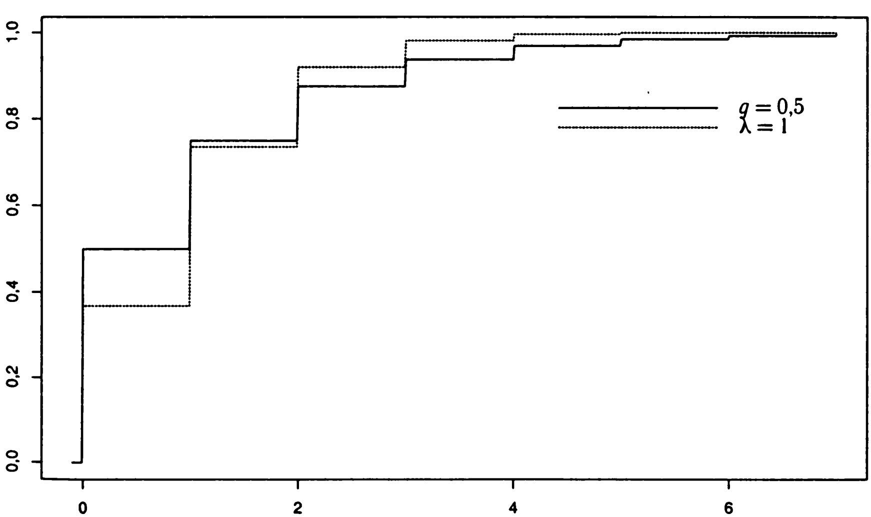


Рис. 2.11. Функции распределения геометрического распределения и распределения Пуассона

Говорят, что случайная величина X имеет равномерное, гауссовское, показательное, гамма-распределение или распределение Коши (с соответствующими параметрами), если ф. р. $F_X(y)$ выражается формулой $\mathsf{P}(X < y) = \int f(x) I(x < y)\, dx$ через соответствующую п. р. в. Например, для равномерного распределения

$$F_X(y) = \begin{cases} 0, & y \leqslant a, \\ (y-a)/(b-a), & a < y < b, \\ 1, & y \geqslant b, \end{cases} \tag{2.1.11а}$$

для гауссовского распределения

$$F_X(y) = \frac{1}{\sqrt{2\pi}\sigma} \int\limits_{-\infty}^{y} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx = \Phi\left(\frac{y-\mu}{\sigma}\right), \quad y \in \mathbb{R}, \tag{2.1.11б}$$

для показательного распределения

$$F_X(y) = \begin{cases} 0, & y \leqslant 0, \\ 1 - e^{-\lambda y}, & y > 0, \end{cases} \tag{2.1.11в}$$

для гамма-распределения

$$F_X(y) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int\limits_{0}^{y} x^{\alpha-1} e^{-\lambda x}\, dx \cdot I(y > 0), \tag{2.1.11г}$$

а для распределения Коши

$$F_X(y) = \frac{1}{\pi}\left(\operatorname{arctg}\left(\frac{y-\alpha}{\tau}\right) + \frac{\pi}{2}\right), \quad y \in \mathbb{R}. \tag{2.1.11д}$$

Соответствующие обозначения таковы: $X \sim \mathrm{U}(a, b)$, $X \sim \mathrm{N}(\mu, \sigma^2)$, $X \sim \sim \mathrm{Exp}(\lambda)$, $X \sim \mathrm{Gam}(\alpha, \lambda)$ и $X \sim \mathrm{Ca}(\alpha, \tau)$.

Ниже приведены некоторые графики этих ф. р. Графики ф. р. нормального распределения приведены на рис. 2.12.

Графики ф. р. показательного распределения показаны на рис. 2.13.

Далее, графики ф. р. гамма-распределения показаны на рис. 2.14.

Наконец, графики ф. р. для распределения Коши показаны на рис. 2.15.

В общем случае говорят, что X имеет п. р. в. f (и пишут $X \sim f$), если $\forall y \in \mathbb{R}$ имеет место равенство

$$\mathsf{P}(X < y) = \int\limits_{-\infty}^{y} f(x)\, dx. \tag{2.1.12}$$

Тогда $\forall a, b \in \mathbb{R}$, $a < b$, выполняется равенство

$$\mathsf{P}(a < X < b) = \int\limits_{a}^{b} f(x)\, dx, \tag{2.1.13}$$

и вообще, для любого измеримого множества $A \subset \mathbb{R}$ справедливо равенство $\mathsf{P}(X \in A) = \int\limits_A f(x)\,dx$.

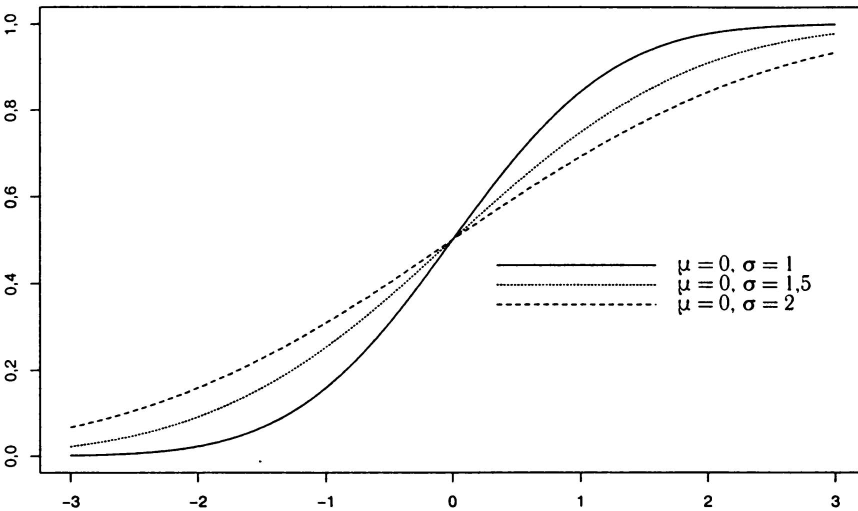


Рис. 2.12. Функции распределения нормального распределения

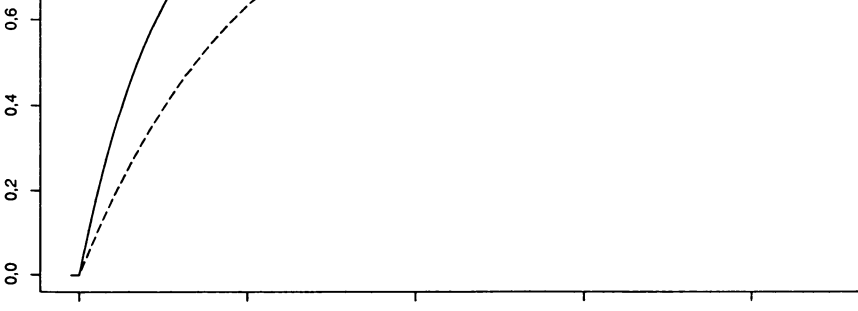


Рис. 2.13. Функции распределения показательного распределения

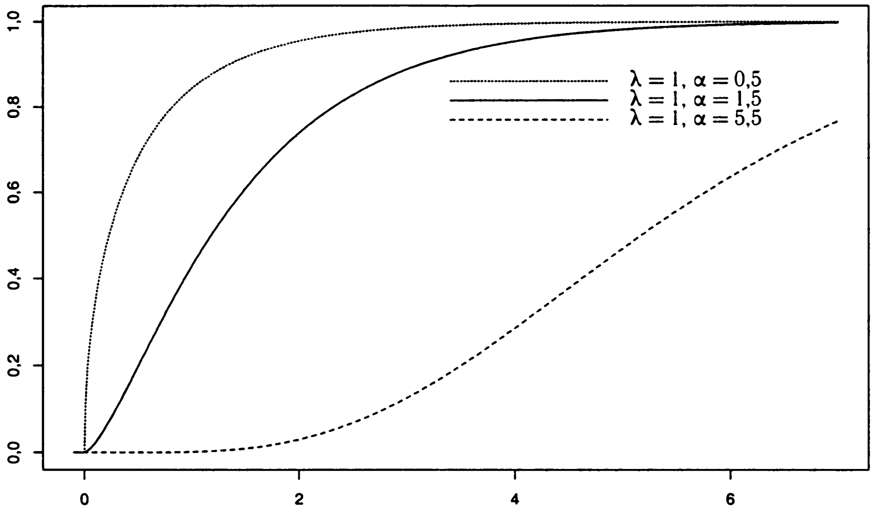


Рис. 2.14. Функции распределения гамма-распределения

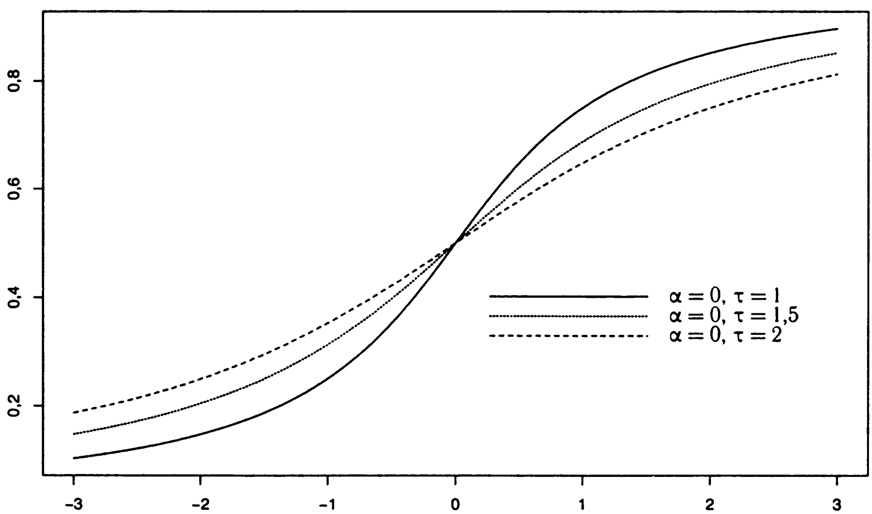


Рис. 2.15. Функции распределения распределения Коши

Заметим, что во всех вычислениях с использованием п. р. в. можно пренебречь такими множествами C, что $\int\limits_C dx = 0$ (множествами меры 0).

Поэтому вероятности $\mathsf{P}(a \leqslant X \leqslant b)$ и $\mathsf{P}(a < X < b)$ совпадают. (Это, вообще говоря, неверно для дискретных с. в.)

Медианой $m(X)$ с. в. X называется такое значение, которое «делит» область значений X на две части одинаковой массы. В терминах ф. р. и п. р. в. это означает следующее:

$$m(X) = \max\left[y\colon \bar{F}_X(y) \geqslant \frac{1}{2}\right] = \max\left[y\colon \int_y^{\infty} f_X(x)\,dx \geqslant \frac{1}{2}\right]. \qquad (2.1.14)$$

Если функция F_X строго монотонна и непрерывна, то, очевидно, медиана $m(X)$ равна единственному значению y, для которого $F_X(y) = 1/2$. Иными словами, $m(X)$ — это единственное значение y, для которого

$$\int_{-\infty}^{y} f_X(x)\,dx = \int_y^{\infty} f_X(x)\,dx.$$

Модой с. в. X с ограниченной п. р. в. f_X называется такое значение x, в котором функция f_X достигает своего максимума; иногда локальные максимумы называют локальными модами.

Задача 2.1.6. Мы с моей мамой планируем принять участие в телевизионном национальном тестировании коэффициента IQ (IQ — intelligence quotient — коэффициент умственного развития), где нужно ответить на большое число вопросов, вместе с отобранными профессорами математики, модельными парикмахерами, музыкантами духовых оркестров и другими представителями различных слоев общества (не забудем и знаменитых личностей, конечно). Коэффициент IQ по-разному определяется для разных возрастных групп. Для меня он равен $-80\ln(1-x)$, где x — доля правильных ответов, x — любое число между 0 и 1. Для мамы коэффициент IQ задается формулой $-70\ln\dfrac{3/4-y}{3/4} = -70\ln(3/4-y) + 70\ln 3/4$, где y — ее доля правильных ответов (в ее возрастной группе не предполагается, что y превзойдет $3/4$).

Каждый из нас хочет получить значение IQ, равное как минимум 110. Какова вероятность того, что у нас это получится? Какова вероятность того, что мой IQ окажется выше, чем мамин?

Решение. Мы вновь используем предположение о равномерности распределения. Исход $\omega = (x_1, x_2)$ предполагается равномерно распределенным на множестве Ω, которое является прямоугольником со сторонами $(0, 1) \times (0, 3/4)$ площади $3/4$. Имеется пара с. в.:

$$X(\omega) = -80\ln(1-x_1) \quad \text{для моего IQ},$$
$$Y(\omega) = -70\ln\frac{3/4-x_2}{3/4} \quad \text{для маминого IQ},$$

и $\forall y > 0$ выполняется соотношение

$$\mathsf{P}(X < y) = \frac{1}{3/4}\int\limits_0^1\int\limits_0^{3/4} I\left(-\ln(1-x_1) < \frac{1}{80}y\right)dx_2\,dx_1 =$$

$$= \int\limits_0^{1-e^{-y/80}} dx_1 = 1 - e^{-y/80}, \quad \text{т. е. } X \sim \mathrm{Exp}(1/80),$$

$$\mathsf{P}(Y < y) = \frac{1}{3/4}\int\limits_0^1\int\limits_0^{3/4} I\left(-\ln\frac{3/4 - x_2}{3/4} < \frac{1}{70}y\right)dx_2\,dx_1 =$$

$$= \frac{1}{3/4}\int\limits_0^{3(1-e^{-y/70})/4} dx_2 = 1 - e^{-y/70}, \quad \text{т. е. } Y \sim \mathrm{Exp}(1/70).$$

Далее, $\mathsf{P}(\min[X, Y] < y) = 1 - \mathsf{P}(\min[X, Y] \geqslant y)$, и

$$\mathsf{P}(\min[X, Y] \geqslant y) = \frac{1}{3/4}\int\limits_0^1\int\limits_0^{3/4} I\left(-\ln(1-x_1) \geqslant \frac{y}{80}, -\ln\frac{3/4-x_2}{3/4} \geqslant \frac{y}{70}\right)dx_2\,dx_1 =$$

$$= e^{-y/80}e^{-y/70} = e^{-3y/112}, \quad \text{т. е. } \min[X, Y] \sim \mathrm{Exp}(3/112). \qquad (2.1.15)$$

Поэтому

$$\mathsf{P}(\text{каждый получит не менее } 110) = e^{-3\cdot 110/112} \approx e^{-3}$$

(довольно маленькая вероятность). Чтобы увеличить эту вероятность, нужно усердно потрудиться и изменить исходное равномерное распределение на другое, смещенное к бо́льшим значениям x_1 и x_2, которые равны долям правильных ответов.

Для того чтобы вычислить $\mathsf{P}(X > Y)$, рекомендуется использовать якобиан $\dfrac{\partial(x_1, x_2)}{\partial(u_1, u_2)}$ «обратной» замены переменных $x_1 = 1 - e^{-u_1/80}$, $x_2 = 3(1 - e^{-u_2/70})/4$ («прямая» замена — это $u_1 = -80\ln(1 - x_1)$, $u_2 = -70 \times \ln\dfrac{3/4 - x_2}{3/4}$). Действительно,

$$\frac{\partial(x_1, x_2)}{\partial(u_1, u_2)} = \frac{3}{4}\frac{1}{80}e^{-u_1/80}\frac{1}{70}e^{-u_2/70}, \qquad u_1, u_2 > 0,$$

и

$$\mathsf{P}(X > Y) = \frac{1}{3/4}\int\limits_0^1\int\limits_0^{3/4} I\left(-80\ln(1-x_1) > -70\ln\frac{3/4 - x_2}{3/4}\right)dx_2\,dx_1 =$$

$$= \frac{1}{3/4}\int\limits_0^\infty\int\limits_0^\infty \frac{3}{4}\frac{1}{80}e^{-u_1/80}\frac{1}{70}e^{-u_2/70} I(u_1 > u_2)\,du_2\,du_1 =$$

$$= \int\limits_0^\infty \frac{1}{70}e^{-u_2/70}\int\limits_{u_2}^\infty \frac{1}{80}e^{-u_1/80}\,du_1\,du_2 = \frac{8}{15}. \quad \square$$

В предыдущих примерах ф. р. F или имела вид $F(y) = \int\limits_{-\infty}^{y} f(x)\,dx$, $y \in \mathbb{R}$, или была кусочно постоянной, с положительными скачками в точках дискретного множества $\mathbb{X} \subset \mathbb{R}$. В первом случае говорят, что соответствующие случайные величины имеют абсолютно непрерывные распределения с функцией плотности f, а во втором — что они имеют дискретное распределение, сосредоточенное на $\mathbb{X}$. Нетрудно проверить, что абсолютная непрерывность влечет непрерывность функции F, но не наоборот. Для дискретного распределения функция распределения кусочно постоянна, т. е., без сомнения, разрывна. Возможна, однако, комбинация этих двух типов распределений (см. рис. 2.16).

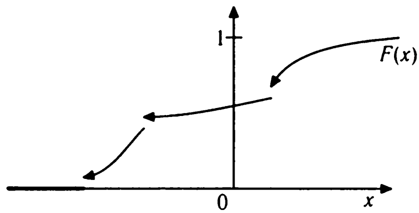


Рис. 2.16

Более того, существуют функции распределения, не принадлежащие ни к одному из указанных типов, но мы не будем обсуждать их здесь (основной пример — *канторовская лестница*, она непрерывна, но возрастает от 0 до 1 на множестве $\mathcal{K}$ нулевой меры, см. рис. 2.17).

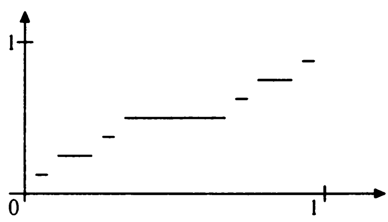


Рис. 2.17

Возвращаясь к случайным величинам, отметим, что во многих случаях не имеет значения, на каком вероятностном пространстве Ω задана случайная величина $X(\omega)$. Например, нормальные с. в. появляются во многих статистических моделях, но при этом важно только, являются ли они совместно гауссовскими или нет. Точно так же, показательное распределение возникает во многих моделях, например, оно описывает время жизни индивидуума или время между последовательными изменениями состояния системы, появляется и в чисто геометрическом контексте. Необходимо уметь работать с такими случайными величинами безотносительно конкретного вероятностного пространства Ω.

С другой стороны, стандартный способ представить действительную с. в. X с заданной плотностью $f(x)$, $x \in \mathbb{R}$, таков. Вы выбираете Ω как *носитель плотности* f, т. е. множество $\{x \in \mathbb{R}: f(x) > 0\}$, определяете вероятность $\mathrm{P}(A)$ как $\int\limits_A f(x)\,dx$ (см. формулу (2.1.6)) и полагаете $X(\omega) = \omega$ (или, если желаете, $X(x) = x$, $x \in \Omega$). Действительно, тогда событие $\{X < y\}$ совпадает с множеством $\{x \in \mathbb{R}: f(x) > 0, x < y\}$, а его вероятность равна

$$\mathrm{P}(X < y) = \int f(x) I(x < y)\,dx = \int\limits_{-\infty}^{y} f(x)\,dx.$$

В последней части решения задачи 2.1.6 мы в точности это и выполнили: замена переменных $u_1 = -80\ln(1 - x_1)$, $u_2 = -70\ln(3/4 - x_2)/(3/4)$ с обратным якобианом

$$\frac{3}{4} \cdot \frac{1}{80} e^{-u_1/80} \cdot \frac{1}{70} e^{-u_2/70}$$

приводит к полупрямым $u_1 > 0$ и $u_2 > 0$ с плотностями $e^{-u_1/80}/80$ и $e^{-u_1/70}/70$ и множителем $I(u_1 > u_2)$, определяющим событие.

Чтобы ясно представлять себе равномерную с. в. на интервале (a, b), рассмотрим модель с $\Omega = (a, b)$ и определим функцию f формулой (2.1.4а); для показательного или гамма-распределения $\Omega = \mathbb{R}_+$ (положительная полуось) и f задается формулами (2.1.4в) и (2.1.4г) соответственно, а для гауссовского распределения или распределения Коши $\Omega = \mathbb{R}$ (вся прямая), а f определяется формулами (2.1.4б) и (2.1.4д) соответственно. Во всех случаях стандартное уравнение $X(x) = x$ определяет с. в. X с соответствующим распределением.

Такое представление с. в. X с заданной плотностью/функцией распределения особенно полезно, когда приходится иметь дело с функцией $Y = g(X)$. См., например, соотношение (2.1.17).

До сих пор рассматривались два типа с. в.: или а) с дискретным множеством значений (конечным или счетным), или б) с плотностью (на множестве $\mathbb{R}$ или $\mathbb{C}$). Эти два типа не исчерпывают всех ситуаций, которые могут возникнуть. В частности, в ряде приложений требуется рассмотреть с. в. X, которая является «смесью» двух вышеупомянутых типов, т. е. положительная часть вероятностной массы сосредоточена в точке или точках, а другая часть распределена с некоторой плотностью на интервале из $\mathbb{R}$. Тогда соответствующая ф. р. F_X имеет в точках x_j, для которых $\mathrm{P}(X = x_j) > 0$, скачки размера $\mathrm{P}(X = x_j)$ и абсолютно непрерывна вне этих точек. Типичный пример — функция распределения F_W времени ожидания W в очереди со случайными временами поступления требований и случайными временами обслуживания (популярная интерпретация — парикмахерская, где посетитель ждет, пока парикмахер обслужит предыдущего клиента). Вы можете удачно выбрать момент прихода, когда очередь в парикмахерской отсутствует, тогда время ожидания равно нулю, а вероятность такого события хоть и мала, но положительна. В противном случае нужно подождать положительное время; при простейших предположениях о временах поступлений требований и временах обслуживания функция распределения такова:

$$F_W(y) = \mathrm{P}(W < y) = \begin{cases} 0, & y \leqslant 0, \\ 1 - \dfrac{\lambda}{\mu} e^{-(\mu-\lambda)y}, & y > 0. \end{cases} \tag{2.1.16}$$

Здесь $\lambda, \mu > 0$ — интенсивности двух показательных распределений: λ — интенсивность поступления требования и μ — интенсивность обслуживания. Формула (2.1.16) имеет смысл, если $\lambda < \mu$, т. е. скорость обслуживания превышает скорость поступления заявок; тогда есть гарантия, что очередь не переполнится со временем. Вероятность $\mathrm{P}(W = 0)$ того, что вам не придется ждать, равна $1 - \lambda/\mu > 0$, а вероятность того, что время ожидания превысит y, равна $\dfrac{\lambda}{\mu} e^{-(\mu-\lambda)y}$.

В этом примере распределение величины W имеет дискретную компоненту (сосредоточенную в точке 0) и абсолютно непрерывную компоненту (сосредоточенную на $(0, \infty)$).

Очень часто нужно найти плотность и функцию распределения случайной величины Y, которая является функцией $g(X)$ другой случайной величины X с заданной плотностью или функцией распределения.

Задача 2.1.7. Площадь круга имеет показательное распределение с параметром λ. Найдите плотность распределения радиуса круга.

Ответ. Искомая плотность имеет вид $f_R(y) = 2\pi\lambda y e^{-\lambda\pi y^2} I(y > 0)$. □

Задача 2.1.8. Радиус круга показательно распределен с параметром λ. Найдите плотность распределения площади этого круга.

Ответ. Искомая плотность имеет вид $g(s) = \dfrac{\lambda}{\sqrt{4\pi s}} e^{-\lambda\sqrt{s/\pi}}$. □

В этих двух задачах присутствуют два взаимно однозначных отображения — квадрат и квадратный корень. Для функции g общего вида ответ возникает в результате непосредственных вычислений при помощи обратного якобиана. Точнее, если $Y = g(X)$, то прямая замена переменных имеет вид $y = g(x)$ и

$$f_Y(y) = I(y \in \mathrm{Range}(g)) \sum_{x:\ g(x)=y} f_X(x) \frac{1}{|g'(x)|}. \tag{2.1.17}$$

Здесь $\mathrm{Range}(g)$ — это множество $\{y\colon\ y = g(x)$ для некоторого $x \in \mathbb{R}\}$ и предполагается, что прообраз точки y — дискретное множество, которое позволяет выполнить суммирование.

Уравнение (2.1.17) имеет место, если его правая часть — это корректно определенная плотность распределения (это значит, что $g'(x) = 0$ лишь на «тонком» множестве значений x).

Задача 2.1.9. Случайная величина X равномерно распределена на отрезке $[0,\ 1]$. Найдите функцию и плотность распределения с. в. Y, где

$$Y = \frac{a + bX}{c - dX},$$

$a,\ b,\ c,\ d > 0$ и $d \leqslant c$.

Решение. Используя формулу (2.1.17), немедленно получаем, что плотность распределения случайной величины Y равна

$$f_Y(y) = \frac{bc + ad}{(b + dy)^2}, \qquad y \in \left(\frac{a}{c},\ \frac{a+b}{c-d}\right). \qquad \square$$

Пример. Если $b,\ c \in \mathbb{R}$ — постоянные, $c \neq 0$, то

$$f_{X+b}(y) = f_X(y - b) \quad \text{и} \quad f_{cX}(y) = \frac{1}{|c|} f_X(y/c).$$

Комбинируя эти две формулы, легко видеть, что нормальное распределение и распределение Коши имеют следующие масштабные свойства: если $X \sim \mathrm{N}(\mu,\ \sigma^2)$, то $\frac{1}{\sigma}(X - \mu) \sim \mathrm{N}(0,\ 1)$, и если $X \sim \mathrm{Ca}(\alpha,\ \tau)$, то $\frac{1}{\tau}(X - \alpha) \sim \mathrm{Ca}(1,\ 0)$. Также если $X \sim \mathrm{N}(\mu,\ \sigma^2)$, то $cX + b \sim \mathrm{N}(c\mu + b,\ c^2\sigma^2)$, и если $X \sim \mathrm{Ca}(\alpha,\ \tau)$, то $cX + b \sim \mathrm{Ca}(c\alpha + b,\ c\tau)$. $\square$

Формула, возникающая в задаче 2.1.7, имеет вид

$$f_{\sqrt{X}}(y) = 2\sqrt{y}\, f_X(y^2) I(y > 0)$$

(в предположении, что с. в. X принимает неотрицательные значения). Аналогично в задаче 2.1.8 мы имеем

$$f_{X^2}(y) = \frac{1}{2\sqrt{y}} (f_X(\sqrt{y}) + f_X(-\sqrt{y})) I(y > 0)$$

(что равно $1/(2\sqrt{y}) f_X(\sqrt{y}) I(y > 0)$, если случайная величина X неотрицательна).

Предполагая, что g — взаимно однозначное отображение, по крайней мере на области значений с. в. X, можно упростить формулу (2.1.17), поскольку суммирование сводится к одной точке $x(y) = g^{-1}(y)$:

$$f_Y(y) = f_X(x(y))\frac{1}{|g'(x(y))|}I(y \in \mathrm{Range}(g)) = f_X(x(y))\,|x'(y)|\,I(y \in \mathrm{Range}(g)). \tag{2.1.18}$$

Рассматривая пару X, Y из общей совокупности случайных величин $X_1, \ldots, X_n$, удобно использовать понятие *совместной плотности распределения* и *совместной функции распределения* (точно так же, как мы использовали совместные вероятности в дискретном случае). Говоря формальным языком, для действительных с. в. $X_1, \ldots, X_n$ мы полагаем

$$F_{\boldsymbol{X}}(y_1, \ldots, y_n) = \mathrm{P}(X_1 < y_1, \ldots, X_n < y_n) \tag{2.1.19}$$

и говорим, что они имеют совместную плотность распределения f, если $\forall y_1, \ldots, y_n \in \mathbb{R}$ выполняется равенство

$$F_{\boldsymbol{X}}(y_1, \ldots, y_n) = \int\limits_{-\infty}^{y_1} \ldots \int\limits_{-\infty}^{y_n} f(\boldsymbol{x})\,d\boldsymbol{x}. \tag{2.1.20}$$

Здесь $\boldsymbol{X} = (X_1, \ldots, X_n)$ — случайный вектор, образованный случайными величинами $X_1, \ldots, X_n$, $\boldsymbol{x} = (x_1, \ldots, x_n)$ — точка из $\mathbb{R}^n$ и $d\boldsymbol{x} = dx_1 \ldots dx_n$ — элемент евклидова объема. Тогда для заданной совокупности полуинтервалов $[a_j, b_j)$, $j = 1, \ldots, n$, мы имеем

$$\mathrm{P}(a_1 < X_1 < b_1, \ldots, a_n < X_n < b_n) = \int\limits_{a_1}^{b_1} \ldots \int\limits_{a_n}^{b_n} f(\boldsymbol{x})\,d\boldsymbol{x},$$

а на самом деле $\mathrm{P}(\boldsymbol{X} \in A) = \int\limits_A f(\boldsymbol{x})\,d\boldsymbol{x}$ для любого измеримого подмножества $A \in \mathbb{R}^n$.

В части В этой книги, изучая курс «Статистики IB», мы будем постоянно использовать обозначение $f_{\boldsymbol{X}}$ для совместной плотности распределения с. в. $X_1, \ldots, X_n$, образующих вектор $\boldsymbol{X}$; в случае двух с. в. X, Y пишем $f_{X,Y}(x, y)$, обозначая совместную плотность распределения, и $F_{X,Y}(x, y)$, обозначая совместную функцию распределения, x, $y \in \mathbb{R}$. Удобная формула для $f_{X,Y}$ — это

$$f_{X,Y}(x, y) = \frac{\partial^2}{\partial x\,\partial y} F_{X,Y}(x, y). \tag{2.1.21}$$

Как и ранее, совместная п. р. в. $f(x, y)$ должна только быть неотрицательной и иметь интеграл $\int\limits_{\mathbb{R}^2} f(x, y)\,dy\,dx = 1$; она может быть неограниченной и разрывной. То же верно для $f_{\boldsymbol{X}}(\boldsymbol{x})$. Мы пишем $(X, Y) \sim f$, если $f = f_{X,Y}$.

В задаче 2.1.6, приведенной выше, совместная функция распределения двух IQ равна

$$F_{X,Y}(x, y) = v\Big(\Big\{(x_1, x_2)\colon 0 \leqslant x_1, x_2 \leqslant 1,\\ -80\ln(1 - x_1) < x,\ -70\ln\frac{3/4 - x_2}{3/4} < y\Big\}\Big) I(x, y > 0) =\\ = (1 - e^{-x/80})(1 - e^{-y/70}) I(x, y > 0) = F_X(x) F_Y(y).$$

Естественно, совместная плотность распределения также является произведением

$$f_{X,Y}(x, y) = \frac{1}{80} e^{-x/80} I(x > 0) \frac{1}{70} e^{-y/70} I(y > 0).$$

Если нам известна совместная п. р. в. $f_{X,Y}$ двух с. в. X и Y, то их маргинальные плотности f_X и f_Y можно найти интегрированием по дополнительной переменной:

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x, y)\, dy, \qquad f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x, y)\, dx. \tag{2.1.22}$$

Это равенство интуитивно понятно: мы применяем непрерывный аналог формулы полной вероятности, интегрируя по всем значениям Y и исключая таким образом случайную величину Y из рассмотрения. Точнее, для любого $y \in R$ справедливо равенство

$$F_X(y) = \mathbf{P}(X < y) = \mathbf{P}(X < y, -\infty < Y < \infty) =\\ = \int_{-\infty}^{y} \int_{-\infty}^{\infty} f_{X,Y}(x, x')\, dx'\, dx = \int_{-\infty}^{y} g_X(x)\, dx, \tag{2.1.23}$$

где

$$g_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, x')\, dx'.$$

С другой стороны, вновь для любого $y \in R$ мы имеем

$$F_X(y) = \int_{-\infty}^{y} f_X(x)\, dx. \tag{2.1.24}$$

Сравнивая правые части равенств (2.1.23) и (2.1.24), получим равенство $f_X = g_X$, что и требовалось.

Пример. Рассмотрим двумерную случайную величину (X, Y), имеющую нормальное распределение. Совместное распределение величин X и Y имеет вид (2.1.5) при $d = 2$. Теперь положительно определенные матрицы

Σ и Σ^{-1} можно записать в виде

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2 r \\ \sigma_1\sigma_2 r & \sigma_2^2 \end{pmatrix}, \quad \Sigma^{-1} = \frac{1}{1-r^2}\begin{pmatrix} 1/\sigma_1^2 & -r/(\sigma_1\sigma_2) \\ -r/(\sigma_1\sigma_2) & 1/\sigma_2^2 \end{pmatrix},$$

где σ_1, σ_2 — ненулевые действительные числа, т. е. $\sigma_1^2, \sigma_2^2 > 0$, а r — действительное число, $|r| < 1$. Равенство (2.1.5) тогда принимает вид

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \exp\Bigl(\frac{-1}{2(1-r^2)} \times \\ \times \Bigl[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2r\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\Bigr]\Bigr). \qquad (2.1.25)$$

Мы хотим проверить, что по отдельности $X \sim \mathrm{N}(\mu_1, \sigma_1^2)$ и $Y \sim \mathrm{N}(\mu_2, \sigma_2^2)$, т. е. вычислить маргинальные плотности распределения. Запишем

$$f_X(x) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \int \exp\Bigl(-\frac{1}{2(1-r^2)}\Bigl[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2r\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \\ + \frac{(y-\mu_2)^2}{\sigma_2^2}\Bigr]\Bigr)\, dy = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \int \exp\Bigl(-\frac{1}{2(1-r^2)}\Bigl[\frac{(1-r^2)(x-\mu_1)^2}{\sigma_1^2} + \\ + \Bigl(\frac{y-\mu_2}{\sigma_2} - r\frac{x-\mu_1}{\sigma_1}\Bigr)^2\Bigr]\Bigr)\, dy = \\ = \frac{1}{\sqrt{2\pi}\sigma_1} \frac{e^{-(x-\mu_1)^2/2\sigma_1^2}}{\sqrt{2\pi}\sigma_2\sqrt{1-r^2}} \int \exp\Bigl[-\frac{1}{2\sigma_2^2(1-r^2)}(y_1 - \nu_1)^2\Bigr]\, dy_1,$$

где

$$y_1 = y - \mu_2, \quad \nu_1 = r\frac{\sigma_2}{\sigma_1}(x - \mu_1).$$

Получим, что

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-(x-\mu_1)^2/2\sigma_1^2},$$

т. е. $X \sim \mathrm{N}(\mu_1, \sigma_1^2)$. Аналогично $Y \sim \mathrm{N}(\mu_2, \sigma_2^2)$. □

Вышеупомянутый прием «стандартного представления», когда действительную величину $X \sim f$ интерпретируют как $X(x) = x$ для $x \in \mathbb{R}$, а вероятности $\mathrm{P}(A)$ задаются формулой (2.1.8), применим также и для совместных распределений. Метод тот же: если у вас есть пара $(X, Y) \sim f$, возьмите $\omega = (x, y) \in \Omega = \mathbb{R}^2$, положите $\mathrm{P}(A) = \int_A f(x, y)\, dy\, dx$ для $A \subset \mathbb{R}^2$ и определите $X(\omega) = x$, $Y(\omega) = y$. Подобный метод работает и в случае произвольной совокупности $X_1, \ldots, X_n$.

Ряд задач, приведенных далее, относится к преобразованиям пары (X, Y) в (U, V), где

$$U = g_1(X, Y), \quad V = g_2(X, Y).$$

В этом случае для вычисления совместной п. р. в. $f_{U,V}$ на основе $f_{X,Y}$ нужно использовать формулу, подобную (2.1.17). А именно, если замена переменных $(x, y) \mapsto (u, v)$, $u = g_1(x, y)$, $v = g_2(x, y)$, взаимно однозначна на области значений случайных величин X и Y, то

$$f_{U,V}(u, v) = I((u, v) \in \mathrm{Range}(g_1, g_2))f_{X,Y}(x(u, v), y(u, v))\left|\frac{\partial(x, y)}{\partial(u, v)}\right|. \quad (2.1.26)$$

Здесь

$$\frac{\partial(x, y)}{\partial(u, v)} = \det\begin{pmatrix} \partial x/\partial u & \partial x/\partial v \\ \partial y/\partial u & \partial y/\partial v \end{pmatrix}$$

— якобиан обратного преобразования $(u, v) \mapsto (x, y)$:

$$\det\begin{pmatrix} \partial x/\partial u & \partial x/\partial v \\ \partial y/\partial u & \partial y/\partial v \end{pmatrix} = \det\begin{pmatrix} \partial u/\partial x & \partial u/\partial y \\ \partial v/\partial x & \partial v/\partial y \end{pmatrix}^{-1}.$$

Наличие абсолютной величины $\left|\frac{\partial(x, y)}{\partial(u, v)}\right|$ гарантирует, что $f_{U,V}(u, v) \geqslant 0$.

В частности, плотность распределения суммы $X + Y$ и произведения XY вычисляется как

$$f_{X+Y}(u) = \int_{\mathbb{R}} f_{X,Y}(x, u - x)\,dx = \int_{\mathbb{R}} f_{X,Y}(u - y, y)\,dy \quad (2.1.27)$$

и

$$f_{XY}(u) = \int_{\mathbb{R}} f_{X,Y}(x, u/x)\frac{1}{|x|}\,dx = \int_{\mathbb{R}} f_{X,Y}(u/y, y)\frac{1}{|y|}\,dy; \quad (2.1.28)$$

ср. с формулами (1.4.9) и (1.4.10). Для отношения X/Y формула имеет вид

$$f_{X/Y}(u) = \int_{\mathbb{R}} |y| f_{X,Y}(yu, y)\,dy. \quad (2.1.29)$$

Вывод формулы (2.1.27) таков. Если $U = X + Y$, то соответствующая замена переменных имеет вид $u = x + y$ и, например, $v = y$, а обратное преобразование имеет вид $x = u - v$, $y = v$. Мы имеем $\frac{\partial(u, v)}{\partial(x, y)} = 1 = \frac{\partial(x, y)}{\partial(u, v)}$, откуда следует, что $f_{X+Y,Y}(u, v) = f_{X,Y}(u - v, v)$. Интегрируя по dv, получим $f_{X+Y}(u) = \int f_{X,Y}(u - v, v)\,dv$, что и дает формулу (2.1.27). Средний интеграл можно получить, используя замену переменных $u = x + y$, $v = x$ (или просто отмечая, что X и Y можно поменять местами).

Вывод формул (2.1.28) и (2.1.29) аналогичен, величины $1/|x|$, $1/|y|$ и $|y|$, возникают как абсолютные значения соответствующих (обратных) якобианов.

Для полноты изложения выведем общую формулу, предполагая, что отображение $u_1 = g_1(\boldsymbol{x}), \ldots, u_n = g_n(\boldsymbol{x})$ взаимно однозначно на области значений случайных величин $X_1, \ldots, X_n$. В этом случае для векторов $\boldsymbol{X} = (X_1, \ldots, X_n)$ и $\boldsymbol{U} = (U_1, \ldots, U_n)$, $U_i = g_i(\boldsymbol{X})$, справедливо равенство

$$f_U(u_1, \ldots, u_n) = \left|\frac{\partial(x_1, \ldots, x_n)}{\partial(u_1, \ldots, u_n)}\right| f_{\boldsymbol{X}}(x_1(\boldsymbol{u}), \ldots, x_n(\boldsymbol{u})) \times \\ \times I(\boldsymbol{u} \in \mathrm{Range}(g_1, \ldots, g_n)). \quad (2.1.30)$$

Пример. Для пары с. в. X, Y, имеющих совместное нормальное распределение, рассмотрим плотность распределения $X+Y$. Из формул (2.1.25), (2.1.26) следует, что

$$f_{X+Y}(u) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \int_{\mathbb{R}} \exp\left(\frac{-1}{2(1-r^2)} \times \right.$$
$$\left. \times \left[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2r\frac{(x-\mu_1)(u-x-\mu_2)}{\sigma_1\sigma_2} + \frac{(u-x-\mu_2)^2}{\sigma_2^2}\right]\right) dx =$$
$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \int_{\mathbb{R}} \exp\left(\frac{-1}{2(1-r^2)}\left[\frac{x_1^2}{\sigma_1^2} - 2r\frac{x_1(u_1-x_1)}{\sigma_1\sigma_2} + \frac{(u_1-x_1)^2}{\sigma_2^2}\right]\right) dx_1,$$

где $x_1 = x - \mu_1$, $u_1 = u - \mu_1 - \mu_2$.

Выделяя полный квадрат, получим

$$\frac{x_1^2}{\sigma_1^2} - 2r\frac{x_1(u_1-x_1)}{\sigma_1\sigma_2} + \frac{(u_1-x_1)^2}{\sigma_2^2} =$$
$$= \left(x_1\frac{\sqrt{\sigma_1^2+2r\sigma_1\sigma_2+\sigma_2^2}}{\sigma_1\sigma_2} - \frac{u_1}{\sigma_2}\frac{\sigma_1+r\sigma_2}{\sqrt{\sigma_1^2+2r\sigma_1\sigma_2+\sigma_2^2}}\right)^2 + \frac{u_1^2(1-r^2)}{\sigma_1^2+2r\sigma_1\sigma_2+\sigma_2^2}.$$

Затем получим, что

$$f_{X+Y}(u) = \frac{\exp\left[-\dfrac{u_1^2}{2(\sigma_1^2+2r\sigma_1\sigma_2+\sigma_2^2)}\right]}{2\pi\sqrt{\sigma_1^2+2r\sigma_1\sigma_2+\sigma_2^2}} \int_{\mathbb{R}} e^{-v^2/2}\, dv =$$
$$= \frac{1}{\sqrt{2\pi(\sigma_1^2+2r\sigma_1\sigma_2+\sigma_2^2)}} e^{-(u-\mu_1-\mu_2)^2/2(\sigma_1^2+2r\sigma_1\sigma_2+\sigma_2^2)}.$$

Здесь переменная интегрирования равна

$$v = \frac{1}{\sqrt{1-r^2}}\left(x_1\frac{\sqrt{\sigma_1^2+2r\sigma_1\sigma_2+\sigma_2^2}}{\sigma_1\sigma_2} - \frac{u_1}{\sigma_2}\frac{\sigma_1+r\sigma_2}{\sqrt{\sigma_1^2+2r\sigma_1\sigma_2+\sigma_2^2}}\right).$$

Видим, что $X+Y \sim \mathrm{N}(\mu_1+\mu_2, \sigma_1^2+2r\sigma_1\sigma_2+\sigma_2^2)$ со средним значением $\mu_1+\mu_2$ и дисперсией $\sigma_1^2+2r\sigma_1\sigma_2+\sigma_2^2$. □

Определение условной вероятности $\mathsf{P}(A \mid B)$ не отличается от дискретного случая: $\mathsf{P}(A \mid B) = \mathsf{P}(A \cap B)/\mathsf{P}(B)$. Например, если X — показательная с. в., то $\forall y,\ w > 0$ мы имеем

$$\mathsf{P}(X \geqslant y + w \mid X \geqslant w) = \frac{\mathsf{P}(X \geqslant y + w)}{\mathsf{P}(X \geqslant w)} = \\ = \frac{\int\limits_{y+w}^{\infty} \lambda e^{-\lambda u}\, du}{\int\limits_{w}^{\infty} \lambda e^{-\lambda v}\, dv} = \frac{e^{-\lambda(y+w)}}{e^{-\lambda w}} = e^{-\lambda y} = \mathsf{P}(X \geqslant y). \qquad (2.1.31)$$

Это свойство называется *свойством отсутствия памяти* показательного распределения. Оно является аналогом соответствующего свойства геометрического распределения (см. формулу (1.5.6)). Неудивительно, что показательное распределение является пределом геометрического при $p = e^{-\lambda/n} \nearrow 1$ $(n \to \infty)$. А именно, если $X \sim \mathrm{Exp}(\lambda)$ и $X^{(n)} \sim \mathrm{Geom}(e^{-\lambda/n})$, то $\forall y > 0$ мы имеем

$$(1 - e^{-\lambda y}) = \mathsf{P}(X < y) = \lim_{n \to \infty} \mathsf{P}(X^{(n)} < ny). \qquad (2.1.32)$$

Говоря об условных вероятностях (в дискретной или непрерывной постановке), полезно иметь в виду *условное распределение вероятностей*. Соответственно, в непрерывной постановке можно говорить об условной плотности распределения, см. соотношение (2.1.34).

Конечно, формула полной вероятности и формула Байеса остаются верными (не только для конечной, но и для счетной совокупности попарно непересекающихся событий B_j, $\mathsf{P}(B_j) > 0$).

Другой замечательный факт — это две леммы Бореля—Кантелли, названных по имени Э. Бореля (1871—1956), известного французского математика (и военно-морского министра на протяжении 15 лет), и Ф. П. Кантелли (1875—1966), итальянского математика (основателя итальянского Института актуариев). Первая лемма состоит в том, что *для последовательности таких (не обязательно непересекающихся) событий* $B_1, B_2, \ldots$, *что* $\sum\limits_j \mathsf{P}(B_j) < \infty$, *вероятность* $\mathsf{P}(A)$ *равна* 0, *где* A — *это пересечение* $\bigcap\limits_{n \geqslant 1} \bigcup\limits_{j \geqslant n} B_j$. Доказательство немедленно получит тот, кто обучен основным манипуляциям с вероятностями: если $A_n = \bigcup\limits_{j \geqslant n} B_j$, то $A_{n+1} \subseteq A_n$ и $A = \bigcap\limits_n A_n$. Тогда $A \subseteq A_n$, и потому $\mathsf{P}(A) \leqslant \mathsf{P}(A_n)$ $\forall n$. Но $\mathsf{P}(A_n) \leqslant \sum\limits_{j \geqslant n} \mathsf{P}(B_j)$, а эта сумма стремится к 0 при $n \to \infty$, потому что $\sum\limits_j \mathsf{P}(B_j) < \infty$. Значит, $\mathsf{P}(A) = 0$.

Первая лемма Бореля—Кантелли имеет весьма поразительную интерпретацию: если $\sum\limits_j \mathsf{P}(B_j) < \infty$, то с вероятностью 1 происходит только конечное число событий B_j. Причина этого в том, что вышеупомянутое пересечение A означает, что «происходит бесконечное число событий B_j». Формально говоря, если $\omega \in A$, то $\omega \in B_j$ для бесконечного числа индексов j.

Вторая лемма Бореля—Кантелли гласит, что *для таких независимых событий* $B_1, B_2, \ldots$, *что* $\sum\limits_j \mathrm{P}(B_j) = \infty$, *выполняется равенство* $\mathrm{P}(A) = 1$. Доказательство вновь получаем непосредственными оценками:

$$\mathrm{P}(A^c) = \mathrm{P}\Big(\bigcup_{n\geqslant 1} A_n^c\Big) \leqslant \sum_{n\geqslant 1} \mathrm{P}(A_n^c).$$

Далее,

$$\mathrm{P}(A_n^c)=\mathrm{P}\Big(\bigcap_{j\geqslant n} B_j^c\Big)=\prod_{j\geqslant n}\mathrm{P}(B_j^c)=\exp\Big(\sum_{j\geqslant n}\ln(1-\mathrm{P}(B_j))\Big)\leqslant\exp\Big(-\sum_{j\geqslant n}\mathrm{P}(B_j)\Big),$$

поскольку $\ln(1-x) \leqslant -x$ для $x \geqslant 0$. В силу того что $\sum\limits_{j\geqslant n} \mathrm{P}(B_j) = \infty$, получаем $\mathrm{P}(A_n^c) = 0$ $\forall n$. Тогда $\mathrm{P}(A^c) = 0$ и $\mathrm{P}(A) = 1$.

Таким образом, если $B_1, B_2, \ldots$ — независимые события и $\sum\limits_j \mathrm{P}(B_j) = \infty$, то «с вероятностью 1 происходит бесконечное число этих событий».

Например, в задачах 1.4.8 и 2.2.12 события «год k является рекордным» независимы и имеют вероятности $1/k$. По второй лемме Бореля—Кантелли с вероятностью 1 будет бесконечное число рекордных лет, если продолжать наблюдения неограниченно.

В непрерывном случае также можно попробовать работать с условными вероятностями вида $\mathrm{P}(A \mid X = x)$ при условии, что с. в. X с плотностью распределения f_X приняла значение x. Известно, что такое событие имеет нулевую вероятность, поэтому даже совершенно корректные выкладки могут не привести к верным результатам. Фокус состоит в том, чтобы рассматривать не одно значение x, а все x из области значений с. в. X сразу. Говоря формальным языком, мы будем работать с непрерывным аналогом формулы полной вероятности:

$$\mathrm{P}(A) = \int \mathrm{P}(A \mid X = x) f_X(x)\, dx. \tag{2.1.33}$$

Эта формула верна при условии, что вероятность $\mathrm{P}(A \mid X = x)$ определена «разумно», что, как правило, возможно во всех естественно возникающих ситуациях. Например, если $f_{X,Y}(x, y)$ — совместная п. р. в. случайных величин X и Y, а A — событие вида $\{a < Y < b\}$, то

$$\mathrm{P}(A \mid X = x) = \int_a^b f_{X,Y}(x, y)\, dy \Big/ \Big(\int_{\mathbb{R}} f_{X,Y}(x, y)\, dy\Big).$$

Следующий шаг — ввести *условную* п. р. в. $f_{Y|X}(y\,|\,x)$ случайной величины Y при условии $\{X=x\}$:

$$f_{Y|X}(y\,|\,x)=\frac{f_{X,Y}(x,\,y)}{f_X(x)} \tag{2.1.34}$$

и записать естественные аналоги формулы полной вероятности и формулы Байеса:

$$f_Y(y)=\int f_{Y|X}(y\,|\,x)f_X(x)\,dx, \qquad f_{Y|X}(y\,|\,x)=f_{X,Y}\Big/\int f_{X|Y}(x\,|\,z)f_Y(z)\,dz. \tag{2.1.35}$$

Как и в дискретном случае, два события A и B называются независимыми, если

$$\mathsf{P}(A\cap B)=\mathsf{P}(A)\,\mathsf{P}(B); \tag{2.1.36}$$

в случае произвольной конечной совокупности $A_1,\,\ldots,\,A_n$, $n>2$, для независимости требуется, чтобы для любого $k=2,\,\ldots,\,n$ и для любых $i_1,\,\ldots,\,i_k$, $1\leqslant i_1<\ldots<i_k\leqslant n$, выполнялись равенства

$$\mathsf{P}\Big(\bigcap_{1\leqslant j\leqslant k}A_{i_j}\Big)=\prod_{1\leqslant j\leqslant k}\mathsf{P}(A_{i_j}). \tag{2.1.37}$$

Аналогично с. в. X и Y, заданные на одном и том же вероятностном пространстве Ω, называются *независимыми*, если

$$\mathsf{P}(X<x,\,Y<y)=\mathsf{P}(X<x)\,\mathsf{P}(Y<y)\quad \forall x,\,y\in\mathbb{R}. \tag{2.1.38}$$

Иными словами, совместная ф. р. $F_{X,Y}(x,\,y)$ разлагается в произведение $F_X(x)F_Y(y)$. Наконец, совокупность случайных величин $X_1,\,\ldots,\,X_n$ (снова заданных на одном и том же вероятностном пространстве Ω), $n>2$, называется независимой (т. е. эти случайные величины независимы в совокупности), если для любого $k=2,\,\ldots,\,n$ и для любых $i_1,\,\ldots,\,i_k$, $1\leqslant i_1<\ldots<i_k\leqslant n$, выполняются равенства

$$\mathsf{P}(X_{i_1}<y_1,\,\ldots,\,X_{i_k}<y_k)=\prod_{1\leqslant j\leqslant k}\mathsf{P}(X_{i_j}<y_j),\qquad y_1,\,\ldots,\,y_k\in\mathbb{R}. \tag{2.1.39}$$

Замечание. Достаточно, чтобы только равенство (2.1.39) выполнялось для всей совокупности $X_1,\,\ldots,\,X_n$, но при этом нужно разрешить y_i принимать значение $+\infty$:

$$\mathsf{P}(X_1<y_1,\,\ldots,\,X_n<y_n)=\prod_{1\leqslant j\leqslant n}\mathsf{P}(X_j<y_j),\qquad y_1,\,\ldots,\,y_n\in\overline{\mathbb{R}}=\mathbb{R}\cup\{\infty\}. \tag{2.1.40}$$

В самом деле, если некоторые случайные величины X_i в формуле (2.1.40) равны ∞, это значит, что соответствующее условие $y_i<\infty$ тривиально выполняется и его можно опустить.

Если рассматриваемые случайные величины имеют общую п. р. в. $f_{X,Y}$, или $f_{\boldsymbol{X}}$, то формулы (2.1.39) и (2.1.40) эквивалентны тому, что эти плотности разлагаются в произведение:

$$f_{X,Y}(x, y) = f_X(x)f_Y(y), \qquad f_{\boldsymbol{X}}(\boldsymbol{x}) = f_{X_1}(x_1)\dots f_{X_n}(x_n). \tag{2.1.41}$$

Формальное доказательство этого факта следующее. Разложение $F_{X,Y}(y_1, y_2) = F_X(y_1)F_Y(y_2)$ означает, что для любых $y_1, y_2 \in \mathbb{R}$ мы имеем

$$\int_{-\infty}^{y_1}\int_{-\infty}^{y_2} f_{X,Y}(x, y)\,dy\,dx = \int_{-\infty}^{y_1}\int_{-\infty}^{y_2} f_X(x)f_Y(y)\,dy\,dx.$$

Поэтому подынтегральные функции должны совпадать, т. е. $f_{X,Y}(x, y) = f_X(x)f_Y(y)$. Обратная импликация очевидна. Рассуждения для произвольной совокупности $\boldsymbol{X} = (X_1, \dots, X_n)$ аналогичны.

Для независимых с. в. X, Y уравнения (2.1.27)—(2.1.29) принимают вид

$$f_{X+Y}(u) = \int f_X(x)f_Y(u-x)\,dx = \int f_X(u-y)f_Y(y)\,dy, \tag{2.1.42}$$

$$f_{XY}(u) = \int_{\mathbb{R}} f_X(x)f_Y(u/x)\frac{1}{|x|}\,dx = \int_{\mathbb{R}} f_X(u/y)f_Y(y)\frac{1}{|y|}\,dy \tag{2.1.43}$$

и

$$f_{X/Y}(u) = \int_{\mathbb{R}} |y|\, f_X(yu)\, f_Y(y)\,dy. \tag{2.1.44}$$

Ср. с задачей 2.1.15, приведенной ниже. Как и в дискретном случае, уравнение (2.1.42) называется уравнением типа свертки (для плотностей).

Понятие н. о. р. с. в., изученное в предыдущей главе, будет по-прежнему играть выдающуюся роль, особенно в § 2.3.

Задача 2.1.10. Случайные величины X и Y независимы и показательно распределены с параметрами λ и μ. Положим

$$U = \max[X, Y], \qquad V = \min[X, Y].$$

Являются ли случайные величины U и V независимыми? Независима ли случайная величина U от события $\{X > Y\}$ (т. е. от с. в. $I(X > Y)$)?

Решение. В терминах стандартного представления рассмотрим

$$\Omega = \mathbb{R}_+^2 = \{(x_1, x_2)\colon 0 < x_1, x_2 < \infty\},$$
$$\mathrm{P}(A) = \int_A \lambda e^{-\lambda x_1}\mu e^{-\mu x_2}\,dx_2\,dx_1.$$

Тогда вероятность $\mathrm{P}(V > y_1, U < y_2)$ равна

$$\mathrm{P}(y_1 < X, Y < y_2) = \mathrm{P}(y_1 < X < y_2)\,\mathrm{P}(y_1 < Y < y_2),$$

что, в свою очередь, равно

$$(e^{-\lambda y_1} - e^{-\lambda y_2})(e^{-\mu y_1} - e^{-\mu y_2}) I(0 < y_1 < y_2).$$

С другой стороны, произведение

$$\begin{aligned} \mathsf{P}(V > y_1)\,\mathsf{P}(U < y_2) = \mathsf{P}(X,\ Y > y_1)\,\mathsf{P}(X,\ Y < y_2) = \\ = e^{-(\lambda+\mu)y_1}(1 - e^{-\lambda y_2})(1 - e^{-\mu y_2}) \end{aligned}$$

явно имеет другое значение. Поэтому случайные величины U и V зависимы.

Далее, $\mathsf{P}(X > Y) = \dfrac{\mu}{\lambda+\mu}$, и произведение

$$\mathsf{P}(X > Y)\,\mathsf{P}(U < y) = \frac{\mu}{\lambda+\mu}(1 - e^{-\lambda y})(1 - e^{-\mu y})$$

отличается от $\mathsf{P}(X > Y,\ U < y) = \mathsf{P}(Y < X < y)$. Значит, U и $\{X > Y\}$ зависимы. □

Задача 2.1.11. Случайные величины X и Y независимы и показательно распределены, каждая с параметром λ. Покажите, что с. в. $X + Y$ и $X/(X + Y)$ независимы, и найдите их распределение.

Решение. Рассмотрим с. в. $U = X + Y$ и $V = X/(X + Y)$ с областью значений $\mathbb{R}_+$ и $(0,\ 1)$ соответственно. Отображение

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x + y \\ x/(x + y) \end{pmatrix} := \begin{pmatrix} u \\ v \end{pmatrix}, \qquad x,\ y \in \mathbb{R},$$

имеет обратное $x = uv$, $y = u(1 - v)$ с областью значений $\{u > 0,\ 0 < v < 1\}$ и якобианом

$$\frac{\partial(x,\ y)}{\partial(u,\ v)} = \det\begin{pmatrix} v & u \\ 1 - v & -u \end{pmatrix} = -u.$$

Тогда совместная п. р. в. $f_{U,V}(u,\ v)$ равна

$$\lambda^2 u e^{-\lambda u} I(u > 0) I(0 < v < 1).$$

С другой стороны,

$$\begin{aligned} f_U(u) &= \int f_{U,V}(u,\ v)\,dv = \lambda^2 u e^{-\lambda u} I(u > 0), \\ f_V(v) &= \int f_{U,V}(u,\ v)\,du = I(0 < v < 1). \end{aligned}$$

Поэтому $f_{U,V}(u,\ v) = f_U(u) f_V(v)$, и случайные величины U и V независимы. □

Задача 2.1.12. Спортсмен стреляет по круговой мишени. Вертикальная и горизонтальная координаты точки попадания пули (при условии, что центр мишени — начало координат) — независимые случайные величины, каждая с распределением $\mathrm{N}(0,\ 1)$.

Покажите, что расстояние от точки попадания до центра имеет п. р. в. $re^{-r^2/2}$ для $r > 0$. Найдите медиану этого распределения.

Решение. Имеем $X \sim \mathrm{N}(0, 1)$, $Y \sim \mathrm{N}(0, 1)$ и $f_{X,Y}(x, y) = \dfrac{1}{2\pi} e^{-(x^2+y^2)/2}$, $x, y \in \mathbb{R}$. Положим $R = \sqrt{X^2 + Y^2}$ и $\Theta = \operatorname{arctg}(Y/X)$. Область значений отображения $\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} r \\ \theta \end{pmatrix}$ — это $\{r > 0,\ \theta \in (-\pi, \pi)\}$, и якобиан

$$\frac{\partial(r, \theta)}{\partial(x, y)} = \det \begin{pmatrix} \partial r/\partial x & \partial r/\partial y \\ \partial\theta/\partial x & \partial\theta/\partial y \end{pmatrix}$$

равен

$$\det \begin{pmatrix} x/r & y/r \\ \dfrac{-y/x^2}{1+(y/x)^2} & \dfrac{1/x}{1+(y/x)^2} \end{pmatrix} = \frac{1}{r}, \quad x, y \in \mathbb{R}, \quad r > 0, \quad \theta \in (-\pi, \pi).$$

Тогда обратное отображение $\begin{pmatrix} r \\ \theta \end{pmatrix} \mapsto (xy)$ имеет якобиан $\dfrac{\partial(x, y)}{\partial(r, \theta)} = r$. Поэтому

$$f_{R,\Theta}(r, \theta) = re^{-r^2/2} I(r > 0) \frac{1}{2\pi} I(-\pi, \pi).$$

Интегрируя по $d\theta$, получаем

$$f_R(r) = re^{-r^2/2} I(r > 0),$$

что и требовалось.

Чтобы найти медиану $m(R)$, рассмотрим уравнение

$$\int_0^y re^{-r^2/2}\, dr = \int_y^\infty re^{-r^2/2}\, dr,$$

из которого следует, что $e^{-y^2/2} = 1/2$. Значит, $m(R) = \sqrt{\ln 4}$. □

Мы завершим теоретическую часть этого параграфа следующим замечанием. Пусть X — с. в. с ф. р. F и y, $0 < y < 1$, — значение, принимаемое функцией F, причем $F(x^*) = y$ в точке $x^* = \inf[x \in \mathbb{R}\colon F(x) \geqslant y]$. Тогда

$$\mathrm{P}(F(X) < y) = y. \tag{2.1.45а}$$

В самом деле, в этом случае событие $\{F(X) < y\} = \{X < x^*\}$ имеет вероятность $\mathrm{P}(F(X) < y) = F(x^*)$. В общем случае, если $g\colon \mathbb{R} \to \mathbb{R}$ — другая функция и существует единственная такая точка $a \in \mathbb{R}$, что $F(a) = g(a)$, $F(x) < g(x)$ для $x < a$ и $F(x) \geqslant g(x)$ для $x \geqslant a$, то

$$\mathrm{P}(F(X) < g(X)) = g(a). \tag{2.1.45б}$$

Задача 2.1.13. Случайная выборка $X_1, \ldots, X_{2n+1}$ взята из распределения с п. р. в. f. Пусть $Y_1, \ldots, Y_{2n+1}$ — значения величин $X_1, \ldots, X_{2n+1}$, расположенные в возрастающем порядке. Найдите распределение каждой с. в. Y_k, $k = 1, \ldots, 2n + 1$.

Решение. Если $X_j \sim f$, то

$$P_{Y_k}(y) = \mathrm{P}(Y_k < y) = \sum_{j=k}^{2n+1} C_{2n+1}^j F(y)^j (1 - F(y))^{2n+1-j}, \qquad y \in \mathbb{R}.$$

Плотности f_{Y_k}, $k = 1, \ldots, 2n + 1$, можно получить следующим образом. Случайная величина Y_k принимает значение x тогда и только тогда, когда $k - 1$ значение из $X_1, \ldots, X_{2n+1}$ меньше чем x, $2n + 1 - k$ значений больше чем x, а одно равно x. Поэтому

$$f_{Y_k}(x) = \frac{(2n+1)!}{(k-1)!\,(2n+1-k)!}[F(x)]^{k-1}[1 - F(x)]^{2n+1-k} f(x).$$

В частности, если $X_i \sim \mathrm{U}(0, 1)$, то плотность *выборочной медианы* Y_{n+1} равна

$$f_{Y_{n+1}}(x) = \frac{(2n+1)!}{n!\,n!}(x(1-x))^n. \qquad \square$$

Задача 2.1.14. а) Пусть X и Y — независимые случайные величины с непрерывными симметричными распределениями с плотностями f_X и f_Y соответственно. Покажите, что плотность распределения с. в. $Z = X/Y$ равна

$$h(a) = 2 \int_0^\infty y f_X(ay) f_Y(y)\, dy.$$

б) Пусть X и Y — независимые нормальные случайные величины с распределениями $\mathrm{N}(0, \sigma^2)$ и $\mathrm{N}(0, \tau^2)$ соответственно. Покажите, что случайная величина $Z = X/Y$ имеет плотность $h(a) = \dfrac{d}{\pi(d^2 + a^2)}$, где $d = \sigma/\tau$ (плотность распределения Коши).

Указание. Используйте формулу

$$F_Z(a) = 2 \int_0^\infty \int_{-\infty}^{ay} f_X(x) f_Y(y)\, dx\, dy.$$

Задача 2.1.15. Пусть X и Y — независимые случайные величины с плотностями f_X и f_Y. Покажите, что случайная величина $Z = Y/X$ имеет плотность

$$h(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(zx) |x|\, dx.$$

Выведите отсюда, что с. в. $T = \operatorname{arctg}(Y/X)$ равномерно распределена на интервале $(-\pi/2, \pi/2)$ тогда и только тогда, когда

$$\int\limits_{-\infty}^{\infty} f_X(x) f_Y(xz)|x|\, dx = \frac{1}{\pi(1+z^2)}$$

для всех $z \in \mathbb{R}$. Проверьте, что это равенство верно, если обе величины X и Y имеют нормальное распределение со средним 0 и ненулевой дисперсией σ^2.

Решение. Функция распределения с. в. Z равна

$$F_Z(u) = \mathsf{P}(Z < u) = \mathsf{P}(Y/X < u,\ X > 0) + \mathsf{P}(Y/X < u,\ X < 0) =$$
$$= \int\limits_0^{\infty} f_X(x) \int\limits_{-\infty}^{ux} f_Y(y)\, dy\, dx + \int\limits_{-\infty}^{0} f_X(x) \int\limits_{ux}^{\infty} f_Y(y)\, dy\, dx. \qquad (2.1.46)$$

Тогда плотность распределения имеет вид

$$f_Z(u) = \frac{d}{du} F_Z(u) = \int\limits_0^{\infty} f_X(x) f_Y(ux) x\, dx - \int\limits_{-\infty}^{0} f_X(x) f_Y(ux) x\, dx =$$
$$= \int\limits_{-\infty}^{\infty} f_X(x) f_Y(ux)|x|\, dx,$$

что согласуется с формулой, полученной с помощью якобиана.

Если случайная величина T равномерно распределена, то

$$F_Z(u) = \mathsf{P}(\operatorname{arctg} Z \leqslant \operatorname{arctg} u) = \frac{\operatorname{arctg} u + \pi/2}{\pi},$$
$$f_Z(u) = \frac{d}{du} F_Z(u) = \frac{1}{\pi(u^2+1)}.$$

Обратно,

$$f_Z(u) = \frac{1}{\pi(1+u^2)} \Rightarrow F_Z(u) = \frac{1}{\pi} \operatorname{arctg} u + \frac{1}{2} \Rightarrow \mathsf{P}(T \leqslant u) = \frac{1}{\pi} u + \frac{1}{2},$$

или $f_T(u) = \dfrac{1}{\pi}$ на $(-\pi/2, \pi/2)$.

Наконец, рассмотрим

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2}, \qquad f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-y^2/2\sigma^2}.$$

Тогда

$$\int\limits_{\mathbb{R}} f_X(x) f_Y(ux)|x|\, dx = \frac{1}{\pi\sigma^2} \int\limits_0^{\infty} e^{-x^2/(2\sigma^2) - x^2u^2/(2\sigma^2)} x\, dx =$$
$$= \frac{1}{\pi\sigma^2} \left(\frac{-\sigma^2}{u^2+1} e^{-x^2(1+u^2)/2\sigma^2} \right) \Bigg|_0^{\infty} = \frac{1}{\pi(1+u^2)}. \qquad \square$$

Задача 2.1.16. Пусть $X_1, X_2, \ldots$ — независимые случайные величины, распределенные по Коши, и каждая имеет плотность

$$f(x) = \frac{d}{\pi(d^2 + x^2)}.$$

Покажите, что случайная величина $(X_1 + X_2 + \ldots + X_n)/n$ имеет то же распределение, что и X_1.

Решение. Пусть сначала $d = 1$, $n = 2$, $f(x) = \dfrac{1}{\pi(1 + x^2)}$. Тогда случайная величина $\dfrac{X_1 + X_2}{2}$ имеет функцию распределения вида $G(x) = \mathsf{P}(X_1 + X_2 < 2x)$ с плотностью распределения

$$g(x) = 2\tilde{g}(2x), \qquad \tilde{g}(y) = \frac{1}{\pi^2} \int_{-\infty}^{\infty} \frac{1}{1 + u^2} \cdot \frac{du}{1 + (y - u)^2}.$$

Теперь используем тождество

$$m \int_{-\infty}^{\infty} \frac{1}{(1 + y^2)(m^2 + (x - y)^2)}\, dy = \frac{\pi(m + 1)}{(m + 1)^2 + x^2},$$

которое является простым упражнением на комплексное интегрирование. Получим

$$\tilde{g}(y) = \frac{2}{\pi} \cdot \frac{1}{4 + y^2},$$

откуда следует, что

$$g(x) = 2\tilde{g}(2x) = \frac{1}{\pi} \cdot \frac{1}{1 + x^2} = f(x).$$

В общем случае, если с. в. имеет вид $Y = qX_1 + (1 - q)X_2$, то ее плотность равна

$$\frac{d}{dx} \int_{\mathbb{R}} f(u) \int_{-\infty}^{(x-qu)/(1-q)} f(y)\, dy\, du = \frac{1}{1 - q} \int_{\mathbb{R}^1} f\Big(\frac{x - qu}{1 - q}\Big) f(u)\, du,$$

что в силу того же тождества (с $m = (1 - q)/q$) равно

$$\frac{1}{q} \cdot \frac{1}{\pi}\Big(\frac{1 - q}{q} + 1\Big) \frac{1}{\dfrac{1}{q^2} + \dfrac{x^2}{q^2}} = f(x).$$

Теперь для $d = 1$ и произвольного n запишем $S_n = \dfrac{n - 1}{n} S_n + \dfrac{1}{n} X_n$. Базой индукции будет предположение, что $S_{n-1} \sim f(x)$. Тогда согласно предыдущим рассуждениям (с двумя слагаемыми S_{n-1} и X_n) $S_n \sim f(x)$.

Наконец, для произвольного d положим $\tilde{X}_i = X_i/d$. Тогда $\tilde{S}_n \sim \tilde{f}(x) = \frac{1}{\pi}\frac{1}{1+x^2}$ и $S_n \sim f(x) = \frac{1}{\pi}\frac{d}{d^2+x^2}$. □

Замечание. С помощью характеристических функций решение можно получить быстрее, см. формулу (2.2.26д). Доказательство см. в задаче 2.3.19, приведенной ниже.

Задача 2.1.17. Докажите, что если X, Y и Z — независимые случайные величины и каждая из них равномерно распределена на $(0, 1)$, то с. в. $(XY)^Z$ также равномерно распределена на $[0, 1]$.

Решение. Рассмотрим величину

$$\ln[(XY)^Z] = Z(\ln X + \ln Y).$$

Доказать, что величина $(XY)^Z$ равномерно распределена на $[0, 1]$, — это то же самое, что доказать, что величина $-Z(\ln X + \ln Y)$ показательно распределена на $[0, \infty)$. Величина $W = -\ln X - \ln Y$ имеет плотность

$$f_W(y) = \begin{cases} ye^{-y}, & 0 < y < \infty, \\ 0, & y \leqslant 0. \end{cases}$$

Совместная плотность $f_{Z,W}(x, y)$ имеет вид

$$f_{Z,W}(x, y) = \begin{cases} ye^{-y}, & \text{если } 0 < x < 1,\ 0 < y < \infty, \\ 0 & \text{в противном случае.} \end{cases}$$

Нас интересует произведение ZW. Удобно перейти от x, y к величинам $u = xy$, $v = y/x$ с помощью обратного якобиана $1/(2v)$. В терминах новых переменных совместная плотность $f_{U,V}$ случайных величин $U = ZW$ и $V = W/Z$ имеет вид

$$f_{U,V}(u, v) = \begin{cases} \dfrac{1}{2v}(uv)^{1/2}e^{-(uv)^{1/2}}, & \text{если } u, v > 0,\ 0 < u/v < 1, \\ 0 & \text{в противном случае.} \end{cases}$$

Плотность распределения U тогда равна

$$f_U(u) = \int f_{U,V}(u, v)\,dv = \int\limits_u^\infty \frac{1}{2v}(uv)^{1/2}e^{-(uv)^{1/2}}\,dv =$$
$$= -\int\limits_u^\infty d(e^{-(uv)^{1/2}}) = -[e^{-(uv)^{1/2}}]_u^\infty = e^{-u}, \quad u > 0,$$

где $f_U(u) = 0$ для $u < 0$. □

Задача 2.1.18. Пусть случайные величины $X_1, \ldots, X_n$ независимы и показательно распределены с одним и тем же параметром λ. Используя индукцию по n (или иным способом), покажите, что случайные величины

$$\max[X_1, \ldots, X_n] \quad \text{и} \quad X_1 + \frac{1}{2}X_2 + \ldots + \frac{1}{n}X_n$$

имеют то же распределение.

Решение. Введем обозначения

$$Y_n = \max[X_1, \ldots, X_n], \qquad Z_n = X_1 + \frac{1}{2}X_2 + \ldots + \frac{1}{n}X_n.$$

Теперь запишем

$$\mathrm{P}(Y_n < y) = (\mathrm{P}(X_1 < y))^n = (1 - e^{-\lambda y})^n.$$

Для $n = 1$ имеет место равенство $Y_1 = Z_1$. Используем индукцию по n:

$$\mathrm{P}(Z_n < y) = \mathrm{P}(Z_{n-1} + \frac{1}{n}X_n < y).$$

Поскольку $X_n/n \sim \mathrm{Exp}(n\lambda)$, последняя вероятность равна

$$\int_0^y (1 - e^{-\lambda z})^{n-1} n\lambda e^{-n\lambda(y-z)}\, dz =$$

$$= n\lambda e^{-n\lambda y} \int_0^y e^{\lambda z}(1 - e^{-\lambda z})^{n-1} e^{(n-1)\lambda z}\, dz = ne^{-n\lambda y} \int_1^{e^{\lambda y}} (u-1)^{n-1}\, du =$$

$$= ne^{-n\lambda y} \int_0^{e^{\lambda y}-1} v^{n-1}\, dv = e^{-n\lambda y}(e^{\lambda y} - 1)^n = (1 - e^{-\lambda y})^n. \quad \square$$

Задача 2.1.19. Предположим, что $\alpha \geqslant 1$ и X_α — положительная действительная случайная величина с плотностью

$$f_\alpha(t) = A_\alpha t^{\alpha-1} \exp(-t^\alpha)$$

для $t > 0$, где A_α — постоянная. Найдите A_α и покажите, что при $\alpha > 1$ и $s, t > 0$ выполняется неравенство

$$\mathrm{P}(X_\alpha \geqslant s + t \mid X_\alpha \geqslant t) < \mathrm{P}(X_\alpha \geqslant s).$$

Как выглядит это соотношение при $\alpha = 1$?

Решение. Должно выполняться равенство $\int\limits_0^\infty f_\alpha(t)\,dt = 1$, откуда следует, что

$$A_\alpha^{-1} = \int\limits_0^\infty t^{\alpha-1}\exp(-t^\alpha)\,dt = \alpha^{-1}\int\limits_0^\infty \exp(-t^\alpha)\,d(t^\alpha) = \alpha^{-1}[e^{-t^\alpha}]_0^\infty = \alpha^{-1}$$

и $A_\alpha = \alpha$. Если $\alpha > 1$, то для любых s, $t > 0$ мы имеем

$$\mathsf{P}(X_\alpha \geqslant s+t \mid X_\alpha \geqslant t) = \frac{\mathsf{P}(X_\alpha \geqslant s+t)}{\mathsf{P}(X_\alpha \geqslant t)} =$$

$$= \frac{\int\limits_{s+t}^\infty \exp(-u^\alpha)\,d(u^\alpha)}{\int\limits_t^\infty \exp(-u^\alpha)\,d(u^\alpha)} = \frac{\exp(-(s+t)^\alpha)}{\exp(-t^\alpha)} = \exp(t^\alpha - (s+t)^\alpha) =$$

$$= \exp(-s^\alpha + \text{отрицательные слагаемые}) < \exp(-s^\alpha) = \mathsf{P}(X_\alpha \geqslant s).$$

Если $\alpha = 1$, то $t^\alpha - (s+t)^\alpha = -s$ и вышеупомянутое неравенство становится равенством, так как $\mathsf{P}(X_\alpha \geqslant t) = \exp(-t)$. (Это отсутствие памяти у показательного распределения.) □

Замечание. Если интерпретировать X_α как время жизни определенного устройства, например электрической лампы, то неравенство $\mathsf{P}(X_\alpha \geqslant s+t \mid X_\alpha \geqslant t) < \mathsf{P}(X_\alpha \geqslant s)$ подчеркивает «возрастной» феномен, когда старое устройство, бывшее в употреблении в течение времени t, с меньшей вероятностью прослужит еще время s, чем новое. Имеются примеры обратного неравенства: качество устройства (или индивидуума) улучшается в процессе работы (службы, жизни).

Задача 2.1.20. Пусть X, Y, Z — независимые с. в., каждая из которых равномерно распределена на отрезке $[0, 1]$.

а) Покажите, что с. в. $X+Y$ имеет п. р. в. треугольной формы:

$$f_{X+Y}(u) = \begin{cases} u, & 0 \leqslant u \leqslant 1,\\ 2-u, & 1 \leqslant u \leqslant 2,\\ 0, & u \notin [0, 2]. \end{cases}$$

б) Покажите, что $\mathsf{P}(Z > X+Y) = 1/6$.

в) Пусть вам даны три стержня длиной X, Y, Z соответственно. Покажите, что вероятность того, что из этих стержней можно образовать треугольник, равна $1/2$.

г) Найдите п. р. в. $f_{X+Y+Z}(s)$ случайной величины $X+Y+Z$ для значений $s \in [0, 1]$. Пусть с. в. W равномерно распределена на $[0, 1]$ и не зависит от X, Y, Z. Покажите, что вероятность того, что стержни длиной W, X, Y, Z могут служить сторонами некоторого четырехугольника, равна $5/6$.

Решение. а) По формуле свертки п. р. в. $f_{X+Y}(u)$ равна

$$\int_0^1 f_X(x)f_Y(u-x)\,dx = \int_0^1 I(0 \leqslant u-x \leqslant 1)\,dx,$$

что приводится к виду

$$\begin{cases} \int_0^u dx = u, & 0 \leqslant u \leqslant 1, \\ \int_{u-1}^1 dx = 2-u, & 1 \leqslant u \leqslant 2, \end{cases}$$

что и требовалось показать.

Замечание. Плотность распределения вероятностей такой треугольной формы соответствует *распределению Симпсона.*

Т. Симпсон (1700—1761), английский ученый, внес значительный вклад в теорию интерполяции и численные методы интегрирования. Он был наиболее выдающимся среди странствующих лекторов и обычно читал лекции в модных лондонских кафе (в Англии в XVIII столетии это был популярный способ распространения научной информации).

б) Для вероятности $\mathsf{P}(Z > X+Y)$ запишем следующее выражение:

$$\int_0^1\int_0^1\int_0^1 I(z > x+y)\,dz\,dy\,dx = \int_0^1\int_0^1 I(x+y<1)\int_{x+y}^1 dz\,dy\,dx =$$
$$= \int_0^1\int_0^{1-x}(1-x-y)\,dy\,dx = \int_0^1 (1-x)^2\,dx - \frac{1}{2}\int_0^1 (1-x)^2\,dx = 1/6;$$

более краткий путь вычислений таков:

$$\mathsf{E}[\mathsf{P}(Z > X+Y \,|\, X+Y)] = \mathsf{E}[(1-(X+Y))I(X+Y<1)] =$$
$$= \int_0^1 (1-u)f_{X+Y}(u)\,du = \int_0^1 (1-u)u\,du = \frac{1}{6}.$$

в) Из трех стержней можно составить треугольник тогда и только тогда, когда ни один из них не длиннее, чем два других, взятых вместе. Таким образом, вероятность дополнительного события равна $3\cdot 1/6 = 1/2$. Отсюда заключаем, что

$$\mathsf{P}(\text{треугольник возможен}) = \frac{1}{2}.$$

г) Как и в п. а), по формуле свертки находим, что функция $f_{X+Y+Z}(s)$ равна

$$\int f_{X+Y}(x) f_Z(s-x)\,dx = \int_0^1 f_{X+Y}(x) I(0 \leqslant s-x \leqslant 1)\,dx = \int_0^s x\,dx = s^2/2$$

при $0 \leqslant s \leqslant 1$. Следовательно,

$$\mathsf{P}(W > X+Y+Z) = \int_0^1 \int_0^w \frac{s^2}{2}\,ds\,dw = \int_0^1 \frac{w^3}{6}\,dw = \frac{1}{24}.$$

Можно рассуждать иначе: вероятность $\mathsf{P}(W > X+Y+Z)$ равна

$$\mathsf{E}([1-(X+Y+Z)] I_{X+Y+Z<1}) = \int_0^1 \frac{s^2}{2}(1-s)\,ds = \frac{1}{6} - \frac{1}{8} = \frac{1}{24}.$$

Наконец, находим

$$\mathsf{P}(\text{четырехугольник невозможен}) = 4 \cdot \frac{1}{24} = \frac{1}{6}. \qquad \square$$

Задача 2.1.21. Пусть точка на плоскости имеет декартовы координаты (X, Y) и полярные координаты (R, Θ). Покажите, что если X и Y являются независимыми одинаково распределенными нормальными случайными величинами с нулевыми средними, то случайная величина R^2 имеет показательное распределение и не зависит от Θ.

Решение. Совместная плотность распределения $f_{X,Y}$ равна $\dfrac{1}{2\pi\sigma^2} \times$ $\times\, e^{-(x^2+y^2)/2\sigma^2}$, а R и Θ определяются как

$$R^2 = T = X^2 + Y^2, \qquad \Theta = \operatorname{arctg} \frac{Y}{X}.$$

Тогда

$$f_{R^2,\Theta}(t, \theta) = f_{X,Y}(x(t,\theta), y(t,\theta)) I(t>0) I(0<\theta<2\pi) \left| \frac{\partial(x,y)}{\partial(t,\theta)} \right|,$$

где обратный якобиан имеет вид

$$\frac{\partial(x,y)}{\partial(t,\theta)} = \left(\frac{\partial(t,\theta)}{\partial(x,y)}\right)^{-1} = \det \begin{pmatrix} 2x & 2y \\ \dfrac{-y}{x^2+y^2} & \dfrac{x}{x^2+y^2} \end{pmatrix}^{-1} = \left(\frac{2x^2+2y^2}{x^2+y^2}\right)^{-1} = \frac{1}{2}.$$

Следовательно,

$$f_{R^2,\Theta}(t,\theta) = \frac{1}{4\pi\sigma^2}e^{-t/2\sigma^2}I(t>0)I(0<\theta<2\pi) = f_{R^2}(t)f_\Theta(\theta),$$

где

$$f_{R^2}(t) = \frac{1}{2\sigma^2}e^{-t/2\sigma^2}I(t>0),\ f_\Theta(\theta) = \frac{1}{2\pi}I(0<\theta<2\pi).$$

Таким образом, $R^2 \sim \mathrm{Exp}(1/2\sigma^2)$, $\Theta \sim \mathrm{U}(0, 2\pi)$, и величины R^2 и Θ независимы. □

Задача 2.1.22. Планета Зог представляет собой шар с центром O. Три космических корабля A, B и C приземляются наугад на поверхность планеты, местоположения их посадок являются независимыми и равномерно распределенными на поверхности планеты. Вычислите плотность распределения угла $\angle AOB$, образованного прямыми OA и OB.

Корабли A и B могут иметь прямую радиосвязь, если $\angle AOB < \pi/2$, и аналогично для кораблей B и C и кораблей A и C. При заданном значении угла $\angle AOB = \gamma < \pi/2$ вычислите вероятность того, что корабль C может связываться непосредственно либо с A, либо с B. При заданном значении угла $\angle AOB = \gamma > \pi/2$ вычислите вероятность того, что корабль C может непосредственно связываться с обоими кораблями A и B. Заключите отсюда (или из иных рассуждений), что вероятность того, что все три космических корабля могут поддерживать связь друг с другом (таким образом, что, например, A может выходить на связь с B через C, если необходимо), равна $(\pi+2)/(4\pi)$.

Решение. Находим

$$\mathrm{P}(\angle AOB \in (\gamma, \gamma+\delta\gamma)) \sim \delta\gamma \sin\gamma.$$

Поскольку $\int\limits_0^\pi \sin\gamma\, d\gamma = 2$, плотность распределения угла $\angle AOB$ равна $f(\gamma) = \frac{1}{2}\sin\gamma$. В зависимости от величины $\angle AOB$ возможны два случая. Если $\angle AOB = \gamma < \frac{\pi}{2}$, то вероятность того, что C может иметь непосредственную связь либо с A, либо с B, равна $(\pi+\gamma)/(2\pi)$. Если $\angle AOB = \gamma > \frac{\pi}{2}$, то вероятность того, что C может иметь непосредственную связь с обоими кораблями A и B, равна $(\pi-\gamma)/(2\pi)$.

Действительно, корабль C находится «между» A и B, и теневая зона, где C может общаться только с A, но не может общаться с B, — это угол $\gamma - \pi/2$. Зона, где C может общаться с B, но не может общаться с A, — это также угол $\gamma - \pi/2$. Тогда допустимая зона, т. е. зона «видимости», — это угол $\gamma - 2(\gamma - \pi/2) = \pi - \gamma$. Это показано на рис. 2.18.

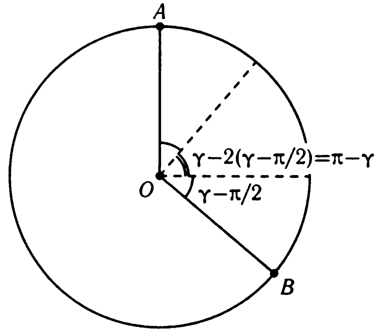


Рис. 2.18

Следовательно,

$$\mathrm{P}(\text{все 3 корабля могут поддерживать радиосвязь друг с другом}) =$$

$$= \int\limits_0^{\pi/2} f(\gamma)\frac{\pi+\gamma}{2\pi}\,d\gamma + \int\limits_{\pi/2}^{\pi} f(\gamma)\frac{\pi-\gamma}{2\pi}\,d\gamma = \int\limits_0^{\pi/2}\left(\frac{\pi+\gamma}{2\pi}+\frac{\gamma}{2\pi}\right)\frac{1}{2}\sin\gamma\,d\gamma =$$

$$= \frac{1}{4}\int\limits_0^{\pi/2}\sin\gamma\,d\gamma + \frac{1}{2\pi}\int\limits_0^{\pi/2}\gamma\sin\gamma\,d\gamma = \frac{\pi+2}{4\pi}. \quad \square$$

Задача 2.1.23. а) Пусть $A_1, A_2, \ldots, A_r$ — такие случайные события, что $A_i \cap A_j = \varnothing$ при $i \neq j$. Покажите, что для числа N наступивших событий выполняется равенство

$$\mathrm{P}(N=0) = 1 - \sum_{i=1}^{r}\mathrm{P}(A_i).$$

б) Планета Зог представляет собой шар с центром O. На поверхность планеты наугад приземляются N космических кораблей, при этом точки их приземления являются независимыми и равномерно распределенными на поверхности сферы. Корабль, находящийся в точке A, имеет непосредственную радиосвязь с другим кораблем в точке B, если $\angle AOB < \pi/2$. Вычислите вероятность того, что каждая точка поверхности имеет непосредственную радиосвязь хотя бы с одним из N кораблей.

Указание. Сечение поверхности сферы плоскостью, проходящей через центр сферы, называют ***большой окружностью***. Полезно использовать тот факт, что с вероятностью 1 случайным образом выбранные N больших окружностей разбивают поверхность сферы на $N(N-1)+2$ непересекающихся областей.

Решение. а) Находим

$$\mathsf{P}(N=0)=\mathsf{P}(A_1^c\cap\ldots\cap A_r^c)=1-\mathsf{P}\left(\bigcup_{l=1}^{r}A_l\right)=1-\sum_{i=1}^{r}\mathsf{P}(A_i).$$

б) Положение космического корабля определяет большую окружность — границу зоны его радиопокрытия. Если такая большая окружность задана, то корабль с равной вероятностью может находиться в одной либо другой из полусфер, ограниченных этой окружностью. Пусть положения N больших окружностей заданы и через $A_1, \ldots, A_{f(N)}$ обозначены

$$f(N)=N(N-1)+2$$

областей, на которые разбивается поверхность сферы этими большими окружностями. Введем событие

$E_i=\{$точки области A_i *не имеют* радиосвязи ни с одним из N кораблей$\}$.

Тогда $\mathsf{P}(E_i)=1/2^N$. Однако

$$\mathsf{P}\{\text{каждая точка имеет радиосвязь хотя бы с одним кораблем}\}=$$
$$=1-\mathsf{P}\left(\bigcup_{i=1}^{f(N)}E_i\right)=1-\sum_{i=1}^{f(N)}\mathsf{P}(E_i)=1-f(N)\frac{1}{2^N}=1-\frac{N(N-1)+2}{2^N},$$

поскольку $E_i\cap E_j=\varnothing$ при $i\neq j$. Отметим, что A_i, A_j должны лежать по разные стороны хотя бы от одной большой окружности.

Краткое доказательство утверждения, сформулированного в указании, заключается в следующем. Пусть в общем случае N больших окружностей разбивают сферу на $f(N)$ непересекающихся областей. Введем в рассмотрение еще одну окружность: в общем случае она будет пересекать каждую из предыдущих окружностей в двух точках. Немедленно получаем, что каждое такое пересечение соответствует разбиению уже существующей области на две. Следовательно,

$$f(N+1)=f(N)+2N,\qquad f(1)=2.$$

Из этого рекуррентного соотношения находим

$$f(N)=2+2(1+\ldots+N-1)=2+N(N-1).$$

Выражение «в общем случае» здесь означает «кроме вырожденных случаев»; ясно, что вырождения не возникает с вероятностью 1. □

§ 2.2. Математическое ожидание, условное математическое ожидание, дисперсия, производящая функция, характеристическая функция

Tales of the expected value[1]

Сказки об ожидаемых величинах

(Из серии «Фильмы, которые не вышли на большой экран».)

They usually have difficult or threatening names such as Bernoulli, De Moivre—Laplace, Chebychev, Poisson. Where are the probabilists with names such as Smith, Brown, or Johnson?

У них обычно трудные или даже пугающие фамилии, такие как Бернулли, Муавр—Лаплас, Чебышёв, Пуассон. Где же вероятностники с такими фамилиями, как Смит, Браун или Джонсон? (...Иванов, Петров, Сидоров)

(Из серии «Почему их не понимают».)

Математическое ожидание, или *среднее*, и *дисперсия* случайной величины X, имеющей плотность распределения f_X, вычисляются аналогично случаю дискретных величин. А именно,

$$\mathbb{E}X = \int x f_X(x)\,dx \quad \text{и} \quad \operatorname{Var} X = \int (x - \mathbb{E}X)^2 f_X(x)\,dx. \tag{2.2.1}$$

Ясно, что

$$\begin{aligned}\operatorname{Var} X &= \int (x^2 - 2x\,\mathbb{E}X + (\mathbb{E}X)^2) f_X(x)\,dx = \\ &= \int x^2 f_X(x)\,dx - 2\,\mathbb{E}X \int x f_X(x)\,dx + (\mathbb{E}X)^2 \int f_X(x)\,dx = \\ &= \mathbb{E}X^2 - 2(\mathbb{E}X)^2 + (\mathbb{E}X)^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2.\end{aligned} \tag{2.2.2}$$

Конечно, эти формулы имеют смысл только тогда, когда интегралы существуют; в случае математического ожидания требуется, чтобы выполнялось равенство $\int |x| f_X(x)\,dx < \infty$. В противном случае, т. е. когда $\int |x| f_X(x)\,dx = \infty$, говорят, что случайная величина X не имеет конечного математического ожидания (или что $\mathbb{E}X$ не существует). Аналогично для $\operatorname{Var} X$.

Замечание. Иногда проводят дальнейшую классификацию, рассматривая вклад, который вносит в интеграл $\int x f_X(x)\,dx$ каждый из интегралов $\int\limits_0^\infty$ и $\int\limits_{-\infty}^0$. Действительно, $\int\limits_0^\infty x f_X(x)\,dx$ соответствует $\mathbb{E}X_+$, а $\int\limits_{-\infty}^0 (-x) f_X(x)\,dx$

[1] Ср. название популярного в США и Великобритании телесериала «Tales of the Unexpected».

соответствует $\mathsf{E}X_-$, где $X_+ = \max[0, X]$ и $X_- = -\min[0, X]$. Работать с интегралами $\int\limits_0^\infty$ и $\int\limits_{-\infty}^0$ проще, так как переменная x сохраняет свой знак на каждом из интервалов $(0, \infty)$ и $(-\infty, 0)$. Тогда говорят, что $\mathsf{E}X = \infty$, если $\int\limits_0^\infty x f_X(x)\,dx = \infty$ и $\int\limits_{-\infty}^0 (-x) f_X(x)\,dx < \infty$. Аналогично $\mathsf{E}X = -\infty$, если $\int\limits_0^\infty x f_X(x)\,dx < \infty$, но $\int\limits_{-\infty}^0 (-x) f_X(x)\,dx = \infty$.

Формулы (2.2.1) и (2.2.2) согласуются со стандартным представлением случайных величин, когда полагают $X(x) = x$ (или $X(\omega) = \omega$). Действительно, мы можем сказать, что для вычисления $\mathsf{E}X$ мы интегрируем величины $X(x)$, а для $\operatorname{Var} X$ — величины $(X(x) - \mathsf{E}X)^2$ относительно плотности $f_X(x)$, что приводит к полной аналогии с формулами для дискретного случая, в котором $\mathsf{E}X = \sum\limits_\omega X(\omega) p_X(\omega)$ и $\operatorname{Var} X = \sum\limits_\omega (X(\omega) - \mathsf{E}X)^2 p(\omega)$.

Пример. Для $X \sim \mathrm{U}(a, b)$ среднее и дисперсия имеют вид

$$\mathsf{E}X = \frac{b+a}{2}, \qquad \operatorname{Var} X = \frac{(b-a)^2}{12}. \tag{2.2.3а}$$

Действительно,

$$\mathsf{E}X = \frac{1}{b-a}\int\limits_a^b x\,dx = \frac{1}{b-a}\left(\frac{x^2}{2}\right)\Big|_a^b = \frac{1}{2}\frac{b^2-a^2}{b-a} = \frac{b+a}{2};$$

это середина интервала (a, b). Далее,

$$\frac{1}{b-a}\int\limits_a^b x^2\,dx = \frac{b^3-a^3}{3(b-a)} = \frac{1}{3}(b^2+ab+a^2).$$

Следовательно, $\operatorname{Var} X = \frac{1}{3}(b^2+ab+a^2) - \frac{1}{4}(b+a)^2 = \frac{(b-a)^2}{12}$. □

Пример. Для $X \sim \mathrm{N}(\mu, \sigma^2)$ мы имеем

$$\mathsf{E}X = \mu, \qquad \operatorname{Var} X = \sigma^2. \tag{2.2.3б}$$

Действительно,

$$\mathsf{E}X = \frac{1}{\sqrt{2\pi}\sigma}\int x \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)dx =$$
$$= \frac{1}{\sqrt{2\pi}\sigma}\int (x-\mu+\mu)\exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)dx =$$

$$= \frac{1}{\sqrt{2\pi}\sigma} \int x \exp\left(-\frac{1}{2\sigma^2}x^2\right) dx + \mu \frac{1}{\sqrt{2\pi}\sigma} \int \exp\left(-\frac{1}{2\sigma^2}x^2\right) dx =$$
$$= 0 + \mu \frac{1}{\sqrt{2\pi}\sigma} \int \exp\left(-\frac{1}{2}x^2\right) dx = \mu$$

и

$$\operatorname{Var} X = \frac{1}{\sqrt{2\pi}\sigma} \int (x-\mu)^2 \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx =$$
$$= \frac{1}{\sqrt{2\pi}\sigma} \int x^2 \exp\left(-\frac{1}{2\sigma^2}x^2\right) dx = \sigma^2 \frac{1}{\sqrt{2\pi}} \int x^2 \exp\left(-\frac{1}{2}x^2\right) dx = \sigma^2.$$

Пример. Для $X \sim \mathrm{Exp}(\lambda)$ мы имеем

$$\mathsf{E}X = \frac{1}{\lambda}, \quad \operatorname{Var} X = \frac{1}{\lambda^2}, \tag{2.2.3в}$$
$$\mathsf{E}X = \lambda \int_0^\infty x e^{-\lambda x}\, dx = \frac{1}{\lambda} \int_0^\infty x e^{-x}\, dx = \frac{1}{\lambda}$$

и

$$\lambda \int_0^\infty x^2 e^{-\lambda x}\, dx = \frac{1}{\lambda^2} \int_0^\infty x^2 e^{-x}\, dx = \frac{2}{\lambda^2};$$

отсюда следует, что $\operatorname{Var} X = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$. □

Пример. Для $X \sim \mathrm{Gam}(\alpha, \lambda)$ мы имеем

$$\mathsf{E}X = \frac{\alpha}{\lambda}, \ \operatorname{Var} X = \frac{\alpha}{\lambda^2}. \tag{2.2.3г}$$

Действительно,

$$\mathsf{E}X = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x x^{\alpha-1} e^{-\lambda x}\, dx = \frac{1}{\lambda\Gamma(\alpha)} \int_0^\infty x^\alpha e^{-x}\, dx = \frac{\Gamma(\alpha+1)}{\lambda\Gamma(\alpha)} = \frac{\alpha}{\lambda}.$$

Далее,

$$\frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^2 x^{\alpha-1} e^{-\lambda x}\, dx = \frac{1}{\lambda^2\Gamma(\alpha)} \int_0^\infty x^{\alpha+1} e^{-x}\, dx = \frac{\Gamma(\alpha+2)}{\lambda^2\Gamma(\alpha)} = \frac{(\alpha+1)\alpha}{\lambda^2},$$

откуда заключаем, что $\operatorname{Var} X = \frac{(\alpha+1)\alpha}{\lambda^2} - \frac{\alpha^2}{\lambda^2} = \frac{\alpha}{\lambda^2}$. □

Пример. Наконец, для $X \sim \mathrm{Ca}(\alpha, \tau)$ находим

$$\int \frac{\tau|x|}{\tau^2 + (x-\alpha)^2}\, dx = \infty, \tag{2.2.3д}$$

что означает, что $\mathsf{E}X$ не существует (не говоря уже о $\operatorname{Var} X$). □

Таблица некоторых вероятностных распределений приведена в конце книги (благодаря любезности Д. Вишика).

В общем случае, когда распределение случайной величины X имеет дискретную и абсолютно непрерывную компоненты, формулы (2.2.1), (2.2.2) необходимо модифицировать. Например, для случайной величины W из формулы (2.1.15) имеем

$$\mathsf{E}\,W = 0 \cdot \mathsf{P}(W = 0) + \frac{\lambda}{\mu}\int\limits_0^\infty xe^{-(\mu-\lambda)x}\,dx = \frac{\lambda}{\mu(\mu-\lambda)^2}.$$

Важным свойством математического ожидания является аддитивность:

$$\mathsf{E}(X+Y) = \mathsf{E}X + \mathsf{E}Y. \tag{2.2.4}$$

Чтобы проверить этот факт, воспользуемся совместной плотностью распределения $f_{X,Y}$:

$$\begin{aligned}\mathsf{E}X + \mathsf{E}Y &= \int\limits_{\mathbb{R}} xf_X(x)\,dx + \int\limits_{\mathbb{R}} yf_Y(y)\,dy = \\ &= \int\limits_{\mathbb{R}} x\int\limits_{\mathbb{R}} f_{X,Y}(x,y)\,dy\,dx + \int\limits_{\mathbb{R}} y\int\limits_{\mathbb{R}} f_{X,Y}(x,y)\,dx\,dy = \\ &= \int\limits_{\mathbb{R}}\int\limits_{\mathbb{R}} (x+y)f_{X,Y}(x,y)\,dy\,dx = \mathsf{E}(X+Y).\end{aligned}$$

Как и в дискретном случае, мы хотим подчеркнуть, что (2.2.4) является общим свойством, которое выполняется для любых случайных величин X и Y. Здесь мы получаем это свойство в предположении, что X и Y имеют совместную плотность распределения (а в § 1.4 это свойство было установлено для дискретных случайных величин), однако свойство аддитивности имеет место и в общем случае, независимо от того, являются ли X и Y дискретными или имеют маргинальные или совместные плотности распределений. Полное доказательство требует привлечения понятия интегрирования по Лебегу и будет дано в одном из последующих томов.

Еще одно свойство состоит в следующем: если c — это некоторая константа, то

$$\mathsf{E}(cX) = c\,\mathsf{E}X. \tag{2.2.5}$$

И вновь для случайной величины, имеющей плотность распределения, доказательство получить легко: для $c \neq 0$ имеем

$$\mathsf{E}(cX) = \int\limits_{\mathbb{R}} xf_{cX}(x)\,dx = \int\limits_{\mathbb{R}} x\frac{1}{c}f_X(x/c)\,dx = c\int\limits_{\mathbb{R}} xf_X(x)\,dx;$$

при $c > 0$ последнее равенство получается немедленно, а при $c < 0$ для получения результата необходимо поменять местами пределы интегрирования. Из формул (2.2.4) и (2.2.5), взятых вместе, получаем линейность математического ожидания:

$$\mathsf{E}(c_1X + c_2Y) = c_1\,\mathsf{E}X + c_2\,\mathsf{E}Y. \tag{2.2.6}$$

Это свойство справедливо и для любого счетного набора случайных величин:

$$\mathsf{E}\Bigl(\sum_i c_i X_i\Bigr) = \sum_i c_i\, \mathsf{E}X_i \tag{2.2.7}$$

при условии, что все средние значения $\mathsf{E}X_i$ существуют и ряд $\sum\limits_i c_i\, \mathsf{E}X_i$ абсолютно сходится. В частности, для таких случайных величин $X_1, X_2, \ldots,$ что $\mathsf{E}X_i = \mu$, находим $\mathsf{E}\Bigl(\sum\limits_{i=1}^{n} X_i\Bigr) = n\mu$.

Удобной (и часто используемой) формулой является *правило незадачливого статистика*

$$\mathsf{E}g(X) = \int\limits_{\mathbb{R}} g(x) f_X(x)\, dx. \tag{2.2.8}$$

Эта формула выражает среднее значение величины $Y = g(X)$, т. е. функции от случайной величины X, через плотность распределения величины X. Формула справедлива при условии, что интеграл в правой части существует, т. е. $\int\limits_{\mathbb{R}} |g(x)| f_X(x)\, dx < \infty$.

Полное доказательство формулы (2.2.8) вновь требует применения интегрирования по Лебегу, но основано оно на простых рассуждениях. Предположим вначале, что функция g дифференцируема и $g' > 0$ (т. е. g монотонно возрастает и, следовательно, обратима: для любого y из области значений g существует единственное такое значение $x = x(y)\ (= g^{-1}(y)) \in \mathbb{R}$, что $g(x) = y$). Благодаря формуле (2.1.17) мы имеем

$$\mathsf{E}g(X) = \int\limits_{\mathbb{R}} y f_{g(X)}(y)\, dy = \int\limits_{\mathrm{Range}(g)} y f_X(x(y)) \frac{1}{g'(x(y))}\, dy.$$

Замена переменных $y = g(x)$, где $x(g(x)) = x$, приводит последний интеграл к виду

$$\int\limits_{\mathbb{R}} g'(x) g(x) f_X(x) \frac{1}{g'(x)}\, dx = \int\limits_{\mathbb{R}} g(x) f_X(x)\, dx,$$

что и дает правую часть формулы (2.2.8). Подобные аргументы применимы и в случае, когда $g' < 0$: знак «минус» можно компенсировать перестановкой пределов интегрирования $-\infty$ и ∞.

Идея доказательства сохраняется и в случае, когда функция g непрерывно дифференцируема и ее производная g' имеет дискретное множество нулей (т. е. конечное или счетное множество, но без точек накопления в $\mathbb{R}$). Это условие охватывает, например, все многочлены. Тогда мы можем разбить прямую $\mathbb{R}$ на интервалы, на которых g' сохраняет знак, и повторить наши рассуждения на каждом таком интервале. Затем суммирование по всем этим интервалам опять приводит к формуле (2.2.8).

Еще одно полезное (и простое) наблюдение состоит в следующем: равенство (2.2.8) выполняется для любой кусочно постоянной функции g, принимающей конечное или счетное число значений y_j: в этом случае

$$\mathsf{E}g(X)=\sum_j y_j\,\mathsf{P}(g(X)=y_j)=\sum_j y_j \int f_X(x)I(x\colon g(x)=y_j)\,dx=\int_{\mathbb{R}} g(x)f_X(x)\,dx.$$

Используя свойство линейности, формулу (2.2.8) можно распространить на класс функций g, непрерывно дифференцируемых на конечном или счетном множестве интервалов, образующих разбиение $\mathbb{R}$, например

$$g(x)=\begin{cases}0, & x<0,\\ x, & x\geqslant 0.\end{cases}$$

Этот класс функций охватывает все приложения, которые рассматриваются в этом томе.

Примерами применения формулы (2.2.8) являются следующие равенства:

$$\mathsf{E}I(a<X<b)=\int_a^b f_X(x)\,dx=\mathsf{P}(a<X<b)$$

и

$$\mathsf{E}XI(a<X<b)=\int_a^b xf_X(x)\,dx.$$

Аналогичная формула справедлива для математического ожидания $\mathsf{E}g(X, Y)$, это математическое ожидание можно выразить в виде двумерного интеграла по совместной п. р. в. $f_{X,Y}$:

$$\mathsf{E}g(X, Y)=\int_{\mathbb{R}^2} g(x, y)f_{X,Y}(x, y)\,dx\,dy. \tag{2.2.9}$$

Важными примерами функций g являются: а) сумма $x+y$, для которой

$$\mathsf{E}(X+Y)=\int_{\mathbb{R}^2}(x+y)f_{X,Y}(x, y)\,dx\,dy=\int_{\mathbb{R}} xf_X(x)\,dx+\int_{\mathbb{R}} yf_Y(y)\,dy=\mathsf{E}X+\mathsf{E}Y$$

(ср. (2.2.4)), и б) произведение $g(x, y)=xy$, для которого получаем

$$\mathsf{E}(XY)=\int_{\mathbb{R}^2} xyf_{X,Y}(x, y)\,dx\,dy. \tag{2.2.10}$$

Отметим, что имеет место неравенство Коши—Шварца (1.4.18):

$$|\,\mathsf{E}(XY)|\leqslant(\mathsf{E}X^2\,\mathsf{E}Y^2)^{1/2},$$

так как при его доказательстве в § 1.4 использовалось лишь свойство линейности математического ожидания.

Ковариация $\mathrm{Cov}(X, Y)$ случайных величин X и Y определяется формулой

$$\mathrm{Cov}(X, Y) = \int_{\mathbb{R}^2} (x - \mathsf{E}X)(y - \mathsf{E}Y) f_{X,Y}(x, y)\, dx\, dy = \mathsf{E}(X - \mathsf{E}X)(Y - \mathsf{E}Y) \tag{2.2.11}$$

и совпадает с

$$\int_{\mathbb{R}} \int_{\mathbb{R}} f_{X,Y}(x, y)(xy - \mathsf{E}X\,\mathsf{E}Y)\, dy\, dx = \mathsf{E}(XY) - \mathsf{E}X\,\mathsf{E}Y. \tag{2.2.12}$$

В силу неравенства Коши—Шварца $|\mathrm{Cov}(X, Y)| \leqslant (\mathrm{Var}\,X)^{1/2}(\mathrm{Var}\,Y)^{1/2}$, при этом равенство имеет место тогда и только тогда, когда X и Y связаны отношением пропорциональности, т. е. $X = cY$, где c — (действительная) константа.

Как и в дискретном случае,

$$\mathrm{Var}(X + Y) = \mathrm{Var}\,X + \mathrm{Var}\,Y + 2\,\mathrm{Cov}(X, Y) \tag{2.2.13}$$

и

$$\mathrm{Var}(cX) = c^2\,\mathrm{Var}\,X. \tag{2.2.14}$$

Если случайные величины X и Y независимы и имеют конечные средние значения, то

$$\mathsf{E}(XY) = \mathsf{E}X\,\mathsf{E}Y. \tag{2.2.15}$$

Доказательство подобно тому, что было приведено для дискретных случайных величин: в силу равенства (2.2.8) мы получаем

$$\mathsf{E}(XY) = \int_{\mathbb{R}^2} yx f_X(x) f_Y(y)\, dx\, dy = \int_{\mathbb{R}} x f_X(x)\, dx \int_{\mathbb{R}} y f_Y(y)\, dy = \mathsf{E}X\,\mathsf{E}Y.$$

Немедленное следствие для независимых случайных величин таково:

$$\mathrm{Cov}(X, Y) = 0, \tag{2.2.16}$$

значит,

$$\mathrm{Var}(X + Y) = \mathrm{Var}\,X + \mathrm{Var}\,Y. \tag{2.2.17}$$

Однако, как и прежде, ни одно из равенств (2.2.15), (2.2.16) не влечет за собой независимость величин (см. задачи 2.2.1 и 2.2.14).

Обобщая, находим, что для любого конечного или счетного набора независимых случайных величин и действительных чисел имеет место формула

$$\mathrm{Var}\Big(\sum_i c_i X_i\Big) = \sum_i c_i^2\,\mathrm{Var}\,X_i \tag{2.2.18}$$

при условии, что все дисперсии $\operatorname{Var} X_i$ существуют и $\sum_i c_i^2 \operatorname{Var} X_i < \infty$. В частности, для таких н. о. р. с. в. $X_1, X_2, \ldots$, что $\operatorname{Var} X_i = \sigma^2$, находим $\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) = n\sigma^2$.

Коэффициент корреляции двух случайных величин X, Y определяется как

$$\operatorname{Corr}(X, Y) = \frac{\operatorname{Cov}(X, Y)}{\sqrt{\operatorname{Var} X}\sqrt{\operatorname{Var} Y}}. \tag{2.2.19}$$

Поскольку $|\operatorname{Cov}(X, Y)| \leqslant (\operatorname{Var} X)^{1/2}(\operatorname{Var} Y)^{1/2}$, справедливы равенства $-1 \leqslant \operatorname{Corr}(X, Y) \leqslant 1$. Более того, $\operatorname{Corr}(X, Y) = 0$, когда X и Y независимы (но не «тогда и только тогда», т. е. утверждения эти не являются равносильными), и $\operatorname{Corr}(X, Y) = 1$ тогда и только тогда, когда $X = cY$ и $c > 0$, и равен -1 тогда и только тогда, когда $X = cY$ и $c < 0$.

Пример. Для пары с. в. X, Y, имеющей двумерное нормальное распределение с совместной плотностью распределения $f_{X,Y}$ вида (2.1.24), параметр $r \in [-1, 1]$ представляет собой коэффициент корреляции $\operatorname{Corr}(X, Y)$. Более точно,

$$\operatorname{Var} X = \sigma_1^2, \quad \operatorname{Var} Y = \sigma_2^2 \quad \text{и} \quad \operatorname{Cov}(X, Y) = r\sigma_1\sigma_2. \tag{2.2.20}$$

На самом деле утверждение о дисперсиях получается немедленно, поскольку $X \sim \mathrm{N}(\mu_1, \sigma_1^2)$ и $Y \sim \mathrm{N}(\mu_2, \sigma_2^2)$.

Далее, для ковариации находим

$$\begin{aligned}
\operatorname{Cov}(X, Y) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \iint (x-\mu_1)(y-\mu_2) \exp\left(-\frac{1}{2(1-r^2)} \times \right.\\
&\qquad \left. \times \left[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2r\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right) dy\, dx = \\
&= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \iint xy \exp\left(-\frac{1}{2(1-r^2)}\left[\frac{x^2}{\sigma_1^2} - 2r\frac{xy}{\sigma_1\sigma_2} + \frac{y^2}{\sigma_2^2}\right]\right) dy\, dx = \\
&= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \int x \int y \exp\left(-\frac{1}{2(1-r^2)}\left[\frac{x^2(1-r^2)}{\sigma_1^2} + \frac{y_1^2}{\sigma_2^2}\right]\right) dy\, dx,
\end{aligned}$$

где теперь

$$y_1 = y - rx\frac{\sigma_2}{\sigma_1}.$$

Следовательно,

$$\begin{aligned}
\operatorname{Cov}(X, Y) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \int x e^{-x^2/2\sigma_1^2} \int \left(y_1 + rx\frac{\sigma_2}{\sigma_1}\right) \times \\
&\qquad \times \exp\left[-\frac{y_1^2}{2\sigma_2^2(1-r^2)}\right] dy_1\, dx = \frac{1}{\sqrt{2\pi}\sigma_1} \int xrx\frac{\sigma_2}{\sigma_1} e^{-x^2/2\sigma_1^2}\, dx = \\
&= r\sigma_1\sigma_2 \frac{1}{\sqrt{2\pi}} \int x^2 e^{-x^2/2}\, dx = r\sigma_1\sigma_2.
\end{aligned}$$

Другими словами, те элементы матриц Σ и Σ^{-1}, что находятся вне диагоналей, совпадают с ковариацией. Следовательно, $\mathrm{Corr}(X, Y) = r$. Легко видеть, что если случайные величины X и Y независимы, то $r = 0$ и обе матрицы Σ и Σ^{-1} являются диагональными. А совместная плотность превращается в произведение маргинальных плотностей:

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2} - \frac{(y-\mu_2)^2}{2\sigma_2^2}\right) = \\ = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-(x-\mu_1)^2/2\sigma_1^2} \frac{1}{\sqrt{2\pi}\sigma_2} e^{-(y-\mu_2)^2/2\sigma_2^2}.$$

Важно отметить, что верно и обратное: если $r = 0$, то матрицы Σ и Σ^{-1} являются диагональными и $f_{X,Y}(x, y)$ распадается в произведение. Следовательно, для совместно нормальных случайных величин $\mathrm{Cov}(X, Y) = 0$ тогда и только тогда, когда X и Y независимы. □

Понятие производящей функции распределения (и производящей функции моментов) также вводится для непрерывных случайных величин следующим образом:

$$\begin{gathered} \varphi_X(s) = \mathsf{E}s^X = \int_{\mathbb{R}} s^x f_X(x)\,dx, \quad s > 0, \\ M_X(\theta) = \mathsf{E}e^{\theta X} = \int_{\mathbb{R}} e^{\theta x} f_X(x)\,dx = \varphi_X(e^\theta), \quad \theta \in \mathbb{R}. \end{gathered} \tag{2.2.21}$$

Производящая функция оказывается особенно полезной для неотрицательных случайных величин. Но даже в этом случае φ_X может не существовать для всех $s > 0$. Например, для $X \sim \mathrm{Exp}(\lambda)$ мы имеем

$$\varphi_X(s) = \lambda \int_0^\infty s^x e^{-\lambda x}\,dx = \frac{\lambda}{\lambda - \ln s}, \quad M_X(\theta) = \frac{\lambda}{\lambda - \theta}.$$

Здесь φ_X существует только тогда, когда $\ln s < \lambda$, т. е. $s < e^\lambda$, и $M_X(\theta)$ существует при $\theta < \lambda$. Для случайной величины $X \sim \mathrm{Ca}(\alpha, \tau)$ производящая функция моментов не существует: $\mathsf{E}e^{\theta X} \equiv \infty$, $\theta \in \mathbb{R}$.

В некоторых приложениях вместо аргумента e^θ используют $e^{-\theta}$; функцию

$$L_X(\theta) = \mathsf{E}e^{-\theta X} = M_X(-\theta) = \int_{\mathbb{R}} e^{-\theta x} f_X(x)\,dx, \quad \theta \in \mathbb{R},$$

называют *преобразованием Лапласа* плотности f_X.

С другой стороны, характеристическая функция $\psi_X(t)$, определяемая формулой

$$\psi_X(t) = \mathsf{E}e^{itX} = \int_{\mathbb{R}} e^{itx} f_X(x)\,dx, \quad t \in \mathbb{R}, \tag{2.2.22}$$

существует при любом $t \in \mathbb{R}$ и любой плотности f_X. Более того, $\psi_X(0) = \mathsf{E}\,1 = 1$, и

$$|\psi_X(t)| \leqslant \int |e^{itx}| f_X(x)\,dx = \int f_X(x)\,dx = 1. \tag{2.2.23}$$

Кроме того, ψ_X как функция от t является (равномерно) непрерывной на всей прямой $\mathbb{R}$. Действительно,

$$|\psi_X(t+\delta) - \psi_X(t)| \leqslant \int |e^{i(t+\delta)x} - e^{itx}| f_X(x)\,dx =$$
$$= \int |e^{itx}| \cdot |e^{i\delta x} - 1| f_X(x)\,dx = \int |e^{i\delta x} - 1| f_X(x)\,dx.$$

Правая часть не зависит от $t \in \mathbb{R}$ (равномерность) и стремится к 0 при $\delta \to 0$.

Последнее утверждение выполняется, поскольку $|e^{i\delta x} - 1| \to 0$ и все подынтегральное выражение $|e^{i\delta x} - 1| f_X(x) \leqslant 2 f_X(x)$ ограничено интегрируемой функцией. Формальное доказательство заключается в следующем: зафиксируем $\varepsilon > 0$ и выберем $A > 0$ настолько большим, чтобы выполнялось неравенство

$$2\int_{-\infty}^{-A} f_X(x)\,dx + 2\int_{A}^{\infty} f_X(x)\,dx < \frac{\varepsilon}{2}.$$

Далее, выбираем δ таким малым, чтобы выполнялось неравенство

$$|e^{i\delta x} - 1| < \frac{\varepsilon}{4A} \quad \forall x,\ -A < x < A;$$

этого можно добиться, поскольку $e^{i\delta x} \to 1$ при $\delta \to 0$ равномерно на $(-A, A)$. Затем разбиваем весь интеграл на сумму и оцениваем каждое слагаемое отдельно:

$$\int |e^{i\delta x} - 1| f_X(x)\,dx = \left(\int_{-\infty}^{-A} + \int_{-A}^{A} + \int_{A}^{\infty}\right) |e^{i\delta x} - 1| f_X(x)\,dx \leqslant$$
$$\leqslant 2\left(\int_{-\infty}^{-A} + \int_{A}^{\infty}\right) f_X(x)\,dx + \int_{-A}^{A} |e^{i\delta x} - 1| f_X(x)\,dx \leqslant \frac{\varepsilon}{2} + \frac{2A\varepsilon}{4A} = \varepsilon.$$

Отметим несколько фактов, которые следуют непосредственно из определения.

1. Если $Y = cX + b$, то $\psi_Y(t) = e^{itb}\psi_X(ct)$.

2. Как и в дискретном случае, среднее $\mathsf{E}X$, дисперсию $\operatorname{Var} X$ и старшие моменты $\mathsf{E}X^k$ можно выразить через значения производных функции ψ_X в точке $t = 0$:

$$\mathsf{E}X = \frac{1}{i}\frac{d}{dt}\psi_X(t)\Big|_{t=0}, \tag{2.2.24}$$

$$\operatorname{Var} X = \left(-\frac{d^2}{dt^2}\psi_X(t) + \left(\frac{d}{dt}\psi_X(t)\right)^2\right)\Big|_{t=0}. \tag{2.2.25}$$

3. Характеристическая функция $\psi_X(t)$ однозначно определяет плотность распределения $f_X(x)$: если $\psi_X(t) \equiv \psi_Y(t)$, то $f_X(x) \equiv f_Y(x)$. (Уточним, что

последнее равенство следует понимать так, что f_X и f_Y могут не совпадать на множестве нулевой меры.) Строгое доказательство будет дано в последующих томах.

4. Для независимых случайных величин X и Y справедливо равенство $\psi_{X+Y}(t) = \psi_X(t)\psi_Y(t)$. Доказательство аналогично приведенному для дискретного случая.

Если $X \sim \mathrm{U}(a, b)$, то

$$\psi_X(t) = \frac{1}{b-a}\int_a^b e^{itx}\,dx = \frac{e^{itb} - e^{ita}}{it(b-a)}. \tag{2.2.26а}$$

Если $X \sim \mathrm{N}(\mu, \sigma^2)$, то

$$\psi_X(t) = e^{it\mu - t^2\sigma^2/2}. \tag{2.2.26б}$$

Для доказательства положим $\mu = 0$, $\sigma^2 = 1$, возьмем производные по t и проинтегрируем по частям:

$$\frac{d}{dt}\psi_X(t) = \frac{d}{dt}\frac{1}{\sqrt{2\pi}}\int e^{itx}e^{-x^2/2}\,dx = \frac{1}{\sqrt{2\pi}}\int (ix)e^{itx}e^{-x^2/2}\,dx =$$
$$= \frac{1}{\sqrt{2\pi}}\left(-ie^{-x^2/2}e^{itx}\right)\Big|_{-\infty}^{\infty} - \frac{1}{\sqrt{2\pi}}\int te^{itx}e^{-x^2/2}\,dx = -t\psi_X(t).$$

Таким образом,

$$(\ln\psi_X(t))' = -t, \quad \text{следовательно,} \quad \ln\psi_X(t) = -\frac{t^2}{2} + c, \quad \text{т. е.} \quad \psi_X(t) = e^c e^{-t^2/2}.$$

Поскольку $\psi_X(0) = 1$, мы получаем, что $c = 0$. Характеристическую функцию для случая общих μ и σ^2 находим из сформулированного выше свойства 1, так как $(X - \mu)/\sigma \sim \mathrm{N}(0, 1)$.

Производящая функция моментов может быть вычислена тем же способом: $\mathsf{E}e^{\theta X} = e^{\theta\mu + \sigma^2\theta^2/2}$.

Замечание. В силу единственности случайной величины с заданной характеристической функцией мы вновь находим подтверждение установленному ранее факту, что если случайные величины $X \sim \mathrm{N}(\mu_1, \sigma_1^2)$ и $Y \sim \mathrm{N}(\mu_2, \sigma_2^2)$ независимы, то $X + Y \sim \mathrm{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Если $X \sim \mathrm{Exp}(\lambda)$, то

$$\psi_X(t) = \lambda\int_0^\infty e^{itx}e^{-\lambda x}\,dx = \frac{\lambda}{\lambda - it}. \tag{2.2.26в}$$

Если $X \sim \mathrm{Gam}(\alpha, \lambda)$, то

$$\psi_X(t) = \left(1 - \frac{it}{\lambda}\right)^{-\alpha}. \tag{2.2.26г}$$

Чтобы доказать это утверждение, мы, как и ранее, продифференцируем по t и проинтегрируем по частям:

$$\frac{d}{dt}\psi_X(t) = \frac{d}{dt}\frac{\lambda^\alpha}{\Gamma(\alpha)}\int\limits_0^\infty e^{itx}x^{\alpha-1}e^{-\lambda x}\,dx = \frac{\lambda^\alpha}{\Gamma(\alpha)}\int\limits_0^\infty ix^\alpha e^{(it-\lambda)x}\,dx =$$

$$= \frac{\lambda^\alpha}{\Gamma(\alpha)}\frac{i\alpha}{\lambda - it}\int\limits_0^\infty x^{\alpha-1}e^{(it-\lambda)x}\,dx = \frac{i\alpha}{\lambda - it}\psi_X(t),$$

т. е.

$$(\ln\psi_X(t))' = (-\alpha\ln(\lambda - it))', \quad \text{откуда} \quad \psi_X(t) = c(\lambda - it)^{-\alpha}.$$

Поскольку $\psi_X(0) = 1$, мы получаем, что $c = \lambda^\alpha$. Это приводит к требуемому результату.

Замечание. Используя вновь свойство единственности случайной величины с заданной характеристической функцией, можно видеть, что если $X \sim \mathrm{Gam}(\alpha, \lambda)$, $Y \sim \mathrm{Gam}(\alpha', \lambda)$ и величины X и Y независимы, то $X + Y \sim \mathrm{Gam}(\alpha + \alpha', \lambda)$. Кроме того, если $X_1, \ldots, X_n$ являются н. о. р. с. в. и $X_i \sim \mathrm{Exp}(\lambda)$, то $X_1 + \ldots + X_n \sim \mathrm{Gam}(n, \lambda)$.

Для того чтобы найти характеристическую функцию распределения Коши, мы сделаем отступление. В математическом анализе функция

$$t \in \mathbb{R} \mapsto \int\limits_{\mathbb{R}} e^{itx}f(x)\,dx, \quad t \in \mathbb{R},$$

называется *преобразованием Фурье* функции f и обозначается $\hat{f}$. *Обратное* преобразование Фурье — это функция

$$x \in \mathbb{R} \mapsto \frac{1}{2\pi}\int\limits_{\mathbb{R}} e^{-itx}\hat{f}(t)\,dt, \quad x \in \mathbb{R}.$$

Если f_X является плотностью распределения, т. е. $f_X \geqslant 0$ и $\int\limits_{\mathbb{R}} f_X(x)\,dx = 1$, и для ее преобразования Фурье $\hat{f} = \psi_X$ выполняется неравенство $\int\limits_{\mathbb{R}} |\hat{f}(t)|\,dt < \infty$, то обратное преобразование Фурье функции ψ_X совпадает с f_X:

$$f_X(x) = \frac{1}{2\pi}\int\limits_{\mathbb{R}} e^{-itx}\psi_X(t)\,dt.$$

Более того, если нам известно, что п. р. в. f_X представляет собой обратное преобразование Фурье некоторой функции $\psi(t)$:

$$f_X(x) = \frac{1}{2\pi}\int\limits_{\mathbb{R}} e^{-itx}\psi(t)\,dx,$$

то $\psi(t) \equiv \psi_X(t)$. (Это последнее утверждение есть не что иное, как другая формулировка утверждения о единственности плотности распределения, соответствующей заданной характеристической функции.)

Преобразование Фурье названо в честь Ж. Б. Ж. Фурье (J. B. J. Fourier) (1768—1830), плодовитого ученого, который играл также активную и выдающуюся роль в политической и административной жизни Франции. Он был префектом нескольких французских департаментов. Упомянем в особенности Гренобль, где Фурье был чрезвычайно популярен и где о нем помнят по сегодняшний день. В 1798 г. Фурье отправился с наполеоновской экспедицией в Египет, где принимал активное участие в научной идентификации многих древних сокровищ. К своему несчастью, он был схвачен военными силами Британии и пребывал некоторое время военнопленным. До конца своей жизни он оставался верным бонапартистом. Однако, когда Наполеон в 1814 г. на своем пути от острова Эльба в Париж проходил Гренобль, Фурье развернул активную кампанию, направленную против императора, поскольку он был убежден, что Франции нужен мир, а не очередная военная авантюра.

В математике Фурье, изучая задачи о передаче тепла, создал целую область, которая сегодня носит название «анализ Фурье». Анализ Фурье имеет чрезвычайное значение и применяется буквально во всех теоретических и прикладных областях. Общепризнанно также, что Фурье является создателем такой области, как современная статистика в социологических науках.

Рассмотрим теперь п. р. в.

$$f(x) = \frac{1}{2}e^{-|x|}, \qquad x \in \mathbb{R}.$$

Соответствующая характеристическая функция равна полусумме характеристических функций «положительной» и «отрицательной» показательных с. в. с параметром $\lambda = 1$:

$$\frac{1}{2}\int\limits_{\mathbb{R}} e^{itx}e^{-|x|}\,dx = \frac{1}{2}\left(\frac{1}{1-it} + \frac{1}{1+it}\right) = \frac{1}{1+t^2},$$

что равно, с точностью до скалярного множителя, плотности распределения Коши с параметрами (1, 0). Следовательно, обратное преобразование Фурье функции $e^{-|t|}$ равно

$$\frac{1}{2\pi}\int\limits_{\mathbb{R}} e^{-itx}e^{-|t|}\,dt = \frac{1}{\pi}\cdot\frac{1}{1+x^2}.$$

Из приведенного выше наблюдения следует, что для с. в. $X \sim \mathrm{Ca}(1, 0)$ характеристическая функция имеет вид $\psi_X(t) = e^{-|t|}$. Для с. в. $X \sim \mathrm{Ca}(\alpha, \tau)$ имеем

$$\psi_X(t) = e^{i\alpha t - \tau|t|}, \tag{2.2.26д}$$

так как $\frac{1}{\tau}(X - \alpha) \sim \mathrm{Ca}(1, 0)$.

Замечание. Тот факт, что у характеристической функции $\psi_X(t)$ случайной величины $X \sim \mathrm{Ca}(\alpha, \tau)$ не существует производной в точке $t = 0$, отражает то свойство, что с. в. X не имеет конечного математического ожидания. С другой стороны, если случайные величины $X \sim \mathrm{Ca}(\alpha_1, \tau_1)$ и $Y \sim \mathrm{Ca}(\alpha_2, \tau_2)$ независимы, то $X + Y \sim \mathrm{Ca}(\alpha_1 + \alpha_2, \tau_1 + \tau_2)$.

Задача 2.2.1. Пусть случайная точка равномерно распределена в круге D, так что вероятность того, что точка попадает в подмножество $A \subseteq D$, пропорциональна площади этого подмножества. Пусть R — это расстояние от точки до центра круга.

Найдите (кумулятивную) функцию распределения $F_R(x) = \mathrm{P}(R < x)$, математическое ожидание $\mathrm{E}R$ и дисперсию $\mathrm{Var}\, R = \mathrm{E}R^2 - (\mathrm{E}R)^2$.

Пусть Θ — это угол, образованный радиусом круга, проведенным через случайную точку, и горизонтальной прямой. Докажите, что R и Θ — независимые с. в.

Введем систему координат с началом в центре круга D. Пусть X и Y — горизонтальная и вертикальная координаты случайной точки.

Найдите ковариацию $\mathrm{Cov}(X, Y) = \mathrm{E}XY - \mathrm{E}X\,\mathrm{E}Y$ и определите, являются ли величины X и Y независимыми.

Вычислите сумму математических ожиданий $\mathrm{E}\dfrac{X}{R} + i\,\mathrm{E}\dfrac{Y}{R}$. Покажите, что ее можно записать как математическое ожидание $\mathrm{E}e^{i\xi}$, и определите с. в. ξ.

Решение. Мы имеем

$$F_R(x) = \mathrm{P}(R < x) = \begin{cases} 0, & \text{если } x \leqslant 0, \\ 1, & \text{если } x \geqslant 1, \end{cases}$$

и

$$F_R(x) = \frac{\text{площадь круга радиуса } x}{\text{площадь всего круга}} = x^2, \quad \text{если } 0 \leqslant x \leqslant 1,$$

а п. р. в. равна

$$f_R(x) = 2xI(0 \leqslant x \leqslant 1).$$

Таким образом,

$$\mathrm{E}R = 2\int_0^1 x^2\,dx = \frac{2}{3}, \quad \mathrm{E}R^2 = 2\int_0^1 x^3\,dx = \frac{1}{2}$$

и

$$\mathrm{Var}\, R = \mathrm{E}R^2 - (\mathrm{E}R)^2 = \frac{1}{18}.$$

Площадь, записанная в полярных координатах (ρ, θ), равна

$$\frac{1}{\pi}\rho I(0 \leqslant \rho \leqslant 1) I(0 \leqslant \theta \leqslant 2\pi)\,d\rho\,d\theta = f_R(\rho)\,d\rho\,g_\Theta(\theta)\,d\theta,$$

откуда следует, что величины R и Θ независимы.

Пусть теперь $X = R\cos\Theta$, $Y = R\sin\Theta$. Находим

$$\mathrm{E}XY = \mathrm{E}(R^2\cos\Theta\sin\Theta) = \mathrm{E}(R^2)\,\mathrm{E}\Big(\frac{1}{2}\sin(2\Theta)\Big) = 0,$$

поскольку

$$\mathsf{E}\sin(2\Theta) = \frac{1}{2\pi}\int_0^{2\pi}\sin(2\theta)\,d\theta = 0.$$

Аналогично $\mathsf{E}X = \mathsf{E}Y = 0$. Например, $\mathsf{E}X = \mathsf{E}R\,\mathsf{E}\cos\Theta = 0$. Следовательно, $\mathrm{Cov}(X, Y) = 0$. Однако величины X и Y не являются независимыми:

$$\mathsf{P}\left(X > \frac{\sqrt{2}}{2},\ Y > \frac{\sqrt{2}}{2}\right) = 0, \quad \text{но} \quad \mathsf{P}\left(X > \frac{\sqrt{2}}{2}\right) = \mathsf{P}\left(Y > \frac{\sqrt{2}}{2}\right) > 0.$$

Наконец,

$$\mathsf{E}\frac{X}{R} + i\,\mathsf{E}\frac{Y}{R} = \mathsf{E}\cos\Theta + i\,\mathsf{E}\sin\Theta = \mathsf{E}e^{i\Theta},$$

следовательно, $\xi = \Theta$. □

Задача 2.2.2. Радиоактивный источник испускает частицы в случайном направлении (при этом все направления равновероятны). Этот источник находится на расстоянии d от фотопластины, которая представляет собой бесконечную вертикальную плоскость.

а) При условии, что частица попадает в плоскость, покажите, что горизонтальная координата точки попадания (если начало координат выбирается в точке, ближайшей к источнику) имеет плотность распределения $d/(\pi(d^2 + x^2))$. (Это распределение известно как распределение Коши.)

б) Можно ли вычислить среднее этого распределения?

Решение. а) Используем сферические координаты r, ψ, θ. Частица попадает в фотопластинку, если луч вдоль направления испускания частицы пересекает полусферу, касающуюся плоскости; вероятность этого равна $1/2$. Условное распределение направления испускания частицы является равномерным распределением в этой полусфере. Следовательно, такой угол ψ, что $x/d = \mathrm{tg}\,\psi$ и $x/r = \sin\psi$, имеет равномерное распределение на интервале $(0, \pi)$, и θ имеет равномерное распределение на интервале $(0, 2\pi)$, см. рис. 2.19.

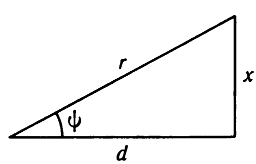


Рис. 2.19

Таким образом, для обратного отображения $g^{-1}\colon \mathbb{R} \to (-\pi/2, \pi/2)$, $x \to \psi(x)$, где

$$\psi(x) = \mathrm{arctg}\left(\frac{x}{d}\right),$$

находим, что абсолютное значение якобиана равно

$$\left|\frac{\partial\psi}{\partial x}\right| = \frac{1}{1+(x/d)^2}\frac{1}{d} = \frac{d}{d^2+x^2}.$$

Тогда п. р. в. горизонтальной координаты точки попадания имеет вид

$$g(x) = \frac{1}{\pi}\left|\frac{\partial\psi}{\partial x}\right| = \frac{d}{\pi(d^2+x^2)}.$$

б) Нет, поскольку соответствующий интеграл расходится и на $+\infty$, и на $-\infty$. □

На самом деле распределение Коши возникает как отношение двух совместно нормальных с. в. А именно, пусть совместная плотность $f_{X,Y}$ имеет вид (2.1.24). Тогда в силу формулы (2.1.28) п. р. в. $f_{X/Y}(u)$ равна

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}\int\limits_{\mathbb{R}}|y|\exp\left[\frac{-y^2}{2\sigma_1^2\sigma_2^2(1-r^2)}(\sigma_2^2u^2-2r\sigma_1\sigma_2u+\sigma_1^2)\right]dy =$$

$$= \frac{1}{\pi\sigma_1\sigma_2\sqrt{1-r^2}}\int\limits_0^\infty y\exp\left[\frac{-y^2}{2(1-r^2)}\frac{(\sigma_2^2u^2-2r\sigma_1\sigma_2u+\sigma_1^2)}{\sigma_1^2\sigma_2^2}\right]dy =$$

$$= \frac{\sigma_1\sigma_2\sqrt{1-r^2}}{\pi(\sigma_2^2u^2-2r\sigma_1\sigma_2u+\sigma_1^2)}\int\limits_0^\infty e^{-y_1}\,dy_1,$$

где

$$y_1 = \frac{y^2}{2(1-r^2)}\frac{\sigma_2^2u^2-2ru\sigma_1\sigma_2+\sigma_1^2}{\sigma_1^2\sigma_2^2}.$$

Мы видим, что

$$f_{X/Y}(u) = \frac{\sqrt{1-r^2}\,\sigma_1/\sigma_2}{\pi((u-r\sigma_1/\sigma_2)^2+(1-r^2)\sigma_1^2/\sigma_2^2)}, \tag{2.2.27}$$

т. е. $X/Y \sim \mathrm{Ca}\left(r\frac{\sigma_1}{\sigma_2}, \sqrt{1-r^2}\frac{\sigma_1}{\sigma_2}\right)$. Если случайные величины независимы, то

$$f_{X/Y}(u) = \frac{\sigma_1/\sigma_2}{\pi(u^2+\sigma_1^2/\sigma_2^2)}, \tag{2.2.28}$$

т. е. $X/Y \sim \mathrm{Ca}\left(0, \frac{\sigma_1}{\sigma_2}\right)$. Ср. с задачей 2.1.14.

Задача 2.2.3. Пусть n приборов одновременно тестируются и времена их безотказной работы — это независимые с. в., имеющие показательное распределение с параметром λ каждая. Определите среднее значение и дисперсию промежутка времени до того момента, пока не выйдут из строя (откажут) r приборов.

Решение. Пусть T_r — это рассматриваемая с. в. Тогда

$$T_r = Y_1 + \ldots + Y_r,$$

где Y_i — это время между $(i-1)$-м и i-м отказом. По свойству отсутствия последействия (потери памяти) показательного распределения $Y_i \sim \mathrm{Exp}((n-i+1)\lambda)$ (поскольку мы ищем минимум оставшихся $n-(i-1)$ независимых экспоненциальных с. в.). Кроме того, случайные величины $Y_1, \ldots, Y_r$ независимы. Следовательно,

$$\mathsf{E}Y_i = \frac{1}{(n-i+1)\lambda}, \ \mathrm{Var}\, Y_i = \frac{1}{(n-i+1)^2\lambda^2},$$

$$\mathsf{E}T_r = \sum_{i=1}^{r} \mathsf{E}Y_i, \ \mathrm{Var}(T_r) = \sum_{i=1}^{r} \mathrm{Var}(Y_i). \qquad \square$$

Эта же задача представляет интерес и для физики и возникает, например, при расчете атомных часов, работающих на эмиссии фотонов. Осложнение состоит в том, что параметр n является случайным (например, $\mathrm{Po}(\lambda)$). Однако основная идея решения все еще применима, см. [KSR].

Задача 2.2.4. Пусть $U_1, U_2, \ldots, U_n$ — н. о. р. с. в. Расположив эти величины в возрастающем порядке, получим последовательность величин $U_{(1)} \leqslant \ldots \leqslant U_{(n)}$, которые называются *порядковыми статистиками*.

а) Пусть $Z_1, Z_2, \ldots, Z_n$ — н. о. р. с. в., имеющие плотность $f(x) = e^{-x}$, $x \geqslant 0$. Покажите, что с. в. $Z_{(1)}$ имеет показательное распределение, и определите ее среднее значение.

б) Пусть $X_1, \ldots, X_n$ — н. о. р. с. в., равномерно распределенные на отрезке $[0, 1]$ с плотностью распределения $f(x) = 1$, $0 \leqslant x \leqslant 1$. Проверьте, что совместная п. р. в. случайных величин $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ имеет вид

$$f_{X_{(1)}, X_{(2)}, \ldots, X_{(n)}}(x_1, \ldots, x_n) = n!\, I(0 \leqslant x_1 \leqslant \ldots \leqslant x_n \leqslant 1).$$

в) Пусть $Z_1, Z_2, \ldots, Z_n$ и $X_1, X_2, \ldots, X_n$ — случайные величины, введенные выше. Докажите, что случайные величины $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ имеют то же самое совместное распределение, что и величины

$$\frac{S_1}{S_{n+1}}, \ \ldots, \ \frac{S_n}{S_{n+1}},$$

где для $1 \leqslant i \leqslant n$ с. в. S_i равна сумме $\sum\limits_{j=1}^{i} Z_j$, а $S_{n+1} = S_n + Z_{n+1}$, $Z_{n+1} \sim \mathrm{Exp}(1)$ и Z_{n+1} не зависит от $Z_1, \ldots, Z_n$.

г) Докажите, что совместное распределение введенных выше с. в. $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ совпадает с совместным условным распределением величин $S_1, S_2, \ldots, S_n$ при условии, что $S_{n+1} = 1$.

Решение. а) Мы имеем $\mathrm{P}(Z_{(1)} > x) = \mathrm{P}(Z_1 > x, \ldots, Z_n > x) = e^{-nx}$. Следовательно, $Z_{(1)}$ — показательная с. в. со средним $1/n$.

б) По определению нам необходимо взять значения $X_1, \ldots, X_n$ и упорядочить их. Фиксированный набор неубывающих величин $x_1, \ldots, x_n$ может быть образован из $n!$ неупорядоченных выборок. Следовательно, совместная п. р. в. $f_{X_{(1)},X_{(2)},\ldots,X_{(n)}}(x_1, \ldots, x_n)$ равна

$$n!\, f(x_1)\ldots f(x_n) I(0 \leqslant x_1 \leqslant \ldots \leqslant x_n \leqslant 1) = n!\, I(0 \leqslant x_1 \leqslant \ldots \leqslant x_n \leqslant 1)$$

для случая равномерного распределения.

в) Совместная п. р. в. $f_{S_1/S_{n+1},\ldots,S_n/S_{n+1},S_{n+1}}$ случайных величин S_1/S_{n+1}, $\ldots, S_n/S_{n+1}$ и S_{n+1} вычисляется следующим образом:

$$\begin{aligned}
&f_{S_1/S_{n+1},\ldots,S_n/S_{n+1},S_{n+1}}(x_1, x_2, \ldots, x_n, t) = \\
&\quad = f_{S_{n+1}}(t)\frac{\partial^n}{\partial x_1 \ldots \partial x_n} \mathrm{P}\left(\frac{S_1}{S_{n+1}} < x_1, \ldots, \frac{S_n}{S_{n+1}} < x_n \;\Big|\; S_{n+1} = t\right) = \\
&\quad = f_{S_{n+1}}(t)\frac{\partial^n}{\partial x_1 \ldots \partial x_n} \mathrm{P}(S_1 < x_1 t, \ldots, S_n < x_n t \mid S_{n+1} = t) = \\
&\quad = t^n f_{S_{n+1}}(t) f_{S_1,\ldots,S_n}(x_1 t, \ldots, x_n t \mid S_{n+1} = t) = \\
&\quad = t^n f_{S_1,\ldots,S_n,S_{n+1}}(t x_1, \ldots, t x_n, t) = \\
&\quad = t^n e^{-t x_1} e^{-t(x_2 - x_1)} \ldots e^{-t(1 - x_n)} I(0 \leqslant x_1 \leqslant \ldots \leqslant x_n \leqslant 1) = \\
&\quad = t^n e^{-t} I(0 \leqslant x_1 \leqslant \ldots \leqslant x_n \leqslant 1).
\end{aligned}$$

Следовательно,

$$f_{S_1/S_{n+1},\ldots,S_n/S_{n+1}}(x_1, x_2, \ldots, x_n) = \int_0^\infty t^n e^{-t}\, dt \cdot I(0 \leqslant x_1 \leqslant \ldots \leqslant x_n \leqslant 1).$$

Это выражение равно $n!\, I(0 \leqslant x_1 \leqslant \ldots \leqslant x_n \leqslant 1)$, т. е. представляет собой совместную п. р. в. величин $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ в силу п. б).

г) Совместная п. р. в. $f_{S_1,\ldots,S_{n+1}}(x_1, \ldots, x_{n+1})$ равна

$$\begin{aligned}
&f_{Z_1,\ldots,Z_{n+1}}(x_1, x_2 - x_1, \ldots, x_{n+1} - x_n) I(0 \leqslant x_1 \leqslant \ldots \leqslant x_n \leqslant x_{n+1}) = \\
&\quad = e^{-x_1} e^{-(x_2 - x_1)} \ldots e^{-(x_{n+1} - x_n)} I(0 \leqslant x_1 \leqslant \ldots \leqslant x_n \leqslant x_{n+1}) = \\
&\quad = e^{-(x_{n+1})} I(0 \leqslant x_1 \leqslant \ldots \leqslant x_n \leqslant x_{n+1}).
\end{aligned}$$

Сумма S_{n+1} имеет п. р. в. $f_{S_{n+1}}(x) = \dfrac{x^n}{n!} e^{-x} I(x \geqslant 0)$. Следовательно, условную плотность можно представить как отношение

$$\frac{f_{(S_1,\ldots,S_{n+1})}(x_1, \ldots, x_n, 1)}{f_{S_{n+1}}(1)} = n!\, I(0 \leqslant x_1 \leqslant \ldots \leqslant x_n \leqslant 1). \qquad \square$$

Задача 2.2.5. Пусть $X_1, X_2, \ldots$ — независимые случайные величины, каждая из которых имеет равномерное распределение на $[0, 1]$. Пусть

$$U_n = \max_{1 \leqslant j \leqslant n} X_j, \qquad V_n = \min_{1 \leqslant j \leqslant n} X_j.$$

Рассматривая

$$\mathsf{P}(v \leqslant X_1 \leqslant v + \delta v,\ v \leqslant X_2, X_3, \ldots, X_{n-1} \leqslant u,\ u \leqslant X_n \leqslant u + \delta u)$$

при $0 < v < u < 1$ (или иным способом), покажите, что пара (U_n, V_n) имеет совместную п. р. в., которая задается формулой

$$f(u, v) = \begin{cases} n(n-1)(u-v)^{n-2}, & \text{если } 0 \leqslant v \leqslant u \leqslant 1, \\ 0 & \text{в противном случае.} \end{cases}$$

Найдите плотности f_{U_n} и f_{V_n}. Покажите, что $\mathsf{P}(1 - 1/n \leqslant U_n \leqslant 1)$ стремится к ненулевому пределу при $n \to \infty$, и найдите этот предел. Что можно сказать о $\mathsf{P}(0 \leqslant V_n \leqslant 1/n)$?

Найдите коэффициент корреляции τ_n случайных величин U_n и V_n. Покажите, что $\tau_n \to 0$ при $n \to \infty$. Почему это можно было бы предвидеть?

Решение. Совместная ф. р. величин U_n и V_n задается формулой

$$\mathsf{P}(U_n < u,\ V_n < v) = \mathsf{P}(U_n < u) - \mathsf{P}(U_n < u,\ V_n \geqslant v) =$$
$$= \mathsf{P}(U_n < u) - \mathsf{P}(v \leqslant X_1 < u, \ldots, v \leqslant X_n < u) = (F(u))^n - (F(u) - F(v))^n.$$

Следовательно, совместная п. р. в. имеет вид

$$f_{U_n, V_n}(u, v) = \frac{\partial^2}{\partial u\, \partial v}[(F(u))^n - (F(u) - F(v))^n] =$$
$$= \begin{cases} n(n-1)(u-v)^{n-2}, & \text{если } 0 \leqslant v \leqslant u \leqslant 1, \\ 0 & \text{иначе.} \end{cases}$$

Маргинальные плотности имеют вид

$$f_{U_n}(u) = \int_0^u n(n-1)(u-v)^{n-2}\, dv = nu^{n-1}$$

и

$$f_{V_n}(v) = \int_v^1 n(n-1)(u-v)^{n-2}\, du = n(1-v)^{n-1}.$$

Тогда вероятность $\mathsf{P}(1 - 1/n \leqslant U_n \leqslant 1)$ равна

$$\int_{1-1/n}^{1} f_{U_n}(u)\, du = n \int_{1-1/n}^{1} u^{n-1}\, du = 1 - \left(1 - \frac{1}{n}\right)^n \to 1 - e^{-1}.$$

Аналогично $\mathsf{P}(0 \leqslant V_n \leqslant 1/n) \to 1 - e^{-1}$.

Далее,

$$\mathsf{E}U_n = \frac{n}{n+1}, \qquad \sqrt{\operatorname{Var} U_n} = \frac{1}{n+1}\sqrt{\frac{n}{n+2}},$$

$$\mathsf{E}V_n = \frac{1}{n+1}, \qquad \sqrt{\operatorname{Var} V_n} = \frac{1}{n+1}\sqrt{\frac{n}{n+2}}.$$

И, наконец, последнее утверждение гласит, что

$$\operatorname{Cov}(U_n, V_n) = \frac{1}{(n+2)(n+1)^2}.$$

Действительно,

$$\operatorname{Cov}(U_n, V_n) = \int_0^1\int_0^u f(u, v)\left(u - \frac{n}{n+1}\right)\left(v - \frac{1}{n+1}\right) dv\,du =$$

$$= \int_0^1\int_0^u n(n-1)(u-v)^{n-2}\left(u - \frac{n}{n+1}\right)\left(v - \frac{1}{n+1}\right) dv\,du =$$

$$= n(n-1)\int_0^1\left(u - \frac{n}{n+1}\right) du \int_0^u (u-v)^{n-2}\left(v - \frac{1}{n+1}\right) dv =$$

$$= n(n-1)\int_0^1\left(u - \frac{n}{n+1}\right)\left[\frac{1}{n(n-1)}u^n - \frac{1}{(n-1)(n+1)}u^{n-1}\right] du =$$

$$= \left[\frac{1}{(n+2)}u^{n+2} - 2\frac{n}{(n+1)^2}u^{n+1} + \frac{n}{(n+1)^2}u^n\right]\Bigg|_0^1 = \frac{1}{(n+1)^2(n+2)}.$$

Следовательно,

$$\operatorname{Corr}(U_n, V_n) = \frac{\operatorname{Cov}(U_n, V_n)}{\sqrt{\operatorname{Var} U_n}\sqrt{\operatorname{Var} V_n}} = O\left(\frac{1}{n}\right) \to 0$$

при $n \to \infty$. □

Задача 2.2.6. Две действительные с. в. X и Y имеют совместную п. р. в.

$$p(x_1, x_2) = \frac{1}{2\pi\sqrt{1-r^2}} \exp[-(1/2(1-r^2))(x_1^2 - 2rx_1x_2 + x_2^2)],$$

где $-1 < r < 1$. Докажите, что каждая из величин X и Y имеет нормальное распределение со средним 0 и дисперсией 1.

Докажите, что число r является коэффициентом корреляции величин X и Y.

Решение. Кумулятивная ф. р. случайной величины X задается формулой

$$\mathsf{P}(X < t) = \int_{-\infty}^{t}\int_{-\infty}^{\infty} p(x_1, x_2)\,dx_2\,dx_1.$$

Интеграл $\int\limits_{-\infty}^{\infty} p(x_1, x_2)\,dx_2$ равен следующему выражению:

$$\frac{1}{2\pi\sqrt{1-r^2}} \int\limits_{-\infty}^{\infty} \exp\left(-\frac{1}{2(1-r^2)}\left[(x_2 - rx_1)^2 + (1-r^2)x_1^2\right]\right) dx_2 =$$

$$= \frac{e^{-x_1^2/2}}{2\pi} \int\limits_{-\infty}^{\infty} \exp\left(-\frac{1}{2}y^2\right) dy, \quad \text{где} \quad y = \frac{x_2 - rx_1}{\sqrt{1-r^2}}.$$

Таким образом,

$$\int\limits_{-\infty}^{\infty} p(x_1, x_2)\,dx_2 = \frac{1}{\sqrt{2\pi}} e^{-x_1^2/2}.$$

Это дает нам п. р. в. величины X. Далее находим, что

$$\mathsf{P}(X < t) = \int\limits_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-x_1^2/2}\,dx_1,$$

т. е. $X \sim \mathrm{N}(0, 1)$. Аналогично $Y \sim \mathrm{N}(0, 1)$.

Следовательно,

$$\mathrm{Corr}(X, Y) = \mathrm{Cov}(X, Y) = \mathsf{E}(XY) = \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} x_1 x_2 p(x_1, x_2)\,dx_2\,dx_1.$$

Теперь

$$\int\limits_{-\infty}^{\infty} x_2 p(x_1, x_2)\,dx_2 =$$

$$= \frac{1}{2\pi\sqrt{1-r^2}} \int\limits_{-\infty}^{\infty} x_2 \exp\left(-\frac{1}{2(1-r^2)}\left[(x_2 - rx_1)^2 + (1-r^2)x_1^2\right]\right) dx_2 = \frac{rx_1 e^{-x_1^2/2}}{\sqrt{2\pi}}.$$

Тогда

$$\mathsf{E}(XY) = \frac{r}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} x_1^2 \exp\left(-\frac{x_1^2}{2}\right) dx_1 = r.$$

Следовательно, $\mathrm{Corr}(X, Y) = r$, что и требовалось показать. □

Задача 2.2.7. Пусть X, Y — независимые с. в. со значениями в полупрямой $[0, \infty)$ и одной и той же п. р. в. $\frac{2}{\sqrt{\pi}} e^{-x^2}$. Пусть $U = X^2 + Y^2$, $V = Y/X$. Вычислите совместную п. р. в. $f_{U,V}$ и докажите, что U и V независимы.

Решение. Совместная п. р. в. случайных величин X и Y равна

$$f_{X,Y}(x, y) = \frac{4}{\pi} e^{-(x^2+y^2)}.$$

Замена переменных $u = x^2 + y^2$, $v = y/x$ образует якобиан

$$\frac{\partial(u, v)}{\partial(x, y)} = \det\begin{pmatrix} 2x & 2y \\ -y/x^2 & 1/x \end{pmatrix} = 2 + 2\frac{y^2}{x^2} = 2(1 + v^2),$$

и обратный якобиан равен

$$\frac{\partial(x, y)}{\partial(u, v)} = \frac{1}{2(1 + v^2)}.$$

Следовательно, совместная п. р. в. имеет вид

$$f_{U,V}(u, v) = f_{X,Y}(x(u, v), y(u, v))\frac{\partial(x, y)}{\partial(u, v)} = \frac{2}{\pi}e^{-u}\frac{1}{1 + v^2}, \qquad u > 0,\ v > 0.$$

Это означает, что случайные величины U и V независимы, $U \sim \mathrm{Exp}(1)$, а $V \sim \mathrm{Ca}(0, 1)$, но величина V сужена на $(0, \infty)$. □

Задача 2.2.8. а) Совместная плотность распределения непрерывных с. в. X и Y задается формулой

$$f(x, y) = \frac{(m + n + 2)!}{m!\,n!}(1 - x)^m y^n$$

при $0 < y \leqslant x < 1$, где m, n — заданные натуральные числа. Проверьте, что f является действительной плотностью распределения (т. е. что интеграл от этой функции равен единице). Найдите маргинальные распределения величин X и Y. Затем вычислите

$$\mathsf{P}\left(Y \leqslant \frac{1}{3} \,\middle|\, X = \frac{2}{3}\right).$$

б) Пусть X и Y — случайные величины. Проверьте, что

$$\mathrm{Cov}(X, Y) = \mathsf{E}[1 - X] \cdot \mathsf{E}[Y] - \mathsf{E}[(1 - X)Y].$$

в) Пусть с. в. X и Y такие, как в п. а). Используя вид функции $f(x, y)$, выразите математические ожидания $\mathsf{E}[1 - X]$, $\mathsf{E}[Y]$ и $\mathsf{E}[(1 - X)Y]$ в терминах факториалов. Используя п. б) (или иным способом), покажите, что ковариация $\mathrm{Cov}(X, Y)$ равна выражению

$$\frac{(m + 1)(n + 1)}{(m + n + 3)^2(m + n + 4)}.$$

Решение. а) Используя формулу $\mathrm{B}(m, n) = \dfrac{\Gamma(m)\Gamma(n)}{\Gamma(m + n)}$, находим

$$\frac{(m + n + 2)!}{m!\,n!}\int_0^1 (1 - x)^m\,dx \int_0^x y^n\,dy = \frac{(m + n + 2)!}{m!\,n!}\int_0^1 (1 - x)^m \frac{x^{n+1}}{n + 1}\,dx =$$

$$= \frac{(m + n + 2)!}{m!\,n!}\frac{1}{n + 1}\mathrm{B}(m + 1, n + 2) = \frac{(m + n + 2)!}{m!\,n!}\frac{1}{n + 1}\frac{m!\,(n + 1)!}{(m + n + 2)!} = 1.$$

Следовательно,

$$\mathsf{P}\left(Y \leqslant \tfrac{1}{3} \;\middle|\; X = \tfrac{2}{3}\right) = \left(\tfrac{1}{3}\right)^m \int\limits_0^{1/3} y^n\,dy \Big/ \left(\left(\tfrac{1}{3}\right)^m \int\limits_0^{2/3} y^n\,dy\right) = \left(\tfrac{1}{2}\right)^{n+1}.$$

б) Прямые преобразования показывают, что

$$\mathsf{E}[1 - X] \cdot \mathsf{E}[Y] - \mathsf{E}[(1 - X)Y] = \mathsf{E}[XY] - \mathsf{E}[X]\,\mathsf{E}[Y].$$

в) Используя определение бета-функции (см. пример 1.1.5 в части В), находим

$$\mathsf{E}[1 - X] = \frac{(m+n+2)!}{m!\,n!}\frac{1}{n+1}\mathrm{B}(m+1,\, n+2) = \frac{m+1}{m+n+3},$$

$$\mathsf{E}[Y] = \frac{(m+n+2)!}{m!\,n!}\frac{1}{n+2}\mathrm{B}(m+1,\, n+3) = \frac{n+1}{m+n+3}$$

и

$$\mathsf{E}[(1 - X)Y] = \frac{(m+n+2)!}{m!\,n!}\frac{1}{n+2}\mathrm{B}(m+2,\, n+3) = \frac{(m+1)(n+1)}{(m+n+3)(m+n+4)}.$$

Следовательно,

$$\mathrm{Cov}(X,\, Y) = \frac{(m+1)(n+1)}{(m+n+3)^2} - \frac{(m+1)(n+1)}{(m+n+3)(m+n+4)} = \\ = \frac{(m+1)(n+1)}{(m+n+3)^2(m+n+4)}. \qquad \square$$

В задаче 2.2.9 используются следующие понятия: *мода* (плотности распределения f) — это точка максимума; и, соответственно, говорят об *унимодальной*, *бимодальной* или *мультимодальной плотностях*.

Задача 2.2.9. Производится выстрел по круглой мишени. Вертикальная и горизонтальная координаты точки попадания (за начало координат выбирается центр мишени) являются независимыми случайными величинами, каждая из которых $\mathrm{N}(0,\, 1)$-распределена.

а) Покажите, что расстояние от точки попадания до центра имеет плотность распределения $re^{-r^2/2}$ при $r > 0$.

б) Покажите, что среднее значение этого расстояния равно $\sqrt{\pi/2}$, медиана равна $\sqrt{\ln 4}$ и мода равна 1.

Решение. (Для п. б); решение п. а) см. в задаче 2.1.12.) По условию задачи $X \sim \mathrm{N}(0,\, 1)$, $Y \sim \mathrm{N}(0,\, 1)$, и совместная п. р. в. имеет вид $f(x,\, y) = \dfrac{1}{2\pi}e^{-(x^2+y^2)/2}$.

Пусть $R = \sqrt{X^2 + Y^2}$, $\Theta = \operatorname{arctg}(y/x)$. Якобиан

$$\frac{\partial(r, \theta)}{\partial(x, y)} = \det\begin{pmatrix} \partial r/\partial x & \partial r/\partial y \\ \partial\theta/\partial x & \partial\theta/\partial y \end{pmatrix}$$

был вычислен в задаче 2.1.12 и равен $1/r$. Следовательно, $\dfrac{\partial(x, y)}{\partial(r, \theta)} = r$. Далее,

$$g(r, \theta) = \frac{r}{2\pi} e^{-r^2/2} I(r > 0) I(0 \leqslant \theta \leqslant 2\pi,\ 0 \leqslant r < \infty).$$

Следовательно,

$$p(r) = re^{-r^2/2} I(r > 0), \qquad \mathsf{E}R = \int_0^\infty r^2 e^{-r^2/2}\, dr = \sqrt{\frac{\pi}{2}}.$$

Наконец, чтобы найти медиану $\hat{x}$, запишем

$$\int_0^{\hat{x}} re^{-r^2/2}\, dr = \int_{\hat{x}}^\infty re^{-r^2/2}\, dr.$$

Тогда

$$1 - e^{-\hat{x}^2/2} = e^{-\hat{x}^2/2}, \qquad \text{значит,} \qquad \hat{x} = \sqrt{\ln 4}.$$

Чтобы найти моду, запишем $f(r) = re^{-r^2/2} I(r > 0)$, и из уравнения $f'(r) = 0$ находим $r = 1$. Следовательно, мода равна 1. □

Задача 2.2.10. Предположим, что $X_1, X_2, \ldots$ образуют последовательность независимых случайных величин, каждая из которых равномерно распределена на $(0, 1)$. Положим

$$N = \min\{n\colon X_1 + X_2 + \ldots + X_n \geqslant 1\}.$$

Покажите, что $\mathsf{E}N = e$.

Решение. Ясно, что N принимает значения 2, 3, … Тогда

$$\mathsf{E}N = \sum_{l \geqslant 2} l\,\mathsf{P}(N = l) = 1 + \sum_{l \geqslant 2} \mathsf{P}(N \geqslant l).$$

Далее, $\mathsf{P}(N \geqslant l) = \mathsf{P}(X_1 + X_2 + \ldots + X_{l-1} < 1)$. Наконец, полагая

$$\mathsf{P}(X_1 + X_2 + \ldots + X_l < y) = q_l(y), \qquad 0 < y < 1,$$

запишем $q_1(y) = y$ и

$$q_l(y) = \int_0^y p_{X_l}(u) q_{l-1}(y - u)\, du = \int_0^y q_{l-1}(y - u)\, du,$$

что приводит к равенствам $q_2(y) = y^2/2!$, $q_3(y) = y^3/3!$, ... Из предположения индукции о том, что $q_{l-1}(y) = y^{l-1}/(l-1)!$, находим

$$q_l(y) = \int_0^y \frac{u^{l-1}}{(l-1)!}\,du = \frac{y^l}{l!}$$

и получаем, что

$$\mathsf{E}N = \sum_{l \geqslant 1} q_l(1) = 1 + \frac{1}{2!} + \frac{1}{3!} + \ldots = e.$$

Или, иным способом, положим $N(x) = \min\{n\colon X_1 + X_2 + \ldots + X_n \geqslant x\}$ и $m(x) = \mathsf{E}N(x)$. Тогда

$$m(x) = 1 + \int_0^x m(u)\,du, \quad \text{следовательно,} \quad m'(x) = m(x).$$

Интегрируя это обыкновенное дифференциальное уравнение с начальным условием $m(0) = 1$, получаем $m(1) = e$. □

Задача 2.2.11. Случайная величина X имеет *логнормальное распределение*, если с. в. $Y = \ln X$ нормально распределена. Если $Y \sim \mathrm{N}(\mu, \sigma^2)$, вычислите, чему равны среднее и дисперсия с. в. X. Логнормальное распределение иногда используется для описания размера малых частиц, претерпевших процесс дробления, или как модель будущих цен на предметы потребления. Когда частичка расщепляется, то образовавшаяся дочерняя частица может представлять собой некоторую часть от родительской частицы; что касается цен, то их колебания могут выражаться в процентном отношении к исходным ценам; когда цены меняются, изменения можно подсчитывать в процентах.

Решение. Мы имеем

$$\mathsf{E}X = e^{\mu + \frac{1}{2}\sigma^2} \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{1/2}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(y - \mu - \sigma^2)^2\right\} dy = e^{\mu + \frac{1}{2}\sigma^2}.$$

Аналогично $\mathsf{E}X^2 = e^{2\mu + 2\sigma^2}$ и $\operatorname{Var} X = e^{2\mu + 2\sigma^2}(1 - e^{-\sigma^2})$.

Или, другим способом, используя производящую функцию моментов, получаем $\mathsf{E}(X^r) = M_Y(r) = e^{r\mu + r^2\sigma^2/2}$. При $r = 1, 2$ мы немедленно находим выражения для $\mathsf{E}X$ и $\mathsf{E}X^2$. □

Задача 2.2.12. Пусть $X_1, X_2, \ldots$ — независимые случайные величины с одной и той же п. р. в. $p(x)$. Будем говорить, что наблюдается рекордное значение в момент времени $n > 1$, если $X_n > \max[X_1, \ldots, X_{n-1}]$. Докажите следующие утверждения.

а) Вероятность того, что рекорд зафиксирован в момент времени n, равна $1/n$.

б) Математическое ожидание числа рекордов до момента времени n равно $\sum\limits_{1<i\leqslant n} 1/i$.

в) Пусть Y_n — это случайная величина, принимающая значение 1, если в момент n зафиксирован рекорд, и значение 0 — в противном случае. Тогда случайные величины $Y_1, Y_2, \ldots$ независимы.

г) Дисперсия числа рекордов до момента времени n равна $\sum\limits_{1<i\leqslant n} (i-1)/i^2$.

е) Если T — момент появления первого рекорда после момента времени 1, то $\mathsf{E}T = \infty$.

Решение. а) (Ср. с задачей 1.4.8.) В силу симметрии

$$\mathsf{P}(\text{рекорд достигнут в момент } n) = \mathsf{P}(X_n = \max[X_1, \ldots, X_n]) = \frac{1}{n}.$$

Тогда для $Y_1, Y_2, \ldots$ получаем

$$\mathsf{P}(Y_n = 1) = \frac{1}{n} = \mathsf{E}Y_n, \qquad \operatorname{Var} Y_n = \frac{1}{n} - \frac{1}{n^2}.$$

Если положить

$$Z_n = \text{ранг } X_n \text{ среди величин } X_1, \ldots, X_n,$$

то

$$\mathsf{P}(Z_i = z) = \frac{1}{i}, \qquad z = 1, \ldots, i.$$

Тогда Y_n — это функция от Z_n. Но величины $Z_1, Z_2, \ldots$ являются независимыми. Действительно, если зафиксировать значения $z_1, \ldots, z_n$ величин $Z_1, \ldots, Z_n$, то тем самым мы упорядочим множество величин (определим некоторую перестановку множества величин) $X_1, \ldots, X_n$, и наоборот. Но все такие перестановки равновероятны, т. е.

$$\mathsf{P}(Z_1 = z_1, \ldots, Z_n = z_n) = \frac{1}{n!} = \mathsf{P}(Z_1 = z_1) \ldots \mathsf{P}(Z_n = z_n),$$

что и означает независимость, так как $\mathsf{P}(Z_1 = z_1) = 1, \ldots, \mathsf{P}(Z_n = z_n) = 1/n$. Таким образом, $Y_1, Y_2, \ldots$ являются независимыми. Поэтому

$$\mathsf{E}(\text{число рекордов до момента времени } n) = \sum_{1<i\leqslant n} \frac{1}{i},$$

$$\operatorname{Var}(\text{число рекордов до момента времени } n) = \sum_{1<i\leqslant n} \left(\frac{1}{i} - \frac{1}{i^2}\right).$$

Наконец,

$$\mathsf{E}T = \sum_{n\geqslant 2} n\,\mathsf{P}(T = n) = 1 + \sum_{n\geqslant 2} \mathsf{P}(T \geqslant n) = \infty.$$

Действительно, из п. а) следует, что $\mathsf{P}(T \geqslant n) = 1/(n-1)$. Или, рассуждая иначе, получаем

$$\mathsf{P}(T \geqslant n) = \mathsf{P}(2, \ldots, n-1 \text{ не являются рекордами}) = \\ = 1 \cdot \frac{1}{2} \cdot \frac{2}{3} \cdot \ldots \cdot \frac{n-2}{n-1} = \frac{1}{n-1}.$$

Отсюда следует, что $\mathsf{E}T = \infty$. Заметим, тем не менее, что с вероятностью 1 появление рекорда будет зафиксировано бесконечно много раз, если наблюдения будут продолжаться бесконечно. □

Задача 2.2.13. Когда говорят, что действительные случайные величины X и Y независимы, то что это означает? И как из совместной плотности распределения $f_{X,Y}$ величин X и Y можно распознать, являются ли эти величины независимыми? Пусть неотрицательные с. в. X и Y имеют совместную п. р. в.

$$f_{X,Y}(x, y) = \frac{1}{2}(x+y)e^{-(x+y)}I(x, y > 0), \quad x, y \in \mathbb{R}.$$

Найдите п. р. в. f_X и выведите отсюда, что с. в. X и Y не являются независимыми. Найдите совместную п. р. в. пары (tX, tY), где t — это положительное число. Предположим теперь, что T — это случайная величина, независимая от X и Y, с п. р. в.

$$p(t) = \begin{cases} 2t, & \text{если } 0 < t < 1, \\ 0 & \text{в противном случае.} \end{cases}$$

Докажите, что TX и TY независимы.

Решение. Ясно, что

$$f_X(x) = \int_0^\infty f_{X,Y}(x, y)\,dy = \frac{1}{2}(1+x)e^{-x}, \quad x > 0.$$

Аналогично $f_Y(y) = (1+y)e^{-y}/2$, $y > 0$. Следовательно, $f_{X,Y} \neq f_X f_Y$, и случайные величины X и Y зависимы.

Далее, для $t, x, y \in \mathbb{R}$ имеем

$$f_{T,X,Y}(t, x, y) = t(x+y)e^{-(x+y)}I(0 < t < 1)I(x, y > 0).$$

Чтобы найти совместную п. р. в. случайных величин TX и TY, мы должны перейти к переменным t, u и v, где $u = tx$, $v = ty$. Находим, что якобиан равен

$$\frac{\partial(t, u, v)}{\partial(t, x, y)} = \det\begin{pmatrix} 1 & x & y \\ 0 & t & 0 \\ 0 & 0 & t \end{pmatrix} = t^2.$$

Следовательно,

$$f_{TX,TY}(u, v) = \int_0^1 t^{-2} f_{T,X,Y}\left(t, \frac{u}{t}, \frac{v}{t}\right) dt = \int_0^1 t^{-2} t \frac{(u+v)}{t} \exp\left(-\frac{u+v}{t}\right) dt.$$

Выполняя замену переменных $t \to \tau = t^{-1}$, получаем

$$f_{TX,TY}(u, v) = \int_1^\infty (u+v) \exp(-\tau(u+v))\, d\tau = e^{-(u+v)}, \quad u, v > 0.$$

Следовательно, случайные величины TX, TY независимы (и имеют показательное распределение со средним 1). □

Задача 2.2.14. Определите ковариацию $\mathrm{Cov}(X, Y)$ случайных величин X и Y. Пусть $\mathrm{Cov}(X, Y)$ — это ковариация случайных величин X и Y. Докажите или опровергните, приведя контрпример, каждое из следующих утверждений.

а) Если $\mathrm{Cov}(X, Y) = 0$, то случайные величины X и Y независимы.

б) Для любых случайных величин X, Y, Z справедливо равенство

$$\mathrm{Cov}(X + Y, Z) = \mathrm{Cov}(X, Z) + \mathrm{Cov}(Y, Z).$$

в) Если X и Y — одинаково распределенные, но не обязательно независимые случайные величины, то

$$\mathrm{Cov}(X + Y, X - Y) = 0.$$

Решение. а) Утверждение ложно. Рассмотрим

$$\mathsf{P}(X = 1, Y = 1) = \frac{1}{4}, \quad \mathsf{P}(X = -1, Y = -1) = \frac{1}{4},$$
$$\mathsf{P}\left(X = 2, Y = -\frac{1}{2}\right) = \frac{1}{4}, \quad \mathsf{P}\left(X = -2, Y = \frac{1}{2}\right) = \frac{1}{4}.$$

Тогда $\mathrm{Cov}(X, Y) = 0$, но $\mathsf{P}(X = 1, Y = 1) = 1/4 \neq \mathsf{P}(X = 1)\,\mathsf{P}(Y = 1) = 1/16$.

б) Утверждение верно.

в) Утверждение верно. Действительно,

$$\mathrm{Cov}(X + Y, X - Y) = \mathsf{E}((X - \mathsf{E}X)^2 - (Y - \mathsf{E}Y)^2) = 0,$$

поскольку X и Y одинаково распределены. □

Задача 2.2.15. Фармацевтическая компания производит порошки лекарств, основанных на химическом веществе аметанол. Эффективность единичной дозы порошка измеряется величиной $-\ln(1 - x)$, где $0 < x < 1$ — доля аметанола в этой единичной дозе и $1 - x$ — доля наполнителя (нейтрального по своему действию)? Вы тестируете выборку из трех единичных доз, которые взяты из трех больших контейнеров. Контейнеры заполнены смесями из аметанола и наполнителя в неизвестных пропорциях.

Определите совместную кумулятивную ф. р. эффективностей единичных доз этих трех фармацевтических смесей. Найдите кумулятивную ф. р. для минимальной эффективности.

Решение. Удобно положить $\omega = (x, y, z)$, где $x, y, z \in [0, 1]$ представляют собой долю аметанола в каждой из единичных доз. Тогда Ω представляет собой единичный куб $\{\omega = (x, y, z)\colon 0 \leqslant x, y, z \leqslant 1\}$. Если X_1, X_2, X_3 — эффективности этих отобранных единичных доз, то

$$X_1(\omega) = -\ln(1 - x), \quad X_2(\omega) = -\ln(1 - y), \quad X_3(\omega) = -\ln(1 - z).$$

Мы предположим, что вероятностная масса распределена равномерно в Ω, т. е. доли X, Y, Z равномерно распределены на $(0, 1)$. Тогда для любого $j = 1, 2, 3$ для кумулятивной ф. р. $F_{X_j}(x) = \mathsf{P}(X_j < y)$ находим

$$\int_0^1 I(-\ln(1 - x) < y)\, dx = \int_0^{1-e^{-y}} dx \cdot I(y > 0) = (1 - e^{-y}) I(y > 0),$$

т. е. $X_j \sim \mathrm{Exp}(1)$. Далее, $\mathsf{P}(\min\limits_j X_j < y) = 1 - \mathsf{P}(\min\limits_j X_j \geqslant y)$,

$$\mathsf{P}(\min_j X_j \geqslant y) = \mathsf{P}(X_1 \geqslant y, X_2 \geqslant y, X_3 \geqslant y) = \prod_j \mathsf{P}(X_j \geqslant y) = e^{-3y}.$$

Таким образом, $\min\limits_j X_j \sim \mathrm{Exp}(3)$.

В этой задаче совместная ф. р. $F_{X_1,X_2,X_3}(y_1, y_2, y_3)$ трех доз лекарства равна объему множества

$$\{(x, y, z)\colon 0 \leqslant x, y, z \leqslant 1, -\ln(1 - x) < y_1, -\ln(1 - y) < y_2, -\ln(1 - z) < y_3\}$$

и совпадает с произведением

$$(1 - e^{-y_1})(1 - e^{-y_2})(1 - e^{-y_3}),$$

т. е. с $F_{X_1}(y_1) F_{X_2}(y_2) F_{X_3}(y_3)$. Совместная п. р. в. также равна произведению:

$$f_{X_1,X_2,X_3}(x_1, x_2, x_3) = e^{-x_1} e^{-x_2} e^{-x_3}. \qquad \square$$

Задача 2.2.16. а) Пусть $X_1, X_2, \ldots$ — это последовательность н. о. р. с. в. с производящей функцией моментов $M(\theta)$, и пусть N — неотрицательная целочисленная случайная величина с производящей функцией $g(s)$; предположим далее, что N не зависит от последовательности $\{X_i\}$. Покажите, что производящая функция моментов случайной величины $Z = X_1 + X_2 + \ldots$ $\ldots + X_n$ равна $g(M(\theta))$.

б) Размеры выплат страховой компании образуют последовательность н. о. р. с. в. с п. р. в. $f(x) = e^{-x}$, $x \geqslant 0$. Число выплат в течение данного года имеет пуассоновское распределение с параметром λ. Покажите, что производящая функция моментов для суммарного объема T выплат в течение года имеет вид

$$h(\theta) = \exp\left[\frac{\lambda\theta}{1-\theta}\right], \quad \theta < 1.$$

Отсюда заключите, что T имеет среднее значение λ и дисперсию 2λ.

Решение. а) Находим производящую функцию моментов:

$$M_Z(\theta) = \mathsf{E}e^{\theta Z} = \mathsf{E}[\mathsf{E}(e^{\theta(X_1+\ldots+X_n)} \mid N=n)] = \sum_{n=0}^{\infty} \mathsf{P}(N=n)(M_{X_1}(\theta))^n = g(M_{X_1}(\theta)),$$

что и требовалось показать.

б) Аналогично находим и производящую функцию моментов для с. в. T:

$$T = X_1 + \ldots + X_N,$$

где с. в. $X_1, \ldots, X_n$ представляют собой размеры выплат, имеют п. р. в. $f_{X_i}(x) = e^{-x}I(x>0)$ и независимы, а N обозначает число выплат и $\mathsf{P}(N=n) = \frac{\lambda^n}{n!}e^{-\lambda}$.

Далее, производящая функция с. в. N равна $g(s) = e^{\lambda(s-1)}$, а производящая функция с. в. X_i — это $M_{X_i}(\theta) = \frac{1}{1-\theta}$. Тогда производящая функция моментов с. в. T равна $h(\theta) = \exp\left[\lambda\left(\frac{1}{1-\theta} - 1\right)\right] = \exp\left[\frac{\lambda\theta}{1-\theta}\right]$, что и требовалось показать. Наконец,

$$h'(0) = \mathsf{E}T = \lambda, \quad h''(0) = \mathsf{E}T^2 = 2\lambda + \lambda^2,$$

и $\operatorname{Var} T = 2\lambda$. □

Задача 2.2.17. Пусть X — показательно распределенная с. в. с плотностью распределения

$$f(x) = \frac{1}{\mu}e^{-x/\mu}, \quad x > 0,$$

где $\mu > 0$. Покажите, что $\mathsf{E}X = \mu$ и $\operatorname{Var} X = \mu^2$.

В некотором эксперименте получены n независимых наблюдений $X_1, \ldots, X_n$ из показательного распределения со средним μ и найдена оценка $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ величины μ. Во втором независимом эксперименте получены m независимых наблюдений $Y_1, \ldots, Y_m$ той же показательной с. в., что и в первом эксперименте, и найдена вторая оценка $\overline{Y} = \frac{1}{m}\sum_{j=1}^{m} Y_j$ величины μ. Эти две оценки затем объединяются в одну оценку вида

$$T_p = p\overline{X} + (1-p)\overline{Y},$$

где $0 < p < 1$.

Найдите $\mathsf{E}T_p$ и $\operatorname{Var}T_p$ и покажите, что для всех $\varepsilon > 0$ выполняется условие

$$\mathsf{P}(|T_p - \mu| > \varepsilon) \to 0 \quad \text{при } m, n \to \infty.$$

Найдите $\bar{p}$ — то значение p, которое минимизирует $\operatorname{Var}T_p$, и проинтерпретируйте $T_{\bar{p}}$. Покажите, что отношение обратных значений дисперсий $\bar{X}$ и $\bar{Y}$ равно $\bar{p}/(1-\bar{p})$.

Решение. Интегрируя по частям, находим

$$\mathsf{E}X = \frac{1}{\mu}\int_0^\infty xe^{-x/\mu}\,dx = (-xe^{-x/\mu})\Big|_0^\infty + \int_0^\infty e^{-x/\mu}\,dx = \mu$$

и

$$\mathsf{E}X^2 = \int_0^\infty x^2\frac{1}{\mu}e^{-x/\mu}\,dx = (-x^2e^{-x/\mu})\Big|_0^\infty + 2\mu\,\mathsf{E}X = 2\mu^2,$$

откуда получаем

$$\operatorname{Var}X = 2\mu^2 - \mu^2 = \mu^2.$$

Далее,

$$\mathsf{E}\bar{X} = \frac{1}{n}\sum_{i=1}^n \mathsf{E}X_i = \mu, \qquad \mathsf{E}\bar{Y} = \frac{1}{m}\sum_{j=1}^m \mathsf{E}Y_j = \mu$$

и

$$\mathsf{E}T_p = p\,\mathsf{E}\bar{X} + (1-p)\,\mathsf{E}\bar{Y} = \mu.$$

Аналогично

$$\operatorname{Var}\bar{X} = \frac{1}{n^2}\sum_{i=1}^n \operatorname{Var}X_i = \frac{\mu^2}{n}, \qquad \operatorname{Var}\bar{Y} = \frac{1}{m^2}\sum_{j=1}^m \operatorname{Var}Y_j = \frac{\mu^2}{m}$$

и

$$\operatorname{Var}T_p = p^2\operatorname{Var}\bar{X} + (1-p)^2\operatorname{Var}\bar{Y} = \mu^2\left(\frac{p^2}{n} + \frac{(1-p)^2}{m}\right).$$

Из неравенства Чебышёва $\mathsf{P}(|Z| \geqslant \varepsilon) \leqslant \frac{1}{\varepsilon^2}\mathsf{E}Z^2$ получаем

$$\mathsf{P}(|T_p - \mu| > \varepsilon) \leqslant \frac{1}{\varepsilon^2}\mathsf{E}(T_p - \mu)^2 = \frac{1}{\varepsilon^2}\operatorname{Var}T_p =$$
$$= \frac{\mu^2}{\varepsilon^2}\left(\frac{p^2}{n} + \frac{(1-p)^2}{m}\right) \to 0 \quad \text{при } n, m \to \infty.$$

Чтобы найти минимум, решим уравнение

$$\frac{d}{dp}\operatorname{Var}T_p = 2\mu^2\left(\frac{p}{n} - \frac{1-p}{m}\right) = 0.$$

Решение имеет вид $\bar{p} = \dfrac{n}{n+m}$. Поскольку

$$\frac{d^2}{dp^2}\operatorname{Var} T_p\Big|_{p=\bar{p}} = 2\mu^2\left(\frac{1}{n} + \frac{1}{m}\right) > 0,$$

мы заключаем, что $\bar{p}$ — точка (глобального) минимума. Тогда

$$T_{\bar{p}} = \frac{n\overline{X} + m\overline{Y}}{n+m}$$

— это среднее значение всех $n+m$ наблюдений. Наконец,

$$\frac{\bar{p}}{1-\bar{p}} = \frac{n}{m} \quad \text{и} \quad \frac{1/\operatorname{Var}\overline{X}}{1/\operatorname{Var}\overline{Y}} = \frac{\mu^2/m}{\mu^2/n} = \frac{n}{m}. \qquad \square$$

Задача 2.2.18. Пусть $X_1, X_2, \ldots$ — н. о. р. с. в. с плотностью распределения

$$f(x) = \begin{cases} \dfrac{\alpha}{x^{\alpha+1}}, & x \geqslant 1, \\ 0, & x < 1, \end{cases}$$

где $\alpha > 0$.

Говорят, что последовательность $\{X_i\}$ *превышает* уровень u в момент j (или что в момент j происходит *выход* последовательности $\{X_i\}$ за уровень u), если $X_j > u$. Пусть $L(u) = \min\{i \geqslant 1 : X_i > u\}$, где $u > 1$ — это момент первого выхода последовательности $\{X_i\}$ за уровень u. Найдите выражение для $\mathsf{P}(L(u) = k)$ при $k = 1, 2, \ldots$ в терминах α и u. Найдите затем математическое ожидание $\mathsf{E}L(u)$ момента первого выхода за уровень u. Покажите, что $\mathsf{E}L(u) \to \infty$ при $u \to \infty$.

Найдите предел при $u \to \infty$ вероятности того, что произойдет выход за уровень u до момента времени $\mathsf{E}L(u)$.

Решение. Легко видеть, что

$$\mathsf{P}(X_i > u) = \int\limits_u^\infty \alpha x^{-\alpha-1}\,dx = (-x^{-\alpha})\Big|_u^\infty = \frac{1}{u^\alpha}.$$

Тогда

$$\mathsf{P}(L(u) = k) = \left(1 - \frac{1}{u^\alpha}\right)^{k-1}\frac{1}{u^\alpha}, \qquad k = 1, 2, \ldots,$$

т. е. $L(u) \sim \mathrm{Geom}(1 - 1/u^\alpha)$ и $\mathsf{E}L(u) = u^\alpha \to \infty$ при $u \to \infty$.

Теперь, обозначая через $[\,\cdot\,]$ целую часть, находим

$$\mathsf{P}(L(u) < \mathsf{E}L(u)) = \mathsf{P}(L(u) \leqslant [u^\alpha]) = 1 - \mathsf{P}(L(u) > [u^\alpha]) =$$

$$= 1 - \sum_{k \geqslant [u^\alpha]+1}\left(1 - \frac{1}{u^\alpha}\right)^{k-1}\frac{1}{u^\alpha} = 1 - \frac{1}{u^\alpha}\left(1 - \frac{1}{u^\alpha}\right)^{[u^\alpha]}\frac{1}{1 - (1 - 1/u^\alpha)} =$$

$$= 1 - \left(1 - \frac{1}{u^\alpha}\right)^{[u^\alpha]} \to 1 - e^{-1} \quad \text{при } u \to \infty. \qquad \square$$

Задача 2.2.19. Пусть A и B — независимые с. в., каждая из которых имеет равномерное распределение на отрезке $[0, 1]$. Пусть $U = \min\{A, B\}$ и $V = \max\{A, B\}$. Найдите среднее значение случайной величины U, а затем ковариацию величин U и V.

Решение. Найдем хвост распределения для $0 \leqslant x \leqslant 1$:

$$1 - F_U(x) = \mathsf{P}(U \geqslant x) = \mathsf{P}(A \geqslant x,\ B \geqslant x) = (1 - x)^2.$$

Отсюда для тех же значений $0 \leqslant x \leqslant 1$ получаем $F_U(x) = 2x - x^2$ и

$$f_U(x) = F'_U(x) I(0 < x < 1) = (2 - 2x) I(0 < x < 1).$$

Поэтому

$$\mathsf{E}U = \int_0^1 x(2 - 2x)\,dx = \left(x^2 - \frac{2}{3}x^3\right)\Big|_0^1 = \frac{1}{3}.$$

Поскольку $U + V = A + B$ и $\mathsf{E}(A + B) = \mathsf{E}A + \mathsf{E}B = 1/2 + 1/2 = 1$, мы имеем

$$\mathsf{E}V = 1 - 1/3 = 2/3.$$

Далее, поскольку $UV = AB$, получаем

$$\mathrm{Cov}(U, V) = \mathsf{E}(UV) - \mathsf{E}U\,\mathsf{E}V = \mathsf{E}(AB) - \frac{2}{9} = \mathsf{E}A\,\mathsf{E}B - \frac{2}{9} = \frac{1}{4} - \frac{2}{9} = \frac{1}{36}. \ \square$$

Задача 2.2.20. Случайные величины X и Y принимают значения из множества $\{0, 1\}$, и их совместное распределение $P(x, y) = \mathsf{P}(X = x, Y = y)$ задается следующим образом:

$$P(0, 0) = a, \quad P(0, 1) = b, \quad P(1, 0) = c, \quad P(1, 1) = d.$$

Найдите необходимые и достаточные условия для того, чтобы X и Y были

а) некоррелированы;

б) независимы.

Являются ли условия, установленные для п. а) и б), эквивалентными?

Решение. а) Находим

$$\mathsf{P}(X = 0) = a + b, \quad \mathsf{P}(Y = 0) = a + c,$$
$$\mathsf{P}(X = 1) = c + d, \quad \mathsf{P}(Y = 1) = b + d.$$

Далее,

$$\mathrm{Cov}(X, Y) = d - (c + d)(b + d).$$

Таким образом, X и Y являются некоррелированными тогда и только тогда, когда $d = (c + d)(b + d)$.

б) Из независимости событий A и B следует независимость событий A и $\overline{B}$, событий $\overline{A}$ и B и событий $\overline{A}$ и $\overline{B}$. Таким образом, X и Y независимы тогда и только тогда, когда $\mathsf{P}(X = 1, Y = 1) = \mathsf{P}(X = 1)\,\mathsf{P}(Y = 1)$, т. е. $d=(c+d)(b+d)$. Таким образом, для заданных величин X и Y (в отличие от общего случая) условия для а) и б) являются эквивалентными. □

Задача 2.2.21. Случайная величина X имеет плотность распределения

$$f(x)=\begin{cases} cx(1-x), & \text{если } 0\leqslant x\leqslant 1,\\ 0 & \text{в противном случае.}\end{cases}$$

Определите c и найдите среднее значение и дисперсию величины X.

Решение. Из равенства

$$1=\int_0^1 cx(1-x)\,dx=c\left(\frac{x^2}{2}-\frac{x^3}{3}\right)\Bigg|_0^1=\frac{c}{6}$$

находим, что $c=6$. Следовательно,

$$\mathsf{E}X=6\int_0^1 x^2(1-x)\,dx=6\frac{1}{12}=\frac{1}{2},$$

$$\mathsf{E}X^2=6\int_0^1 x^3(1-x)\,dx=6\frac{1}{20}=\frac{3}{10}$$

и

$$\operatorname{Var}X=\frac{3}{10}-\frac{1}{4}=\frac{1}{20}.$$

Замечание. Интеграл $\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx$ равен значению $\mathrm{B}(\alpha,\beta)$ бета-функции. См. пример 1.1.5 в части В. □

Задача 2.2.22. Пусть случайные величины X_1 и X_2 независимы и имеют каждая показательное распределение с параметром λ. Найдите совместную плотность распределения величин

$$Y_1=X_1+X_2,\qquad Y_2=X_1/X_2$$

и покажите, что Y_1 и Y_2 независимы. Чему равна плотность распределения величины Y_2?

Решение. Рассмотрим отображение

$$T\colon (x_1,x_2)\mapsto(y_1,y_2),\qquad \text{где } y_1=x_1+x_2,\ y_2=\frac{x_1}{x_2}$$

и $x_1, x_2, y_1, y_2\geqslant 0$. Обратное отображение T^{-1} действует следующим образом:

$$T^{-1}\colon (y_1,y_2)\mapsto(x_1,x_2),\qquad \text{где } x_1=\frac{y_1y_2}{1+y_2},\ x_2=\frac{y_1}{1+y_2},$$

и имеет якобиан

$$J(y_1, y_2) = \det \begin{pmatrix} \dfrac{y_2}{1+y_2} & \dfrac{y_1}{1+y_2} - \dfrac{y_1 y_2}{(1+y_2)^2} \\ \dfrac{1}{1+y_2} & -\dfrac{y_1}{(1+y_2)^2} \end{pmatrix} =$$

$$= -\frac{y_1 y_2}{(1+y_2)^3} - \frac{y_1}{(1+y_2)^3} = -\frac{y_1}{(1+y_2)^2}.$$

Тогда совместная п. р. в. имеет вид

$$f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}\left(\frac{y_1 y_2}{1+y_2}, \frac{y_1}{1+y_2}\right)\left|-\frac{y_1}{(1+y_2)^2}\right|.$$

Подставляя $f_{X_1,X_2}(x_1, x_2) = \lambda e^{-\lambda x_1}\lambda e^{-\lambda x_2}$, $x_1, x_2 \geqslant 0$, находим

$$f_{Y_1,Y_2}(y_1, y_2) = (\lambda^2 y_1 e^{-\lambda y_1})\left(\frac{1}{(1+y_2)^2}\right), \quad y_1, y_2 \geqslant 0.$$

Маргинальные п. р. в. равны

$$f_{Y_1}(y_1) = \int_0^\infty f_{Y_1,Y_2}(y_1, y_2)\, dy_2 = \lambda^2 y_1 e^{-\lambda y_1}, \quad y_1 \geqslant 0,$$

и

$$f_{Y_2}(y_2) = \int_0^\infty f_{Y_1,Y_2}(y_1, y_2)\, dy_1 = \frac{1}{(1+y_2)^2}, \quad y_2 \geqslant 0.$$

Так как $f_{Y_1,Y_2}(y_1, y_2) = f_{Y_1}(y_1) f_{Y_2}(y_2)$, случайные величины Y_1 и Y_2 независимы и $f_{Y_2}(y_2) = 1/(1+y_2)^2$, $y_2 \geqslant 0$. □

§ 2.3. Нормальное распределение. Сходимость случайных величин и распределений. Центральная предельная теорема

Probabilists do it. After all, it's only normal.

Вероятностники делают это.
В конце концов, это только нормально.

(Из серии «Как они делают это».)

Мы уже изучили некоторые свойства нормального распределения. Важное значение этого распределения было признано еще на ранних этапах развития теории вероятности Лапласом, Пуассоном и, конечно же, Гауссом. Однако дальнейшее осознание особой природы нормального распределения происходило постепенно и требовало привлечения фактов и методов из других разделов математики, включая главы анализа и математической физики (особенно комплексного анализа и дифференциальных уравнений в частных производных). В наши дни особое значение придается многомерным нормальным распределениям, которые играют фундаментальную роль всюду, где бы ни использовались вероятностные понятия. Несмотря на появление большого разнообразия других интересных примеров, многомерные нормальные распределения продолжают служить прочным базисом для дальнейших разработок. В частности, нормальное распределение лежит в основании таких областей, как статистика и финансовая математика.

Напомним свойства гауссовского распределения, которые мы уже установили ранее.

1. Плотность распределения $\mathrm{N}(\mu, \sigma^2)$-с. в. X равна

$$\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right), \quad x \in \mathbb{R}, \tag{2.3.1}$$

среднее значение и дисперсия таковы:

$$\mathsf{E}X = \mu, \quad \operatorname{Var} X = \sigma^2, \tag{2.3.2}$$

а производящая функция моментов и характеристическая функция имеют вид

$$\mathsf{E}e^{\theta X} = e^{\theta\mu + (1/2)\theta^2\sigma^2}, \quad \mathsf{E}e^{itX} = e^{it\mu - (1/2)t^2\sigma^2}, \quad \theta, t \in \mathbb{R}. \tag{2.3.3}$$

Если $X \sim \mathrm{N}(\mu, \sigma^2)$, то $\dfrac{X-\mu}{\sigma} \sim \mathrm{N}(0, 1)$ и для любых $b, c \in \mathbb{R}$ мы имеем $cX + b \sim \mathrm{N}(c\mu + b, c^2\sigma^2)$.

2. Две совместно нормальные с. в. X и Y независимы тогда и только тогда, когда $\operatorname{Cov}(X, Y) = \operatorname{Corr}(X, Y) = 0$.

3. Сумма $X + Y$ двух совместно нормальных с. в. $X \sim \mathrm{N}(\mu_1, \sigma_1^2)$ и $Y \sim \mathrm{N}(\mu_2, \sigma_2^2)$ с коэффициентом корреляции $\operatorname{Corr}(X, Y) = r$ является нормальной с. в. со средним $\mu_1 + \mu_2$ и дисперсией $\sigma_1^2 + 2r\sigma_1\sigma_2 + \sigma_2^2$ (см. формулу (2.1.30)). В частности, если X и Y независимы, то $X + Y \sim \mathrm{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. В общем случае, если независимые с. в. $X_1, X_2, \ldots$ таковы, что $X_i \sim \sim \mathrm{N}(\mu_i, \sigma_i^2)$, то их линейная комбинация удовлетворяет условию

$$\sum_i c_i X_i \sim \mathrm{N}\left(\sum_i c_i\mu_i, \sum_i c_i^2\sigma_i^2\right).$$

Задача 2.3.1. Каким должен быть размер выборки из нормального распределения, чтобы с вероятностью не меньше чем 0,99 можно было утверждать, что выборочное среднее находится в пределах одного стандартного отклонения от (теоретического) среднего значения распределения?

Указание. Используйте равенство $\Phi(2{,}58) = 0{,}995$.

Замечание. (Ср. с задачей 1.6.3.) Заметим, что знание того, что распределение нормальное, позволяет значительно уменьшить размер выборки. □

Как уже было отмечено ранее, основной факт, стимулирующий наш интерес к гауссовскому распределению, состоит в том, что это распределение появляется в знаменитой центральной предельной теореме (ЦПТ). Ранняя версия этой теоремы — теорема Муавра—Лапласа — была установлена в § 1.6. Утверждение теоремы Муавра—Лапласа можно распространить на общий случай н. о. р. с. в. Русский математик А. М. Ляпунов (1857—1918) доказал в 1900—1901 гг. следующую теорему.

Пусть $X_1, X_2, \ldots$ *— н. о. р. с. в. с конечным средним* $\mathsf{E}X_j = a$ *и дисперсией* $\operatorname{Var} X_j = \sigma^2$. *Если* $S_n = X_1 + \ldots + X_n$ *— случайная величина со средним* $\mathsf{E}S_n = na$ *и дисперсией* $\operatorname{Var} S_n = n\sigma^2$, *то для любого* $y \in \mathbb{R}$ *выполняется равенство*

$$\lim_{n\to\infty} \mathsf{P}\left(\frac{S_n - \mathsf{E}S_n}{\sqrt{\operatorname{Var} S_n}} < y\right) = \Phi(y). \tag{2.3.4}$$

На самом деле *сходимость в формуле* (2.3.4) *является равномерной по* y:

$$\lim_{n\to\infty} \sup_{y\in\mathbb{R}} \left|\mathsf{P}\left(\frac{S_n - \mathsf{E}S_n}{\sqrt{\operatorname{Var} S_n}} < y\right) - \Phi(y)\right| = 0.$$

А. М. Ляпунов и А. А. Марков были современниками и близкими друзьями. Несмотря на то что Ляпунов был моложе только на год, он считал себя учеником Маркова. Имя Ляпунова стало известным благодаря введенному им классу функций (названных в его честь функциями Ляпунова), широко использующемуся при изучении сходимости к равновесному состоянию в различных случайных и детерминированных системах. Ляпунов погиб трагически, он покончил с собой после смерти своей возлюбленной супруги, испытав лишения и террор во время гражданской войны в России.

Как и в § 1.6, предельное соотношение (2.3.4) принято называть *интегральной* ЦПТ, и его часто записывают как $\dfrac{S_n - \mathsf{E}S_n}{\sqrt{\operatorname{Var} S_n}} \sim \mathrm{N}(0, 1)$. Здесь мы сформулировали ЦПТ для н. о. р. с. в., но современные методы позволяют распространить эту теорему на более общие случаи и получить также точные ограничения на скорость сходимости.

Доказательство ЦПТ для более общих н. о. р. с. в. требует привлечения специальных методов. Один распространенный метод основан на применении характеристических функций, и при этом используется следующий результат, который мы приведем здесь без доказательства. (Это доказательство будет дано в одном из последующих томов.) Если $Y, Y_1, Y_2, \ldots$ — последовательность с. в. с функциями распределения $F_Y, F_{Y_1}, F_{Y_2}, \ldots$ и характеристическими функциями $\psi_Y, \psi_{Y_1}, \psi_{Y_2}, \ldots$, то из сходимости характеристических функций $\psi_{Y_n}(t) \to \psi_Y(t)$ $\forall t \in \mathbb{R}$ следует, что $F_{Y_n}(y) \to F_Y(y)$ во всех точках $y \in \mathbb{R}$, где ф. р. в. F_Y непрерывна.

В нашем случае $Y \sim \mathrm{N}(0, 1)$ с функцией распределения $F_Y = \Phi$, которая всюду на $\mathbb{R}$ непрерывна. Положив $Y_j = (S_n - an)/(\sqrt{n}\sigma)$, мы должны проверить, что $\psi_{Y_n}(t) \to e^{-t^2/2}$ $\forall t \in \mathbb{R}$. Это следует из прямых вычислений, которые мы проводим ниже. Поскольку $(S_n - an)/\sigma = \sum\limits_{1 \leqslant j \leqslant n} (X_j - a)/\sigma$ и случайные величины $(X_j - a)/\sigma$ являются н. о. р., находим

$$\psi_{(S_n - an)/(\sqrt{n}\sigma)}(t) = \left(\psi_{(X_j - a)/\sigma}\left(\frac{t}{\sqrt{n}}\right)\right)^n;$$

в последнем равенстве используется то, что X_j — это н. о. р. с. в.

Случайная величина $(X_j - a)/\sigma$ имеет нулевое среднее и единичную дисперсию. Поэтому ее характеристическая функция допускает следующее разложение в ряд Тейлора в окрестности точки 0:

$$\psi(u) = \psi(0) + u\psi'(0) + \frac{u^2}{2}\psi''(0) + o(u^2) = 1 - \frac{u^2}{2} + o(u^2).$$

(Здесь и далее мы опускаем индекс $\dfrac{X_j - a}{\sigma}$.) Это приводит к соотношению

$$\psi\left(\frac{t}{\sqrt{n}}\right) = 1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right).$$

Но тогда

$$\left(\psi\left(\frac{t}{\sqrt{n}}\right)\right)^n = \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n \to e^{-t^2/2}. \tag{2.3.5}$$

Для краткого доказательства формулы (2.3.5) необходимо положить

$$A\ (= A_n) = 1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right), \quad B\ (= B_n) = 1 - \frac{t^2}{2n},$$

тогда сразу можно заметить, что $B^n \to e^{-t^2/2}$. Далее,

$$A^n - B^n = A^{n-1}(A - B) + A^{n-2}(A - B)B + \ldots + (A - B)B^{n-1},$$

откуда следует, что

$$|A^n - B^n| \leqslant (\max[1, A, B])^{n-1}|A - B|,$$

а это выражение стремится к 0 при $n \to \infty$.

Задача 2.3.2. Ежегодно университет присуждает уровень A приблизительно 16% выпускников-математиков, уровень B и уровень C охватывают примерно по 34%, и к уровню D провалившихся причисляют оставшиеся 16% учащихся. Эти цифры повторяются из года в год, как бы ни менялась действительная успеваемость студентов в каждом году.

Одна выпускница пытается осмыслить такую практику. Она предполагает, что итоговый балл, набранный каждым из кандидатов, — это независимые величины $X_1, \ldots, X_n$, которые отличаются друг от друга только своими средними значениями $\mathsf{E}X_j$, так что «центрированные» величины $X_j - \mathsf{E}X_j$ являются одинаково распределенными. Далее, она считает, что распределение суммарного балла — приблизительно нормальное $\mathrm{N}(\mu, \sigma^2)$. Она приходит к догадке, что вышеописанная практика основана на некотором стандартном разбиении значений суммарного итогового балла студентов на четыре категории. Уровень A присуждается, когда итоговый балл превышает определенное пороговое значение, скажем a, уровень B — когда итоговый балл находится между значениями b и a, уровень C — когда этот балл попадает между c и b, и оставшийся уровень — когда итоговый балл ниже, чем c. Очевидно, что пороговые значения c, b и a могут зависеть от μ и σ.

Спустя некоторое время (поразмыслив и воспользовавшись таблицами) она убеждается, что ее догадка верна, и ей удается найти простые формулы, дающие разумные приближения значений a, b и c. Можете ли вы воспроизвести ее ответ?

Решение. Пусть X_j — итоговый балл кандидата j, $S_n = \sum\limits_{1 \leqslant j \leqslant n} X_j$ — суммарный балл, $S_n - \mathsf{E}S_n = \sum\limits_{1 \leqslant j \leqslant n} (X_j - \mathsf{E}X_j)$ — «центрированный» суммарный балл. Предположим, что $X_j - \mathsf{E}X_j$ — н. о. р. с. в. (что, в частности, означает, что $\operatorname{Var} X_j = \sigma^2$ не зависит от j и $\operatorname{Var} S_n = n\sigma^2$). Тогда, воспользовавшись ЦПТ, получим $\dfrac{S_n - \mathsf{E}S_n}{\sqrt{n}\sigma} \sim \mathrm{N}(0, 1)$. Для суммарного среднего балла S_n/n при больших n должно выполняться условие

$$\frac{S_n}{n} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right) \sim \frac{\sigma}{\sqrt{n}} Y + \mu,$$

где $Y \sim \mathrm{N}(0, 1)$ и

$$\mu = \frac{1}{n} \mathsf{E}S_n = \frac{1}{n} \sum_{1 \leqslant j \leqslant n} \mathsf{E}X_j.$$

Найдем такое значение α, что $\mathsf{P}(Y > \alpha) = 1 - \Phi(\alpha) \approx 0{,}16$; это приводит к значению $\alpha = 1$. Ясно, что $\mathsf{P}(Y > 1) = \mathsf{P}(S_n/n > \sigma/\sqrt{n} + \mu)$. Аналогично из равенства $\mathsf{P}(\beta < Y < 1) = \Phi(1) - \Phi(\beta) \approx 0{,}34$ находим $\beta = 0$. Вполне естественно заключить, что $a = \mu + \sigma$, $b = \mu$ и $c = \mu - \sigma$.

Чтобы обосновать эту догадку, запишем $X_j \sim \mathsf{E}X_j + \sigma Y$. Это влечет за собой следующие соотношения:

$$\frac{S_n}{n} \sim \frac{1}{\sqrt{n}}(X_j - \mathsf{E}X_j) + \mu$$

и

$$\mathsf{P}\left(\frac{1}{\sqrt{n}}(X_j - \mathsf{E}X_j) + \mu > \frac{\sigma}{\sqrt{n}} + \mu\right) = \mathsf{P}(X_j > \mathsf{E}X_j + \sigma) = 0{,}16.$$

Мы не знаем, чему равно $\mathsf{E}X_j$, и используем вместо него μ. Таким образом, категории определяются следующим образом: уровень А присуждается, когда итоговый балл кандидата превосходит $\mu + \sigma$, уровень В — когда он находится между μ и $\mu + \sigma$, С — когда он расположен между $\mu - \sigma$ и μ, и D, или провал, — когда итоговый балл меньше чем $\mu - \sigma$. Здесь значения μ и σ^2 могут быть приближенно вычислены как средние $\dfrac{S_n}{n}$ и $\dfrac{1}{n-1}\sum\limits_{1\leqslant j\leqslant n}\left(X_j - \dfrac{S_n}{n}\right)^2$ (см. задачу 1.4.5). □

Задача 2.3.3. (продолжение задачи 2.3.2). Предположим теперь, что все та же студентка пожелала оценить, насколько точной является ее аппроксимация среднего значения ожидаемого балла μ величиной S_n/n. Предположив, что стандартное отклонение σ индивидуального итогового балла X_j не превосходит 10 баллов, определите, каким должно быть n, чтобы гарантировать, что с вероятностью, не превышающей 0,01, величина S_n/n отклоняется от μ не более чем на 5 баллов.

Решение. Необходимо, чтобы выполнялось неравенство

$$\mathsf{P}\left(\left|\frac{S_n}{n} - \frac{\mathsf{E}S_n}{n}\right| \geqslant 5\right) \leqslant 0{,}01.$$

Положив, как и ранее, $Y \sim \mathrm{N}(0,\ 1)$, получаем, что эта вероятность приближенно равна

$$\mathsf{P}\left(|Y| \geqslant 5\frac{\sqrt{n}}{\sigma}\right).$$

Таким образом, должно выполняться неравенство

$$5\frac{\sqrt{n}}{\sigma} \geqslant \Phi^{-1}(0{,}995), \quad \text{т. е.} \quad n \geqslant \sigma^2\left(\frac{\Phi^{-1}(0{,}995)}{5}\right)^2,$$

где $\sigma^2 \leqslant 100$. Так как $\Phi^{-1}(0{,}995) = 2{,}58$, находим, что заведомо достаточным будет значение

$$n = \frac{100}{25}(\Phi^{-1}(0{,}995))^2 \approx 27.$$ □

Statisticians do all the standard deviations.

Статистики совершают все стандартные отклонения.

(Из серии «Как они делают это».)

В последующих вычислениях будет удобно использовать другое обозначение для скалярного произведения:

$$(\boldsymbol{x} - \boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) = \langle \boldsymbol{x} - \boldsymbol{\mu},\ \Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\rangle = \sum_{i,j=1}^{n}(x_i - \mu_i)\Sigma_{ij}^{-1}(x_j - \mu_j).$$

Задача 2.3.4. Предположим, что $X_1, \ldots, X_n$ — независимые случайные величины, $X_j \sim \mathrm{N}(\mu_j, \sigma^2)$, имеющие одинаковую дисперсию. Рассмотрим случайные величины $Y_1, \ldots, Y_n$, определяемые как

$$Y_j = \sum_{i=1}^{n} a_{ij} X_i, \quad j = 1, \ldots, n,$$

где $\boldsymbol{A} = (a_{ij})$ — действительная ортогональная матрица размера $n \times n$. Докажите, что с. в. $Y_1, \ldots, Y_n$ независимы, и найдите их распределения.

Прокомментируйте случай, когда дисперсии $\sigma_1^2, \ldots, \sigma_n^2$ различны.

(Действительная матрица $\boldsymbol{A}$ размера $n \times n$ называется *ортогональной*, если $\boldsymbol{A} = (\boldsymbol{A}^{-1})^T$.)

Решение. В векторных обозначениях

$$\boldsymbol{Y} := \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \boldsymbol{A}^T \boldsymbol{X}, \quad \text{где} \quad \boldsymbol{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}.$$

В силу ортогональности матрицы $\boldsymbol{A}$ выполняются равенства $\boldsymbol{A}^T = \boldsymbol{A}^{-1}$, $(\boldsymbol{A}^T)^{-1} = \boldsymbol{A}$. Кроме того, якобианы взаимно обратных линейных отображений

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \mapsto \boldsymbol{y} = \boldsymbol{A}^T \boldsymbol{x}, \quad \boldsymbol{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \mapsto \boldsymbol{x} = \boldsymbol{A}\boldsymbol{y}, \quad \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n,$$

равны $+1$ или -1 (и равны между собой). Действительно,

$$\frac{\partial(y_1, \ldots, y_n)}{\partial(x_1, \ldots, x_n)} = |\det \boldsymbol{A}^T|, \quad \frac{\partial(x_1, \ldots, x_n)}{\partial(y_1, \ldots, y_n)} = |\det \boldsymbol{A}|$$

и $\det \boldsymbol{A}^T = \det \boldsymbol{A} = \pm 1$.

Плотность распределения $f_{Y_1, \ldots, Y_n}(\boldsymbol{y})$ равна плотности $f_{X_1, \ldots, X_n}(\boldsymbol{A}\boldsymbol{y})$, которая, в свою очередь, равна

$$\left(\frac{1}{(2\pi)^{1/2}\sigma}\right)^n \prod_{1 \leqslant j \leqslant n} \exp\left[-\frac{1}{2\sigma^2}\left(\sum_{1 \leqslant i \leqslant n} a_{ji} y_i - \mu_j\right)^2\right].$$

Положим

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} \quad \text{и} \quad \sigma^{-2}\boldsymbol{I} = \begin{pmatrix} \sigma^{-2} & 0 & \ldots & 0 \\ 0 & \sigma^{-2} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \sigma^{-2} \end{pmatrix},$$

где I — единичная матрица размера $n \times n$. Если обозначить j-ю компоненту вектора Ay через $(Ay)_j$, то произведение экспонент, записанное выше, можно привести к виду

$$\exp\left[-\frac{1}{2\sigma^2}\sum_{1\leqslant j\leqslant n}((Ay)_j-\mu_j)^2\right]=\exp\left[-\frac{1}{2}(Ay-\mu)^T(\sigma^{-2}I)(Ay-\mu)\right]=$$
$$=\exp\left[-\frac{1}{2}(y-A^T\mu)^TA^T(\sigma^{-2}I)A(y-A^T\mu)\right].$$

Для тройного произведения матриц находим $A^T(\sigma^{-2}I)A=\sigma^{-2}A^TA=\sigma^{-2}I$. Следовательно, последнее выражение равно

$$\exp\left[-\frac{1}{2}(y-A^T\mu)^T\sigma^{-2}I(y-A^T\mu)\right]=\exp\left[-\frac{1}{2\sigma^2}\sum_{1\leqslant j\leqslant n}(y_j-(A^T\mu)_j)^2\right]=$$
$$=\prod_{1\leqslant j\leqslant n}\exp\left[-\frac{1}{2\sigma^2}(y_j-(A^T\mu)_j)^2\right],$$

где, как и ранее, $(A^T\mu)_j=\sum\limits_{1\leqslant i\leqslant n}a_{ij}\mu_i$ обозначает j-ю компоненту вектора $A^T\mu$. Таким образом, случайные величины $Y_1, \ldots, Y_n$ независимы, и $Y_j \sim \mathrm{N}((A^T\mu)_j, \sigma^2)$.

В общем случае, когда дисперсии σ_i^2 различны, матрицу $\sigma^{-2}I$ нужно заменить диагональной матрицей

$$\Sigma^{-1}=\begin{pmatrix}\sigma_1^{-2} & 0 & \ldots & 0\\ 0 & \sigma_2^{-2} & \ldots & 0\\ \ldots & \ldots & \ldots & \ldots\\ 0 & 0 & \ldots & \sigma_n^{-2}\end{pmatrix}. \tag{2.3.6}$$

Случайные величины $Y_1, \ldots, Y_n$ будут независимы тогда и только тогда, когда $A^T\Sigma^{-1}A$ — диагональная матрица. Например, если A коммутирует с Σ^{-1}, т. е. $\Sigma^{-1}A=A\Sigma^{-1}$, то $A^T\Sigma^{-1}A=A^TA\Sigma^{-1}=\Sigma^{-1}$, и в этом случае $Y_j \sim \mathrm{N}((A\mu)_j, \sigma_j^2)$, т. е. дисперсия не изменяется. □

Задача 2.3.4 приведет нас к такому общему утверждению: случайные величины $X_1, \ldots, X_n$, имеющие совместное нормальное распределение, независимы тогда и только тогда, когда $\mathrm{Cov}(X_i, X_j)=0$ для любых i, j, $1\leqslant i<j\leqslant n$. Это утверждение является частью свойства 4, которое мы формулируем ниже. Ранее мы уже установили, что две случайные величины X, Y, имеющие совместное нормальное распределение, независимы тогда и только тогда, когда $\mathrm{Cov}(X, Y)=0$ (см. формулу (2.1.25)).

Напомним (см. формулу (2.1.5)), что п. р. в. $f_{\boldsymbol{X}}(\boldsymbol{x})$ общего многомерного нормального вектора $\boldsymbol{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$ имеет вид

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(\sqrt{2\pi})^n (\det \Sigma)^{1/2}} \exp\Big(-\frac{1}{2}\langle \boldsymbol{x} - \boldsymbol{\mu}, \Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\rangle\Big), \qquad (2.3.7)$$

где Σ — обратимая положительно определенная (и, следовательно, симметричная) действительная матрица размера $n \times n$ и

$$\det \Sigma^{-1} = (\det \Sigma)^{-1}.$$

Если $\boldsymbol{X}$ — нормальный вектор, то будем писать $\boldsymbol{X} \sim \mathrm{N}(\boldsymbol{\mu}, \Sigma)$.

Продолжая список свойств 1—3, приведенный в начале этого параграфа, мы установим еще одно свойство гауссовского распределения.

4. Если $\boldsymbol{X} \sim \mathrm{N}(\boldsymbol{\mu}, \Sigma)$ и п. р. в. $f_{\boldsymbol{X}}$ задана формулой (2.3.7), то

а) $X_i \sim \mathrm{N}(\mu_i, \Sigma_{ii})$, $i = 1, \ldots, n$, т. е. каждая компонента имеет среднее μ_i и дисперсию Σ_{ii} — диагональный элемент матрицы Σ;

б) элементы матрицы Σ_{ij}, расположенные вне диагонали, равны ковариациям $\mathrm{Cov}(X_i, X_j)$ $\forall i, j$, $1 \leqslant i < j \leqslant n$. Таким образом, матрицы Σ и Σ^{-1} являются диагональными, и, следовательно, п. р. в. $f_{\boldsymbol{X}}(\boldsymbol{x})$ распадается в произведение тогда и только тогда, когда $\mathrm{Cov}(X_i, X_j) = 0$, $1 \leqslant i < j \leqslant n$. Другими словами, совместно нормальные с. в. $X_1, \ldots, X_n$ независимы тогда и только тогда, когда $\mathrm{Cov}(X_i, X_j) = 0$, $1 \leqslant i < j \leqslant n$.

Вполне естественно, что вектор $\boldsymbol{\mu}$ называют вектором средних значений и матрицу Σ — ковариационной матрицей многомерного случайного вектора $\boldsymbol{X}$.

Доказательство утверждений 4 а) и 4 б) использует непосредственно вид (2.3.7) совместной многомерной нормальной п. р. в. $f_{\boldsymbol{X}}(\boldsymbol{x})$. Обсудим вначале некоторые предварительные факты из алгебры. (На самом деле из них мы выведем еще больше свойств многомерного нормального распределения.)

В § 2.1 было установлено, что матрица Σ, как положительно определенная матрица, может быть приведена к диагональному виду (если базис составлен из собственных векторов этой матрицы), т. е. существует такая ортогональная матрица $\boldsymbol{B} = (b_{ij})$, что $\boldsymbol{B} = (\boldsymbol{B}^{-1})^T$, $\boldsymbol{B}\boldsymbol{D}\boldsymbol{B}^{-1} = \Sigma$ и $\boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{B}^{-1} = \Sigma^{-1}$, где

$$\boldsymbol{D} = \begin{pmatrix} \lambda_1^2 & 0 & \ldots & 0 \\ 0 & \lambda_2^2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \lambda_n^2 \end{pmatrix}, \quad \boldsymbol{D}^{-1} = \begin{pmatrix} \lambda_1^{-2} & 0 & \ldots & 0 \\ 0 & \lambda_2^{-2} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \lambda_n^{-2} \end{pmatrix}$$

и $\lambda_1^2, \lambda_2^2, \ldots, \lambda_n^2$ — (положительные) собственные числа матрицы Σ. Отметим, что $\det\Sigma = \prod_{i=1}^n \lambda_i^2$ и $(\det\Sigma)^{1/2} = \prod_{i=1}^n \lambda_i$.

Если выполнить ортогональную замену переменных $\boldsymbol{x} \mapsto \boldsymbol{y} = \boldsymbol{B}^T\boldsymbol{x}$, где обратное отображение имеет вид $\boldsymbol{y} \mapsto \boldsymbol{x} = \boldsymbol{B}\boldsymbol{y}$ и якобиан $\dfrac{\partial(x_1, \ldots, x_n)}{\partial(y_1, \ldots, y_n)}$ равен $\det\boldsymbol{B} = \pm 1$, то новые случайные величины

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \boldsymbol{B}^T \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$$

имеют п. р. в. $f_{\boldsymbol{Y}}(\boldsymbol{y}) = f_{\boldsymbol{X}}(\boldsymbol{B}\boldsymbol{y})$. Более подробно,

$$\begin{aligned}
f_{\boldsymbol{Y}}(\boldsymbol{y}) &= \left(\prod_{i=1}^n \frac{1}{\sqrt{2\pi}\lambda_i}\right)\exp\left(-\frac{1}{2}(\boldsymbol{B}\boldsymbol{y} - \boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{B}\boldsymbol{y} - \boldsymbol{\mu})\right) = \\
&= \left(\prod_{i=1}^n \frac{1}{\sqrt{2\pi}\lambda_i}\right)\exp\left(-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{B}^T\boldsymbol{\mu})^T\boldsymbol{B}^T\boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{B}^{-1}\boldsymbol{B}(\boldsymbol{y} - \boldsymbol{B}^T\boldsymbol{\mu})\right) = \\
&= \left(\prod_{i=1}^n \frac{1}{\sqrt{2\pi}\lambda_i}\right)\exp\left(-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{B}^T\boldsymbol{\mu})^T\boldsymbol{D}^{-1}(\boldsymbol{y} - \boldsymbol{B}^T\boldsymbol{\mu})\right) = \\
&= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\lambda_i}\exp\left[-\frac{1}{2\lambda_i^2}(y_i - (\boldsymbol{B}^T\boldsymbol{\mu})_i)^2\right].
\end{aligned}$$

Здесь $(\boldsymbol{B}^T\boldsymbol{\mu})_i$ — это i-я компонента вектора $\boldsymbol{B}^T\boldsymbol{\mu}$. Мы видим, что случайные величины $Y_1, \ldots, Y_n$ независимы и $Y_i \sim \mathrm{N}((\boldsymbol{B}^T\boldsymbol{\mu})_i, \lambda_i^2)$. Таким образом, ковариация равна

$$\mathrm{Cov}(Y_i, Y_j) = \mathsf{E}[Y_i - (\boldsymbol{B}^T\boldsymbol{\mu})_i][Y_j - (\boldsymbol{B}^T\boldsymbol{\mu})_j] = \begin{cases} 0, & i \neq j, \\ \lambda_j^2, & i = j. \end{cases}$$

Фактически переменные Y_i помогут нам производить вычисления с величинами X_j. Например, для среднего значения с. в. X_j находим

$$\begin{aligned}
\mathsf{E}X_j &= \int_{\mathbb{R}^n} \frac{x_j}{(2\pi)^{n/2}\sqrt{\det\Sigma}}\exp\left(-\frac{1}{2}\langle(\boldsymbol{x} - \boldsymbol{\mu}), \Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\rangle\right)d\boldsymbol{x} = \\
&= \int_{\mathbb{R}^n} \frac{(\boldsymbol{B}\boldsymbol{y})_j}{(2\pi)^{n/2}\sqrt{\det\Sigma}}\exp\left(-\frac{1}{2}\langle(\boldsymbol{y} - \boldsymbol{B}^T\boldsymbol{\mu}), \boldsymbol{D}^{-1}(\boldsymbol{y} - \boldsymbol{B}^T\boldsymbol{\mu})\rangle\right)d\boldsymbol{y} = \\
&= \sum_{1 \leqslant i \leqslant n}\int_{\mathbb{R}^n} \frac{y_i b_{ji}}{(2\pi)^{n/2}\sqrt{\det\Sigma}}\exp\left(-\frac{1}{2}\langle(\boldsymbol{y} - \boldsymbol{B}^T\boldsymbol{\mu}), \boldsymbol{D}^{-1}(\boldsymbol{y} - \boldsymbol{B}^T\boldsymbol{\mu})\rangle\right)d\boldsymbol{y} = \\
&= \sum_{1 \leqslant i \leqslant n} \frac{b_{ji}}{(2\pi)^{1/2}\lambda_i}\int_{\mathbb{R}} y\exp\left(-\frac{1}{2\lambda_i^2}(y - (\boldsymbol{B}^T\boldsymbol{\mu})_i)^2\right)dy = \sum_{1 \leqslant i \leqslant n}(\boldsymbol{B}^T\boldsymbol{\mu})_i b_{ji} = \mu_j.
\end{aligned}$$

Аналогично находим и ковариацию $\mathrm{Cov}(X_i, X_j)$:

$$\int\limits_{\mathbb{R}^n} \frac{(x_i-\mu_i)(x_j-\mu_j)}{(2\pi)^{n/2}\sqrt{\det\Sigma}} \exp\Big(-\frac{1}{2}\langle(\boldsymbol{x}-\boldsymbol{\mu}), \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\rangle\Big)\, d\boldsymbol{x} =$$

$$= \int\limits_{\mathbb{R}^n} \frac{(\boldsymbol{B}(\boldsymbol{y}-\boldsymbol{B}^T\boldsymbol{\mu}))_i(\boldsymbol{B}(\boldsymbol{y}-\boldsymbol{B}^T\boldsymbol{\mu}))_j}{(2\pi)^{n/2}\sqrt{\det\Sigma}} \exp\Big(-\frac{1}{2}\langle(\boldsymbol{y}-\boldsymbol{B}^T\boldsymbol{\mu}), \boldsymbol{D}^{-1}(\boldsymbol{y}-\boldsymbol{B}^T\boldsymbol{\mu})\rangle\Big)\, d\boldsymbol{y} =$$

$$= \sum_{1\leqslant l\leqslant n}\ \sum_{1\leqslant m\leqslant n} \mathrm{Cov}(Y_l, Y_m) b_{il} b_{jm} = \sum_{1\leqslant m\leqslant n} \lambda_m^2 b_{im} b_{jm} = (\boldsymbol{B}\boldsymbol{D}\boldsymbol{B}^{-1})_{ij} = \Sigma_{ij}.$$

Это доказывает утверждение 4 б). При $i=j$ получаем дисперсию $\mathrm{Var}\, X_i$.

На самом деле еще более мощным средством является *совместная характеристическая функция* $\psi_{\boldsymbol{X}}(\boldsymbol{t})$, определяемая формулой

$$\psi_{\boldsymbol{X}}(\boldsymbol{t}) = \mathsf{E} e^{i\langle \boldsymbol{t}, \boldsymbol{X}\rangle} = \mathsf{E} \exp\Big(i\sum_{j=1}^n t_j X_j\Big), \qquad \boldsymbol{t}^T = (t_1, \ldots, t_n) \in \mathbb{R}^n. \tag{2.3.8}$$

Совместная характеристическая функция имеет многие свойства маргинальной характеристической функции. В частности, совместная характеристическая функция однозначно определяет случайный вектор $\boldsymbol{X}$. Для многомерного нормального вектора $\boldsymbol{X}$ совместная характеристическая функция может быть вычислена в явном виде. В самом деле,

$$\mathsf{E} e^{i\boldsymbol{t}^T\boldsymbol{X}} = \mathsf{E} e^{i\boldsymbol{t}^T\boldsymbol{B}\boldsymbol{Y}} = \mathsf{E} e^{i(\boldsymbol{B}^T\boldsymbol{t})^T\boldsymbol{Y}} = \prod_{j=1}^n \mathsf{E}\exp(i(\boldsymbol{B}^T\boldsymbol{t})_j Y_j).$$

Здесь каждый сомножитель представляет собой маргинальную характеристическую функцию:

$$\mathsf{E}\exp(i(\boldsymbol{B}^T\boldsymbol{t})_j Y_j) = \exp(i(\boldsymbol{B}^T\boldsymbol{t})_j(\boldsymbol{B}^{-1}\boldsymbol{\mu})_j - (\boldsymbol{B}^T\boldsymbol{t})_j^2\lambda_j^2/2).$$

Так как $\boldsymbol{B}\boldsymbol{D}\boldsymbol{B}^{-1} = \boldsymbol{\Sigma}$, все произведение в целом равно

$$\exp\Big(i\boldsymbol{t}^T\boldsymbol{B}\boldsymbol{B}^{-1}\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{t}^T\boldsymbol{B}\boldsymbol{D}\boldsymbol{B}^{-1}\boldsymbol{t}\Big) = \exp\Big(i\boldsymbol{t}^T\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{t}^T\Sigma\boldsymbol{t}\Big).$$

Следовательно,

$$\psi_{\boldsymbol{X}}(\boldsymbol{t}) = e^{i\boldsymbol{t}^T\boldsymbol{\mu} - \boldsymbol{t}^T\Sigma\boldsymbol{t}/2}. \tag{2.3.9}$$

Отметим существенное различие в матрицах, которые входят в выражения для многомерной нормальной п. р. в. и многомерной нормальной характеристической функции: в формуле (2.3.9) это ковариационная матрица Σ, а в формулу (2.3.7) входит Σ^{-1}.

Теперь, чтобы получить маргинальную характеристическую функцию (х. ф.) $\psi_{X_j}(t)$, подставим вектор $\boldsymbol{t} = (0, \ldots, 0, t, 0, \ldots, 0)$ (t на j-й позиции) в правую часть формулы (2.3.9):

$$\psi_{X_j}(t) = \exp(i\mu_j t - t^2\Sigma_{jj}/2).$$

В силу единственности п. р. в. с заданной х. ф. находим $X_j \sim \mathrm{N}(\mu_j, \Sigma_{jj})$, что и требуется для утверждения 4 а).

Из вышеприведенных рассуждений немедленно следует еще одно свойство.

4в) Если $\boldsymbol{X}$ — вектор, $\boldsymbol{X} \sim \mathrm{N}(\boldsymbol{\mu}, \Sigma)$, то любое подмножество его компонент (X_{j_l}) также имеет совместное нормальное распределение с вектором средних значений (μ_{j_l}) и ковариационной матрицей $(\Sigma_{j_l j_{l'}})$.

Для характеристической функции мы получили следующее свойство.

5. Совместная х. ф. $\psi_{\boldsymbol{X}}(\boldsymbol{t})$ случайного вектора $\boldsymbol{X} \sim \mathrm{N}(\boldsymbol{\mu}, \Sigma)$ имеет вид $\exp\left(i\langle \boldsymbol{t}, \boldsymbol{\mu}\rangle - \frac{1}{2}\langle \boldsymbol{t}, \Sigma \boldsymbol{t}\rangle\right)$.

Наконец, понятия и методы, развитые выше, позволяют проверить, что справедлив следующий результат

6. Линейная комбинация $\sum\limits_{i=1}^{n} c_i X_i$ совместно нормальных с. в. $\boldsymbol{X} \sim \mathrm{N}(\boldsymbol{\mu}, \Sigma)$ имеет нормальное распределение со средним $\langle \boldsymbol{c}, \boldsymbol{\mu}\rangle = \sum\limits_{1\leqslant i\leqslant n} c_i\mu_i$ и дисперсией $\langle \boldsymbol{c}, \Sigma\boldsymbol{c}\rangle = \sum\limits_{1\leqslant i,j\leqslant n} c_i\Sigma_{ij}c_j$. В общем случае, если $\boldsymbol{Y}$ — это случайный вектор вида $\boldsymbol{Y} = \boldsymbol{A}^T\boldsymbol{X}$, полученный из вектора $\boldsymbol{X}$ при помощи линейного преобразования с обратимой матрицей $\boldsymbol{A}$, то $\boldsymbol{Y} \sim \mathrm{N}(\boldsymbol{A}^T\boldsymbol{\mu}, \boldsymbol{A}^T\Sigma\boldsymbol{A})$. См. пример 1.1.1 в части В. (Это последнее утверждение можно распространить также и на случай преобразований с необратимой матрицей $\boldsymbol{A}$, но мы оставим этот вопрос для последующих томов.)

Задача 2.3.5. Найдите распределение суммы n независимых случайных величин, каждая из которых имеет распределение Пуассона с параметром λ. Используя центральную предельную теорему, докажите, что

$$e^{-n}\left(1 + \frac{n}{1!} + \frac{n^2}{2!} + \ldots + \frac{n^n}{n!}\right) \to \frac{1}{2}$$

при $n \to \infty$.

Решение. Пусть $X_1, \ldots, X_n$ — н. о. р. Po(1)-с. в. Производящая функция имеет вид

$$\psi_{X_i}(s) = \mathsf{E}s^{X_i} = \sum_{l\geqslant 0} s^l \frac{1}{l!}e^{-1} = e^{s-1}, \quad s \in \mathbb{R}^1.$$

В общем случае, если $Y \sim \mathrm{Po}(\lambda)$, то

$$\psi_Y(s) = \sum_{l\geqslant 0} \frac{s^l\lambda^l}{l!}e^{-\lambda} = e^{\lambda(s-1)}.$$

Далее, если $S_n = X_1 + \ldots + X_n$, то

$$\psi_{S_n}(s) = \psi_{X_1}(s) \ldots \psi_{X_n}(s),$$

откуда следует, что $S_n \sim \mathrm{Po}(n)$ и $\mathsf{E}S_n = \mathrm{Var}(S_n) = n$. В силу ЦПТ мы имеем $T_n = \dfrac{S_n - n}{\sqrt{n}} \sim \mathrm{N}(0, 1)$ при больших n. Однако

$$e^{-n}\left(1 + \frac{n}{1!} + \frac{n^2}{2!} + \ldots + \frac{n^n}{n!}\right) = \mathsf{P}(S_n \leqslant n) =$$

$$= \mathsf{P}(T_n \leqslant 0) \to \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{0} e^{-y^2/2}\, dy = \frac{1}{2}. \quad \square$$

Задача 2.3.6 (вопрос из курса алгебры). Пусть $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathbb{R}^n$ — это линейно независимые векторы-столбцы. Покажите, что матрица размера $n \times n$ вида

$$\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^T$$

является обратимой.

Решение. Достаточно показать, что матрица $\sum\limits_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^T$ не переводит в нулевой вектор ни один ненулевой вектор. Предположим, что

$$\left(\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^T\right) \boldsymbol{c} = 0,$$

где $\boldsymbol{c} = (c_1, \ldots, c_n)^T \in \mathbb{R}^n$. Это означает, что

$$\sum_{k=1}^{n} \sum_{i=1}^{n} x_{ij} x_{ik} c_k = 0, \qquad 1 \leqslant j \leqslant n,$$

или что

$$\sum_{i=1}^{n} x_{ij} \langle \boldsymbol{x}_i, \boldsymbol{c} \rangle = 0, \qquad 1 \leqslant j \leqslant n.$$

Здесь $\boldsymbol{x}_i = \begin{pmatrix} x_{i1} \\ \vdots \\ x_{in} \end{pmatrix}$, $1 \leqslant i \leqslant n$. Последнее равенство означает, что

$$\sum_{i=1}^{n} \boldsymbol{x}_i \langle \boldsymbol{x}_i, \boldsymbol{c} \rangle = 0.$$

Поскольку векторы $\boldsymbol{x}_i'$ являются линейно независимыми, коэффициенты $\langle \boldsymbol{x}_i, \boldsymbol{c} \rangle$ равны 0, $1 \leqslant i \leqslant n$. Но это означает, что $\boldsymbol{c} = 0$. $\square$

Три следующие задачи были позаимствованы из углубленных курсов по статистике; их можно опустить при первом чтении, но они окажутся полезными для тех читателей, которые намереваются достичь лучшего понимания уже на данном этапе.

Задача 2.3.7. Пусть $X_1, \ldots, X_n$ — н. о. р. выборка из $\mathrm{N}(\mu_1, \sigma_1^2)$ и $Y_1, \ldots$ $\ldots, Y_m$ — н. о. р. выборка из $\mathrm{N}(\mu_2, \sigma_2^2)$, и предположим, что эти две выборки взаимно независимы. Найдите совместное распределение разности $\overline{X} - \overline{Y}$ и суммы $\frac{1}{\sigma_1^2}\sum\limits_{i=1}^{n} X_i + \frac{1}{\sigma_2^2}\sum\limits_{j=1}^{m} Y_j$?

Решение. Находим $\overline{X} \sim \mathrm{N}\left(\mu_1, \frac{1}{n}\sigma_1^2\right)$, $\overline{Y} \sim \mathrm{N}\left(\mu_2, \frac{1}{m}\sigma_2^2\right)$. Далее,

$$f_{\overline{X},\overline{Y}}(x, y) = \left(\frac{n}{2\pi\sigma_1^2}\right)^{1/2}\left(\frac{m}{2\pi\sigma_2^2}\right)^{1/2}\exp\left(-n\frac{(x-\mu_1)^2}{2\sigma_1^2} - m\frac{(y-\mu_2)^2}{2\sigma_2^2}\right).$$

Мы видим, что обе величины $U = \overline{X} - \overline{Y}$ и $S = \frac{1}{\sigma_1^2}\sum\limits_{i=1}^{n} X_i + \frac{1}{\sigma_2^2}\sum\limits_{j=1}^{m} Y_j = \frac{n}{\sigma_1^2}\overline{X} +$ $+ \frac{m}{\sigma_2^2}\overline{Y}$ представляют собой линейные комбинации (независимых) нормальных случайных величин, а следовательно, и сами являются нормальными. Прямые вычисления показывают, что

$$\mathrm{E}U = \mu_1 - \mu_2, \quad \mathrm{E}S = \frac{n}{\sigma_1^2}\mu_1 + \frac{m}{\sigma_2^2}\mu_2, \quad \mathrm{Var}\,U = \frac{m\sigma_1^2 + n\sigma_2^2}{mn}, \quad \mathrm{Var}\,S = \frac{n}{\sigma_1^2} + \frac{m}{\sigma_2^2}.$$

Таким образом,

$$f_U(u) = \frac{(mn)^{1/2}}{(2\pi(m\sigma_1^2 + n\sigma_2^2))^{1/2}}\exp\left\{-\frac{[u - (\mu_1 - \mu_2)]^2}{2(m\sigma_1^2 + n\sigma_2^2)/(mn)}\right\}, \quad u \in \mathbb{R},$$

и

$$f_S(s) = \frac{\sigma_1\sigma_2}{(2\pi(m\sigma_1^2 + n\sigma_2^2))^{1/2}}\exp\left\{-\left[s - \left(\frac{n}{\sigma_1^2}\mu_1 + \frac{m}{\sigma_2^2}\mu_2\right)\right]^2 \Big/ \frac{2(m\sigma_1^2 + n\sigma_2^2)}{\sigma_1^2\sigma_2^2}\right\}, \quad s \in \mathbb{R}.$$

Наконец, величины U и S являются независимыми, поскольку $\mathrm{Cov}(U, S) = 0$. Следовательно, совместная п. р. в. удовлетворяет условию $f_{U,S}(u, s) = f_U(u) f_S(s)$. □

Замечание. Из формул для $f_U(u)$ и $f_S(s)$ следует, что пара (U, S) образует так называемую «достаточную статистику» для пары неизвестных параметров (μ_1, μ_2). См. § 1.3 в части В.

Задача 2.3.8. Пусть $X_1, X_2, \ldots, X_n$ — случайная выборка из распределения $\mathrm{N}(\theta, \sigma^2)$, и предположим, что априорное распределение параметра θ — это распределение $\mathrm{N}(\mu, \tau^2)$, где σ^2, μ и τ^2 известны. Определите апостериорное распределение θ при заданных $X_1, X_2, \ldots, X_n$.

Решение. Априорная п. р. в. является гауссовской:

$$\pi(\theta) = \frac{1}{\sqrt{2\pi}\,\tau} e^{-\frac{(\theta-\mu)^2}{2\tau^2}},$$

и таковой является и (совместная) п. р. в. вектора $\boldsymbol{X}$ при заданном значении θ:

$$f(\boldsymbol{x};\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x_i-\theta)^2}{2\sigma^2}}.$$

Таким образом (знак $\propto$ означает «пропорционально»),

$$\pi(\theta) f(\boldsymbol{x};\theta) \propto \exp\Big(-\frac{(\theta-\mu)^2}{2\tau^2} - \sum \frac{(x_i-\theta)^2}{2\sigma^2}\Big) =$$

$$= \exp\Big(-\frac{1}{2\tau^2}(\theta^2 - 2\mu\theta + \mu^2) - \frac{1}{2\sigma^2}\sum(x_i^2 - 2\theta\sum x_j + \theta^2)\Big) =$$

$$= \exp\Big(-\frac{1}{2}\Big[\theta^2\Big(\frac{1}{\tau^2} + \frac{n}{\sigma^2}\Big) - 2\theta\Big(\frac{\mu}{\tau^2} + \frac{n\bar{x}}{\sigma^2}\Big)\Big]\Big),$$

если отбросить члены, в которые θ не входит. Тогда апостериорное распределение имеет вид

$$\pi(\theta \mid \boldsymbol{x}) = \frac{\pi(\theta) f(\boldsymbol{x};\theta)}{\int \pi(\theta') f(\boldsymbol{x};\theta')\, d\theta'} = \frac{1}{\sqrt{2\pi}\,\tau_n} \exp\Big[-\frac{(\theta-\mu_n)^2}{2\tau_n^2}\Big],$$

где

$$\frac{1}{\tau_n^2} = \frac{1}{\tau^2} + \frac{n}{\sigma^2}, \qquad \mu_n = \frac{\mu/\tau^2 + n\bar{x}/\sigma^2}{1/\tau^2 + n/\sigma^2}. \qquad \square$$

Задача 2.3.9. Пусть $\Theta, X_1, X_2, \ldots$ — случайные величины. Предположим, что при условии $\Theta = \theta$ случайные величины $X_1, X_2, \ldots$ являются независимыми и величина X_k нормально распределена со средним θ и дисперсией σ_k^2. Предположим, что маргинальная плотность Θ равна

$$\pi(\theta) = \frac{1}{\sqrt{2\pi}} e^{-\theta^2/2}, \qquad \theta \in \mathbb{R}.$$

Вычислите среднее и дисперсию величины Θ при условии, что $X_1 = x_1, \ldots$ $\ldots, X_n = x_n$.

Решение. Прямые вычисления показывают, что условная плотность $f_{\Theta \mid X_1,\ldots,X_n}(\theta) = f(\theta \mid x_1, \ldots, x_n)$ является кратной выражению

$$\exp\Big[-\frac{1}{2}\Big(1 + \sum_i \frac{1}{\sigma_i^2}\Big)\Big(\theta - \frac{\sum_i x_i/\sigma_i^2}{1 + \sum_i 1/\sigma_i^2}\Big)^2\Big]$$

с коэффициентом, который зависит от значений $x_1, \ldots, x_n$ величин $X_1, \ldots$ $\ldots, X_n$. Отсюда следует, что условное среднее имеет вид

$$\mathsf{E}(\Theta \mid X_1, \ldots, X_n) = \frac{1}{1 + \sum\limits_i 1/\sigma_i^2} \sum_i \frac{X_i}{\sigma_i^2}$$

и условная дисперсия вычисляется по формуле

$$\mathrm{Var}(\Theta \mid X_1, \ldots, X_n) = \frac{1}{1 + \sum\limits_i 1/\sigma_i^2},$$

т. е. не зависит от $X_1, \ldots, X_n$. □

Задача 2.3.10. Пусть X и Y — н. о. р. с. в. со стандартной нормальной п. р. в.

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad x \in \mathbb{R}.$$

Найдите совместную п. р. в. величин $U = X + Y$ и $V = X - Y$. Покажите, что величины U и V независимы, и выпишите их маргинальные распределения. Пусть

$$Z = \begin{cases} |Y|, & \text{если } X > 0, \\ -|Y|, & \text{если } X < 0. \end{cases}$$

Вычислив $\mathsf{P}(Z \leqslant z)$ при $z < 0$ и $z > 0$, покажите, что величина Z имеет стандартное нормальное распределение. Кратко объясните, почему совместное распределение величин X и Z не является двумерным нормальным распределением.

Решение. Запишем $(U, V) = T(X, Y)$. Тогда для п. р. в. имеем

$$f_{U,V}(u, v) = f_{X,Y}(T^{-1}(u, v)) \left| \frac{\partial(x, y)}{\partial(u, v)} \right|.$$

Обратное отображение здесь таково:

$$T^{-1}(u, v) = \left(\frac{u+v}{2}, \frac{u-v}{2} \right), \quad \text{где} \quad \frac{\partial(x, y)}{\partial(u, v)} = -\frac{1}{2}.$$

Следовательно,

$$f_{U,V}(u, v) = \frac{1}{2\pi} e^{-\frac{1}{2}\left(\frac{(u+v)^2}{4} + \frac{(u-v)^2}{4} \right)} \cdot \frac{1}{2} = \frac{1}{4\pi} e^{-\frac{1}{4}(u^2+v^2)},$$

т. е. величины U, V являются независимыми. Далее, если $z \geqslant 0$, то

$$\mathsf{P}(Z < z) = \frac{1}{2}(1 + \mathsf{P}(|Y| < z)) = \mathsf{P}(Y < z),$$

а если $z < 0$, то

$$\mathsf{P}(Z < z) = \frac{1}{2} \mathsf{P}(-|Y| < z) = \mathsf{P}(Y > |z|) = \mathsf{P}(Y < z).$$

Таким образом, величина Z имеет то же самое нормальное распределение, что и Y. Но совместное распределение пары (X, Z) приписывает нулевые массы второму и четвертому квадранту, следовательно, Z и X не являются независимыми. □

Задача 2.3.11. Пусть X_1 и Y_1 — независимые нормальные с. в. со средним 0 и дисперсией $\sigma^2 > 0$, и пусть $X_2 = X_1$, $Y_2 = (X_1 + Y_1)/\sqrt{2}$.

а) Выпишите совместные п. р. в. f_{X_1,Y_1} и f_{X_2,Y_2}.

б) Случайная величина N не зависит от X_1, Y_1 и принимает значения 1 и 2 с вероятностью 1/2. Чему равна совместная плотность пары (X_N, Y_N)? Найдите $\mathsf{E}X_N$, $\mathsf{E}Y_N$, $\operatorname{Var}X_N$, $\operatorname{Var}Y_N$ и $\operatorname{Cov}(X_N, Y_N) = \mathsf{E}(X_N - \mathsf{E}X_N)(Y_N - \mathsf{E}Y_N)$.

в) Покажите, что каждая из случайных величин X_N и Y_N имеет нормальное распределение. Будет ли пара (X_N, Y_N) иметь совместное нормальное распределение?

Решение. а) Прямая замена переменных имеет вид

$$x_2 = x_1, \quad y_2 = \frac{1}{\sqrt{2}}(x_1 + y_1),$$

а обратная замена

$$x_1 = x_2, \quad y_1 = \sqrt{2}y_2 - x_2,$$

имеет якобиан

$$\det\begin{pmatrix} 1 & 0 \\ -1 & \sqrt{2} \end{pmatrix} = \sqrt{2}.$$

Следовательно,

$$f_{X_1,Y_1}(x, y) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{1}{2\sigma^2}(x^2 + y^2)\right)$$

и

$$f_{X_2,Y_2}(x, y) = \frac{\sqrt{2}}{2\pi\sigma^2}\exp\left(-\frac{1}{2\sigma^2}(x^2 + (\sqrt{2}y - x)^2)\right) = \\ = \frac{1}{\sqrt{2}\,\pi\sigma^2}\exp\left(-\frac{1}{2\sigma^2}(2x^2 - 2\sqrt{2}xy + 2y^2)\right).$$

б) Далее, $f = \frac{1}{2}(f_{X_1,Y_1} + f_{X_2,Y_2})$ является п. р. в., так как $f \geqslant 0$ и

$$\int f = \frac{1}{2}\left(\int f_{X_1,Y_1} + \int f_{X_2,Y_2}\right) = \frac{1}{2} + \frac{1}{2} = 1.$$

Рассматривая соответствующую пару (X, Y), удобно использовать случайную величину

$$N = \begin{cases} 1 & \text{с вероятностью } 1/2, \\ 2 & \text{с вероятностью } 1/2 \end{cases}$$

и положить $(X, Y) = (X_N, Y_N)$. Тогда $\mathsf{E}X = \frac{1}{2}(\mathsf{E}X_1 + \mathsf{E}X_2) = 0$; аналогично $\mathsf{E}Y = 0$, $\operatorname{Var}X = \frac{1}{2}(\operatorname{Var}X_1 + \operatorname{Var}X_2) = \sigma^2$, и то же самое для $\operatorname{Var}Y$,

$$\operatorname{Cov}(X, Y) = \frac{1}{2}(\operatorname{Cov}(X_1, Y_1) + \operatorname{Cov}(X_2, Y_2)) = \frac{1}{2}(0 + \operatorname{Cov}(X_2, Y_2)) =$$
$$= \frac{1}{2\sqrt{2}}\mathsf{E}X_1(X_1 + Y_1) = \frac{1}{2\sqrt{2}}(\mathsf{E}X_1^2 + \mathsf{E}X_1Y_1) = \frac{1}{2\sqrt{2}}\sigma^2.$$

в) Интегрирование $f(x, y)$ по x или y приводит к $\mathrm{N}(0, \sigma^2)$-п. р. в.:

$$f_{X_N}(x) = \frac{1}{2}(f_{X_1}(x) + f_{X_2}(x)) = f_{X_1}(x) \sim \mathrm{N}(0, \sigma^2),$$

и аналогично для $f_{Y_N}(y)$. Можно рассуждать и по-другому: рассмотрим

$$\mathsf{E}e^{tY_N} = \frac{1}{2}\mathsf{E}e^{tY_1} + \frac{1}{2}\mathsf{E}e^{tY_2} = e^{t^2\sigma^2/2}.$$

Это выражение является в точности производящей функцией моментов для с. в. $\mathrm{N}(0, \sigma^2)$, поскольку

$$\mathsf{E}e^{tY_2} = \mathsf{E}e^{t(X_1+Y_1)/\sqrt{2}} = \mathsf{E}e^{tX_1/\sqrt{2}}\,\mathsf{E}e^{tY_1/\sqrt{2}} = e^{t^2\sigma^2/2}.$$

Но пара (X_N, Y_N) не является совместно нормальной, так как $f_{X_N,Y_N}(x, y)$ не имеет вид плотности двумерного нормального распределения. □

Задача 2.3.12. Проверьте, что стандартная нормальная плотность распределения $p(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ удовлетворяет уравнению

$$\int_y^\infty xp(x)\,dx = p(y), \qquad y > 0.$$

Используя это уравнение и равенство $\sin x = \int_0^x \cos y\,dy$ (или иным способом), докажите, что если X — это $\mathrm{N}(0, 1)$-с. в., то

$$(\mathsf{E}\cos X)^2 \leqslant \operatorname{Var}(\sin X) \leqslant \mathsf{E}(\cos X)^2.$$

Решение. Запишем

$$\frac{1}{\sqrt{2\pi}}\int_y^\infty xe^{-x^2/2}\,dx = \frac{1}{\sqrt{2\pi}}\int_y^\infty e^{-x^2/2}\,d\left(\frac{x^2}{2}\right) = \frac{1}{\sqrt{2\pi}}e^{-y^2/2}.$$

Теперь

$$\mathsf{E}\sin X = \frac{1}{\sqrt{2\pi}}\int e^{-x^2/2}\sin x\,dx = 0,$$

поскольку $e^{-x^2/2}\sin x$ является нечетной функцией. Таким образом,

$$\mathrm{Var}(\sin X) = \mathsf{E}(\sin X)^2 = \int p(x)(\sin x)^2\,dx = \int p(x)\left(\int\limits_0^x \cos y\,dy\right)^2 dx.$$

В силу неравенства Коши—Шварца последний интеграл не превосходит

$$\int p(x)|x| \int\limits_0^x (\cos y)^2\,dy\,dx =$$

$$= -\int\limits_{-\infty}^0 (\cos y)^2 \int\limits_{-\infty}^{y} xp(x)\,dx\,dy + \int\limits_0^\infty (\cos y)^2 \int\limits_y^\infty xp(x)\,dx\,dy =$$

$$= \int\limits_{-\infty}^0 p(y)(\cos y)^2\,dy + \int\limits_0^\infty p(y)(\cos y)^2\,dy = \int p(y)(\cos y)^2\,dy = \mathsf{E}(\cos X)^2.$$

С другой стороны, поскольку $\mathsf{E}X^2 = 1$, мы получаем

$$\mathrm{Var}(\sin X) = \mathsf{E}(\sin X)^2 = \mathsf{E}X^2\,\mathsf{E}(\sin X)^2 \geqslant$$

$$\geqslant (\mathsf{E}(X\sin X))^2 = \left(\int xp(x)\,dx \int\limits_0^x \cos y\,dy\right)^2 =$$

$$= \left(-\int\limits_{-\infty}^0 \cos y \int\limits_{-\infty}^{y} xp(x)\,dx\,dy + \int\limits_0^\infty \cos y \int\limits_y^\infty xp(x)\,dx\,dy\right)^2 =$$

$$= \left(\int p(y)\cos y\,dy\right)^2 = (\mathsf{E}\cos X)^2. \quad \square$$

Задача 2.3.13. В условиях задачи 2.2.6 докажите, что X и Y являются независимыми тогда и только тогда, когда $r = 0$.

Решение. В общем случае, если X и Y независимы, то $r = \mathsf{E}XY = \mathsf{E}X\,\mathsf{E}Y = 0$. В гауссовском случае верно и обратное: если $r = 0$, то совместная п. р. в. имеет вид

$$f_{X,Y}(x,\,y) = \frac{1}{2\pi}e^{-(x^2+y^2)/2} = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}\frac{1}{\sqrt{2\pi}}e^{-y^2/2} = f_X(x)f_Y(y),$$

т. е. X и Y являются независимыми. $\square$

Задача 2.3.14. Докажите центральную предельную теорему для независимых одинаково распределенных действительных случайных величин со средним μ и дисперсией σ^2.

Предположим, что $X_1, X_2, \ldots$ — это н. о. р. с. в., каждая из которых равномерно распределена на отрезке $[0,\,1]$. Вычислите среднее и дисперсию $\ln X_1$.

Пусть $0 \leqslant a < b$. Покажите, что

$$\mathsf{P}\left((X_1X_2\ldots X_n)^{n^{-1/2}}e^{n^{1/2}} \in [a,\,b]\right)$$

стремится к пределу, и найдите, чему равен этот предел.

Решение. Пусть $X_1, X_2, \ldots$ — н. о. р. с. в. со средним $\mathsf{E}X_i = \mu$ и дисперсией $X_i = \sigma^2$. Центральная предельная теорема утверждает, что $\forall a, b$, $-\infty \leqslant a < b \leqslant \infty$, выполняется равенство

$$\lim_{n\to\infty} \mathsf{P}\left(a < \frac{X_1 + \ldots + X_n - n\mu}{\sqrt{n}\sigma} < b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\, dx.$$

Более того, если $X \sim \mathrm{U}(0, 1)$, то

$$\mathsf{E}\ln X_i = \int_0^1 \ln x\, dx = -\int_\infty^0 y(de^{-y}) = (-ye^{-y})\Big|_\infty^0 + \int_\infty^0 e^{-y}\, dy = -1,$$

$$\mathsf{E}(\ln X_i)^2 = \int_0^1 (\ln x)^2\, dx = \int_\infty^0 y^2\, d(e^{-y}) = (y^2e^{-y})\Big|_\infty^0 - 2\int_\infty^0 e^{-y}y\, dy = 2$$

и

$$\mathrm{Var}(\ln X_i) = 2 - (-1)^2 = 1.$$

Наконец, в силу ЦПТ получаем

$$\mathsf{P}((X_1X_2\ldots X_n)^{n^{-1/2}}e^{n^{1/2}} \in [a, b]) = \mathsf{P}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \ln X_i + \sqrt{n} \in [\ln a, \ln b]\right) =$$

$$\mathsf{P}\left(\frac{1}{\sqrt{n}}\left(\sum_{i=1}^n \ln X_i + n\right) \in [\ln a, \ln b]\right) \to \frac{1}{\sqrt{2\pi}} \int_{\ln a}^{\ln b} e^{-x^2/2}\, dx. \quad \square$$

Задача 2.3.15. Случайные величины X_i, $i = 1, 2$, имеют нормальное распределение со средним μ_i и дисперсией σ_i^2 и являются независимыми. Найдите распределение с. в. $Z = a_1X_1 + a_2X_2$, где $a_1, a_2 \in \mathbb{R}$.

Указание. Вы можете воспользоваться тем, что $\mathsf{E}e^{\theta X_i} = e^{\theta\mu_i + \frac{1}{2}\theta^2\sigma_i^2}$.

Решение. Производящая функция моментов имеет вид

$$\varphi_{a_1X_1+a_2X_2}(\theta) = \mathsf{E}(e^{\theta(a_1X_1+a_2X_2)}) = \mathsf{E}(e^{\theta a_1X_1}e^{\theta a_2X_2}) =$$
$$= \mathsf{E}(e^{\theta a_1X_1})\,\mathsf{E}(e^{\theta a_2X_2}) = \varphi_{X_1}(a_1\theta)\varphi_{X_2}(a_2\theta)$$

в силу независимости. Далее,

$$\varphi_{X_1}(a_1\theta)\varphi_{X_2}(a_2\theta) = \exp\left(a_1\theta\mu_1 + \frac{1}{2}a_1^2\theta^2\sigma_1^2 + a_2\theta\mu_2 + \frac{1}{2}a_2^2\theta^2\sigma_2^2\right) =$$
$$= \exp\left[\theta(a_1\mu_1 + a_2\mu_1) + \frac{1}{2}\theta^2(a_1^2\sigma_1^2 + a_2^2\sigma_2^2)\right] = M_Z(\theta),$$

где $Z \sim \mathrm{N}((a_1\mu_1 + a_2\mu_1), (a_1^2\sigma_1^2 + a_2^2\sigma_2^2))$. В силу единственности п. р. в. с заданной производящей функцией моментов получаем, что

$$(a_1X_1 + a_2X_2) \sim \mathrm{N}((a_1\mu_1 + a_2\mu_1), (a_1^2\sigma_1^2 + a_2^2\sigma_2^2)). \quad \square$$

Задача 2.3.16. а) Пусть X — нормально распределенная с. в. со средним 0 и дисперсией 1. Вычислите $\mathsf{E}X^r$ при $r = 0, 1, 2, 3, 4$. Пусть Y — нормально распределенная с. в. со средним μ и дисперсией σ^2. Вычислите $\mathsf{E}Y^r$ при $r = 0, 1, 2, 3, 4$. Что можно сказать о сумме двух независимых случайных величин? (Сформулируйте без доказательства.)

б) Госпожа президент Статистического общества обычно отдыхает, занимаясь рыбной ловлей в чистых водах озера Чебышёва. Количество пойманных рыбок является пуассоновской случайной величиной с параметром λ. Вес каждой из рыбок, плавающих в озере Чебышёва, является независимой нормально распределенной случайной величиной со средним μ и дисперсией σ^2. (Если считать, что μ намного больше, чем σ, то рыбки с отрицательным весом будут попадаться крайне редко и будут очень цениться гурманами.) Пусть Z — это общий вес всего улова. Вычислите $\mathsf{E}Z$ и $\mathsf{E}Z^2$.

Покажите, цитируя все необходимые вам результаты, что вероятность того, что улов президента будет весить меньше чем $\lambda\mu/2$, не превосходит $4(\mu^2 + \sigma^2)\lambda^{-1}\mu^{-2}$.

Решение. а) В силу симметрии $\mathsf{E}X^0 = \mathsf{E}1 = 1$, $\mathsf{E}X^1 = \mathsf{E}X^3 = 0$. Далее, находим $\mathsf{E}X^2$ и $\mathsf{E}X^4$, применив интегрирование по частям:

$$\frac{1}{\sqrt{2\pi}} \int x^2 e^{-x^2/2}\, dx = \frac{1}{\sqrt{2\pi}} \left[(-xe^{-x^2/2})\Big|_{-\infty}^{\infty} + \int e^{-x^2/2}\, dx \right] = 1,$$

$$\frac{1}{\sqrt{2\pi}} \int x^4 e^{-x^2/2}\, dx = \frac{1}{\sqrt{2\pi}} \left[(-x^3e^{-x^2/2})\Big|_{-\infty}^{\infty} + 3\int x^2 e^{-x^2/2}\, dx \right] = 3.$$

Затем находим

$$\mathsf{E}Y^0 = \mathsf{E}1 = 1,$$

$$\mathsf{E}Y^1 = \mathsf{E}(\sigma X + \mu) = \sigma\,\mathsf{E}X + \mu\,\mathsf{E}1 = \mu,$$

$$\mathsf{E}Y^2 = \mathsf{E}(\sigma^2X^2 + 2\mu\sigma X + \mu^2) = \sigma^2 + \mu^2,$$

$$\mathsf{E}Y^3 = \mathsf{E}(\sigma^3X^3 + 3\mu\sigma^2X^2 + 3\mu^2\sigma X + \mu^3) = 3\mu\sigma^2 + \mu^3$$

и

$$\mathsf{E}Y^4 = \mathsf{E}(\sigma^4X^4 + 4\mu\sigma^3X^3 + 6\mu^2\sigma^2X^2 + 4\mu^3\sigma X + \mu^4) = 3\sigma^4 + 6\mu^2\sigma^2 + \mu^4.$$

Теперь если X_1, X_2 — независимые нормальные с. в. со средними μ_1, μ_2 и дисперсиями σ_1^2, σ_2^2 соответственно, то $X_1 + X_2$ является $\mathrm{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$-с. в.

б) Таким образом, если Y_r — это вес r рыбок, то $Y_r \sim \mathrm{N}(r\mu, r\sigma^2)$. Наконец,

$$\mathsf{E}Z = \sum_{r=0}^{\infty} \mathsf{P}(\text{поймала } r \text{ рыбок})\,\mathsf{E}Y_r = \sum \frac{\lambda^r e^{-\lambda}}{r!} r\mu = \lambda\mu,$$

и аналогично

$$\mathsf{E}Z^2 = \sum \frac{\lambda^r e^{-\lambda}}{r!} \mathsf{E}Y_r^2 = \sum \frac{\lambda^r e^{-\lambda}}{r!}(r\sigma^2 + r^2\mu^2) =$$
$$= \lambda\sigma^2 + \mu^2 \sum \frac{\lambda^r e^{-\lambda}}{r!} r(r-1) + \mu^2 \sum \frac{\lambda^r e^{-\lambda}}{r!} r = \lambda(\sigma^2 + \mu^2) + \lambda^2\mu^2.$$

Отсюда следует, что

$$\operatorname{Var} Z = \mathsf{E}Z^2 - (\mathsf{E}Z)^2 = \lambda(\sigma^2 + \mu^2).$$

Тогда в силу неравенства Чебышёва можно записать

$$\mathsf{P}\left(Z < \frac{\lambda\mu}{2}\right) \leqslant \mathsf{P}\left(|Z - \lambda\mu| > \frac{\lambda\mu}{2}\right) \leqslant \frac{\operatorname{Var} Z}{(\lambda\mu/2)^2} = \frac{4(\sigma^2 + \mu^2)}{\lambda\mu^2}. \qquad \square$$

Задача 2.3.17. Пусть X и Y — независимые и нормально распределенные с. в. с одной и той же плотностью распределения

$$\frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Найдите плотности распределения следующих величин:

а) $X + Y$;

б) X^2;

в) $X^2 + Y^2$.

Решение. а) Для кумулятивной функции распределения запишем

$$F_{X+Y}(t) = \frac{1}{2\pi} \iint e^{-(x^2+y^2)/2} I(x + y < t)\, dy\, dx.$$

Поскольку п. р. в. $e^{-(x^2+y^2)/2}/2\pi$ симметрична относительно вращений, последнее выражение равно

$$\frac{1}{2\pi} \int e^{-(x^2+y^2)/2} I\left(x < \frac{t}{\sqrt{2}}\right) dy\, dx = \frac{1}{2\sqrt{\pi}} \int_{-\infty}^{t} e^{-u^2/4}\, du,$$

откуда заключаем, что п. р. в. равна

$$f_{X+Y}(x) = \frac{1}{2\sqrt{\pi}} e^{-x^2/4}.$$

б) Аналогично при $t \geqslant 0$ получаем

$$F_{X^2}(t) = \frac{1}{2\pi} \iint e^{-(x^2+y^2)/2} I(x^2 < t)\, dy\, dx = \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{t}}^{\sqrt{t}} e^{-x^2/2}\, dx = \frac{1}{\sqrt{2\pi}} \int_0^t \frac{1}{\sqrt{u}} e^{-u/2}\, du,$$

откуда следует, что

$$f_{X^2}(x) = \frac{1}{\sqrt{2\pi x}} e^{-x/2} I(x \geqslant 0).$$

в) Наконец,

$$F_{X^2+Y^2}(t) = \frac{1}{2\pi}\iint e^{-(x^2+y^2)/2} I(x^2+y^2<t)\,dy\,dx =$$

$$= \frac{1}{2\pi}\int_0^\infty\int_0^{2\pi} re^{-r^2/2} I(r^2<t)\,d\theta\,dr = \frac{1}{2}\int_0^t e^{-u/2}\,du,$$

и п. р. в. имеет вид

$$f_{X^2+Y^2}(x) = \frac{1}{2}e^{-x/2}I(x \geqslant 0). \qquad \square$$

Задача 2.3.18. а) Плотность распределения вероятностей t-распределения с q степенями свободы имеет вид

$$f(x;\ q) = \frac{\Gamma\left(\frac{q+1}{2}\right)}{\Gamma\left(\frac{q}{2}\right)\sqrt{\pi q}}\left(1+\frac{x^2}{q}\right)^{-(q+1)/2}, \quad -\infty < t < \infty.$$

(См. пример 1.1.3 в части В.) Используя свойства показательной функции и тот факт, что

$$\Gamma\left(\frac{q}{2}+b\right) \to \sqrt{2\pi}\left(\frac{q}{2}\right)^{q/2+b-1/2}\exp\left(-\frac{q}{2}\right)$$

при $q\to\infty$, докажите, что $f(x;\ q)$ стремится к плотности распределения вероятностей нормальной случайной величины N(0, 1).

Указание. Можно записать

$$\left(1+\frac{t^2}{q}\right)^{-(q+1)/2} = \left[\left(1+\frac{t^2}{q}\right)^q\right]^{-(1/2+1/2q)}.$$

б) Найдите плотность распределения вероятностей случайной величины $Y = Z^2$, где Z — это N(0, 1)-с. в.

в) Плотность распределения вероятностей для F-распределения с (1, q) степенями свободы имеет вид

$$g(x;\ q) = \frac{\Gamma\left(\frac{q+1}{2}\right)}{\Gamma\left(\frac{q}{2}\right)\sqrt{\pi q}}x^{-1/2}\left(1+\frac{x}{q}\right)^{-(q+1)/2}, \quad 0 < x < \infty.$$

Используя полученные выше предельные результаты, покажите, что $g(x;\ q)$ стремится к плотности распределения с. в. Y при $q\to\infty$.

Решение. б) Плотность распределения вероятностей случайной величины Y равна $\frac{1}{\sqrt{2\pi x}}e^{-x/2}I(x>0)$.

в) Применив предельный переход к отношению гамма-функций, находим

$$\left(1+\frac{v}{q}\right)^{-(q+1)/2} \to \exp\left(-\frac{v}{2}\right).$$

Следовательно, п. р. в. для F-распределения стремится к

$$\frac{\sqrt{q/2}}{\sqrt{\pi q}}x^{-1/2}\exp\left(-\frac{x}{2}\right)=\frac{1}{\sqrt{2\pi x}}\exp\left(-\frac{x}{2}\right),$$

что и требовалось показать.

Этот результат получается вполне естественно. Действительно, из примера 1.1.4 из части В следует, что $F_{1,q}$-распределение связано с отношением $X_1^2 \Big/ \left(\sum\limits_{j=1}^{q} Y_j^2/q\right)$, где $X_1, Y_1, \ldots, Y_q$ — н. о. р. N(0, 1)-с. в. Знаменатель $\sum\limits_{j=1}^{q} Y_j^2/q$ стремится к 1 при $q\to\infty$ по закону больших чисел. □

Задача 2.3.19. Пусть $X_1, X_2, \ldots$ — независимые с. в., каждая из которых имеет распределение Коши с п. р. в.

$$f(x)=\frac{d}{\pi(d^2+x^2)}.$$

Покажите, что случайная величина $(X_1+X_2+\ldots+X_n)/n$ имеет то же самое распределение, что и с. в. X_1.

Противоречит ли это слабому закону больших чисел или центральной предельной теореме?

Указание. Характеристическая функция случайной величины X_1 равна $e^{-|t|}$, и эту же самую характеристическую функцию имеет и с. в. $(X_1+X_2+\ldots+X_n)/n$. Это следует из того, что характеристическая функция однозначно определяет плотность распределения.

Для выполнения ЗБЧ и ЦПТ требуется существование математического ожидания и дисперсии. □

Задача 2.3.20. Пусть $X\sim\mathrm{N}(\mu,\sigma^2)$, и предположим, что $h(x)$, $x\in\mathbb{R}$, — гладкая ограниченная функция. Докажите формулу Стейна

$$\mathsf{E}[(X-\mu)h(X)]=\sigma^2\mathsf{E}[h'(X)].$$

Решение. Не теряя общности, можно положить $\mu=0$. Тогда

$$\mathsf{E}[Xh(X)]=\frac{1}{\sqrt{2\pi}\sigma}\int\limits_{-\infty}^{\infty}xh(x)e^{-x^2/2\sigma^2}\,dx=$$

$$=\frac{1}{\sqrt{2\pi}\sigma}\int\limits_{-\infty}^{\infty}h(x)\,d(-\sigma^2e^{-x^2/2\sigma^2})=\frac{1}{\sqrt{2\pi}\sigma}\int\limits_{-\infty}^{\infty}h'(x)\sigma^2e^{-x^2/2\sigma^2}\,dx,$$

что выполняется, поскольку интегралы сходятся абсолютно. □

Задача 2.3.21. Пусть $X\sim\mathrm{N}(\mu,\sigma^2)$ и Φ — кумулятивная функция распределения случайной величины N(0, 1). Предположим, что $h(x)$, $x\in\mathbb{R}$, — гладкая ограниченная функция. Докажите, что для любых действительных чисел θ, α, β выполняются следующие равенства:

$$\mathsf{E}[e^{\theta X}h(X)] = e^{\mu\theta+\sigma^2\theta^2/2}\,\mathsf{E}[h(X+\theta\sigma^2)],$$

$$\mathsf{E}[\Phi(\alpha X+\beta)] = \Phi\Big(\frac{\alpha\mu+\beta}{\sqrt{1+\alpha^2\sigma^2}}\Big).$$

Решение. Опять, не теряя общности, можно предположить, что $\mu = 0$. Тогда

$$\mathsf{E}[e^{\theta X}h(X)] = \frac{1}{\sqrt{2\pi}\sigma}\int\limits_{-\infty}^{\infty} e^{\theta x - x^2/2\sigma^2}h(x)\,dx =$$

$$= \frac{1}{\sqrt{2\pi}\sigma}\int\limits_{-\infty}^{\infty} e^{-(x-\sigma^2\theta)^2/2\sigma^2}e^{\sigma^2\theta^2/2}h(x)\,dx =$$

$$= e^{\sigma^2\theta^2/2}\frac{1}{\sqrt{2\pi}\sigma}\int\limits_{-\infty}^{\infty} e^{-(x-\sigma^2\theta)^2/2\sigma^2}h(x)\,dx = e^{\sigma^2\theta^2/2}\,\mathsf{E}[h(X+\theta\sigma^2)].$$

Все интегралы здесь сходятся абсолютно.

При доказательстве второй формулы будем рассматривать случай общего μ, для того чтобы изложение было более прозрачным. Если $Z \sim \mathrm{N}(0,\ 1)$ и Z не зависит от X, то

$$\mathsf{E}[\Phi(\alpha X+\beta)] = \mathsf{P}(Z \leqslant \alpha X+\beta) = \mathsf{P}(Z - \alpha(X-\mu) \leqslant \alpha\mu+\beta) =$$

$$= \mathsf{P}\Big(\frac{Z-\alpha(X-\mu)}{\sqrt{1+\alpha^2\sigma^2}} \leqslant \frac{\alpha\mu+\beta}{\sqrt{1+\alpha^2\sigma^2}}\Big) = \Phi\Big(\frac{\alpha\mu+\beta}{\sqrt{1+\alpha^2\sigma^2}}\Big). \quad \square$$

Задача 2.3.22. Предположим, что пара $(X,\ Y)$ имеет совместное нормальное распределение и $h(x)$ — гладкая ограниченная функция. Докажите следующие соотношения:

$$\mathsf{E}[(Y-\mathsf{E}Y)X] = \frac{\mathrm{Cov}[X,\ Y]}{\mathrm{Var}\,X}(X-\mathsf{E}X)$$

и

$$\mathrm{Cov}[h(X),\ Y] = \mathsf{E}[h'(X)]\,\mathrm{Cov}[X,\ Y].$$

Решение. Опять предположим, что обе с. в. X и Y имеют нулевые средние. Для совместной плотности $f_{X,Y}(x,\ y)$ запишем

$$\propto \exp\Big[-\frac{1}{2}(x,\ y)\Sigma^{-1}\binom{x}{y}\Big],$$

где $\Sigma = (\Sigma_{ij})$ — это ковариационная матрица размера 2×2. Тогда при условии $X = x$ для плотности $f_Y(y\,|\,x)$ можно записать

$$\propto \exp\Big[-\frac{1}{2}((\Sigma^{-1})_{22}y^2 + 2xy(\Sigma^{-1})_{12})\Big].$$

Напомним, что Σ^{-1} — матрица, обратная ковариационной матрице Σ. Это показывает, что условная плотность $f_Y(y\,|\,x)$ является гауссовской

плотностью, у которой среднее — линейная функция по x. Таким образом,
$$\mathsf{E}[Y\,|\,X] = \gamma X.$$
Для того чтобы найти γ, умножим обе части на X и вычислим математическое ожидание. В левой части получим $\mathsf{E}(XY) = \mathrm{Cov}[X,\ Y]$, а правая часть равна $\gamma \operatorname{Var} X$. Из этого результата в силу формулы Стейна следует второе равенство:
$$\mathrm{Cov}[h(X),\ Y] = \mathsf{E}[h(X)Y] = \mathsf{E}[h(X)\,\mathsf{E}(Y\,|\,X)] = \frac{\mathrm{Cov}[X,\ Y]}{\operatorname{Var} X}\,\mathsf{E}[Xh(X)] = \\ = \mathsf{E}[h'(X)]\,\mathrm{Cov}[X,\ Y]. \quad \square$$

В заключение части А заметим, что формула Стирлинга следует из неравенства Чебышёва. Пусть $X_1, \ldots, X_n$ — независимые одинаково распределенные случайные величины и $X_i \sim \mathrm{Exp}(1)$. Ясно, что в этом случае $\mathsf{E}X_i = \operatorname{Var} X_i = 1$. В силу неравенства Чебышёва для любого $a > 0$ имеем
$$\mathsf{P}\left(\frac{1}{\sqrt{n}}\left|\sum_{i=1}^{n} X_i - n\right| \leqslant a\right) = \mathsf{P}\left(\frac{1}{\sqrt{n}}\left|\sum_{i=1}^{n}(X_i - \mathsf{E}X_i)\right| \leqslant a\right) \geqslant \\ \geqslant 1 - \frac{1}{na^2} n \operatorname{Var} X_1 = 1 - \frac{1}{a^2}. \quad (2.3.10)$$
С другой стороны, $\sum\limits_{i=1}^{n} X_i \sim \mathrm{Gam}(n,\ 1)$ с плотностью
$$f(x) = \frac{1}{(n-1)!} x^{n-1} e^{-x} I(x > 0).$$

Поэтому для достаточно больших n левая часть неравенства (2.3.10) равна, а правая часть не превосходит
$$\frac{1}{(n-1)!} \int\limits_{-\sqrt{n}a+n}^{\sqrt{n}a+n} x^{n-1} e^{-x}\, dx = \frac{n^{n-1/2} e^{-n}}{(n-1)!} \int\limits_{-a}^{a} \left(1 + \frac{y}{\sqrt{n}}\right)^{n-1} e^{-\sqrt{n}y}\, dy = \\ = \frac{n^{n-1/2} e^{-n}}{(n-1)!} \int\limits_{-a}^{a} \exp\left(-\frac{y^2}{2} + A_n(y)\right) dy < 1.$$

Здесь $A_n(y) = (n-1)\ln\left(1 + \frac{y}{\sqrt{n}}\right) - \sqrt{n}y + \frac{y^2}{2} \to 0$ при $n \to \infty$ равномерно по y, $-a \leqslant y \leqslant a$.

Следовательно, правая часть неравенства (2.3.10) равна
$$1 - \frac{1}{a^2} \leqslant \liminf_{n\to\infty} \frac{n^{n+1/2} e^{-n}}{n!} \int\limits_{-a}^{a} e^{-y^2/2}\, dy \leqslant \limsup_{n\to\infty} \frac{n^{n+1/2} e^{-n}}{n!} \int\limits_{-a}^{a} e^{-y^2/2}\, dy \leqslant 1.$$

При $a \to \infty$ получаем $\int\limits_{-a}^{a} e^{-y^2/2}\, dy \to \sqrt{2\pi}$ и $\lim\limits_{n\to\infty} \sqrt{2\pi} \frac{n^{n+1/2} e^{-n}}{n!} = 1$, что и доказывает формулу (1.6.20).

Часть B

Основы статистики

Глава 1

Оценивание параметров

§ 1.1. Предварительные сведения. Некоторые важные вероятностные распределения

Все модели ошибочны, но некоторые из них полезны.

Дж. Бокс (1919—), британский статистик

Model without a Cause

Модель, основанная ни на чем

(Из серии «Фильмы, которые не вышли на большой экран».)

Во второй части этого тома рассмотрен материал, относящийся к курсу «Статистика» второго года обучения. Этот материал является естественным продолжением курса «Вероятность». Статистика, которую на факультете математики Кембриджского университета называют «applicable» (т. е. дисциплиной, тяготеющей к приложениям), в настоящее время занимает место где-то между «чистыми» и «прикладными» дисциплинами в общей схеме курсов Кембриджа. Согласно одному из современных определений, статистика — это *набор процедур и принципов, применяемых для сбора и обработки информации с целью принятия решений в условиях неопределенности*. Интересно сравнить это определение с более ранними интерпретациями термина «статистика» и связанных с ним терминов. Традиционно слова «статистический» (англ. «statistic») и «статистика» (англ. «statistics») происходят от слова «state», взятого в значении «политическая форма правления». Действительно, слово «statist» появляется в трагедии Шекспира «Гамлет», акт 5, сцена 2[1]:

Hamlet: Being thus benetted round with villainies, —
Ere I could make a prologue to my brains,
They had begun the play, — I sat me down,
Devis'd a new commission; wrote it fair:
I once did hold it, as our *statists* do,

[1] Перевод М. Лозинского.

A baseness to write fair, and labour'd much
How to forget that learning; but, sir, now
It did me yeoman's service. Wilt thou know
Th' effect of what I wrote?

Гамлет: Итак, кругом опутан негодяйством, —
Мой ум не сочинил еще пролога,
Как приступил к игре, — я сел, составил
Другой приказ; переписал красиво;
Когда-то я считал, как наша *знать*,
Стыдом писать красиво и старался
Забыть искусство это; но теперь
Оно мне удружило. Хочешь знать,
Что написал я?

а затем в пьесе Шекспира «Цимбелин», акт 2, сцена 4[1]:

Posthumus: I do believe,
Statist though I am none, nor like to be,
That this will prove a war; ...

Постум: Хоть не *политик* я и им не стану,
Но полагаю — быть войне. ...

Слово «statist», по-видимому, обозначает лицо, выполняющее государственную функцию. (Глоссарий к книге «The Complete Works by William Shakespeare. The Alexander Text» (London и Glasgow: Collins, 1990) попросту определяет значение этого слова как «государственный деятель».) В подобном значении это же слово используется в произведении Мильтона «Обретенный рай», книга четвертая, строка 355.

Многочисленные определения статистики, появившиеся до 1935 г., можно найти в [Wil]; их смысл по сути сводится к такому: «описание прошлой или настоящей политической или финансовой ситуации данного государства». Характерно, что Наполеон описывал статистику как «бюджет вещей».

Определение статистики занимало многие умы и позднее, после 1935 г., см. [NiFY], широкий спектр мнений по этому вопросу был представлен различными авторами (а иногда и одним и тем же автором в разные периоды времени). Политические и идеологические факторы вносили свою лепту в эту запутанную ситуацию: авторы советской эпохи сообща атаковали западных за описание статистики скорее как методологической, чем матери-

[1]Перевод Н. Мелковой.

альной науки. Пределом абсурда явилось провозглашение существования «пролетарской статистики», которая противопоставлялась «буржуазной статистике». Первая из них помогала «борьбе рабочего класса против его эксплуататоров», тогда как вторая была «служанкой монополистического капитала».

Особым предметом разногласий стал вопрос о месте и роли математической статистики. Например, Дж. Бокс (1919—), американский математик британского происхождения, начинавший свою карьеру студентом-химиком и служивший затем как статистик-практик в британской армии во время Второй мировой войны, писал, что было «ошибкой изобретать термин „математическая статистика“. Этот серьезный промах послужил источником бесчисленных трудностей».

Интересно сравнить это с двумя довольно различными высказываниями Дж. У. Тьюки (1915—2000), одного из величайших специалистов всех времен в статистике и во многих областях прикладной математики, которому приписывают, в числе многих других вещей, изобретение терминов «бит» (англ. «bit» — сокращение для «binary digit») и «software» — программное обеспечение. Тьюки, по образованию чистый математик (его докторская степень была получена в области топологии), говорил: «Статистика — это часть сложной и запутанной сети, связывающей математику, научную философию и другие отрасли науки, включая выборочные обследования, с процедурами анализа и, иногда, сбора данных». С другой стороны, Тьюки выразил следующее мнение: «Статистика — это часть прикладной математики, которая имеет дело со случайными процессами (хотя и не только с ними)». Эта точка зрения на статистику была поддержана в значительной части университетов (в том числе в Кембридже), в которых многие сотрудники статистических факультетов или подразделений (включая заведующих кафедрами математической статистики) являются на самом деле специалистами по случайным процессам.

В этой части книги, посвященной статистике, изучаются различные способы обработки данных (результатов наблюдений) и выводы, сделанные на их основе: точечное оценивание, интервальное оценивание, проверка гипотез, регрессионный анализ. Некоторые из применяемых методов будут опираться на четкое логическое обоснование, а некоторые другие возникли *ad hoc*, т. е. специально для решения конкретных задач, и взяты на вооружение лишь потому, что они дают ответы на важные практические вопросы.

Может показаться, что после десятилетий стараний и усилий (особенно в 1940—1970 гг.) все попытки свести современную статистику к единому и строгому обоснованию сейчас почти забыты в большей части академических кругов. (Возможно, это и преувеличение, но именно так зачастую выглядит ситуация с точки зрения неспециалистов.) Однако такой авторитетный ученый, как Рао (с именем которого связаны теорема Рао—Блекуэлла и неравенство Крамера—Рао; см. ниже), подчеркивает, что связи между статистикой и математикой становятся все прочнее и многообразнее.

С другой стороны, в течение последних 30 лет наблюдается замечательное распространение статистических методов буквально во все области научного анализа, и главным основанием для их применения служит то, что эти методы работают, и работают успешно. С приходом современных вычислительных технологий (включая пакеты SPSS, MINITAB и SPLUS) появилась возможность обрабатывать и анализировать огромные массивы данных и отображать результаты в удобной форме. Можно сказать, что *компьютеры освободили статистиков от жесткой необходимости использовать только простые модели, допускающие анализ с помощью ручки и бумаги* [We].

Необходимо подчеркнуть, что даже (и особенно) на начальном этапе изучения статистики тщательные вычисления «вручную» являются чрезвычайно важными для успешной сдачи экзаменов, и мы советуем кандидатам уделять серьезное внимание вычислительной работе.

Для изучения курса «Статистика» следует освежить знания некоторых ключевых фактов из курса «Вероятность». Эти факты охватывают следующие основные понятия: вероятностные распределения, плотности распределения вероятностей, случайные величины, математическое ожидание, совместные распределения, ковариация, независимость. В случае дискретных случайных величин удобно рассматривать дискретную функцию распределения (д. ф. р.), а в случае непрерывных величин — плотность распределения вероятностей (п. р. в.). Традиционно статистические курсы начинаются с изучения некоторых важных семейств д. ф. р./п. р. в., зависящих от параметра (или нескольких параметров, образующих вектор). Например, д. ф. р. распределения Пуассона $\mathrm{Po}(\lambda)$ зависит от параметра $\lambda > 0$, так же как и п. р. в. показательного распределения $\mathrm{Exp}(\lambda)$. Нормальные п. р. в. параметризованы парой (μ, σ^2), где $\mu \in \mathbb{R}$ — это среднее и $\sigma^2 > 0$ — дисперсия. «Истинное» значение параметра (или нескольких параметров) считается неизвестным, и нам необходимо разработать средства для его оценивания.

Значительная часть курса имеет дело с н. о. р. нормальными $\mathrm{N}(0, 1)$-с. в. $X_1, X_2, \ldots$ и функциями от них. Простейшими функциями являются линейные комбинации $\sum\limits_{i=1}^{n} a_i X_i$.

Пример 1.1.1. *Линейные комбинации независимых нормальных с. в.* Мы уже обсуждали свойства линейности нормальных с. в. в § 2.3

части А; здесь мы напомним их с некоторыми видоизменениями. Предположим, что с. в. $X_1, \ldots, X_n$ являются независимыми и $X_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$. Совместная п. р. в. этих величин имеет вид

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{1}{2}(x_i - \mu_i)^2/\sigma_i^2\right), \quad \boldsymbol{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n. \tag{1.1.1}$$

Тогда для любых действительных $a_1, \ldots, a_n$ мы имеем

$$\sum_i a_i X_i \sim \mathrm{N}\left(\sum_i a_i \mu_i, \sum_i a_i^2 \sigma_i^2\right). \tag{1.1.2}$$

В частности, если $a_1 = \ldots = a_n = 1/n$, $\mu_1 = \ldots = \mu_n = \mu$ и $\sigma_1 = \ldots = \sigma_n = \sigma$, то

$$\frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right). \tag{1.1.3}$$

В общем случае

$$\sum_{i=1}^{n}(X_i - \mu_i) \Big/ \left(\sum_{i=1}^{n} \sigma_i^2\right)^{1/2} \sim \mathrm{N}(0, 1).$$

Далее, предположим, что $A = (A_{ij})$ — это действительная обратимая матрица размера $n \times n$, т. е. $\det A \neq 0$, и $A^{-1} = (A'_{ij})$ — обратная матрица. Запишем

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \quad \text{и} \quad \boldsymbol{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$$

и рассмотрим взаимно обратные линейные преобразования $\boldsymbol{Y} = A^T \boldsymbol{X}$ и $\boldsymbol{X} = (A^{-1})^T \boldsymbol{Y}$, так что

$$Y_j = (A^T \boldsymbol{X})_j = \sum_{i=1}^{n} X_i A_{ij}, \quad X_i = ((A^{-1})^T \boldsymbol{Y})_i = \sum_{j=1}^{n} Y_j A'_{ji}.$$

Тогда с. в. $Y_1, \ldots, Y_n$ являются совместно нормальными. Более точно, совместная п. р. в. $f_{\boldsymbol{Y}}(\boldsymbol{y})$ вычисляется как

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \frac{1}{|\det A|} f_{\boldsymbol{X}}[(A^{-1})^T \boldsymbol{y}] = \frac{1}{|\det A|} \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left[-\frac{1}{2\sigma_j^2}\left(\sum_{i=1}^{n} y_i A'_{ij} - \mu_j\right)^2\right] =$$

$$= \frac{1}{(2\pi)^{n/2}(\det(A^T \Sigma A))^{1/2}} \exp\left[-\frac{1}{2}\left\langle(\boldsymbol{y} - A^T \boldsymbol{\mu}), (A^T \Sigma A)^{-1}(\boldsymbol{y} - A^T \boldsymbol{\mu})\right\rangle\right].$$

Здесь, как и ранее, $\langle\cdot,\cdot\rangle$ обозначает скалярное произведение в $\mathbb{R}^n$,

$$\mu=\begin{pmatrix}\mu_1\\ \vdots\\ \mu_n\end{pmatrix},\quad y=\begin{pmatrix}y_1\\ \vdots\\ y_n\end{pmatrix}\in\mathbb{R}^n\quad\text{и}\quad \Sigma=\begin{pmatrix}\sigma_1^2 & 0 & \dots & 0\\ 0 & \sigma_2^2 & \dots & 0\\ \dots & \dots & \dots & \dots\\ 0 & 0 & \dots & \sigma_n^2\end{pmatrix}.$$

Напомним, что μ и Σ — это вектор средних значений и ковариационная матрица вектора $\boldsymbol{X}$ соответственно.

Легко видеть, что вектор средних значений вектора $\boldsymbol{Y}$ равен $A^T\mu$, а ковариационная матрица равна $A^T\Sigma A$:

$$\mathsf{E}Y_j=\sum_{i=1}^n A_{ij}\mu_i,\qquad \mathrm{Cov}(Y_i,\,Y_j)=\sum_{k=1}^n A_{ki}\sigma_{kk}A_{kj}.$$

Предположим теперь, что A — это действительная ортогональная матрица размера $n\times n$ и $\sum\limits_k A_{ki}A_{kj}=\delta_{ij}$, т. е. A^TA — единичная матрица размера $n\times n$. Тогда $\det A=\pm1$. Предположим, что введенные выше с. в. X_i имеют одинаковую дисперсию: $\sigma_1^2=\ldots=\sigma_n^2=\sigma^2$. Тогда

$$\mathrm{Cov}(Y_i,\,Y_j)=\mathrm{Cov}[(A^T\boldsymbol{X})_i,\,(A^T\boldsymbol{X})_j]=\sum_{k,l=1}^n A_{ki}A_{lj}\,\mathrm{Cov}(X_k,\,X_l)=$$
$$=\sigma^2\sum_{k,l}A_{ki}A_{lj}\delta_{k,l}=\sigma^2\sum_k A_{ki}A_{kj}=\sigma^2\delta_{i,j}.$$

Это означает, что случайный вектор $\boldsymbol{X}^TA=\boldsymbol{Y}^T$ имеет независимые компоненты $Y_1,\,\ldots,\,Y_n$ и $Y_j\sim\mathrm{N}((A^T\mu)_j,\,\sigma^2)$. □

Пример 1.1.2. *Суммы квадратов*: *распределение* χ^2. Еще один пример, который будет неоднократно использоваться в дальнейшем, — это сумма квадратов. Пусть $X_1,\,X_2,\,\ldots$ — это н. о. р. $\mathrm{N}(0,\,1)$-с. в. Распределение суммы

$$\sum_{i=1}^n X_i^2$$

называют *распределением хи-квадрат*, или χ_n^2-*распределением с* n *степенями свободы*, или, сокращенно, χ_n^2-распределением. Как будет установлено ниже, это распределение имеет п. р. в. $f_{\chi_n^2}$, которая сконцентрирована на положительной полуоси $(0,\,\infty)$:

$$f_{\chi_n^2}(x)\propto x^{n/2-1}e^{-x/2}I(x>0),$$

где коэффициент пропорциональности равен $1/[\Gamma(n/2)2^{n/2}]$. Здесь

$$\Gamma(n/2)=\int_0^\infty\frac{1}{2^{n/2}}x^{n/2-1}e^{-x/2}\,dx.$$

Нетрудно заметить, что χ^2_n-распределение является $\mathrm{Gam}(\alpha, \lambda)$-распределением с $\alpha = n/2$, $\lambda = 1/2$. С другой стороны, если $X_1, \ldots, X_n$ — это н. о. р. $\mathrm{N}(\mu, \sigma^2)$-с. в., то

$$\sum_{i=1}^{n}(X_i - \mu)^2 \sim \mathrm{Gam}\left(\frac{n}{2}, \frac{1}{2\sigma^2}\right) \quad \text{и} \quad \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2 \sim \chi^2_n. \qquad (1.1.4)$$

Среднее значение χ^2-распределения равно n, а дисперсия равна $2n$. Все χ^2-п. р. в. являются унимодальными. Графики некоторых п. р. в. $f_{\chi^2_n}$ представлены на рис. 1.1.

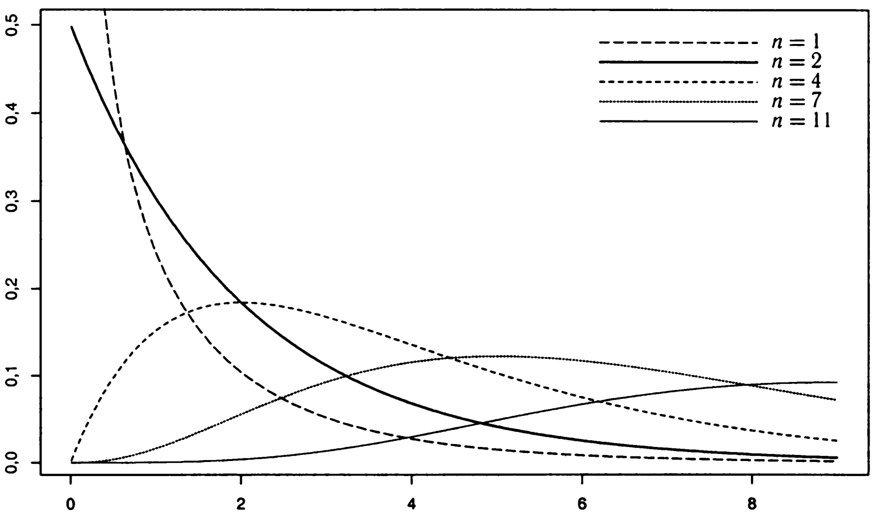


Рис. 1.1. Плотности χ^2-распределения при $n = 1$, $n = 2$, $n = 4$, $n = 7$, $n = 11$

Полезным свойством семейства χ^2-распределений является его замкнутость относительно независимого суммирования. Это означает, что если с. в. $Z \sim \chi^2_n$ и $Z' \sim \chi^2_{n'}$ независимы, то $Z + Z' \sim \chi^2_{n+n'}$. Конечно, χ^2-распределения унаследовали это свойство от гамма-распределений.

Можно легко проверить, что

$$f_{\chi^2_n}(x) = \frac{1}{\Gamma(n/2)} \frac{1}{2^{n/2}} x^{n/2-1} e^{-x/2} I(x > 0), \qquad (1.1.5)$$

если воспользоваться производящей функцией моментов. Производящая функция моментов $M_{X_i^2}(\theta) = \mathsf{E}e^{\theta X_i^2}$ равна

$$\frac{1}{\sqrt{2\pi}} \int e^{\theta x^2} e^{-x^2/2}\, dx = \frac{1}{\sqrt{2\pi}} \int e^{-(1-2\theta)x^2/2}\, dx =$$
$$= \frac{1}{\sqrt{1-2\theta}} \frac{1}{\sqrt{2\pi}} \int e^{-y^2/2}\, dy = \frac{1}{\sqrt{1-2\theta}}, \quad \theta < \frac{1}{2},$$

т. е. она является производящей функцией моментов для распределения Gam(1/2, 1/2). Далее, производящая функция моментов $M_{Y_n}(t)$ величины $Y_n = \sum_{i=1}^{n} X_i^2$ равна степени $(M_{X_i^2}(\theta))^n = (1-2\theta)^{-n/2}$. Последнее выражение представляет собой производящую функцию моментов для распределения Gam(n/2, 1/2). Следовательно, $f_{Y_n} \sim$ Gam(n/2, 1/2), что и требовалось показать. □

Пример 1.1.3. *Распределение Стьюдента* (t-*распределение*). Как и выше, пусть $X_1, X_2, \ldots$ — это н. о. р. N(0, 1)-с. в., тогда распределение отношения

$$\frac{X_{n+1}}{\left(\sum_{i=1}^{n} X_i^2/n\right)^{1/2}}$$

называется *распределением Стьюдента с n степенями свободы* или, кратко, t_n-распределением. Это распределение имеет п. р. в. f_{t_n}, которая определена на всей оси $\mathbb{R}$ и является симметричной (четной) относительно инверсии $x \mapsto -x$:

$$f_{\mathrm{t}_n}(t) \propto \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2},$$

где коэффициент пропорциональности равен

$$\frac{1}{\sqrt{\pi n}} \frac{\Gamma((n+1)/2)}{\Gamma(n/2)}.$$

Очевидно, при $n > 1$ это распределение имеет среднее значение 0. При $n > 2$ дисперсия равна $n/(n-2)$. Все п. р. в. распределения Стьюдента унимодальны. Образцы графиков п. р. в. f_{t_n} показаны на рис. 1.2.

Эти п. р. в. похожи на нормальные п. р. в. (и, как показано в задаче 2.3.18 части А, $f_{\mathrm{t}_n}(t)$ приближается к $e^{-t^2/2}/\sqrt{2\pi}$ при $n \to \infty$). Однако при конечных значениях n «хвосты» плотности f_{t_n} более «тяжелые», чем у нормальных п. р. в. В частности, производящая функция моментов для t-распределения не существует (за исключением точки $\theta = 0$): если $X \sim \mathrm{t}_n$, то $\mathrm{E}e^{\theta X} = \infty$ для любого $\theta \in \mathbb{R} \setminus \{0\}$.

Отметим, что если $n = 1$, то t_1-распределение совпадает с распределением Коши.

Чтобы получить формулу для п. р. в. f_{t_n} для t_n-распределения, заметим, что эта плотность совпадает с п. р. в. отношения $T = X/\sqrt{Y/n}$, где с. в. X и Y независимы и $X \sim$ N(0, 1), $Y \sim \chi_n^2$. Тогда

$$f_{X,Y}(x, y) = \frac{1}{\sqrt{2\pi}} \frac{2^{-n/2}}{\Gamma(n/2)} e^{-x^2/2} y^{n/2-1} e^{-y/2} I(y > 0).$$

Якобиан $\dfrac{\partial(t, u)}{\partial(x, y)}$ преобразования переменных

$$t = \frac{x}{\sqrt{y/n}}, \quad u = y$$

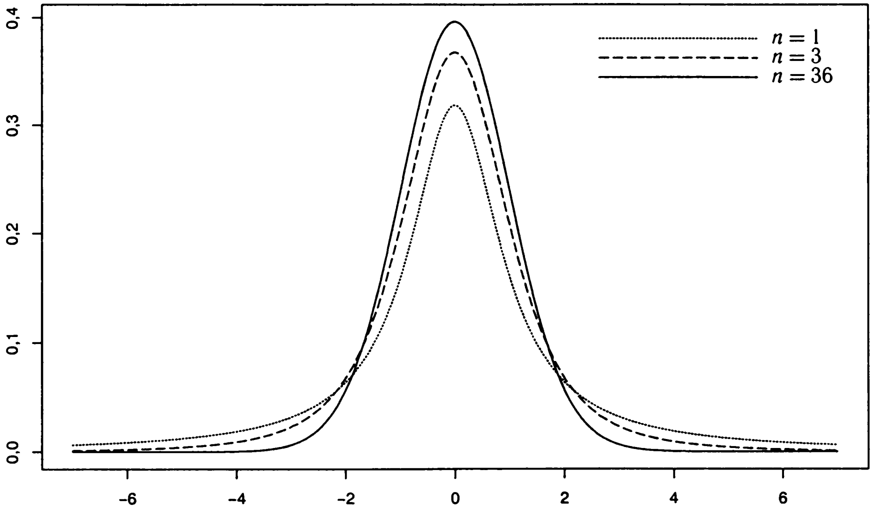


Рис. 1.2. Плотности распределения Стьюдента при $n = 1$, $n = 2$, $n = 36$

равен $(n/y)^{1/2}$, а обратный якобиан равен $\dfrac{\partial(x, y)}{\partial(t, u)} = (u/n)^{1/2}$. Следовательно,

$$f_{\mathrm{t}_n}(t) = f_T(t) = \int\limits_0^\infty f_{X,Y}(t(u/n)^{1/2}, u)\left(\frac{u}{n}\right)^{1/2} du =$$

$$= \frac{1}{\sqrt{2\pi}} \frac{2^{-n/2}}{\Gamma(n/2)} \int\limits_0^\infty e^{-t^2 u/2n} u^{n/2-1} e^{-u/2} \left(\frac{u}{n}\right)^{1/2} du =$$

$$= \frac{1}{\sqrt{2\pi}} \frac{2^{-n/2}}{\Gamma(n/2) n^{1/2}} \int\limits_0^\infty e^{-(1+t^2/n)u/2} u^{(n+1)/2-1}\, du.$$

Последнее подынтегральное выражение равно п. р. в. распределения $\mathrm{Gam}\left(\dfrac{n+1}{2}, \dfrac{1}{2} + \dfrac{t^2}{2n}\right)$. Следовательно,

$$f_{\mathrm{t}_n}(t) = \frac{1}{\sqrt{\pi n}} \frac{\Gamma((n+1)/2)}{\Gamma(n/2)} \left(\frac{1}{1 + t^2/n}\right)^{(n+1)/2}, \tag{1.1.6}$$

что и дает вышеприведенную формулу. □

Пример 1.1.4. F-*распределение Фишера*. Пусть теперь $X_1, \ldots, X_m$ и $Y_1, \ldots, Y_n$ — это н. о. р. N(0, 1)-с. в. Отношение

$$\frac{\sum\limits_{i=1}^m X_i^2/m}{\sum\limits_{j=1}^n Y_j^2/n}$$

имеет распределение, называемое *распределением Фишера*, или F-*распределением с параметрами* (*степенями свободы*) m, n, или, кратко, $\mathrm{F}_{m,n}$-распределением. Соответствующая п. р. в. $f_{\mathrm{F}_{m,n}}$ сконцентрирована на положительной полуоси:

$$f_{\mathrm{F}_{m,n}}(x) \propto x^{m/2-1}\left(1+\frac{m}{n}x\right)^{-(m+n)/2} I(x>0), \tag{1.1.7}$$

где коэффициент пропорциональности равен

$$\frac{\Gamma((m+n)/2)}{\Gamma(m/2)\Gamma(n/2)}\left(\frac{m}{n}\right)^{m/2}.$$

Среднее значение F-распределения равно $n/(n-2)$ (при $n>2$, независимо от значения m), а дисперсия равна

$$\frac{2n^2(m+n-2)}{m(n-2)^2(n-4)} \quad (\text{для } n>4).$$

Заметим, что

$$\text{если } Z\sim \mathrm{t}_n, \text{ то } Z^2\sim \mathrm{F}_{1,n}, \qquad \text{а если } Z\sim \mathrm{F}_{m,n}, \text{ то } Z^{-1}\sim \mathrm{F}_{n,m}.$$

Примеры графиков п. р. в. $f_{\mathrm{F}_{m,n}}$ представлены на рис. 1.3.

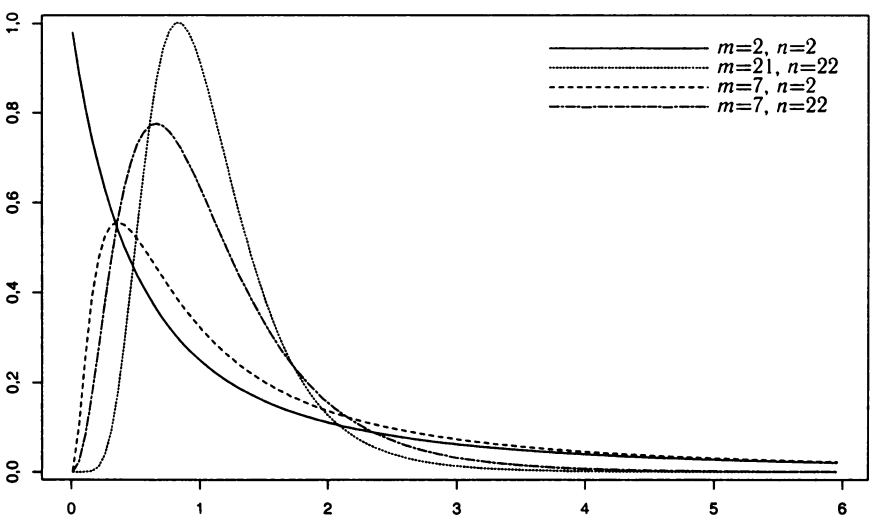


Рис. 1.3. Плотности распределения Фишера при $m=2$, $n=2$; $m=7$, $n=2$; $m=21$, $n=22$; $m=7$, $n=22$

Распределение Фишера часто называют распределением Снедекора—Фишера. Вышеприведенная формула для п. р. в. $f_{\mathrm{F}_{m,n}}$ может быть установлена аналогично формуле для f_{t_n}; мы опускаем соответствующие вычисления. □

Пример 1.1.5. *Бета-распределение* — это вероятностное распределение на интервале (0, 1) с п. р. в.

$$f(x) \propto x^{\alpha-1}(1-x)^{\beta-1} I(0 < x < 1), \tag{1.1.8}$$

где α, $\beta > 0$ — параметры. Коэффициент пропорциональности равен

$$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} := \frac{1}{\mathrm{B}(\alpha, \beta)},$$

где $\mathrm{B}(\alpha, \beta)$ — это *бета-функция*. Мы будем записывать $X \sim \mathrm{Bet}(\alpha, \beta)$, если с. в. X имеет указанную выше п. р. в. Бета-распределение используется для описания различных случайных дробей. У этого распределения

$$\mathsf{E}X = \frac{\alpha}{\alpha+\beta}, \quad \mathrm{Var}\,X = \frac{\alpha\beta}{(\alpha+\beta)(\alpha+\beta+1)}.$$

Графики п. р. в. для бета-распределения представлены на рис. 1.4.

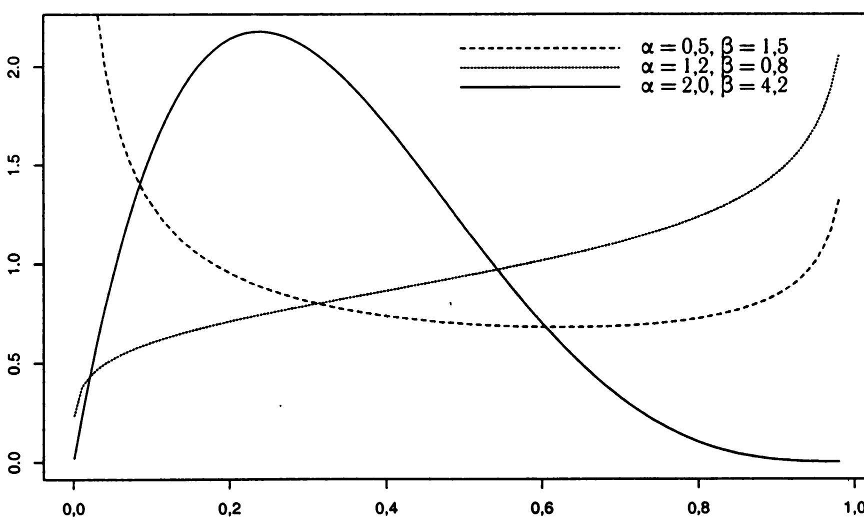


Рис. 1.4. Плотности бета-распределения при $\alpha = 0{,}5$, $\beta = 1{,}5$; $\alpha = 1{,}2$, $\beta = 0{,}8$; $\alpha = 2{,}0$, $\beta = 4{,}2$

Интересно отметить, что если $X \sim \mathrm{F}_{m,n}$, то $\dfrac{(m/n)X}{1+(m/n)X} = \dfrac{mX}{n+mX} \sim$ $\sim \mathrm{Bet}(m/2, n/2)$. □

Дальнейшие примеры читатель может найти в таблице вероятностных распределений.

Важное значение будет придаваться работе с *квантилями* этих (и других) распределений. Для заданного $\gamma \in (0, 1)$ *верхняя* γ-*квантиль* (квантиль порядка γ), или *верхняя* γ-*точка*, $a_+(\gamma)$ для д. ф. р./п. р. в. f, определяется из уравнения

$$\sum_{x \geqslant a_+(\gamma)} f(x) = \gamma \quad \text{или} \quad \int_{a_+(\gamma)}^{\infty} f(x)\,dx = \gamma.$$

Аналогично *нижняя* γ-*квантиль* (*нижняя* γ-*точка*) $a_-(\gamma)$ определяется из уравнения

$$\sum_{x \leqslant a_-(\gamma)} f(x) = \gamma \quad \text{или} \quad \int_{-\infty}^{a_-(\gamma)} f(x)\,dx = \gamma.$$

Очевидно, что для квантилей п. р. в.

$$a_-(\gamma) = a_+(1-\gamma), \quad 0 < \gamma < 1. \tag{1.1.9}$$

Для случая д. ф. р. уравнение (1.1.9) необходимо модифицировать, принимая во внимание, достигается ли значение $a_-(\gamma)$ или нет (т. е. выполняется ли соотношение $f(a_-(\gamma)) > 0$ или $f(a_-(\gamma)) = 0$).

Если γ измеряется в процентах, то говорят о *процентилях* (*процентных точках*) заданного распределения. Квантили и процентили нормального распределения, χ^2-, t- и F-распределений можно найти в стандартных статистических таблицах. Современные пакеты дают возможность вычислять их с высокой точностью для практически любого заданного распределения.

Некоторые основные таблицы процентилей приведены с любезного разрешения Р. Вебера (см. с. 295—297).

Эти таблицы дают такие значения x, что определенный процент распределения находится слева от x. Например, если $X \sim t_3$, то $P(X \leqslant 5{,}84) = 0{,}995$ и $P(-5{,}84 \leqslant X \leqslant 5{,}84) = 0{,}99$. Если $X \sim F_{8,5}$, то $P(X \leqslant 4{,}82) = 0{,}95$.

Таблицы 1.1—1.4 можно использовать для проверки гипотез с уровнями 0,01, 0,05 и 0,10. Для F-распределения представлены только 95-процентные точки; этого достаточно для применения одностороннего критерия с уровнем 0,05. Таблицы точек с другими процентными значениями могут быть найдены в любой книге по статистике или вычислены при помощи компьютерного программного обеспечения. (Эти таблицы были построены при помощи функций, доступных в Microsoft Excel.)

Отметим, что последовательности процентных точек для t_n-распределения сходятся к соответствующим точкам N(0, 1)-распределения при $n \to \infty$.

§ 1.2. Оценки. Несмещенность

License to Sample
You Only Estimate Twice
The Estimator[1]

(Из серии «Фильмы, которые не вышли на большой экран».)

Мы начнем этот параграф с введения понятий несмещенности и достаточности. Основная модель, рассматриваемая в гл. 1 и 2, состоит в следующем: наблюдается *выборка* заданного объема n, состоящая из значений н. о. р. действительных с. в. $X_1, \ldots, X_n$ с общей д. ф. р./п. р. в. $f(x;\,\theta)$. Обозначение $f(x;\,\theta)$ вводится, чтобы подчеркнуть, что рассматриваемая д. ф. р./п. р. в. зависит от параметра θ, значения которого изменяются в пределах некоторой заданной области Θ. Совместная д. ф. р./п. р. в. случайного вектора $\boldsymbol{X}$ обозначается $f_{\boldsymbol{X}}(\boldsymbol{x};\,\theta)$ или $f(\boldsymbol{x};\,\theta)$ и равна произведению

$$f_{\boldsymbol{X}}(\boldsymbol{x};\,\theta)=f(x_1;\,\theta)\ldots f(x_n;\,\theta),\qquad \boldsymbol{X}=\begin{pmatrix}X_1\\ \vdots\\ X_n\end{pmatrix},\qquad \boldsymbol{x}=\begin{pmatrix}x_1\\ \vdots\\ x_n\end{pmatrix}. \tag{1.2.1}$$

Здесь и далее вектор $\boldsymbol{x}$ обозначает выборочное значение вектора $\boldsymbol{X}$. (Сложилась традиция обозначать прописными буквами случайные величины, а строчными буквами — их выборочные значения.) Распределение вероятностей, порождаемое функцией $f_{\boldsymbol{X}}(\cdot\,;\,\theta)$, обозначается P_θ, а математическое ожидание и дисперсия относительно P_θ обозначаются E_θ и Var_θ.

В непрерывной модели, когда мы имеем дело с п. р. в., аргумент x может принимать значения в $\mathbb{R}$; а точнее, в области, где $f(x;\,\theta)>0$ хотя бы для одного $\theta\in\Theta$. Аналогично $\boldsymbol{x}\in\mathbb{R}^n$ — это вектор из множества, на котором $f_{\boldsymbol{X}}(\boldsymbol{x};\,\theta)>0$ хотя бы для одного значения $\theta\in\Theta$. В дискретной модели, когда $f(x;\,\theta)$ — это д. ф. р., x принимает значения в определенном дискретном множестве $\mathbb{V}\subset\mathbb{R}$ (например, в множестве целых неотрицательных чисел $\mathbb{Z}_+=\{0,\,1,\,2,\,\ldots\}$ в случае распределения Пуассона). Тогда $\boldsymbol{x}\in\mathbb{V}^n$ — это вектор, компоненты которого лежат в $\mathbb{V}$.

Индекс $\boldsymbol{X}$ в обозначении $f_{\boldsymbol{X}}(\boldsymbol{x};\,\theta)$ далее часто будем опускать.

Точное значение параметра θ неизвестно; наша цель — «оценить» его на основании выборки $\boldsymbol{x}=(x_1,\,\ldots,\,x_n)$. Это означает, что мы хотим определить функцию $\theta^*(\boldsymbol{x})$, зависящую от выборки $\boldsymbol{x}$, но не зависящую от θ, которую мы могли бы рассматривать как предполагаемое значение θ. Такую функцию мы будем называть *оценкой* параметра θ.

Замечание. В англоязычной литературе используют два термина: «estimator» и «estimate». Первый имеет смысл случайной величины, которая является функцией от выборки (рассматриваемой как случайный

[1] Ср. названия фильмов: «Licence to Kill», «You Only Live Twice» и «The Terminator».

вектор), во втором случае рассматривается значение этой функции для конкретного значения выборки (реализации случайного вектора), т. е. конкретное число (вектор). В русском языке вместо этих двух терминов употребляют один, «оценка», при этом из контекста всегда видно, идет ли речь об оценке как случайной величине либо о ее конкретном значении для каждой заданной выборки. Некоторые авторы в англоязычной литературе используют термин «an estimate» вместо «an estimator», а другие используют термин «a point estimator» или даже «a point estimate» (точечная оценка).

Например, в простом случае, если бы параметр θ принимал только два значения θ_0 и θ_1, оценка $\theta^*(\boldsymbol{x})$ приписывала бы значение θ_0 либо θ_1 каждой наблюдаемой выборке $\boldsymbol{x}$. Это вводило бы разбиение выборочного пространства (множества исходов) на две области: одну, на которой оценка принимает значение θ_0, и другую, на которой оценка равна θ_1. В общем случае, как уже сказано выше, мы предполагаем, что θ принадлежит заданному множеству значений Θ. (Например, значения θ и θ^* могут быть векторами.)

Например, хорошо известно, что число взмахов крыльями (прыжков) птицы перед взлетом описывается геометрическим распределением. Далее, эмиссия альфа-частиц радиоактивным источником описывается распределением Пуассона (это легко получить, если предположить, что механизм излучения частиц не зависит от времени). Однако параметр распределения может быть разным, в зависимости от вида птиц или от материала, используемого в эксперименте с эмиссией (и зависеть также и от других факторов). Важно уметь оценить неизвестное значение параметра (q или λ) по наблюдаемой выборке $x_1, \ldots, x_n$, где x_i — это число выпущенных частиц за i-й интервал наблюдений.

В 1930-х гг., когда методы проведения экспериментов были довольно примитивными, испущенные частицы просто подсчитывали визуально. В Кембриджском университете все еще помнят, как Э. Резерфорд (1871—1937), известный физик и директор лаборатории им. Кавендиша, набирая на работу новых сотрудников, задавал два прямых вопроса: «Вы закончили университет с отличием?» и «Умеете ли Вы считать?» Ответ «Да» на оба вопроса был необходимым условием для поступления на работу.

В принципе, любая функция от $\boldsymbol{x}$ может рассматриваться в качестве оценки, но на практике нам хотелось бы, чтобы она была «разумной». Нам, следовательно, необходимо выработать критерии, согласно которым оценка считается хорошей либо плохой. Соответствующая область статистики называется *параметрическим оцениванием*.

Пример 1.2.1. Пусть $X_1, \ldots, X_n$ — н. о. р. с. в. и $X_i \sim \mathrm{Po}(\lambda)$. Оценкой для $\lambda = \mathsf{E}X_i$ является *выборочное среднее* $\overline{X}$, где

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i. \tag{1.2.2}$$

Отметим, что $n\overline{X} \sim \mathrm{Po}(n\lambda)$. Мы немедленно получаем следующие полезные свойства выборочного среднего.

1. Случайная величина $\overline{X} = \sum\limits_{i=1}^{n} X_i/n$ сосредоточена вблизи истинного значения параметра:

$$\mathsf{E}\overline{X} = \frac{1}{n}\sum_i \mathsf{E}X_i = \mathsf{E}X_1 = \lambda. \tag{1.2.3а}$$

Это свойство называется *несмещенностью* и будет подробно обсуждаться ниже.

2. Величина $\overline{X}$ стремится к истинному значению при $n \to \infty$:

$$\mathsf{P}\big(\lim_{n\to\infty} \overline{X} = \lambda\big) = 1 \quad \text{(усиленный ЗБЧ).} \tag{1.2.3б}$$

Это свойство называют *состоятельностью.*

Несмещенный и состоятельный статистик? Это дополнение к событию вероятности 1.

(Из серии «Почему их не понимают».)

3. Для больших n выполняется условие

$$\sqrt{\frac{n}{\lambda}}(\overline{X} - \lambda) \sim \mathrm{N}(0,\ 1) \quad \text{(ЦПТ).} \tag{1.2.3в}$$

Это свойство часто называют *асимптотической нормальностью.*

Даже если статистики и нормальны, то в большинстве случаев они лишь асимптотически нормальны.

(Из серии «Почему их не понимают».)

Имеется еще одно важное свойство величины $\overline{X}$.

4. Величина $\overline{X}$ имеет минимальную *среднеквадратическую ошибку* в широком классе оценок λ^*:

$$\mathsf{E}_\lambda(\overline{X} - \lambda)^2 \leqslant \mathsf{E}_\lambda(\lambda^*(\boldsymbol{X}) - \lambda)^2. \quad \square \tag{1.2.3г}$$

Пример 1.2.2. Пусть $X_1, \ldots, X_n$ — н. о. р. с. в. и $X_i \sim \mathrm{Bin}(k,\ p)$. Напомним, что $\mathsf{E}X_i = kp$, $\mathrm{Var}\,X_i = kp(1-p)$. Предположим, что k известно, а значение $p = \mathsf{E}X_i/k \in (0,\ 1)$ неизвестно. Оценкой параметра p является такая величина $\overline{X}/k$, что $n\overline{X} \sim \mathrm{Bin}(kn,\ p)$. При этом, как и выше, имеют место следующие свойства:

1) $\mathsf{E}\overline{X}/k = p$ (несмещенность);

2) $\mathsf{P}(\lim\limits_{n\to\infty} \overline{X}/k = p) = 1$ (состоятельность);

3) $\sqrt{kn/[p(1-p)]}(\bar{X}/k-p) \sim \mathrm{N}(0, 1)$ при больших n (асимптотическая нормальность);

4) $\bar{X}/k$ имеет минимальную среднеквадратическую ошибку в широком классе оценок.

Теперь предположим, что мы знаем p, а величина $k = 1, 2, \ldots$ неизвестна. Можно рассматривать величину $\bar{X}/p$ в качестве оценки для k (не беспокоясь о том, что она принимает нецелочисленные значения!). И опять можно проверить, что свойства 1—3 выполняются. □

Пример 1.2.3. Часто рассматривается пример, когда $X_1, \ldots, X_n$ являются н. о. р. с. в. и $X_i \sim \mathrm{N}(\mu, \sigma^2)$. Рассматривая нормальные выборки, различают обычно три случая:

(I) среднее $\mu \in \mathbb{R}$ неизвестно, а дисперсия σ^2 известна (скажем, $\sigma^2 = 1$),

(II) μ известно (скажем, равно 0), а дисперсия $\sigma^2 > 0$ неизвестна,

(III) оба параметра μ и σ неизвестны.

В случаях (I) и (III) оценкой для μ является выборочное среднее

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \text{при этом} \quad \mathrm{E}\bar{X} = \frac{1}{n}\sum_i \mathrm{E}X_i = \mathrm{E}X_1 = \mu. \tag{1.2.4}$$

Из примера 1.1.1 нам известно, что $\bar{X} \sim \mathrm{N}(\mu, \sigma^2/n)$ (равенство (1.1.3)). В случае (II) оценкой для σ^2 является величина $\bar{\Sigma}^2/n$, где

$$\bar{\Sigma}^2 = \sum_i (X_i - \mu)^2, \quad \mathrm{E}\bar{\Sigma}^2 = \sum_i \mathrm{E}(X_i - \mu)^2 = n \operatorname{Var} X_1 = n\sigma^2, \tag{1.2.5}$$

и $\mathrm{E}(\bar{\Sigma}^2/n) = \sigma^2$. Из примера 1.1.2 следует, что $\bar{\Sigma}^2/\sigma^2 \sim \chi^2_n$.

Что касается оценки для σ^2 в случае (III), то, полагая

$$S_{XX} = \sum_{i=1}^{n} (X_i - \bar{X})^2, \tag{1.2.6}$$

получим

$$\mathrm{E}\left(\frac{1}{n-1} S_{XX}\right) = \frac{1}{n-1}\,\mathrm{E}S_{XX} = \sigma^2. \tag{1.2.7}$$

См. задачу 1.4.5 части A, где этот факт был подтвержден для н. о. р. случайных величин с произвольным распределением. Следовательно, $S_{XX}/(n-1)$ является несмещенной оценкой дисперсии σ^2.

Можно установить, что распределение S_{XX}/σ^2 — это χ^2_{n-1}-распределение; это утверждение является частью теоремы Фишера (которая будет приведена ниже).

Таким образом, в случае (III) пару

$$\left(\overline{X}, \frac{S_{XX}}{n-1}\right)$$

можно взять в качестве оценки для вектора (μ, σ^2), и мы получим аналог свойства 1 (совместная несмещенность):

$$\left(\mathsf{E}\overline{X}, \mathsf{E}\frac{S_{XX}}{n-1}\right) = (\mu, \sigma^2).$$

Далее, при $n \to \infty$ обе величины $\overline{X}$ и $S_{XX}/(n-1)$ стремятся к оцениваемым значениям μ и σ^2:

$$\mathsf{P}\left(\lim_{n\to\infty} \overline{X} = \mu, \lim_{n\to\infty} \frac{S_{XX}}{n-1} = \sigma^2\right) = 1 \quad \text{(вновь усиленный ЗБЧ).}$$

Тем самым получаем аналог свойства 2 (совместная состоятельность). Для $\overline{X}$ это свойство можно вывести непосредственно из того факта, что $\overline{X} \sim \mathrm{N}(\mu, \sigma^2/n)$, а для $S_{XX}/(n-1)$ — из того, что $S_{XX}/\sigma^2 \sim \chi^2_{n-1}$. Эти последние утверждения помогают также проверить аналог свойства 3 (совместная асимптотическая нормальность) при $n \to \infty$:

$$\frac{\sqrt{n}}{\sigma}(\overline{X} - \mu) \sim \mathrm{N}(0, 1), \qquad \frac{\sqrt{n-1}}{\sigma^2}\left(\frac{S_{XX}}{n-1} - \sigma^2\right) \sim \mathrm{N}(0, 2)$$

причем эти нормальные величины независимы. Другими словами, пара

$$\left(\frac{\sqrt{n}}{\sigma}(\overline{X} - \mu), \frac{\sqrt{n-1}}{\sigma^2}\left(\frac{S_{XX}}{n-1} - \sigma^2\right)\right)$$

является асимптотически двумерным нормальным вектором с вектором средних значений $\begin{pmatrix}0\\0\end{pmatrix}$ и ковариационной матрицей $\begin{pmatrix}1 & 0\\0 & 2\end{pmatrix}$. При проверке этого факта необходимо установить, что дисперсия S_{XX} равна $2(n-1)\sigma^4$.

Аналог свойства 4 также имеет место для данного примера, однако следует быть внимательным при определении минимальной среднеквадратической ошибки для векторной оценки. □

Очевидно, конкретный вид распределения с. в. X_i не играет существенной роли при установлении свойств 1—4: следует ожидать, что эти свойства выполняются во многих ситуациях. Фактически каждое из них — признанное направление статистической теории. Здесь и далее мы прежде всего сосредоточимся на свойстве 1 и будем называть оценку θ^* $(= \theta^*(\boldsymbol{x}))$ параметра θ *несмещенной*, если

$$\mathsf{E}_\theta \theta^*(\boldsymbol{X}) = \theta \quad \text{для любых } \theta \in \Theta. \tag{1.2.8}$$

Будут обсуждаться также свойства среднеквадратических ошибок.

Таким образом, подводя итог этого параграфа, заключаем, что для вектора $\boldsymbol{X}$ н. о. р. действительных с. в. $X_1, \ldots, X_n$ справедливы следующие утверждения.

(I) Выборочное среднее

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{1.2.9}$$

всегда является несмещенной оценкой среднего значения $\mathsf{E}X_1$:

$$\mathsf{E}\overline{X} = \frac{1}{n}\sum_{i=1}^{n} \mathsf{E}X_i = \mathsf{E}X_1. \tag{1.2.10}$$

(II) В случае, когда среднее значение $\mathsf{E}X_1$ известно, величина

$$\frac{1}{n}\overline{\Sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mathsf{E}X_i)^2 \tag{1.2.11}$$

является несмещенной оценкой дисперсии $\operatorname{Var} X_1$:

$$\mathsf{E}\left(\frac{1}{n}\overline{\Sigma}^2\right) = \frac{1}{n}\sum_{i=1}^{n} \mathsf{E}(X_i - \mathsf{E}X_1)^2 = \mathsf{E}(X_1 - \mathsf{E}X_1)^2 = \operatorname{Var} X_1. \tag{1.2.12}$$

(III) В случае неизвестного среднего величина

$$\frac{1}{n-1}S_{XX} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \tag{1.2.13}$$

является несмещенной оценкой дисперсии $\operatorname{Var} X_1$:

$$\mathsf{E}\left(\frac{1}{n-1}S_{XX}\right) = \operatorname{Var} X_1, \tag{1.2.14}$$

что было установлено в задаче 1.4.5 части А.

Оценки $\overline{\Sigma}^2/n$ и $S_{XX}/(n-1)$ называют иногда *выборочными дисперсиями.*

Statisticians stubbornly insist that the n justifies the means[1].

(Из серии «Почему их не понимают».)

[1] Непереводимая игра слов — «...the end justifies the means» (цель оправдывает средства).

§ 1.3. Достаточные статистики. Критерий факторизации

Есть два вида статистики: та, которую вы изучаете,
и та, которую вы творите сами.

Р. Стаут (1886—1975), американский писатель, автор детективов

В широком смысле *статистика* (или выборочная статистика) — это произвольная функция от вектора выборочных значений $\boldsymbol{x}$ или от соответствующего случайного вектора $\boldsymbol{X}$. В параметрической постановке, которую мы рассматриваем, мы будем называть функцию T от $\boldsymbol{x}$ (возможно, векторнозначную) *достаточной статистикой* для параметра $\theta \in \Theta$, если условное распределение случайного вектора $\boldsymbol{X}$ относительно $T(\boldsymbol{X})$ не зависит от θ. Это означает, что для любого множества $D \subset \mathbb{R}^n$

$$\text{значение } \mathsf{P}_\theta(\boldsymbol{X} \in D \mid T(\boldsymbol{X})) = \mathsf{E}(I(\boldsymbol{X} \in D) \mid T(\boldsymbol{X})) \text{ одинаково } \forall\theta \in \Theta. \quad (1.3.1)$$

Смысл этого понятия состоит в том, что достаточная статистика содержит в себе всю информацию о выборке $\boldsymbol{x}$, необходимую для построения «хорошей» оценки параметра θ.

В примере 1.2.1 выборочное среднее $\overline{X}$ является достаточной статистикой для λ. Действительно, для любого такого неотрицательного вектора с целочисленными значениями $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{Z}_+^n$, что $\sum\limits_i x_i = nt$, условная вероятность $\mathsf{P}_\lambda(\boldsymbol{X} = \boldsymbol{x} \mid \overline{X} = t)$ равна

$$\frac{\mathsf{P}_\lambda(\boldsymbol{X} = \boldsymbol{x}, \overline{X} = t)}{\mathsf{P}_\lambda(\overline{X} = t)} = \frac{\mathsf{P}_\lambda(\boldsymbol{X} = \boldsymbol{x})}{\mathsf{P}_\lambda(\overline{X} = t)} = \frac{e^{-n\lambda}\prod\limits_i(\lambda^{x_i}/x_i!)}{e^{-n\lambda}(n\lambda)^{nt}/(nt)!} = \frac{(nt)!}{n^{nt}}\prod_i \frac{1}{x_i!},$$

а это значение не зависит от $\lambda > 0$. Мы воспользовались здесь тем фактом, что события

$$\{\boldsymbol{X} = \boldsymbol{x}, \overline{X} = t\} \quad \text{и} \quad \{\boldsymbol{X} = \boldsymbol{x}\}$$

совпадают (в силу очевидного равенства $\overline{X} = t$), а также тем, что $n\overline{X} \sim$ $\sim \mathrm{Po}(n\lambda)$.

Таким образом, в общем случае можно записать

$$\mathsf{P}_\lambda(\boldsymbol{X} = \boldsymbol{x} \mid \overline{X} = t) = \frac{(nt)!}{n^{nt}}\prod_i \frac{1}{x_i!} I\left(\sum_{i=1}^n x_i = nt\right).$$

Конечно, $n\bar{x} = \sum\limits_i x_i$ является еще одной достаточной статистикой, и статистики $\bar{x}$ и $n\bar{x}$ (или соответствующие им случайные векторы $\overline{X}$ и $n\overline{X}$) являются, по сути, эквивалентными (так как являются образами друг друга при взаимно однозначном отображении).

Аналогично в примере 1.2.2 величина $\bar{x}$ является достаточной статистикой для p при неизвестном k. В этом случае для любого вектора $\boldsymbol{x} \in \mathbb{Z}^n$ с компонентами $x_i = 0, 1, \ldots, k$ и суммой $\sum\limits_i x_i = nt$ условная вероятность $\mathsf{P}_p(\boldsymbol{X} = \boldsymbol{x} \mid \bar{X} = t)$ равна

$$\frac{\mathsf{P}_p(\boldsymbol{X} = \boldsymbol{x})}{\mathsf{P}_p(\bar{X} = t)} = \frac{\prod\limits_i \dfrac{k!}{x_i!\,(k - x_i)!} p^{x_i}(1 - p)^{k - x_i}}{\dfrac{(nk)!}{(nt)!\,(nk - nt)!} p^{nt}(1 - p)^{nk - nt}} = \frac{(k!)^n}{(nk)!} \frac{(nt)!\,(nk - nt)!}{\prod\limits_i x_i!\,(k - x_i)!},$$

а эта величина не зависит от $p \in (0, 1)$. Как и ранее, если $\sum\limits_i x_i \neq nt$, то $\mathsf{P}_p(\boldsymbol{X} = \boldsymbol{x} \mid \bar{X} = t) = 0$, что опять не зависит от p.

Рассмотрим теперь пример 1.2.3, где $X_1, \ldots, X_n$ — это н. о. р. с. в. и $X_i \sim \mathrm{N}(\mu, \sigma^2)$. Тогда

(I) если значение σ^2 известно, то достаточной статистикой для μ является

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i,$$

(II) если значение μ известно, то достаточной статистикой для σ^2 является

$$\bar{\Sigma}^2 = \sum_i (X_i - \mu)^2,$$

(III) если оба значения μ и σ^2 неизвестны, то достаточной статистикой для (μ, σ^2) является

$$\left(\bar{X}, \sum_{i=1}^{n} X_i^2\right).$$

Наиболее эффективным средством для проверки этих фактов является критерий факторизации.

Критерий факторизации — это утверждение общего характера о достаточной статистике. Этот критерий гласит следующее.

Статистика T является достаточной для параметра θ тогда и только тогда, когда д. ф. р./п. р. в. $f_X(\boldsymbol{x}; \theta)$ может быть представлена в виде произведения $g(T(\boldsymbol{x}), \theta)h(\boldsymbol{x})$, где g и h — некоторые функции.

Доказательство для *дискретного* случая, когда $f_X(\boldsymbol{x}; \theta) = \mathsf{P}_\theta(\boldsymbol{X} = \boldsymbol{x})$, весьма простое. Действительно, для доказательства достаточности предположим, что указанная факторизация имеет место. Тогда для такого выборочного вектора $\boldsymbol{x} \in \mathbb{V}^n$, что $T(\boldsymbol{x}) = t$, условная вероятность $\mathsf{P}_\theta(\boldsymbol{X} = \boldsymbol{x} \mid T = t)$

равна

$$\frac{\mathsf{P}_\theta(\boldsymbol{X}=\boldsymbol{x},\ T=t)}{\mathsf{P}_\theta(T=t)} = \frac{\mathsf{P}_\theta(\boldsymbol{X}=\boldsymbol{x})}{\mathsf{P}_\theta(T=t)} = \frac{g(T(\boldsymbol{x}),\ \theta)h(\boldsymbol{x})}{\sum\limits_{\tilde{\boldsymbol{x}}\in\mathbb{V}^n:\ T(\tilde{\boldsymbol{x}})=t} g(T(\tilde{\boldsymbol{x}}),\ \theta)h(\tilde{\boldsymbol{x}})} =$$

$$= \frac{g(t,\ \theta)h(\boldsymbol{x})}{g(t,\ \theta)\sum\limits_{\tilde{\boldsymbol{x}}\in\mathbb{V}^n:\ T(\tilde{\boldsymbol{x}})=t} h(\tilde{\boldsymbol{x}})} = \frac{h(\boldsymbol{x})}{\sum\limits_{\tilde{\boldsymbol{x}}\in\mathbb{V}^n:\ T(\tilde{\boldsymbol{x}})=t} h(\tilde{\boldsymbol{x}})}.$$

Последнее выражение не зависит от θ.

С другой стороны, если условие совместимости $T(\boldsymbol{x})=t$ не выполняется (т. е. $T(\boldsymbol{x})\neq t$), то $\mathsf{P}_\theta(\boldsymbol{X}=\boldsymbol{x}\mid T=t)=0$, что опять не зависит от θ. Общая формула имеет вид

$$\mathsf{P}_\theta(\boldsymbol{X}=\boldsymbol{x}\mid T=t) = \frac{h(\boldsymbol{x})}{\sum\limits_{\tilde{\boldsymbol{x}}\in\mathbb{V}^n:\ T(\tilde{\boldsymbol{x}})=t} h(\tilde{\boldsymbol{x}})} I(T(\boldsymbol{x})=t).$$

Поскольку правая часть не зависит от θ, величина T является достаточной статистикой.

Для доказательства необходимости предположим, что $\mathsf{P}_\theta(\boldsymbol{X}=\boldsymbol{x}\mid T=t)$ не зависит от θ. Тогда вновь для такого $\boldsymbol{x}\in\mathbb{V}^n$, что $T(\boldsymbol{x})=t$, запишем

$$f_X(\boldsymbol{x}) = \mathsf{P}_\theta(\boldsymbol{X}=\boldsymbol{x}) = \mathsf{P}_\theta(\boldsymbol{X}=\boldsymbol{x}\mid T=t)\mathsf{P}_\theta(T=t).$$

Множитель $\mathsf{P}_\theta(\boldsymbol{X}=\boldsymbol{x}\mid T=t)$ не зависит от θ; мы обозначим его $h(\boldsymbol{x})$. Затем обозначим множитель $\mathsf{P}_\theta(T=t)$ через $g(t,\ \theta)$ и получим тем самым указанную факторизацию.

В *непрерывном* случае доказательство проводится по аналогичной схеме (однако для формально безупречного доказательства необходимо привлечь некоторые факты из теории меры). А именно, мы запишем условную п. р. в. $f_{X\mid T}(\boldsymbol{x}\mid t)$ в виде отношения

$$\frac{f(\boldsymbol{x};\ \theta)}{f_T(t;\ \theta)} I(T(\boldsymbol{x})=t)$$

и представим п. р. в. $f_T(t;\ \theta)$ в виде интеграла

$$f_T(t;\ \theta) = \int\limits_{\{\tilde{\boldsymbol{x}}\in\mathbb{R}^n:\ T(\tilde{\boldsymbol{x}})=t\}} f(\tilde{\boldsymbol{x}};\ \theta)\,d\tilde{\boldsymbol{x}}$$

по поверхности уровня $\{\tilde{\boldsymbol{x}}\in\mathbb{R}^n\colon T(\tilde{\boldsymbol{x}})=t\}$ с элементом площади $d(\tilde{\boldsymbol{x}}\mid t)$ на этой поверхности. Тогда для доказательства достаточности мы опять используем представление $f(\boldsymbol{x})=g(T(\boldsymbol{x}),\ \theta)h(\boldsymbol{x})$ и получим равенство

$$f_{X\mid T}(\boldsymbol{x}\mid t) = \frac{h(\boldsymbol{x})}{\int\limits_{\{\tilde{\boldsymbol{x}}\in\mathbb{R}^n:\ T(\tilde{\boldsymbol{x}})=t\}} h(\tilde{\boldsymbol{x}})\,d\tilde{\boldsymbol{x}}} I(T(\boldsymbol{x})=t),$$

где правая часть не зависит от θ. Для доказательства необходимости мы просто перепишем $f(\boldsymbol{x};\ \theta)$ в виде $f_{X\,|\,T}(\boldsymbol{x}\,|\,t)f_T(t;\ \theta)$, где $t = T(\boldsymbol{x})$, и положим, как и ранее,

$$h(\boldsymbol{x}) = f_{X\,|\,T}(\boldsymbol{x}\,|\,t) \quad \text{и} \quad g(T(\boldsymbol{x}),\ \theta) = f_T(t;\ \theta).$$

Критерий факторизации означает, что T является достаточной статистикой, если из равенства $T(\boldsymbol{x}) = T(\boldsymbol{x}')$ следует, что отношение $f_X(\boldsymbol{x};\ \theta)/f_X(\boldsymbol{x}';\ \theta)$ остается одним и тем же для любого $\theta \in \Theta$. Следующим шагом является рассмотрение *минимальной* достаточной статистики, для которой равенство $T(\boldsymbol{x}) = T(\boldsymbol{x}')$ эквивалентно тому, что отношение $f_X(\boldsymbol{x};\ \theta)/f_X(\boldsymbol{x}';\ \theta)$ остается одним и тем же для любого $\theta \in \Theta$. Это понятие удобно, поскольку любая достаточная статистика является функцией от минимальной статистики. Другими словами, минимальная достаточная статистика имеет самые «широкие» множества уровня, где она принимает постоянное значение, а тогда количество информации, которую мы должны знать о выборке $\boldsymbol{x}$, минимально. Всякое дальнейшее сужение информации о выборке привело бы к потере достаточности.

Во всех примерах, представленных ниже, достаточные статистики являются минимальными.

Идея, лежащая в основе критерия факторизации, восходит к статье 1925 г., написанной Р. А. Фишером (1890—1962), выдающимся британским прикладным математиком, статистиком и генетиком, имя которого будет часто упоминаться в этой части книги. (Некоторые авторы усматривают появление критерия факторизации в его более ранней работе 1912 г.) Это понятие получило дальнейшее развитие в работе Фишера 1934 г. Важную роль сыграла также статья, написанная в 1935 г. Ю. Нейманом (1894—1981), американским статистиком, который родился в Молдавии, получил образование на Украине, работал в Польше и Великобритании и позднее в США. Имя Неймана также будет появляться в этой части книги по многим поводам, главным образом в связи с леммой Неймана—Пирсона (см. далее).

Пример 1.3.1. 1. Пусть $X_i \sim \mathrm{U}(0,\ \theta)$ где $\theta > 0$ неизвестно. Тогда величина $T(\boldsymbol{x}) = \max x_i$, $\boldsymbol{x} = (x_1,\ \ldots,\ x_n) \in \mathbb{R}^n_+$, является достаточной статистикой для θ.

2. Рассмотрим теперь $X_i \sim \mathrm{U}(\theta,\ \theta + 1)$, $\theta \in \mathbb{R}$. В этом случае выборочная п. р. в. имеет вид

$$f_X(\boldsymbol{x};\ \theta) = \prod_i I(\theta \leqslant x_i \leqslant \theta + 1) = I(\min x_i \geqslant \theta) I(\max x_i \leqslant \theta + 1), \qquad \boldsymbol{x} \in \mathbb{R}^n.$$

Здесь имеет место факторизация, причем $T(\boldsymbol{x})$ является двумерным вектором $(\min x_i,\ \max x_i)$, $g((y_1,\ y_2),\ \theta) = I(y_1 \geqslant \theta) I(y_2 \leqslant \theta + 1)$ и $h(\boldsymbol{x}) \equiv 1$. Следовательно, пара $(\min x_i,\ \max x_i)$ является достаточной статистикой для θ.

3. Пусть $X_1, \ldots, X_n$ — это случайная выборка из распределения Пуассона с неизвестным средним значением θ. Найдите одномерную достаточную статистику для θ.

(Ответ. $T(\boldsymbol{X}) = \sum\limits_i X_i$.)

4. Предположим, что $X_i \sim \mathrm{N}(\mu, \sigma^2)$, где оба параметра $\mu \in \mathbb{R}$ и $\sigma^2 > 0$ неизвестны. Тогда $T(\boldsymbol{x}) = \left(\sum\limits_i X_i, \sum\limits_i X_i^2\right)$ — достаточная статистика для $\theta = (\mu, \sigma^2)$. □

§ 1.4. Оценки максимального правдоподобия

Robin Likelihood — Prince of Liars[1]

(Из серии «Фильмы, которые не вышли на большой экран».)

Понятие *оценки максимального правдоподобия* (о. м. п.) лежит в основе мощного (и красивого) метода построения хороших оценок, применяемого повсеместно (и называемого *методом максимального правдоподобия*). При этом д. ф. р./п. р. в. $f_X(\boldsymbol{x}; \theta)$ понимается как функция от $\theta \in \Theta$, которая зависит от наблюдаемой выборки $\boldsymbol{x}$ как от параметра. Затем ищется значение θ, которое максимизирует эту функцию на множестве Θ:

$$\hat{\theta}\ (= \hat{\theta}(\boldsymbol{x})) = \arg\max[f(\boldsymbol{x}; \theta) \colon \theta \in \Theta]. \tag{1.4.1}$$

В этом контексте $f(\boldsymbol{x}; \theta)$ часто называют *функцией правдоподобия* (или, кратко, правдоподобием) для выборки $\boldsymbol{x}$. Вместо того чтобы находить максимум функции $f(\boldsymbol{x}; \theta)$, часто предпочитают находить максимум логарифма этой функции $\ell(\boldsymbol{x}; \theta) = \ln f(\boldsymbol{x}; \theta)$, который называют *функцией логарифма правдоподобия* или, кратко, логарифмом правдоподобия (л. п.). Тогда о. м. п. определяется как

$$\hat{\theta} = \arg\max[\ell(\boldsymbol{x}; \theta) \colon \theta \in \Theta].$$

Идея о. м. п. возникла в 1921 г. благодаря Фишеру.

Часто точка максимума оказывается единственной (хотя и может лежать на границе множества допустимых значений Θ). Если $\ell(\boldsymbol{x}; \theta)$ является гладкой функцией от $\theta \in \Theta$, то можно рассматривать стационарные точки, в которых первая производная обращается в нуль:

$$\frac{d}{d\theta}\ell(\boldsymbol{x}; \theta) = 0 \quad \left(\text{или } \frac{\partial}{\partial\theta_j}\ell(\boldsymbol{x}; \theta) = 0,\ j = 1, \ldots, d, \text{ если } \theta = (\theta_1, \ldots, \theta_d)\right). \tag{1.4.2}$$

Конечно, среди корней уравнения (1.4.2) необходимо отобрать точки локального максимума (проверяя знаки производных второго порядка или

[1] Ср. название фильма «Robin Hood: Prince of Thieves».

каким-либо иным способом) и определить, какая из этих точек является точкой глобального максимума. К счастью, в нижеследующих примерах стационарная точка (если она существует) всегда единственна. В случае, когда параметрическое множество Θ неограниченно (например, действительная прямая), для того чтобы проверить, что (единственная) стационарная точка является точкой глобального максимума, достаточно проверить, что $\ell(\cdot;\,\theta)\to-\infty$ при $|\theta|\to\infty$.

Пример 1.4.1. Найдем о. м. п. для некоторых моделей.

1. Оценка максимального правдоподобия параметра λ для $X_i\sim\mathrm{Po}(\lambda)$. В этом случае для любого $\boldsymbol{x}=(x_1,\,\ldots,\,x_n)\in\mathbb{Z}_+^n$ с неотрицательными целочисленными компонентами $x_i\in\mathbb{Z}_+$ л. п. имеет вид

$$\ell(\boldsymbol{x};\,\lambda)=-n\lambda+\sum_i x_i\ln\lambda-\sum_i\ln(x_i!),\qquad\lambda>0.$$

После дифференцирования по λ находим

$$\frac{\partial}{\partial\lambda}\ell(\boldsymbol{x};\,\lambda)=-n+\frac{1}{\lambda}\sum_i x_i=0\quad\text{и}\quad\hat{\lambda}=\frac{1}{n}\sum_i x_i=\bar{x}.$$

При этом $\dfrac{\partial^2}{\partial\lambda^2}\ell(\boldsymbol{x};\,\lambda)=-\dfrac{1}{\lambda^2}\sum\limits_i x_i<0$. Таким образом, $\bar{x}$ является точкой (глобального) максимума, откуда видно, что о. м. п. $\hat{\lambda}$ параметра λ совпадает с выборочным средним. В частности, эта оценка является несмещенной.

2. Если $X_i\sim\mathrm{U}(0,\,\theta)$, $\theta>0$, то $\hat{\theta}(\boldsymbol{x})=\max x_i$ является о. м. п. для θ.

3. Пусть $X_i\sim\mathrm{U}(\theta,\,\theta+1)$ (ср. с примером 2 в § 1.3). Чтобы найти о. м. п. для θ, снова рассмотрим функцию правдоподобия:

$$f(\boldsymbol{x};\,\theta)=I(\max x_i-1\leqslant\theta\leqslant\min x_i),\qquad\theta\in\mathbb{R}.$$

Если $\max x_i-1<\min x_i$ (в согласии с предположением, что выборка $\boldsymbol{x}$ порождается н. о. р. $\mathrm{U}(\theta,\,\theta+1)$-с. в.), то в качестве о. м. п. для θ можно взять любое значение, лежащее между $\max x_i-1$ и $\min x_i$.

Нетрудно догадаться, что несмещенной о. м. п. для θ будет точка, лежащая посередине:

$$\hat{\theta}=\frac{1}{2}(\max x_i-1+\min x_i)=\frac{1}{2}(\max x_i+\min x_i)-\frac{1}{2}.$$

В самом деле,

$$\mathsf{E}\hat{\theta}=\frac{1}{2}(\mathsf{E}\max X_i+\mathsf{E}\min X_i)-\frac{1}{2}=$$
$$=\frac{1}{2}\left(\int\limits_{\theta}^{\theta+1}x p_{\min X_i}(x)\,dx+\int\limits_{\theta}^{\theta+1}x p_{\max X_i}(x)\,dx\right)-\frac{1}{2},$$

где

$$p_{\min X_i}(x) = -\frac{d}{dx}\,\mathsf{P}(\min X_i > x) = -\frac{d}{dx}(\mathsf{P}(X_i > x))^n =$$

$$= -\frac{d}{dx}\left(\int\limits_x^{\theta+1} dy\right)^n = n(\theta+1-x)^{n-1}, \qquad \theta < x < \theta+1,$$

и аналогично

$$p_{\max X_i}(x) = n(x-\theta)^{n-1}, \qquad \theta < x < \theta+1.$$

Тогда

$$\mathsf{E}\min X_i = n\int\limits_\theta^{\theta+1}(\theta+1-x)^{n-1}x\,dx = -n\int\limits_\theta^{\theta+1}(\theta+1-x)^n\,dx +$$

$$+(\theta+1)n\int\limits_\theta^{\theta+1}(\theta+1-x)^{n-1}\,dx = -n\int\limits_0^1 x^n\,dx +$$

$$+(\theta+1)n\int\limits_0^1 x^{n-1}\,dx = -\frac{n}{n+1}+\theta+1 = \theta+\frac{1}{n+1} \qquad (1.4.3)$$

и аналогично

$$\mathsf{E}\max X_i = \frac{n}{n+1}+\theta, \qquad (1.4.4)$$

откуда и следует, что $\mathsf{E}\hat\theta = \theta$. □

Оценки максимального правдоподобия имеют ряд полезных свойств.

1. Если T — достаточная статистика для θ, то $\ell(\boldsymbol{x};\,\theta) = \ln g(T(\boldsymbol{x}),\,\theta) + \ln h(\boldsymbol{x})$. Тогда отыскание максимума правдоподобия относительно θ сводится к отысканию максимума функции $g(T(\boldsymbol{x}),\,\theta)$ или ее логарифма. Это означает, что о. м. п. $\hat\theta$ будет функцией от $T(\boldsymbol{x})$.

2. При незначительных ограничениях на распределение величин X_i о. м. п. являются асимптотически несмещенными (или могут быть выбраны таковыми) при $n\to\infty$. Более того, о. м. п. $\hat\theta$ часто являются асимптотически нормальными. В случае скалярного параметра это означает, что $\sqrt{n}(\hat\theta-\theta) \sim \mathrm{N}(0,\,v)$, где дисперсия v является минимальной среди всех возможных дисперсий несмещенных оценок параметра θ.

3. Принцип инвариантности для о. м. п.: если $\hat\theta$ является о. м. п. для θ, то при переходе от параметра θ к параметру $\eta = u(\theta)$, где отображение u является взаимно однозначным, величина $u(\hat\theta)$ будет о. м. п. параметра η.

§ 1.5. Нормальные выборки. Теорема Фишера

Слышали ли вы о статистике, которого заключили в тюрьму? У него сейчас 0 степеней свободы.

(Из серии «Почему их не понимают».)

Пример 1.5.1. В этом примере рассмотрим о. м. п. для пары (μ, σ^2) в случае н. о. р. нормальной выборки. Пусть $X_i \sim \mathrm{N}(\mu, \sigma^2)$, тогда логарифм правдоподобия равен

$$\ell(\boldsymbol{x}; \mu, \sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2,$$

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n, \quad \mu \in \mathbb{R}, \quad \sigma^2 > 0,$$

откуда следует, что

$$\frac{\partial}{\partial\mu}\ell(\boldsymbol{x}; \mu, \sigma^2) = \frac{1}{\sigma^2}\sum_i (x_i - \mu),$$

$$\frac{\partial}{\partial\sigma^2}\ell(\boldsymbol{x}; \mu, \sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_i (x_i - \mu)^2.$$

Стационарная точка, для которой $\partial\ell/\partial\mu = \partial\ell/\partial\sigma^2 = 0$, единственна:

$$\hat{\mu} = \bar{x}, \quad \hat{\sigma}^2 = \frac{1}{n}S_{xx},$$

причем, как и в уравнении (1.2.6),

$$S_{xx} \ (= S_{xx}(\boldsymbol{x})) = \sum_i (x_i - \bar{x})^2. \tag{1.5.1}$$

(Иногда вместо S_{xx} употребляют обозначение S^2_{xx} или даже $\bar{S}^2_{xx}$.) Точка $(\hat{\mu}, \hat{\sigma}^2) = (\bar{x}, S_{xx}/n)$ является точкой глобального максимума. Это видно, например, из того, что $\ell(\boldsymbol{x}; \mu, \sigma^2) \to -\infty$ при $|\mu| \to \infty$ и $\sigma^2 \to \infty$, а также $\ell(\boldsymbol{x}; \mu, \sigma^2) \to -\infty$ при $\sigma^2 \to 0$ при каждом μ. Поэтому $(\bar{x}, S_{xx}/n)$ не может быть точкой минимума (или седловой точкой). Следовательно, она является точкой глобального максимума. При этом оценка $\bar{X}$ является несмещенной, а оценка S_{XX}/n смещена: $\mathrm{E}S_{XX}/n = (n-1)\sigma^2/n < \sigma^2$. См. уравнение (1.2.5) и задачу 1.4.5 части А. Однако при $n \to \infty$ смещение пропадает: $\mathrm{E}S_{XX}/n \to \sigma^2$. (Несмещенной оценкой для σ^2 является, конечно, $S_{XX}/(n-1)$.) □

Важный факт содержится в следующем утверждении, которое часто называют *теоремой Фишера*.

Для н. о. р. нормальной выборки о. м. п. $(\hat{\mu}, \hat{\sigma}^2) = (\overline{X}, S_{XX}/n)$ *образована независимыми с. в.* $\overline{X}$ *и* S_{XX}/n, *где*

$$\overline{X} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right) \quad (\textit{т. е. } \sqrt{n}(\overline{X} - \mu) \sim \mathrm{N}(0, \sigma^2))$$

и

$$\frac{S_{XX}}{\sigma^2} \sim \chi^2_{n-1} \quad \left(\textit{т. е. } \frac{S_{XX}}{\sigma^2} \sim \sum_{i=1}^{n-1} Y_i^2, \textit{ где } Y_i \sim \mathrm{N}(0, 1) \textit{ — независимые с. в.}\right).$$

Из теоремы Фишера следует, что с. в. $(S_{XX} - (n-1)\sigma^2)/(\sigma^2\sqrt{2n})$ является асимптотически $\mathrm{N}(0, 1)$-с. в. Тогда, очевидно,

$$\sqrt{n}\left(\frac{1}{n}S_{XX} - \sigma^2\right) \sim \mathrm{N}(0, 2\sigma^4).$$

Чтобы доказать теорему Фишера, запишем сначала равенство

$$\sum_{i=1}^{n}(X_i - \mu)^2 = \sum_{i=1}^{n}(X_i - \overline{X} + \overline{X} - \mu)^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2 + n(\overline{X} - \mu)^2,$$

поскольку $\sum\limits_{i=1}^{n}(X_i - \overline{X})(\overline{X} - \mu) = (\overline{X} - \mu)\sum\limits_{i=1}^{n}(X_i - \overline{X}) = 0$. Другими словами, $\sum\limits_{i}(X_i - \mu)^2 = S_{XX} + n(\overline{X} - \mu)^2$.

Затем используем общий факт: если компоненты вектора

$$\boldsymbol{X} - \mu\mathbf{1} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} - \mu\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

являются н. о. р. с. в., причем $X_i - \mu \sim \mathrm{N}(0, \sigma^2)$, то для любой ортогональной матрицы A размера $n \times n$ компоненты вектора

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = A^T(\boldsymbol{X} - \mu\mathbf{1})$$

будут опять н. о. р. с. в. $Y_i \sim \mathrm{N}(0, \sigma^2)$ (см. задачу 2.3.4 части А).

Возьмем любую ортогональную матрицу A, у которой первый столбец имеет вид

$$\begin{pmatrix} 1/\sqrt{n} \\ \vdots \\ 1/\sqrt{n} \end{pmatrix};$$

получим ее, дополнив этот столбец такими столбцами, которые образуют ортогональный базис в $\mathbb{R}^n$. Например, для этой цели подойдет семейство

$\mathbf{e}_2, \ldots, \mathbf{e}_n$, где у k-го столбца $\mathbf{e}_k$ первые $k-1$ компонент равны $\frac{1}{\sqrt{k(k-1)}}$, k-я компонента равна $\frac{-(k-1)}{\sqrt{k(k-1)}}$, а затем следуют $n-k$ нулевых:

$$\mathbf{e}_k = \begin{pmatrix} \left.\begin{matrix} \frac{1}{\sqrt{k(k-1)}} \\ \vdots \\ \frac{1}{\sqrt{k(k-1)}} \end{matrix}\right\} k-1 \\ \frac{-(k-1)}{\sqrt{k(k-1)}} \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad k = 2, \ldots, n.$$

Тогда $Y_1 = (A^T(\mathbf{X} - \mu\mathbf{1}))_1 = \sqrt{n}(\overline{X} - \mu)$ и $Y_2, \ldots, Y_n$ не зависят от Y_1.

Поскольку ортогональная матрица сохраняет длину вектора, мы получаем

$$\sum_{i=1}^{n} Y_i^2 = \sum_{i=1}^{n} (X_i - \mu)^2 = n(\overline{X} - \mu)^2 + \sum_{i=1}^{n} (X_i - \overline{X})^2 = Y_1^2 + S_{XX}^2,$$

т. е. $S_{XX} = \sum\limits_{i=2}^{n} Y_i^2$. Следовательно, $\frac{S_{XX}}{\sigma^2} = \sum\limits_{i=2}^{n} \frac{Y_i^2}{\sigma^2} \sim \chi^2_{n-1}$, и эта оценка не зависит от Y_1.

Замечание. Некоторые авторы называют статистику

$$s_{XX} = \sqrt{\frac{S_{XX}}{n-1}} \tag{1.5.2}$$

выборочным стандартным отклонением. Термин *стандартная ошибка* часто используют для величины $\frac{s_{XX}}{\sqrt{n}}$, которая является оценкой для $\frac{\sigma}{\sqrt{n}}$. Мы будем следовать этой традиции.

Статистики делают все стандартные отклонения.
Статистики делают все стандартные ошибки.

(Из серии «Как они делают это».)

§ 1.6. Среднеквадратические ошибки. Теорема Рао—Блекуэлла. Неравенство Крамера—Рао

Статистика свидетельствует, что из тех, кто приобрел привычку есть, лишь немногие выживают.

В. Ирвин (1876—1959), американский писатель и издатель

Рассматривая оценку θ^* параметра θ, полезно вычислить ее *среднеквадратическую ошибку* (с. к. о.), которая определяется как

$$\mathsf{E}_\theta(\theta^*(\boldsymbol{X}) - \theta)^2; \tag{1.6.1}$$

для несмещенной оценки эта величина представляет собой дисперсию $\mathrm{Var}_\theta\,\theta^*(\boldsymbol{X})$. В общем случае

$$\begin{aligned}\mathsf{E}_\theta(\theta^*(\boldsymbol{X})-\theta)^2&=\mathsf{E}_\theta(\theta^*(\boldsymbol{X})-\mathsf{E}_\theta\theta^*(\boldsymbol{X})+\mathsf{E}_\theta\theta^*(\boldsymbol{X})-\theta)^2=\\&=\mathsf{E}_\theta(\theta^*(\boldsymbol{X})-\mathsf{E}_\theta\theta^*(\boldsymbol{X}))^2+(\mathsf{E}_\theta\theta^*(\boldsymbol{X})-\theta)^2+2(\mathsf{E}_\theta\theta^*(\boldsymbol{X})-\theta)\,\mathsf{E}_\theta(\theta^*(\boldsymbol{X})-\mathsf{E}_\theta\theta^*(\boldsymbol{X}))=\\&=\mathrm{Var}_\theta\,\theta^*(\boldsymbol{X})+(\mathrm{Bias}_\theta\,\theta^*(\boldsymbol{X}))^2,\end{aligned} \tag{1.6.2}$$

где $\mathrm{Bias}_\theta\,\theta^* = \mathsf{E}_\theta\theta^* - \theta$.

Вообще говоря, существует простой способ уменьшить с. к. о. заданной оценки. Этот способ основан на использовании *теоремы Рао—Блекуэлла* (Р—Б).

Если T — достаточная статистика и θ^* — оценка параметра θ, то для оценки $\hat{\theta}^* = \mathsf{E}(\theta^* \mid T)$ выполняется неравенство

$$\mathsf{E}(\hat{\theta}^* - \theta)^2 \leqslant \mathsf{E}(\theta^* - \theta)^2, \quad \theta \in \Theta. \tag{1.6.3}$$

Более того, если $\mathsf{E}(\theta^*)^2 < \infty$ для некоторого $\theta \in \Theta$, то для этого θ выполняется строгое неравенство, за исключением тех случаев, когда θ^* является функцией от T.

Доказательство короткое: поскольку $\mathsf{E}\hat{\theta}^* = \mathsf{E}[\mathsf{E}(\theta^* \mid T)] = \mathsf{E}\theta^*$, обе оценки θ^* и $\hat{\theta}^*$ имеют одинаковое смещение. По формуле для условной дисперсии находим

$$\mathrm{Var}\,\theta^* = \mathsf{E}[\mathrm{Var}(\theta^* \mid T)] + \mathrm{Var}[\mathsf{E}(\theta^* \mid T)] = \mathsf{E}[\mathrm{Var}(\theta^* \mid T)] + \mathrm{Var}\,\hat{\theta}^*.$$

Следовательно, $\mathrm{Var}\,\theta^* \geqslant \mathrm{Var}\,\hat{\theta}^*$, и, таким образом, $\mathsf{E}(\theta^* - \theta)^2 \geqslant \mathsf{E}(\hat{\theta}^* - \theta)^2$. Равенство достигается тогда и только тогда, когда $\mathrm{Var}(\theta^* \mid T) = 0$.

1. Величина $\mathsf{E}(\theta^* \mid T)$ зависит от значения $\theta^*(\boldsymbol{x})$, но не зависит от θ. Следовательно, оценка $\hat{\theta}^*$ определена корректно.

2. Если θ^* — несмещенная оценка, то таковой является и $\hat{\theta}^*$.

3. Если θ^* сама является функцией от T, то $\hat{\theta}^* = \theta^*$.

Теорема Р—Б носит имена двух выдающихся ученых. Д. Блекуэлл (1919—) — американский математик, один из основных сторонников применения методов теории игр в статистике и других дисциплинах. Блекуэлл — один из первых американских математиков африканского происхождения, который был принят на работу в ведущий университет США. С. Р. Рао (1920—) — индийский математик, который обучался в Индии и Британии (он получил степень доктора наук в Кембриджском университете и был единственным официальным аспирантом Фишера по статистике), долгое время работал в Индии и в настоящее время проживает и работает в США.

Пример 1.6.1. 1. Пусть $X_i \sim \mathrm{U}(0, \theta)$. Тогда величина $2\bar{x} = 2\sum_i x_i/n$ является несмещенной оценкой для θ и

$$\mathrm{Var}(2\bar{X}) = \frac{4}{n}\,\mathrm{Var}\,X_1 = \frac{\theta^2}{3n}.$$

Мы знаем, что достаточная статистика T имеет вид $T(\boldsymbol{X}) = \max_i X_i$ с п. р. в.

$$f_T(x) = n\frac{x^{n-1}}{\theta^n}I(0 < x < \theta).$$

Следовательно,

$$\hat{\theta}^* = \mathrm{E}(\theta^* \mid T) = \frac{2}{n}\sum_i \mathrm{E}(X_i \mid T) = 2\,\mathrm{E}(X_1 \mid T) =$$

$$= 2\left(\max X_i \cdot \frac{1}{n} + \frac{\max X_i}{2}\cdot\frac{n-1}{n}\right) = \frac{n+1}{n}T,$$

и с. к. о. оценки $\hat{\theta}^*$ должна быть не больше, чем у θ^*. Как это ни удивительно, но, отбрасывая значительную часть информации, содержащуюся в выборке, мы улучшаем с. к. о.! Действительно, $\mathrm{Var}\,\hat{\theta}^* = \dfrac{(n+1)^2(\mathrm{Var}\,T)}{n^2}$, где

$$\mathrm{Var}\,T = \int_0^\theta n\frac{x^{n-1}}{\theta^n}x^2\,dx - \left(\int_0^\theta n\frac{x^{n-1}}{\theta^n}x\,dx\right)^2 =$$

$$= \theta^2\left[\frac{n}{n+2} - \left(\frac{n}{n+1}\right)^2\right] = \theta^2\frac{n}{(n+1)^2(n+2)}.$$

Таким образом, $\mathrm{Var}\,\hat{\theta}^* = \dfrac{\theta^2}{[n(n+2)]}$, а эта величина меньше $\dfrac{\theta^2}{3n}$ при $n \geqslant 2$ и стремится к 0 при $n \to \infty$.

2. Пусть $X_i \sim \mathrm{U}(\theta, \theta+1)$. Тогда $\hat{\theta} = \frac{1}{n}\sum_i X_i - \frac{1}{2}$ является несмещенной оценкой для θ; при этом

$$\mathrm{Var}\,\bar{X} = \frac{1}{n}\,\mathrm{Var}\,X_1 = \frac{1}{12n}.$$

Такая оценка появится, если приравнять величину $\hat{\theta} + 1/2$ к выборочному среднему $\sum_{i=1}^n \frac{X_i}{n}$. Подобный трюк лежит в основе так называемого *метода моментов* в статистике. Метод моментов был популярным в прошлом, но в настоящее время уступил место методу максимального правдоподобия.

Нам известно, что достаточная статистика T имеет вид $(\min_i X_i, \max_i X_i)$. Следовательно,

$$\hat{\theta}^* = \mathsf{E}(\hat{\theta} \mid T) = \frac{1}{n} \sum_i \mathsf{E}(X_i \mid T) - \frac{1}{2} = \mathsf{E}(X_1 \mid T) - \frac{1}{2} = \frac{1}{2}(\min X_i + \max X_i - 1).$$

Оценка $\hat{\theta}^*$ является несмещенной:

$$\mathsf{E}\hat{\theta}^* = \frac{1}{2}\left(\frac{n}{n+1} + \theta + \theta + \frac{1}{n+1} - 1\right) = \theta,$$

и ее с. к. о. должна быть не больше, чем у θ^*. Опять игнорируя дополнительную информацию относительно $\boldsymbol{X}$, уменьшаем с. к. о. Действительно, дисперсия $\operatorname{Var}\hat{\theta}^* = \frac{1}{4}\operatorname{Var}(\min X_i + \max X_i)$ равна

$$\frac{1}{4}\mathsf{E}(\min X_i + \max X_i)^2 - \frac{1}{4}(\mathsf{E}\min X_i + \mathsf{E}\max X_i)^2 =$$

$$= \frac{1}{4} \int\limits_{\theta}^{\theta+1} \int\limits_{x}^{\theta+1} f_{\min X_i, \max X_i}(x, y)(x+y)^2\,dy\,dx - \frac{1}{4}(2\theta+1)^2.$$

Ср. с равенствами (1.4.3) и (1.4.4). Здесь

$$f_{\min X_i, \max X_i}(x, y) = -\frac{\partial^2}{dx\,dy}(y-x)^n = n(n-1)(y-x)^{n-2}, \quad \theta < x < y < \theta + 1.$$

Обозначив

$$I = \int\limits_{\theta}^{\theta+1} \int\limits_{x}^{\theta+1} (y-x)^{n-2}(x+y)^2\,dy\,dx,$$

находим, что

$$\frac{1}{4}n(n-1)I = \theta^2 + \theta + \frac{n^2+3n+4}{4(n+1)(n+2)}.$$

При $n \geqslant 3$ получим соотношение

$$\operatorname{Var}\hat{\theta}^* = \frac{n^2+3n+4}{4(n+1)(n+2)} - \frac{1}{4} = \frac{1}{2(n+1)(n+2)} < \frac{1}{12n} = \operatorname{Var}\overline{X}.$$

В самом деле, введенный выше интеграл I равен

$$\int\limits_{\theta}^{\theta+1} \int\limits_{x}^{\theta+1} (y-x)^{n-2}[(y-x)^2 + 4x(y-x) + 4x^2)\,dy\,dx =$$

$$= \int\limits_{\theta}^{\theta+1} \left[\frac{(\theta+1-x)^{n+1}}{n+1} + \frac{4x(\theta+1-x)^n}{n} + \frac{4x^2(\theta+1-x)^{n-1}}{n-1}\right] dx =$$

$$= \frac{1}{(n-1)n(n+1)(n+2)} n(n-1) + 4\theta(n-1)(n+2) +$$

$$+ 4(n-1) + 4\theta^2(n+1)(n+2) + 8\theta(n+2) + 8].$$

Следовательно,

$$\frac{1}{4}n(n-1)I = \theta^2 + \theta + \frac{n^2+3n+4}{4(n+1)(n+2)},$$

что и утверждалось выше. □

Пример 1.6.2. Предположим, что $X_1, \ldots, X_n$ — независимые случайные величины, равномерно распределенные на $(\theta, 2\theta)$. Требуется найти двумерную достаточную статистику $T(X)$ для параметра θ. Покажем, что несмещенной оценкой параметра θ является $\hat{\theta} = 2X_1/3$.

Найдем несмещенную оценку параметра θ, которая является функцией от $T(X)$ и ее с. к. о. не больше, чем у $\hat{\theta}$.

В этом примере функция правдоподобия равна

$$f(\boldsymbol{x}; \theta) = \prod_{i=1}^{n} \frac{1}{\theta} I(\theta < x_i < 2\theta) = \frac{1}{\theta^n} I\left(\min_i x_i > \theta,\ \max_i x_i < 2\theta\right),$$

и, следовательно, в силу критерия факторизации статистика

$$T = \left(\min_i X_i,\ \max_i X_i\right)$$

является достаточной. Очевидно, $\mathsf{E}X_1 = 3\theta/2$, таким образом, величина $2X_1/3$ является несмещенной оценкой. Определим оценку

$$\hat{\theta} = \mathsf{E}\left[\frac{2}{3}X_1 \mid \min_i X_i = a,\ \max_i X_i = b\right] = \frac{2a}{3n} + \frac{2b}{3n} + \frac{n-2}{n}\frac{2}{3}\frac{a+b}{2} = \frac{a+b}{3}.$$

Последнее равенство можно обосновать так: X_1 равно a либо b с вероятностями, равными $1/n$ для каждого значения; в противном случае (т. е. когда $X_1 \neq a, b$, что выполняется с вероятностью $(n-2)/n$) условное математическое ожидание величины X_1 равно $(a+b)/2$ в силу симметрии.

Следовательно, по теореме Р—Б величина

$$\hat{\theta} = \frac{1}{3}\left(\min_i X_i + \max_i X_i\right)$$

является искомой оценкой. □

Конечно, хотелось бы получать оценки с минимальной с. к. о. (*минимальные с. к. о.-оценки*). Эффективное средство для отыскания таких оценок предоставляет нам *неравенство Крамера—Рао* (К—Р), или *граница К—Р*.

Предположим, что п. р. в./д. ф. р. $f(\cdot; \theta)$ является гладкой функцией параметра θ и выполняется следующее условие: для любых $\theta \in \Theta$ справедливо равенство

$$\int \frac{\partial}{\partial\theta} f(x; \theta)\, dx = 0 \quad \text{или} \quad \sum_{x \in \mathbf{V}} \frac{\partial}{\partial\theta} f(x; \theta) = 0. \tag{1.6.4}$$

Рассмотрим нормальные одинаково распределенные наблюдения $\boldsymbol{X} = (X_1, \ldots, X_n)$ *с совместной п. р. в./д. ф. р.* $f(\boldsymbol{x}; \theta) = f(x_1; \theta) \ldots f(x_n; \theta)$. *Пусть* $\theta^*(\boldsymbol{X})$ — *несмещенная оценка параметра* θ, *удовлетворяющая следующему условию: для любых* $\theta \in \Theta$ *справедливо равенство*

$$\int_{\mathbb{R}^n} \theta^*(\boldsymbol{x}) \frac{\partial}{\partial\theta} f(\boldsymbol{x}; \theta)\, d\boldsymbol{x} = 1 \quad \text{или} \quad \sum_{\boldsymbol{x} \in \mathbb{V}^n} \theta^*(\boldsymbol{x}) \frac{\partial}{\partial\theta} f(\boldsymbol{x}; \theta) = 1. \tag{1.6.5}$$

Тогда для любой такой оценки имеет место неравенство

$$\operatorname{Var} \theta^* \geqslant \frac{1}{nA(\theta)}, \tag{1.6.6}$$

где

$$A(\theta) = \int \frac{(\partial f(x; \theta)/\partial\theta)^2}{f(x; \theta)}\, dx \quad \text{или} \quad \sum_{x \in \mathbb{V}} \frac{(\partial f(x; \theta)/\partial\theta)^2}{f(x; \theta)}. \tag{1.6.7}$$

Величину $A(\theta)$ часто называют *количеством информации по Фишеру*, и она возникает во многих областях теории вероятностей и статистики.

Замечание. На практике условие (1.6.4) означает, что мы можем поменять местами дифференцирование $\frac{\partial}{\partial\theta}$ и интегрирование/суммирование в (очевидном) равенстве

$$\frac{\partial}{\partial\theta} \int f(x; \theta)\, dx = 0 \quad \text{или} \quad \frac{\partial}{\partial\theta} \sum_{x \in \mathbb{V}} f(x; \theta) = 0.$$

Это равенство выполняется, поскольку

$$\int f(x; \theta)\, dx = 1 \quad \text{или} \quad \sum_{x \in \mathbb{V}} f(x; \theta) = 1.$$

Достаточным условием для такой перестановки операций является следующее условие:

$$\int \left| \frac{\partial}{\partial\theta} f(x; \theta) \right| dx < \infty \quad \text{или} \quad \sum_{x \in \mathbb{V}} \left| \frac{\partial}{\partial\theta} f(x; \theta) \right| < \infty,$$

которое часто предполагается выполненным по умолчанию. Аналогично неравенство (1.6.5) означает, что можно менять порядок дифференцирования $\frac{\partial}{\partial\theta}$ и интегрирования/суммирования в равенстве

$$1 = \frac{\partial}{\partial\theta} \theta = \frac{\partial}{\partial\theta} \mathsf{E} \theta^*(\boldsymbol{X}),$$

поскольку

$$\mathsf{E}\theta^*(\boldsymbol{X}) = \int_{\mathbb{R}^n} \theta^*(\boldsymbol{x}) f(\boldsymbol{x}; \theta)\, d\boldsymbol{x} \quad \text{или} \quad \sum_{\boldsymbol{x} \in \mathbb{V}^n} \theta^*(\boldsymbol{x}) f(\boldsymbol{x}; \theta).$$

Заметим, что производную $\dfrac{\partial f(\boldsymbol{x};\,\theta)}{\partial\theta}$ можно записать в виде

$$f(\boldsymbol{x};\,\theta)\sum_{i=1}^{n}\frac{\partial f(x_i;\,\theta)/\partial\theta}{f(x_i;\,\theta)}=f(\boldsymbol{x};\,\theta)\sum_{i=1}^{n}\frac{\partial\ln f(x_i;\,\theta)}{\partial\theta}. \tag{1.6.8}$$

Мы докажем неравенство (1.6.6) только в непрерывном случае (для доказательства в дискретном случае потребуется заменить интегралы суммами). Положим

$$D(x,\,\theta)=\frac{\partial}{\partial\theta}\ln f(x;\,\theta).$$

В силу условия (1.6.4) выполняется равенство

$$\int f(x;\,\theta)D(x,\,\theta)\,dx=0,$$

и мы получаем, что $\forall\theta\in\Theta$

$$\int f(\boldsymbol{x};\,\theta)\sum_{i=1}^{n}D(x_i,\,\theta)\,d\boldsymbol{x}=n\int\frac{\partial}{\partial\theta}f(x;\,\theta)\,dx=0.$$

С другой стороны, в силу соотношения (1.6.8) равенство (1.6.5) можно записать в виде

$$\int_{\mathbb{R}^n}\theta^*(\boldsymbol{x})f(\boldsymbol{x};\,\theta)\sum_{i=1}^{n}D(x_i,\,\theta)\,d\boldsymbol{x}=1,\quad \theta\in\Theta.$$

Из последних двух равенств следует, что

$$\int_{\mathbb{R}^n}(\theta^*(\boldsymbol{x})-\theta)\sum_{i=1}^{n}D(x_i,\,\theta)f(\boldsymbol{x};\,\theta)\,d\boldsymbol{x}=1,\quad \theta\in\Theta.$$

В дальнейшем будем опускать область интегрирования $\mathbb{R}^n$ в n-мерных интегралах по $d\boldsymbol{x}$ (интегралы по dx будут, конечно, одномерными). Последнее равенство можно интерпретировать как $\langle g_1,\,g_2\rangle=1$, где

$$\langle g_1,\,g_2\rangle=\int g_1(\boldsymbol{x})g_2(\boldsymbol{x})\eta(\boldsymbol{x})\,d\boldsymbol{x}$$

обозначает скалярное произведение,

$$g_1(\boldsymbol{x})=\theta^*(\boldsymbol{x})-\theta,\ g_2(\boldsymbol{x})=\sum_{i=1}^{n}D(x_i,\,\theta)\quad\text{и}\quad\eta(\boldsymbol{x})=f(\boldsymbol{x};\,\theta)\geqslant 0$$

($\eta(\boldsymbol{x})$ — весовая функция, которая определяет скалярное произведение).

Используем теперь неравенство Коши—Шварца $(\langle g_1,\,g_2\rangle)^2\leqslant\langle g_1,\,g_1\rangle\times$ $\times\langle g_2,\,g_2\rangle$. Мы получим, что

$$1\leqslant\left[\int(\theta^*(\boldsymbol{x})-\theta)^2f(\boldsymbol{x};\,\theta)\,d\boldsymbol{x}\right]\left[\int\left(\sum_{i=1}^{n}D(x_i,\,\theta)\right)^2f(\boldsymbol{x};\,\theta)\,d\boldsymbol{x}\right].$$

Здесь первый сомножитель представляет собой $\operatorname{Var}\theta^*(\boldsymbol{X})$. Второй сомножитель равен в точности $nA(\theta)$. Действительно,

$$\int\left(\sum_{i=1}^{n} D(x_i, \theta)\right)^2 f(\boldsymbol{x}; \theta)\, d\boldsymbol{x} = \sum_{i=1}^{n} \int D(x_i, \theta)^2 f(\boldsymbol{x}; \theta)\, d\boldsymbol{x} + \\ + 2 \sum_{1 \leqslant i_1 < i_2 \leqslant n} \int D(x_{i_1}, \theta) D(x_{i_2}, \theta) f(\boldsymbol{x}; \theta)\, d\boldsymbol{x}.$$

Каждое слагаемое в первой сумме равно $A(\theta)$, поскольку

$$\int D(x_i, \theta)^2 f(\boldsymbol{x}; \theta)\, d\boldsymbol{x} = \int D(x, \theta)^2 f(x; \theta)\, dx.$$

Каждое слагаемое во второй сумме равно нулю, так как $\int D(x, \theta) f(x; \theta)\, dx = 0$:

$$\int D(x_{i_1}, \theta) D(x_{i_1}, \theta) f(\boldsymbol{x}; \theta)\, d\boldsymbol{x} = \left(\int D(x, \theta) f(x; \theta)\, dx\right)^2 = 0.$$

Это и завершает доказательство.

В заключение этого параграфа расскажем кратко о пребывании Рао в Кембридже. Рао прибыл в Кембридж в 1945 г. для проведения в университетском музее археологии и антропологии работ по анализу объектов (человеческих черепов и костей), насчитывающих тысячи лет, которые были доставлены британской экспедицией из Северной Африки с раскопок на стоянке первобытного человека. Он должен был тщательно измерять их и, привлекая к расчетам расстояние по Махаланобису, делать выводы об антропологических характеристиках древнего населения. Рао было тогда 25 лет, и у него насчитывалось 18 работ, уже опубликованных или принятых к публикации. Его первая статья, которая тогда как раз была напечатана, уже содержала в себе и теорему Р—Б и неравенство К—Р. Вскоре после прибытия в Кембридж Рао встретил Фишера и начал работать также и в генетической лаборатории, руководимой Фишером. Изучал он различные характеристики мышей; он должен был разводить мышей, спаривать их определенным образом и фиксировать генетические параметры производимого потомства (форма хвостов, гладкость шерсти и пропорции туловища должны были меняться от поколения к поколению). Как описывается в биографии Рао [Кг], его заключительной обязанностью в каждом эксперименте было ликвидировать мышей, которые были больше не нужны для дальнейшей работы. Молодых мышей усыпляли эфиром, зрелым особям разбивали головы ударом об стол (практика, которая в наши дни несомненно вызвала бы протесты со стороны активистов общества защиты животных). Рао был слишком чувствителен, чтобы выполнять эту часть работы, но у него были друзья, которые делали это за него; во всем остальном он был чрезвычайно доволен своей работой (и своей жизнью) в Кембридже. Одним из его друзей был Абдус Салам, будущий нобелевский лауреат в области физики, а в то время студент колледжа Святого Иоанна. Салам сомневался в будущих результатах своих исследований и, наблюдая твердый выбор Рао, проявил живой интерес к статистике. Однако именно Рао убедил его не менять область исследований.

Работа в лаборатории Фишера послужила основой для докторской диссертации Рао, которую он успешно защитил в 1948 г. К тому времени у него уже насчитывалось 40 опубликованных работ. После получения докторской степени в таком престижном университете, как Кембридж, он становился желанным женихом для многих богатых семейств в Индии. Однако, спустя месяц по возвращении домой из Кембриджа он был помолвлен со своей будущей женой, которой было тогда 23 года, тоже имеющей ученую степень. Брак был устроен матерью Рао, которая была очень прогрессивной женщиной и не возражала против высокообразованной жены, хотя в то время общепринято было считать таких невест не очень желанными в семьях с престижными сыновьями. Их брак оказался в высшей степени счастливым, что заставляет задуматься, почему никем не устраиваемые браки других известных статистиков в Европе и Америке (включая тех, имена которых неоднократно упоминаются в этой книге) были неудачными.

Вклад Рао в статистику сейчас повсеместно признан. Такое признание пришло не сразу, особенно в отношении теоремы Р—Б. Даже в 1953 г., спустя восемь лет после появления статьи Рао и пять лет после выхода статьи Блекуэлла, некоторые статистики ссылались на метод, предоставляемый этой теоремой, как на «блекуэллизацию» («Blackwellisation»). Когда Рао подчеркнул, что он первым открыл этот результат, один из лекторов сообщил, что такой термин легче произносить, чем термин «раоизация» («Raoisation»). Однако в более поздней статье упомянутый статистик предложил использовать термин «рао-блекуэллизация» («Rao-Blackwellisation»), который и используется в настоящее время. Что же касается неравенства Крамера—Рао (термин был предложен Нейманом), Рао вспоминает случай, когда в одном из полетов пропала часть его багажа и звонивший ему служащий тегеранского аэропорта сказал: «У нас хорошие новости, господин Крамер Рао, мы нашли вашу сумку».

Рао нравится подходить с юмором к серьезным ситуациям. В Индии, как и во многих странах, важным является контроль за рождаемостью, и одна из постоянных задач местных и центральных властей — убедить женщин в необходимости ограничивать количество детей. В одной из своих статей по этому вопросу Рао отмечает, что каждый четвертый новорожденный в мире — китаец, и делает следующее заключение для своей индийской аудитории: «Задумайтесь, прежде чем вы решитесь завести еще одного ребенка, имея уже троих. Ведь четвертый будет китайцем!» Надеемся, что читатели найдут это поучительным при использовании выборочного среднего...

§ 1.7. Экспоненциальные (показательные) семейства

Sex, Lie, and Exponential Families[1]

(Из серии «Фильмы, которые не вышли на большой экран».)

Интересно выяснить, когда же достигается граница, указанная в неравенстве К—Р (1.6.6). Здесь вновь решающую роль играет неравенство Коши—Шварца. Известно, что для достижения равенства необходимо, чтобы функции g_1 и g_2 были линейно зависимыми: $g_1(\boldsymbol{x}) = \lambda g_2(\boldsymbol{x})$. Тогда

$$\int g_1(\boldsymbol{x})g_2(\boldsymbol{x})\eta(\boldsymbol{x})\,d\boldsymbol{x} = \lambda \int g_2(\boldsymbol{x})^2\eta(\boldsymbol{x})\,d\boldsymbol{x}.$$

В нашем случае $\int g_1(\boldsymbol{x})g_2(\boldsymbol{x})\eta(\boldsymbol{x})\,d\boldsymbol{x} = 1$ и, следовательно,

$$\lambda = \left(\int g_2(\boldsymbol{x})^2\eta(\boldsymbol{x})\,d\boldsymbol{x}\right)^{-1} = \frac{1}{nA(\theta)}.$$

Таким образом, получаем соотношение

$$\theta^*(\boldsymbol{X}) - \theta = \frac{1}{nA(\theta)}\sum_{i=1}^{n} D(x_i, \theta),$$

[1] Ср. название фильма «Sex, Lie, and Video Tape».

или

$$\theta^*(\boldsymbol{X}) = \theta + \frac{1}{nA(\theta)} \sum_{i=1}^{n} D(x_i, \theta) = \frac{1}{n} \sum_{i=1}^{n} \left(\theta + \frac{D(x_i, \theta)}{A(\theta)} \right).$$

Левая часть последнего равенства не зависит от θ. Следовательно, и каждое слагаемое $\theta + D(x_i, \theta)/A(\theta)$ не должно зависеть от θ:

$$\theta + \frac{D(x, \theta)}{A(\theta)} = C(x).$$

Другими словами,

$$D(x, \theta) = A(\theta)(C(x) - \theta), \quad \theta \in \Theta, \tag{1.7.1}$$

и оценка θ^* имеет вид суммы:

$$\theta^*(\boldsymbol{x}) = \frac{1}{n} \sum_{i=1}^{n} C(x_i). \tag{1.7.2}$$

Теперь, решая уравнение (1.7.1), т. е. уравнение вида

$$\frac{\partial}{\partial\theta} \ln f(x; \theta) = A(\theta)[C(x) - \theta],$$

где $A(\theta) = B''(\theta)$, мы получаем

$$\ln f(x; \theta) = B'(\theta)[C(x) - \theta] + B(\theta) + H(x).$$

Поэтому

$$f(x; \theta) = \exp[B'(\theta)[C(x) - \theta] + B(\theta) + H(x)]. \tag{1.7.3}$$

Такие семейства (п. р. в. или д. ф. р.) называют *экспоненциальными*.

Следовательно, имеет место такое утверждение.

В неравенстве К—Р знак равенства достигается тогда и только тогда, когда семейство $\{f(x; \theta)\}$ *является экспоненциальным, т. е. задается формулой* (1.7.3), *где* $B''(\theta) > 0$. *В этом случае оценка параметра* θ *с минимальной с. к. о. задается формулой* (1.7.2) *и ее дисперсия равна* $1/(nB''(\theta))$. *Таким образом, информация по Фишеру равна* $B''(\theta)$.

Пример 1.7.1. Пусть X_i — это н. о. р. $\mathrm{N}(\mu, \sigma^2)$-с. в. с заданной дисперсией σ^2 и неизвестным μ. Обозначим $c = -[\ln(2\pi\sigma^2)]/2$ (что является константой при фиксированном σ^2). Тогда

$$\ln f(x; \mu) = -\frac{(x-\mu)^2}{2\sigma^2} + c = \frac{\mu(x-\mu)}{\sigma^2} + \frac{\mu^2}{2\sigma^2} - \frac{x^2}{2\sigma^2} + c.$$

Мы видим, что семейство п. р. в. $f(\cdot;\mu)$ является экспоненциальным; в этом случае

$$C(x)=x,\quad B(\mu)=\frac{\mu^2}{2\sigma^2},\quad B'(\mu)=\frac{\mu}{\sigma^2},$$
$$A(\mu)=B''(\mu)=\frac{1}{\sigma^2},\quad H(x)=-\frac{x^2}{2\sigma^2}+c.$$

Отметим, что здесь $A(\mu)$ не зависит от μ.

Следовательно, в классе таких несмещенных оценок $\theta^*(\boldsymbol{x})$, что

$$\sum_{i=1}^{n}\int\theta^*(\boldsymbol{x})\frac{(x_i-\mu)}{\sigma^2(2\pi\sigma^2)^{n/2}}\exp\left(-\sum_j\frac{(x_j-\mu)^2}{2\sigma^2}\right)d\boldsymbol{x}=1,$$

оценкой с минимальной с. к. о. для параметра μ является

$$\bar{x}=\frac{1}{n}\sum_i x_i,$$

т. е. выборочное среднее с дисперсией $\operatorname{Var}\bar{X}=\sigma^2/n$. □

Пример 1.7.2. Если X_i — это н. о. р. $\mathrm{N}(\mu,\sigma^2)$-с. в. с заданным средним μ и неизвестной дисперсией σ^2, то вновь семейство п. р. в. $f(\cdot;\sigma^2)$ является экспоненциальным:

$$\ln f(x;\sigma^2)=-\frac{1}{2\sigma^2}((x-\mu)^2-\sigma^2)-\frac{1}{2}\ln(2\pi\sigma^2)-\frac{1}{2},$$

где

$$C(x)=(x-\mu)^2,\quad B(\sigma^2)=-\frac{1}{2}\ln(\sigma^2),\quad B'(\sigma^2)=-\frac{1}{2\sigma^2},$$
$$A(\sigma^2)=B''(\sigma^2)=\frac{1}{2}\left(\frac{1}{\sigma^2}\right)^2$$

и $H(x)=-(\ln(2\pi)+1)/2$. Мы приходим к заключению, что в классе таких несмещенных оценок $\theta^*(\boldsymbol{x})$, что

$$\sum_{i=1}^{n}\int_{\mathbb{R}^n}\theta^*(\boldsymbol{x})\frac{(x_i-\mu)^2-\sigma^2}{2\sigma^4(2\pi\sigma^2)^{n/2}}\exp\left(-\sum_j\frac{(x_j-\mu)^2}{2\sigma^2}\right)d\boldsymbol{x}=1,$$

оценкой для σ^2 с минимальной с. к. о. является Σ^2/n, где

$$\Sigma^2=\sum_i(x_i-\mu)^2$$

и $\operatorname{Var}(\Sigma^2/n)=2\sigma^4/n$. □

Пример 1.7.3. Пуассоновское семейство, для которого $f_\lambda(k)=e^{-\lambda}\lambda^k/k!$, также является экспоненциальным:

$$\ln f_\lambda(k)=k\ln\lambda-\ln(k!)-\lambda=(k-\lambda)\ln\lambda+\lambda(\ln\lambda-1)-\ln(k!)$$

и

$$C(k) = k, \quad B(\lambda) = \lambda(\ln\lambda - 1), \quad B'(\lambda) = \ln\lambda, \quad A(\lambda) = B''(\lambda) = \frac{1}{\lambda},$$

$H(k) = -\ln(k!)$. Таким образом, оценка с минимальной с. к. о. имеет вид

$$\bar{\lambda} = \frac{1}{n}\sum_{i=1}^{n} k_i, \quad \mathrm{Var}\,\bar{\lambda} = \frac{\lambda}{n}.$$

Предоставляем читателю определить, в каком классе оценок обеспечивается минимальная с. к. о. □

Пример 1.7.4. Семейство с. в. $X_i \sim \mathrm{Exp}(\lambda)$ является экспоненциальным, но относительно $\theta = 1/\lambda$ (что и неудивительно, поскольку $\mathsf{E}X_i = 1/\lambda$). Действительно, в этом случае

$$f(x;\,\theta) = \left[\exp\left(-\frac{1}{\theta}(x-\theta) - \ln\theta - 1\right)\right] I(x > 0)$$

и

$$C(x) = x, \quad B(\theta) = -\ln\theta, \quad B'(\theta) = -\frac{1}{\theta}, \quad A(\theta) = B''(\theta) = \frac{1}{\theta^2}, \quad H(x) = -1.$$

Следовательно, в классе несмещенных оценок $\theta^*(\boldsymbol{x})$ параметра $1/\lambda$, для которых

$$\sum_{i=1}^{n} \int\limits_{\mathbb{R}_+^n} \theta^*(\boldsymbol{x})(\lambda^2 x_i - \lambda)\lambda^n \exp\left(-\lambda\sum_j x_j\right) d\boldsymbol{x} = 1,$$

выборочное среднее $\bar{x} = \sum\limits_i x_i/n$ является оценкой с минимальной с. к. о., и ее дисперсия $\mathrm{Var}\,\bar{X} = \theta^2/n = 1/(n\lambda^2)$. □

Важным свойством является то, что для экспоненциальных семейств оценка с минимальной с. к. о. $\theta^*(\boldsymbol{x})$ совпадает с о. м. п. $\hat{\theta}(\boldsymbol{x})$, т. е.

$$\hat{\theta}(\boldsymbol{x}) = \frac{1}{n}\sum_i C(x_i). \tag{1.7.4}$$

Точнее, запишем уравнение для стационарной точки в виде

$$\frac{\partial}{\partial\theta}\ell(\theta;\,\boldsymbol{x}) = \sum_{i=1}^{n} D(x_i,\,\theta) f(\boldsymbol{x};\,\theta) = 0.$$

При условии $f(\boldsymbol{x};\,\theta) > 0$ оно превращается в условие

$$\sum_{i=1}^{n} D(x_i,\,\theta) = 0. \tag{1.7.5}$$

В случае экспоненциального семейства, для которого

$$f(x;\,\theta) = \exp\bigl(B'(\theta)[C(x) - \theta] + B(\theta) + H(x)\bigr),$$

получаем

$$D(x, \theta) = \frac{\partial}{\partial\theta} \ln f(x; \theta) = B''(\theta)(C(x) - \theta).$$

Видим, что для $\theta^* = \theta^*(\boldsymbol{x}) = \sum\limits_{i=1}^{n} C(x_i)/n$ выполняется равенство

$$D(x_i, \theta^*) = B''(\theta^*)[C(x_i) - \theta^*] = B''(\theta^*)\left[C(x_i) - \frac{1}{n}\sum_j C(x_j)\right],$$

следовательно,

$$\sum_{i=1}^{n} D(x_i, \theta^*) = B''(\theta^*)\sum_{i=1}^{n}\left[C(x_i) - \frac{1}{n}\sum_j C(x_j)\right] = 0.$$

Таким образом, оценка с минимальной с. к. о. θ^* является решением уравнения для стационарной точки. Значит, если экспоненциальное семейство $\{f(x; \theta)\}$ таково, что любое решение стационарного уравнения (1.4.2) является точкой глобального максимума для функции правдоподобия, то оценка θ^* является о. м. п.

Неравенство К—Р названо в честь Г. Крамера (1893—1985), выдающегося шведского теоретика анализа, теории чисел, а также вероятностника и статистика, и в честь С. Р. Рао. Одна из историй гласит, что неравенство в своей окончательной форме было доказано Рао в то время, когда он был молодым и неопытным лектором в Индийском статистическом институте в 1943 г., причем за одну ночь, в ответ на вопрос студента о некоторых неясных местах лекции, прочитанной накануне.

§ 1.8. Доверительные интервалы

Статистикам можно доверять на 95%, когда они делают это.

(Из серии «Как они делают это».)

До сих пор мы развивали идеи, относящиеся к *точечному оцениванию*. Другая полезная идея состоит в рассмотрении *интервальных оценок*, когда работают с *доверительными интервалами* (*д. и.*) (в случае векторного параметра $\theta \in \mathbb{R}^d$ — с доверительными множествами, такими как квадраты, кубы, круги или эллипсы, шары и т. п.). При заданном $\gamma \in (0, 1)$ определим 100γ%-д. и. для скалярного параметра $\theta \in \Theta \subseteq \mathbb{R}$ как любую пару таких функций $a(\boldsymbol{x})$ и $b(\boldsymbol{x})$, что для любого $\theta \in \Theta$ выполняется равенство

$$\mathsf{P}_\theta(a(\boldsymbol{X}) \leqslant \theta \leqslant b(\boldsymbol{X})) = \gamma. \qquad (1.8.1)$$

Мы хотим подчеркнуть следующее: а) случайными здесь являются конечные точки интервала $a(\boldsymbol{x}) < b(\boldsymbol{x})$, но не θ, б) a и b не должны зависеть от θ. Доверительный интервал может выбираться разными способами, но естественно, что всегда интересуются наиболее «короткими» интервалами.

Пример 1.8.1. Первый стандартный пример — это доверительный интервал для неизвестного среднего нормального распределения с известной дисперсией. Предположим, что $X_i \sim \mathrm{N}(\mu, \sigma^2)$, где дисперсия σ^2 известна и необходимо оценить $\mu \in \mathbb{R}$. Мы хотим найти 99%-д. и. для μ. Нам известно, что $\sqrt{n}(\overline{X} - \mu)/\sigma \sim \mathrm{N}(0, 1)$. Следовательно, если мы выберем $z_- < z_+$ такими, что $\Phi(z_+) - \Phi(z_-) = 0{,}99$, то равенство

$$\mathrm{P}\left(z_- < \frac{\sqrt{n}}{\sigma}(\overline{X} - \mu) < z_+\right) = 0{,}99$$

можно переписать в виде

$$\mathrm{P}\left(\overline{X} - \frac{z_+\sigma}{\sqrt{n}} < \mu < \overline{X} - \frac{z_-\sigma}{\sqrt{n}}\right) = 0{,}99,$$

т. е. мы получаем интервал

$$\left(\overline{X} - \frac{z_+\sigma}{\sqrt{n}}, \overline{X} - \frac{z_-\sigma}{\sqrt{n}}\right)$$

с центром в точке $\overline{X} - (z_- + z_+)\sigma/(2\sqrt{n})$, имеющий ширину $(z_+ - z_-)\sigma/\sqrt{n}$.

У нас есть разные возможности для выбора точек z_- и z_+; чтобы получить самый короткий интервал, мы хотели бы выбрать $z_+ = -z_- := z$, поскольку п. р. в. $\mathrm{N}(0, 1)$-с. в. симметрична и имеет максимум в начале координат. Тогда интервал приобретает вид

$$\left(\overline{X} - \frac{z\sigma}{\sqrt{n}}, \overline{X} + \frac{z\sigma}{\sqrt{n}}\right)$$

и z — это верхняя 0,005 точка стандартного нормального распределения, так что $\Phi(z) = 1 - 0{,}005 = 0{,}995$. Из таблиц процентных точек нормального распределения находим $z = 2{,}5758$. Следовательно, получаем ответ

$$\left(\overline{X} - \frac{2{,}5758\sigma}{\sqrt{n}}, \overline{X} + \frac{2{,}5758\sigma}{\sqrt{n}}\right). \qquad \square$$

Пример 1.8.2. В данном примере необходимо определить д. и. для неизвестной дисперсии нормального распределения, когда среднее значение известно. Предположим, что $X_i \sim \mathrm{N}(\mu, \sigma^2)$, где μ известно и требуется оценить $\sigma^2 > 0$. Мы хотим найти 98%-д. и. для σ^2. Известно, что $\sum\limits_i (X_i - \mu)^2/\sigma^2 \sim \chi^2_n$. Обозначим через $F_{\chi^2_n}$ ф. р. $\mathrm{P}(X < x)$ случайной величины $X \sim \chi^2_n$. Выберем $h^- < h^+$ такими, что $F_{\chi^2_n}(h^+) - F_{\chi^2_n}(h^-) = 0{,}98$. Тогда условие

$$\mathrm{P}\left(h^- < \frac{1}{\sigma^2}\left(\sum_i (X_i - \mu)^2\right) < h^+\right) = 0{,}98$$

можно переписать в виде

$$\mathrm{P}\left(\frac{\sum\limits_i (X_i-\mu)^2}{h^+} < \sigma^2 < \frac{\sum\limits_i (X_i-\mu)^2}{h^-}\right) = 0{,}98,$$

что приводит к интервалу

$$\left(\frac{1}{h^+}\sum_i (X_i-\mu)^2,\ \frac{1}{h^-}\sum_i (X_i-\mu)^2\right).$$

Опять у нас есть возможности для выбора h^- и h^+; намереваясь получить симметричный (или имеющий равные «хвосты») интервал, мы возьмем $F_{\chi^2_n}(h^-) = 1 - F_{\chi^2_n}(h^+) = (1-0{,}98)/2 = 0{,}01$. Из таблиц процентных точек распределения χ^2 для $n = 38$ получаем $h^- = 20{,}69$, $h^+ = 61{,}16$. См., например, [LiS, с. 40—41] (это стандартный источник статистических таблиц, который используется во многих примерах и задачах, рассматриваемых ниже). □

В примерах 1.8.1 и 1.8.2 нам удалось найти с. в. $Z = Z(T(\boldsymbol{X}), \theta)$, которая является функцией от достаточной статистики и неизвестного параметра θ (μ в примере 1.8.1 и σ^2 в примере 1.8.2) и распределение которой не зависит от этого параметра. А именно,

$$Z = \sqrt{n}(\overline{X}-\mu)/\sigma \sim \mathrm{N}(0,\ 1) \quad \text{в примере 1.8.1}$$

и

$$Z = \sum_i (X_i-\mu)^2/\sigma^2 \sim \chi^2_n \quad \text{в примере 1.8.2.}$$

Затем мы нашли такие значения a_- и a_+, что $\mathrm{P}(a_- < Z < a_+) = \gamma$, и корни $\theta_i = \theta_i(T,\ a_\pm)$ уравнения $Z(T,\ \theta) = a_\pm$. Не всегда удается найти их непосредственно в явном виде, в некоторых случаях могут быть полезными их приближения.

Пример 1.8.3. В этом (более сложном) примере все та же идея развивается далее. Пусть $X_i \sim \mathrm{Po}(\lambda)$ и мы хотим найти $100\gamma\%$-д. и. для λ. Нам известно, что $n\overline{X} \sim \mathrm{Po}(n\lambda)$, где все еще сохраняется зависимость от λ. Функция распределения $F = F_{\overline{X}}$ случайной величины $\overline{X}$ имеет скачки в точках k/n, $k = 0,\ 1,\ \ldots$, и представима в виде

$$F(x;\ \lambda) = I(x \geqslant 0)\sum_{r=0}^{]nx[} e^{-n\lambda}\frac{(n\lambda)^r}{r!}, \tag{1.8.2}$$

где для значений $y > 0$ мы обозначили

$$]y[= \begin{cases} y-1, & \text{если } y \text{ целое,} \\ [y] \text{ (целая часть числа } y), & \text{если } y \text{ не является целым.} \end{cases}$$

Эта функция дифференцируема и монотонно убывает по λ:

$$\frac{\partial}{\partial\lambda}F\left(\frac{k}{n}, \lambda\right) = -ne^{-n\lambda}\frac{(n\lambda)^{k-1}}{(k-1)!} < 0.$$

Таким образом, отыскание пределов $a < \lambda < b$ эквивалентно решению функциональных неравенств $F(x, b) < F(x; \lambda) < F(x; a)$ для всех положительных x вида k/n, $k = 1, 2, \ldots$ Мы хотим, чтобы a и b были функциями от $\boldsymbol{x}$, а точнее, от выборочного среднего $\bar{x}$. Симметричный, или имеющий равные «хвосты», $100\gamma\%$-д. и. имеет такие концы $a(\bar{X})$ и $b(\bar{X})$, что

$$\mathsf{P}(\lambda \leqslant a(\bar{X})) = \mathsf{P}(\lambda \geqslant b(\bar{X})) = \frac{1-\gamma}{2}.$$

Запишем

$$\mathsf{P}(\lambda > b(\bar{X})) = \mathsf{P}(F(\bar{X}; \lambda) < F(\bar{X}; b(\bar{X}))$$

и используем следующий факт (который является некоторой модификацией формул (2.1.45а ,б) из гл. 2 части А).

Если для ф. р. F с. в. X и функции $g\colon \mathbb{R} \to \mathbb{R}$ существует единственная такая точка $c \in \mathbb{R}$, что $F(c) = g(c)$, $F(y) < g(y)$ при $y < c$ и $F(y) \geqslant g(y)$ при $y \geqslant c$, то

$$\mathsf{P}(F(X) < g(X)) = g(c). \tag{1.8.3}$$

Далее, в силу равенства (1.8.2) получаем

$$\mathsf{P}(F(\bar{X}; \lambda) < F(\bar{X}; b(\bar{X}))) = F\left(\frac{k}{n}; b\left(\frac{k}{n}\right)\right)$$

при условии, что существует такое k, что $F(k/n; \lambda) = F(k/n; b(k/n))$. Другими словами, если выбрать $b = b(\bar{x})$ так, чтобы выполнялось равенство

$$\sum_{l=0}^{]n\bar{x}[} e^{-nb}\frac{(nb)^l}{l!} = \frac{1-\gamma}{2}, \tag{1.8.4}$$

то $\mathsf{P}(\lambda > b(\bar{X})) \leqslant (1-\gamma)/2$.

Аналогично выбор такого $a = a(\bar{x})$, что

$$\sum_{l=n\bar{x}}^{\infty} e^{-na}\frac{(na)^l}{l!} = \frac{1-\gamma}{2}, \tag{1.8.5}$$

будет гарантировать, что $\mathsf{P}(\lambda < a(\bar{X})) \leqslant (1-\gamma)/2$. Следовательно, интервал $(a(\bar{X}), b(\bar{X}))$ является $100\gamma\%$-д. и. для λ (имеющим равные «хвосты»).

Распределения величин $a(\bar{X})$ и $b(\bar{X})$ могут быть определены в терминах квантилей распределения χ^2. Действительно, $\forall k = 0, 1, \ldots$ мы имеем

$$\frac{d}{ds}\sum_{k\leqslant l<\infty} e^{-s}\frac{s^l}{l!} = \sum_{k\leqslant l<\infty}\left(-e^{-s}\frac{s^l}{l!} + e^{-s}\frac{s^{l-1}}{(l-1)!}\right) = e^{-s}\frac{s^{k-1}}{(k-1)!} = 2f_{Y_1}(2s),$$

где $Y_1 \sim \chi^2_{2k}$. Это означает, что $\sum\limits_{k \leqslant l < \infty} e^{-s} s^l / l! = \mathrm{P}(Y_1 < 2s)$. Мы видим, что

$$2na = h^-_{m^-}((1-\gamma)/2),$$

а это нижняя $(1-\gamma)/2$-квантиль χ^2-распределения с $m^- = 2n\overline{X}$ степенями свободы.

Аналогично

$$\frac{d}{ds} \sum_{0 \leqslant l \leqslant k} e^{-s} \frac{s^l}{l!} = \sum_{0 \leqslant l \leqslant k} \left(-e^{-s} \frac{s^l}{l!} + e^{-s} \frac{s^{l-1}}{(l-1)!} \right) = -e^{-s} \frac{s^k}{k!} = -2 f_{Y_2}(2s)$$

для $Y_2 \sim \chi^2_{2k+2}$. Таким образом,

$$\mathrm{P}(Y_2 > 2s) = \sum_{0 \leqslant l \leqslant k} e^{-s} \frac{s^l}{l!}.$$

Это означает, что

$$2nb = h^+_{m^+}((1-\gamma)/2),$$

а это верхняя $(1-\gamma)/2$-квантиль χ^2-распределения с $m^+ = 2n\overline{X} + 2$ степенями свободы.

Такой ответ выглядит несколько громоздко. Однако для больших n можно считать, что

$$\overline{X} \sim \mathrm{N}(\lambda, \lambda/n), \quad \text{т. е.} \quad \sqrt{n/\lambda}\,(\overline{X} - \lambda) \sim \mathrm{N}(0, 1).$$

Затем, выбрав $\gamma = 0{,}99$ и, как и ранее, $-a_- = a_+ = 2{,}5758$, мы получим, что

$$\mathrm{P}\left(a_- < \frac{\sqrt{n}}{\sqrt{\lambda}} (\overline{X} - \lambda) < a_+ \right) \approx 0{,}99.$$

Это уравнение можно решить относительно λ (или, точнее, относительно $\sqrt{\lambda}$). Действительно, $(\overline{X} - \lambda)/\sqrt{\lambda}$ убывает по λ, и из уравнения

$$\frac{\overline{X} - \lambda}{\sqrt{\lambda}} = \frac{a_\pm}{\sqrt{n}}$$

находим

$$\sqrt{\lambda_\mp} = \sqrt{\frac{a_\pm^2}{4n} + \overline{X}} - \frac{a_\pm}{2\sqrt{n}}, \quad \text{т. е.} \quad \lambda_\mp = \left(\sqrt{\frac{a_\pm^2}{4n} + \overline{X}} - \frac{a_\pm}{2\sqrt{n}} \right)^2$$

(отрицательные корни нужно отбросить). Следовательно, вероятность

$$\mathrm{P}\left(\left(\sqrt{\frac{2{,}5758^2}{4n} + \overline{X}} - \frac{2{,}5758}{2\sqrt{n}} \right)^2 < \lambda < \left(\sqrt{\frac{2{,}5758^2}{4n} + \overline{X}} + \frac{2{,}5758}{2\sqrt{n}} \right)^2 \right)$$

будет приближенно равна 0,99. Значит, интервал

$$\left(\left(\sqrt{\frac{2{,}5758^2}{4n}+\bar{X}}-\frac{2{,}5758}{2\sqrt{n}}\right)^2,\ \left(\sqrt{\frac{2{,}5758^2}{4n}+\bar{X}}+\frac{2{,}5758}{2\sqrt{n}}\right)^2\right)$$

дает приближенный ответ, точность которого возрастает вместе с n.

Доверительным интервалам для среднего значения распределения Пуассона уделяется особое внимание во многих книгах, начиная с [Bio]. В этой книге используется термин «доверительная полоса», чтобы подчеркнуть зависимость интервала от n и $\bar{X}$ одновременно. □

§ 1.9. Байесовское оценивание

Bayesian Instinct, Trading Priors[1]

(Из серии «Фильмы, которые не вышли на большой экран».)

Полезной альтернативой вышеизложенному методу оценивания является такой подход, при котором параметр θ интерпретируется как случайная величина со значениями в Θ, имеющая некоторую (заданную) ***априорную*** п. р. в. или д. ф. р. $\pi(\theta)$. Получив в результате наблюдений выборку $\boldsymbol{x}$, можно найти ***апостериорную*** п. р. в. или д. ф. р. $\pi(\theta\,|\,\boldsymbol{x})$. На основании теоремы Байеса $\pi(\theta\,|\,\boldsymbol{x})$ определяется как

$$\pi(\theta\,|\,\boldsymbol{x}) \propto \pi(\theta) f(\boldsymbol{x};\,\theta). \tag{1.9.1}$$

Более точно,

$$\pi(\theta\,|\,\boldsymbol{x}) = \frac{1}{f(\boldsymbol{x})}\pi(\theta) f(\boldsymbol{x};\,\theta), \tag{1.9.2}$$

где

$$f(\boldsymbol{x}) = \int_{\Theta} \pi(\theta) f(\boldsymbol{x};\,\theta)\,d\theta \quad \text{или} \quad \sum_{\theta\in\Theta} \pi(\theta) f(\boldsymbol{x};\,\theta).$$

Образно говоря, (апостериорная п. р. в.) $\propto$ (априорная п. р. в.) $\times$ (правдоподобие), где коэффициент пропорциональности выбирается так, чтобы нормировать общую массу к 1.

Замечание. Отметим, что функция правдоподобия $f(\boldsymbol{x};\,\theta)$ в формулах (1.9.1) и (1.9.2) рассматривается как условная п. р. в./д. ф. р. величины $\boldsymbol{X}$ при заданном θ. Такова байесовская интерпретация функции правдоподобия, что противоположно фишеровской интерпретации, когда функция правдоподобия рассматривается как функция от θ при фиксированном $\boldsymbol{x}$.

В примерах 1.9.1—1.9.4 мы вычислим апостериорные распределения для некоторых моделей.

[1] Ср. названия фильмов: «Basic Instinct» и «Trading Places».

Пример 1.9.1. Пусть $X_i \sim \mathrm{Bin}(m, \theta)$ и априорное распределение для θ — это $\mathrm{Bet}(a, b)$-распределение с некоторыми известными a, b:

$$\pi(\theta) \propto \theta^{a-1}(1-\theta)^{b-1} I(0 < \theta < 1)$$

(см. пример 1.1.5). Тогда для апостериорного распределения получим

$$\pi(\theta \mid \boldsymbol{x}) \propto \theta^{\sum x_i + a - 1}(1-\theta)^{nm - \sum x_i + b - 1} I(0 < \theta < 1),$$

а это выражение представляет собой $\mathrm{Bet}\Big(\sum\limits_i x_i + a,\ nm - \sum\limits_i x_i + b\Big) = \mathrm{Bet}(n\bar{x} + a,\ n(m - \bar{x}) + b)$. Другими словами, априорное бета-распределение порождает апостериорное бета-распределение. Говорят, что семейство бета-распределений является *сопряженным* для биномиальных выборок. □

The Unbelievable Conjugacy of Beta[1]

(Из серии «Фильмы, которые не вышли на большой экран».)

Пример 1.9.2. Семейство бета-распределений является сопряженным также и для выборок из отрицательного биномиального распределения, когда $X_i \sim \mathrm{NegBin}(r, \theta)$ с известным r. Действительно, в этом случае апостериорное распределение

$$\pi(\theta \mid \boldsymbol{x}) \propto \theta^{nr + a - 1}(1-\theta)^{n\bar{x} + b - 1}$$

представляет собой $\mathrm{Bet}(nr + a,\ n\bar{x} + b)$-распределение. □

Пример 1.9.3. Еще один распространенный пример сопряженного семейства — это гамма-распределения для пуассоновских или показательных (экспоненциальных) выборок. Действительно, если $X_i \sim \mathrm{Po}(\lambda)$ и $\pi(\lambda) \propto \lambda^{\tau-1} e^{-\lambda/\alpha}$, то апостериорное распределение

$$\pi(\lambda \mid \boldsymbol{x}) \propto \lambda^{n\bar{x} + \tau - 1} e^{-\lambda(\alpha n + 1)/\alpha}$$

имеет п. р. в. $\mathrm{Gam}(\tau + n\bar{x},\ (n\alpha + 1)/\alpha)$.

Аналогично если $X_i \sim \mathrm{Exp}(\lambda)$ и $\pi(\lambda) \propto \lambda^{\tau-1} e^{-\lambda\alpha}$, то апостериорное распределение

$$\pi(\lambda \mid \boldsymbol{x}) \propto \lambda^{n + \tau - 1} e^{-\lambda(n\bar{x} + \alpha)}$$

опять имеет п. р. в. $\mathrm{Gam}(\tau + n,\ n\bar{x} + \alpha)$. □

Пример 1.9.4. Нормальные распределения также проявляют себя как сопряженное семейство. А именно, предположим, что $X_i \sim \mathrm{N}(\mu, \sigma^2)$, где σ^2 известно и $\pi(\mu) \propto \exp(-(\mu - a)^2/(2b^2))$ с некоторыми известными $a \in \mathbb{R}$ и $b > 0$. Тогда, воспользовавшись решениями задач 2.3.8 и 2.3.9 из части А,

[1] Ср. с названием фильма «The Unbearable Lightness of Being» («Невыносимая легкость бытия»).

получим, что апостериорная плотность $\pi(\mu \mid \boldsymbol{x})$ пропорциональна

$$\propto \exp\left[-\frac{1}{2}\left(\frac{n}{\sigma^2}+\frac{1}{b^2}\right)\mu^2+\left(\frac{a}{b^2}+\frac{n\bar{x}}{\sigma^2}\right)\mu\right] \propto \exp\left[-\frac{1}{2b_1^2}(\mu-a_1)^2\right],$$

где

$$b_1^2=\left(\frac{1}{b^2}+\frac{n}{\sigma^2}\right)^{-1} \quad \text{и} \quad a_1=b_1^2\left(\frac{a}{b^2}+\frac{n\bar{x}}{\sigma^2}\right). \qquad \square$$

Следующим шагом является введение *функции потерь* (ф. п.) для измерения потерь, к которым приводят ошибки при оценивании заданного параметра θ. Это такая функция $L(\theta, a)$, $\theta, a \in \Theta$, которая зависит от истинного значения параметра θ и от его предполагаемого значения a. Например, *квадратическая* ф. п. $L(\theta, a) = (\theta - a)^2$, ф. п., равная *абсолютной ошибке* $L(\theta, a) = |\theta - a|$, и т. д.

Рассмотрим *апостериорные средние потери*

$$R(\boldsymbol{x}, a) = \int_\Theta \pi(\theta \mid \boldsymbol{x}) L(\theta, a)\, d\theta \quad \text{или} \quad \sum_{\theta\in\Theta} \pi(\theta \mid \boldsymbol{x}) L(\theta, a) \tag{1.9.3}$$

при принятии значения a. Мы хотим выбрать значение $\hat{a} = \hat{a}(\boldsymbol{x})$, минимизирующее $R(\boldsymbol{x}, a)$:

$$\hat{a} = \arg\min_a R(\boldsymbol{x}, a). \tag{1.9.4}$$

Эта точка минимума $\hat{a}$ называется *оптимальной байесовской оценкой* или, кратко, оптимальной оценкой. (Некоторые авторы употребляют термин «оптимальная точечная оценка».) Для квадратической функции потерь

$$R(\boldsymbol{x}, a) = \int_\Theta \pi(\theta \mid \boldsymbol{x})(\theta - a)^2\, d\theta.$$

Продифференцировав, находим, что $R(\boldsymbol{x}, a)$ достигает минимума в точке

$$\hat{a} = \int_\Theta \theta\pi(\theta \mid \boldsymbol{x})\, d\theta,$$

которая представляет собой *апостериорное среднее* $\mathsf{E}(\theta \mid \boldsymbol{x})$. Более того, минимальная величина апостериорных средних потерь равна

$$\min_a R(\boldsymbol{x}, a) = R(\boldsymbol{x}, \hat{a}) = \int [\theta - \mathsf{E}(\theta \mid \boldsymbol{x})]^2 \pi(\theta \mid \boldsymbol{x})\, d\theta,$$

т. е. равна апостериорной дисперсии $\mathsf{E}[(\theta - \mathsf{E}(\theta \mid \boldsymbol{x}))^2 \mid \boldsymbol{x}]$. Для функции потерь, равной абсолютной ошибке,

$$R(\boldsymbol{x}, a) = \int_\Theta \pi(\theta \mid \boldsymbol{x})|\theta - a|\, d\theta = \int_{-\infty}^{a} \pi(\theta \mid \boldsymbol{x})(a - \theta)\, d\theta + \int_a^{\infty} \pi(\theta \mid \boldsymbol{x})(\theta - a)\, d\theta,$$

а это выражение минимизируется *апостериорной медианой*, т. е. значением $\hat{a}$, для которого

$$\int_{-\infty}^{a} \pi(\theta \,|\, \boldsymbol{x})\, d\theta = \int_{\hat{a}}^{\infty} \pi(\theta \,|\, \boldsymbol{x})\, d\theta = 1/2.$$

Важное замечание, немедленно вытекающее из изложенного, состоит в том, что в общем случае $\hat{a}(\boldsymbol{x})$ минимизирует также и безусловные средние потери на множестве *всех* оценок $d\colon \boldsymbol{x} \mapsto \Theta$. В данной ситуации полезно несколько изменить терминологию: оценку можно трактовать как *решающее правило* (вы наблюдаете выборку $\boldsymbol{x}$ и принимаете решение: значение параметра равно $d(\boldsymbol{x})$). При заданном значении θ и решающем правиле d величина

$$r(\theta, d) = \mathsf{E}_\theta L(\theta, d(\boldsymbol{X})) = \begin{cases} \sum\limits_{\boldsymbol{x}} L(\theta, d(\boldsymbol{x})) f(\boldsymbol{x}; \theta), \\ \int L(\theta, d(\boldsymbol{x})) f(\boldsymbol{x}; \theta)\, d\boldsymbol{x} \end{cases}$$

представляет собой *риск* при принятии решения d, если истинное значение параметра равно θ. Мы хотим минимизировать *байесовский риск*

$$r^B(d) = \int_\Theta r(\theta, d)\pi(\theta)\, d\theta \quad \text{или} \quad \sum_\Theta r(\theta, d)\pi(\theta). \tag{1.9.5}$$

Таким образом, наше замечание сводится к тому, что

$$\hat{a} = \arg\min_d r^B(d). \tag{1.9.6}$$

По этой причине оптимальная байесовская оценка называется также оптимальным правилом.

Формально в уравнении (1.9.4) находят минимизирующее значение при каждом заданном (фиксированном) $\boldsymbol{x}$, тогда как в уравнении (1.9.6) минимизируют сумму или интеграл, взятые по всем значениям $\boldsymbol{x}$. Важно понять, что обе процедуры приводят к одинаковому ответу. В вышеприведенном замечании мы утверждаем, что точка минимума в уравнении (1.9.4) будет точкой минимума и для $r^B(d)$. А что можно сказать об обратном утверждении? Будет ли каждое решающее правило, минимизирующее $r^B(d)$, совпадать с оптимальной байесовской оценкой? Оказывается, и это верно: меняя порядок суммирования/интегрирования по $\boldsymbol{x}$ и θ, запишем

$$r(\theta, d) = \int \mathsf{E}_{\pi(\theta \,|\, \boldsymbol{x})} L(\theta, d(\boldsymbol{x})) \bar{f}(\boldsymbol{x})\, d\boldsymbol{x} \quad \text{или} \quad \sum_{\boldsymbol{x}} \mathsf{E}_{\pi(\theta \,|\, \boldsymbol{x})} L(\theta, d(\boldsymbol{x})) \bar{f}(\boldsymbol{x}). \tag{1.9.7}$$

Здесь $\bar{f}(\boldsymbol{x})$ — это маргинальная п. р. в./д. ф. р. вектора $\boldsymbol{X}$:

$$\bar{f}(\boldsymbol{x}) = \sum_\Theta \pi(\theta) f(\boldsymbol{x}; \theta) = \int_\Theta \pi(\theta) f(\boldsymbol{x}; \theta)\, d\theta,$$

а $\pi(\theta \,|\, \boldsymbol{x})$ — это апостериорная п. р. в./д. ф. р. для θ при фиксированном $\boldsymbol{x}$.

Поскольку $\bar{f}(\boldsymbol{x}) \geqslant 0$, минимум по d суммы/интеграла в правой части равенства (1.9.7) может достигаться только тогда, когда суммируемые величины или подынтегральное выражение $\mathsf{E}_{\pi(\theta\,|\,\boldsymbol{x})}L(\theta,\ d(\boldsymbol{x}))$ достигают своего минимума по d. А это и означает, что минимизирующее решающее правило равно $\hat{a}$.

Пример 1.9.5. В этом примере мы вычислим байесовские оценки для некоторых моделей. Стандартным примером является случай, когда $X_i \sim \mathrm{N}(\mu,\ \sigma^2)$ и дисперсия σ^2 известна, а μ имеет априорное распределение $\mathrm{N}(0,\ \tau^2)$ с известным $\tau^2 > 0$. Для апостериорного распределения

$$\pi(\mu\,|\,\boldsymbol{x}) \propto \pi(\mu) f(\boldsymbol{x};\ \mu) \propto \exp\left[-\frac{1}{2\sigma^2}\sum_i (x_i-\mu)^2\right]\cdot\exp\left[-\frac{\mu^2}{2\tau^2}\right] \propto$$
$$\propto \exp\left[-\frac{1}{2}\left(\frac{n}{\sigma^2}+\tau^{-2}\right)\left(\mu - \frac{\sum\limits_i x_i/\sigma^2}{n/\sigma^2+\tau^{-2}}\right)^2\right].$$

Это означает, что

$$\pi(\mu\,|\,\boldsymbol{x}) \sim \mathrm{N}\left(\frac{\sum\limits_i x_i/\sigma^2}{n/\sigma^2+1/\tau^2},\ \frac{1}{n/\sigma^2+1/\tau^2}\right).$$

Среднее значение нормального распределения равно его медиане. Таким образом, для обеих ф. п., и квадратической, и абсолютной ошибки, получаем, что оптимальная байесовская оценка для μ равна

$$\frac{\sum\limits_i x_i/\sigma^2}{n/\sigma^2+1/\tau^2}.$$

Если априорным распределением является $\mathrm{N}(\nu,\ \tau^2)$ с произвольным средним ν, то байесовская оценка для μ для обеих ф. п., и квадратической, и абсолютной ошибки, равна

$$\frac{\nu/\tau^2+\sum\limits_i x_i/\sigma^2}{n/\sigma^2+1/\tau^2}. \qquad \square$$

Пример 1.9.6. Далее, пусть $X_i \sim \mathrm{Po}(\lambda)$, где априорное распределение для λ является показательным с известной интенсивностью $\tau > 0$. Апостериорное распределение

$$\pi(\lambda\,|\,\boldsymbol{x}) \propto e^{-n\lambda}\lambda^{\sum_i x_i} e^{-\tau\lambda},$$

т. е. это распределение $\mathrm{Gam}\left(\sum\limits_i x_i + 1,\ n+\tau\right)$. При квадратической ф. п. оптимальная байесовская оценка равна

$$\frac{\sum\limits_i x_i + 1}{n+\tau}.$$

С другой стороны, для ф. п., равной абсолютной ошибке, оптимальная байесовская оценка равна значению $\hat{\lambda} > 0$, при котором

$$\frac{(n+\tau)^{\sum x_i+1}}{\left(\sum\limits_i x_i\right)!}\int\limits_0^{\hat{\lambda}} \lambda^{\sum\limits_i x_i} e^{-(n+\tau)\lambda}\, d\lambda = \frac{1}{2}. \qquad \square$$

Пример 1.9.7. Часто предполагается, что априорным распределением является равномерное распределение на заданном интервале. Например, пусть $X_1, \ldots, X_n$ — н. о. р. с. в. с равномерным распределением на интервале $(\theta - 1, \theta + 1)$, и пусть априорное распределение параметра θ является равномерным распределением на (a, b). Тогда апостериорное распределение имеет вид

$$\pi(\theta \mid \boldsymbol{x}) \propto I(a < \theta < b) I(\max x_i - 1 < \theta < \min x_i + 1) =$$
$$= I\big(a \vee (\max x_i - 1) < \theta < b \wedge (\min x_i + 1)\big),$$

где $\alpha \vee \beta = \max[\alpha, \beta]$, $\alpha \wedge \beta = \min[\alpha, \beta]$. Таким образом, апостериорное распределение является равномерным на этом интервале. Тогда квадратическая ф. п. и ф. п., равная абсолютной ошибке, приводят к одной и той же байесовской оценке

$$\hat{\theta} = \frac{1}{2}[a \vee (\max x_i - 1) + b \wedge (\min x_i + 1)].$$

В качестве еще одного примера рассмотрим параметр θ с равномерным распределением на $(0, 1)$, и пусть с. в. $X_1, \ldots, X_n$ принимают два значения 0 и 1 с вероятностями θ и $1 - \theta$. Здесь апостериорное распределение имеет вид

$$\pi(\theta \mid \boldsymbol{x}) \propto \theta^{\sum x_i}(1-\theta)^{n-\sum x_i},$$

т. е. является бета-распределением $\mathrm{Bet}\left(\sum\limits_i x_i + 1,\ n - \sum\limits_i x_i + 1\right)$:

$$\pi(\theta \mid \boldsymbol{x}) \propto \pi(\theta) f(\boldsymbol{x}; \theta) = \frac{\theta^{\sum x_i}(1-\theta)^{n-\sum x_i}}{\int\limits_0^1 \bar{\theta}^{\sum x_i}(1-\bar{\theta})^{n-\sum x_i}\, d\bar{\theta}}, \qquad 0 < \theta < 1.$$

Таким образом, для квадратических потерь $\ell(\theta, d) = (\theta - d)^2$ оптимальная байесовская оценка определяется как

$$\hat{d} = \frac{\int\limits_0^1 \theta^{t+1}(1-\theta)^{n-t}\, d\theta}{\int\limits_0^1 \theta^{t}(1-\theta)^{n-t}\, d\theta} = \frac{(t+1)!\,(n-t)!\,(n+1)!}{(n+2)!\,t!\,(n-t)!} = \frac{t+1}{n+2}.$$

Здесь $t := \sum_i x_i$, и мы использовали тождество

$$\int_0^1 x^{m-1}(1-x)^{n-1}\,dx = \frac{(m-1)!\,(n-1)!}{(m+n-1)!},$$

которое выполняется для всех целых m и n. □

Следующее замечание касается еще одного класса функций потерь, а именно *потерь типа* 0-1, когда $L(\theta, a) = 1 - \delta_{\theta,a}$, т. е. $L(\theta, a) = 0$ при $\theta = a$ и 1 в противном случае. Такую функцию естественно использовать в случае, когда множество Θ возможных значений параметра θ конечно. Например, предположим, что Θ состоит только из двух значений, скажем, 0 и 1. Пусть априорные вероятности равны p_0 и p_1 и соответствующие д. ф. р./п. р. в. равны f_0 и f_1. Апостериорные вероятности $\pi(\cdot\,|\,\boldsymbol{x})$ равны

$$\pi(0\,|\,\boldsymbol{x}) = \frac{p_0 f_0(\boldsymbol{x})}{p_0 f_0(\boldsymbol{x}) + p_1 f_1(\boldsymbol{x})}, \qquad \pi(1\,|\,\boldsymbol{x}) = \frac{p_1 f_1(\boldsymbol{x})}{p_0 f_0(\boldsymbol{x}) + p_1 f_1(\boldsymbol{x})},$$

и байесовская оценка $\hat{a}$, принимающая значения 0 и 1, должна минимизировать средние апостериорные потери $\pi(1 - a\,|\,\boldsymbol{x})$. Это означает, что

$$\hat{a}\ (= \hat{a}(\boldsymbol{x})) = \begin{cases} 0, & \text{если } \pi(0\,|\,\boldsymbol{x}) > \pi(1\,|\,\boldsymbol{x}), \\ 1, & \text{если } \pi(1\,|\,\boldsymbol{x}) > \pi(0\,|\,\boldsymbol{x}); \end{cases}$$

в случае, когда $\pi(0\,|\,\boldsymbol{x}) = \pi(1\,|\,\boldsymbol{x})$, любой выбор будет давать одно и то же значение 1/2 для средних апостериорных потерь. Другими словами,

$$\hat{a} = \begin{cases} 1, & \text{если } \dfrac{\pi(1\,|\,\boldsymbol{x})}{\pi(0\,|\,\boldsymbol{x})} = \dfrac{p_1 f_1(\boldsymbol{x})}{p_0 f_0(\boldsymbol{x})} > 1, \text{ т. е.} \dfrac{f_1(\boldsymbol{x})}{f_0(\boldsymbol{x})} > k = \dfrac{p_0}{p_1}, \\ 0 & \text{в противном случае.} \end{cases} \tag{1.9.8}$$

Здесь появляется «отношение правдоподобия» $f_1(\boldsymbol{x})/f_0(\boldsymbol{x})$; мы будем встречать его неоднократно и в следующей главе.

Для завершения темы байесовского оценивания рассмотрим следующую модель. Пусть с. в. $X_1, \ldots, X_n$ таковы, что $X_i \sim \mathrm{Po}(\theta_i)$:

$$f(x;\,\theta_i) = \frac{\theta_i^x}{x!} e^{-\theta_i}, \qquad x = 0, 1, \ldots$$

Здесь параметр θ_i сам является случайным и имеет ф. р. B (на $\mathbb{R}_+ = (0, \infty)$). Мы хотим оценить θ_i. Классическим приложением этой модели является случай, когда X_i — это число дорожных происшествий за указанный год, в которые был вовлечен водитель i.

Пусть вначале нам известна функция B, тогда оценка $T_i(X_1, \ldots, X_n)$ параметра θ_i, минимизирующая среднеквадратическую ошибку $\mathsf{E}(T_i - \theta_i)^2$, не зависит от X_j при $j \neq i$:

$$T_i(X_1, \ldots, X_n) = T(X_i).$$

Здесь

$$T(x) = \int \theta f(x;\, \theta)\, dB(\theta) \Big/ g(x)$$

и

$$g(x) = \int f(x;\, \theta)\, dB(\theta).$$

Подставляя пуассоновское распределение вместо $f(x;\, \theta)$, получаем

$$T(x) = (x+1)\frac{g(x+1)}{g(x)}, \quad x = 0,\, 1,\, \ldots$$

Следовательно,

$$T_i(X_1,\, \ldots,\, X_n) = (X_i + 1)\frac{g(X_i + 1)}{g(X_i)}. \tag{1.9.9}$$

А что будет в случае, когда мы не знаем функцию B и ее необходимо оценить на основании выборки $(X_1,\, \ldots,\, X_n)$?

Естественный ответ — заменить B ее выборочной гистограммой

$$\hat{B}(\theta) = \frac{1}{n}\,\mathrm{card}\{i\colon X_i < \theta\}, \quad \theta \geqslant 0,$$

т. е. функцией, имеющей скачки величины $\hat{b}(x) = n_x/n$ в целых точках x, где

$$n_x = \mathrm{card}\{i\colon X_i = x\}, \quad x = 0,\, 1,\, \ldots \tag{1.9.10}$$

Затем подставим $\hat{b}(x)$ вместо $g(x)$. Это приводит к следующей оценке для θ_i:

$$\hat{T}_i(X_1,\, \ldots,\, X_n) = (X_i + 1)\frac{n_{X_i+1}}{n_{X_i}}. \tag{1.9.11}$$

Эта идея (которую некоторые источники приписывают британскому математику А. Тьюрингу (1912—1954)) работает на удивление хорошо (см. [E2], [LS]). Удивляет это потому, что оценка $\hat{T}_i$ построена с использованием наблюдений X_j, $j \neq i$, т. е. без привлечения информации относительно X_i (за исключением, разумеется, того факта, что все наблюдения имеют пуассоновское распределение и одну и ту же априорную ф. р.).

Это наблюдение послужило исходным пунктом для так называемой эмпирической байесовской методологии, развитой Г. Роббинсом (1915—2001), еще одним выдающимся американским статистиком, который начинал свою карьеру как чистый математик. Подобно Дж. Тьюки, Роббинс написал свою докторскую диссертацию в области топологии. В 1941 г. вместе с Р. Курантом он опубликовал книгу [CouR], которая и по сей день необходима всем тем, кто интересуется математикой. В 1955 г. Роббинс напечатал статью [Rob], где он обосновал оценку $\frac{1}{12n+1} < \theta(n) < \frac{1}{12n}$ для остаточного члена в формуле Стирлинга; см. формулу (1.6.20) в части А.

Роббинсу приписывают также множество афоризмов и шуток (одна из них та, что «ни одно доброе дело не остается безнаказанным»).

Роббинс использовал формулу (1.9.11) для получения надежной оценки числа S_0 происшествий, которые произойдут в следующем году с теми n_0 водителями, у которых не было ни одной аварии за данный наблюдаемый год. Ясно, что число 0 было бы здесь недооценкой для числа S_0 (поскольку это предполагает, что будущее идентично прошлому). С другой стороны, число $n_0 = \sum\limits_{i\in\mathbb{Z}_+} in_i$ является переоценкой, поскольку X_0 не вносит никакого вклада в сумму $\sum\limits_i in_i$. Хорошей оценкой для S_0 является X_1, т. е. число водителей, для которых зафиксировано только одно происшествие. В общем случае величина $(i+1)n_{i+1}$ является точной оценкой для числа S_i происшествий, которые произойдут в будущем году с n_i водителями, для которых зарегистрировано i происшествий в данном наблюдаемом году.

Роббинс ввел понятие «число пропущенных объектов». Например, можно подсчитать число n_x слов, использовавшихся в точности x раз в известном нам собрании сочинений Шекспира, $x = 1, 2, \ldots$; эти данные приведены в следующей таблице (см. [ETh1]).

$x \setminus n_x$	1	2	3	4	5	6	7	8	9	10
0+	14376	4343	2292	1463	1043	837	638	519	430	364
10+	305	259	242	223	187	181	179	130	127	128
20+	104	105	99	112	93	74	83	76	72	63

Таким образом, 14376 слов появились лишь по одному разу, 4343 слов появились дважды, и т. д. А 2387 слов появлялись более чем по 30 раз каждое. Общее число всех различных слов, использовавшихся в этих произведениях, равно 31534 (при этом, например, для слов «tree» (дерево) и «trees» (деревья) число их появлений подсчитывалось отдельно). Пропущенные объекты здесь — это слова, которые Шекспир знал, но не использовал; пусть их число равно n_0. Тогда общее число слов, которые знал Шекспир, равно, очевидно, $n = 31534 + n_0$. Длина всего собрания сочинений (т. е. число всех слов, подсчитанных столько раз, сколько они появлялись в тексте) равна $N = 884647$.

Предположим, как и ранее, что число появлений слова i в текстах всех произведений $X_i \sim \mathrm{Po}(\theta_i)$, где параметр θ_i является случайным. И вновь, если ф. р. B параметра θ известна, то апостериорное среднее значение $\mathsf{E}(\theta_i \mid X_i = x)$ для θ_i дает оценку, минимизирующую среднеквадратическую ошибку, и равно

$$(x+1)\frac{g(x+1)}{g(x)}, \tag{1.9.12}$$

где

$$g(x) = \int \frac{\theta^x}{x!} e^{-\theta}\, dB(\theta), \qquad x = 0, 1, \ldots$$

Если B неизвестно, то мы подставляем (пока неизвестную) гистограмму

$$\hat{B}(\theta) = \frac{1}{n}\,\mathrm{card}\{i\colon \theta_i < \theta\}, \qquad \theta \geqslant 0.$$

Оценка для θ_i тогда принимает вид

$$\hat{\mathsf{E}}(\theta_i \mid X_i = x) = (x+1)\frac{n_{x+1}}{n_x}. \tag{1.9.13}$$

Для $x = 1$ это дает следующее значение для математического ожидания:

$$\hat{\mathsf{E}}(\theta_i \mid X_i = 1) = 2 \cdot \frac{4343}{14376} = 0{,}604.$$

Отсюда мы немедленно заключаем, что слова, появлявшиеся лишь один раз, здесь представлены избыточно, в том смысле, что если-бы где-либо существовало новое собрание сочинений Шекспира, равное по объему рассмотренному, то 14376 слов, которые появляются только один раз в рассмотренном случае, в новом собрании сочинений появятся всего лишь $0{,}604 \cdot 14376 = 8683$ раз.

Далее, положим

$$r_0 = \sum_{i=1}^{n} \theta_i I(X_i = 0) \Big/ \sum_{i=1}^{n} \theta_i.$$

Числитель оценим выражением

$$\hat{\mathsf{E}}(\theta_i \mid X_i = 0) n_0 = \frac{n_1}{n_0} n_0 = n_1 = 14376,$$

а знаменатель — числом $N = 884647$. Тогда

$$\hat{r}_0 = \frac{14376}{884647} = 0{,}016.$$

Таким образом, полет фантазии приводит к заключению, что если бы появился новый шекспировский текст, то вероятность того, что первое слово из этого текста отсутствовало в уже существовавших ранее текстах, равна 0,016; это же утверждение верно и для второго слова, и т. д. В 1985 г. Боделианская библиотека в Оксфорде действительно объявила, что найдена прежде неизвестная поэма, авторство которой некоторые эксперты возводят к Шекспиру. Вышеизложенный анализ был применен к этому случаю (см. [ETh2]) и привел к интересным заключениям.

Таблица 1.1. Процентные точки распределения t_n

n	0,995	0,99	0,975	0,95
1	63,66	31,82	12,71	6,31
2	9,92	6,96	4,30	2,92
3	5,84	4,54	3,18	2,35
4	4,60	3,75	2,78	2,13
5	4,03	3,36	2,57	2,02
6	3,71	3,14	2,45	1,94
7	3,50	3,00	2,36	1,89
8	3,36	2,90	2,31	1,86
9	3,25	2,82	2,26	1,83
10	3,17	2,76	2,23	1,81
11	3,11	2,72	2,20	1,80
12	3,05	2,68	2,18	1,78
13	3,01	2,65	2,16	1,77
14	2,98	2,62	2,14	1,76
15	2,95	2,60	2,13	1,75
16	2,92	2,58	2,12	1,75
17	2,90	2,57	2,11	1,74
18	2,88	2,55	2,10	1,73
19	2,86	2,54	2,09	1,73
20	2,85	2,53	2,09	1,72
21	2,83	2,52	2,08	1,72
22	2,82	2,51	2,07	1,72
23	2,81	2,50	2,07	1,71
24	2,80	2,49	2,06	1,71
25	2,79	2,49	2,06	1,71
26	2,78	2,48	2,06	1,71
27	2,77	2,47	2,05	1,70
28	2,76	2,47	2,05	1,70
29	2,76	2,46	2,05	1,70
30	2,75	2,46	2,04	1,70
40	2,70	2,42	2,02	1,68
60	2,66	2,39	2,00	1,67
120	2,62	2,36	1,98	1,66

Таблица 1.2. Процентные точки распределения N(0, 1)

0,995	0,99	0,975	0,95	0,90
2,58	2,33	1,96	1,645	1,282

Таблица 1.3. Процентные точки распределения χ^2_n

n	0,99	0,975	0,95	0,9
1	6,63	5,02	3,84	2,71
2	9,21	7,38	5,99	4,61
3	11,34	9,35	7,81	6,25
4	13,28	11,14	9,49	7,78
5	15,09	12,83	11,07	9,24
6	16,81	14,45	12,59	10,64
7	18,48	16,01	14,07	12,02
8	20,09	17,53	15,51	13,36
9	21,67	19,02	16,92	14,68
10	23,21	20,48	18,31	15,99
11	24,73	21,92	19,68	17,28
12	26,22	23,34	21,03	18,55
13	27,69	24,74	22,36	19,81
14	29,14	26,12	23,68	21,06
15	30,58	27,49	25,00	22,31
16	32,00	28,85	26,30	23,54
17	33,41	30,19	27,59	24,77
18	34,81	31,53	28,87	25,99
19	36,19	32,85	30,14	27,20
20	37,57	34,17	31,41	28,41
30	50,89	46,98	43,77	40,26
40	63,69	59,34	55,76	51,81
50	76,15	71,42	67,50	63,17
60	88,38	83,30	79,08	74,40
70	100,4	95,02	90,53	85,53
80	112,3	106,6	101,8	96,58
90	124,1	118,1	113,1	107,5
100	135,8	129,5	124,3	118,5

Таблица 1.4. 95%-точки распределения $F_{m,n}$

n \ m	1	2	3	4	5	6	8	12	16	20	30	40	50
1	161,4	199,5	215,7	224,5	230,1	233,9	238,8	243,9	246,4	248,0	250,1	251,1	251,7
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,43	19,45	19,46	19,47	19,48
3	10,13	9,55	9,28	9,12	9,01	8,94	8,85	8,74	8,69	8,66	8,62	8,59	8,58
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,84	5,80	5,75	5,72	5,70
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,60	4,56	4,50	4,46	4,44
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,92	3,87	3,81	3,77	3,75
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,49	3,44	3,38	3,34	3,32
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,20	3,15	3,08	3,04	3,02
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,99	2,94	2,86	2,83	2,80
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,83	2,77	2,70	2,66	2,64
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,70	2,65	2,57	2,53	2,51
12	4,75	3,89	3,49	3,26	3,11	3,00	2,85	2,69	2,60	2,54	2,47	2,43	2,40
13	4,67	3,81	3,41	3,18	3,03	2,92	2,77	2,60	2,51	2,46	2,38	2,34	2,31
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,44	2,39	2,31	2,27	2,24
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,38	2,33	2,25	2,20	2,18
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,33	2,28	2,19	2,15	2,12
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,29	2,23	2,15	2,10	2,08
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,25	2,19	2,11	2,06	2,04
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,21	2,16	2,07	2,03	2,00
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,18	2,12	2,04	1,99	1,97
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,13	2,07	1,98	1,94	1,91
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	2,09	2,03	1,94	1,89	1,86
26	4,23	3,37	2,98	2,74	2,59	2,47	2,32	2,15	2,05	1,99	1,90	1,85	1,82
28	4,20	3,34	2,95	2,71	2,56	2,45	2,29	2,12	2,02	1,96	1,87	1,82	1,79
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,99	1,93	1,84	1,79	1,76
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,90	1,84	1,74	1,69	1,66
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	1,95	1,85	1,78	1,69	1,63	1,60
60	4,00	3,15	2,76	2,53	2,37	2,25	2,10	1,92	1,82	1,75	1,65	1,59	1,56
70	3,98	3,13	2,74	2,50	2,35	2,23	2,07	1,89	1,79	1,72	1,62	1,57	1,53
80	3,96	3,11	2,72	2,49	2,33	2,21	2,06	1,88	1,77	1,70	1,60	1,54	1,51
100	3,94	3,09	2,70	2,46	2,31	2,19	2,03	1,85	1,75	1,68	1,57	1,52	1,48

Глава 2

Проверка гипотез

§ 2.1. Вероятности ошибок I и II рода. Наиболее мощные критерии

Статистики делают это с вероятностью лишь 5% быть отвергнутыми.

(Из серии «Как они делают это».)

Проверка статистических гипотез, или просто *проверка гипотез*, является еще одним способом сделать заключение о распределении (п. р. в. или д. ф. р.) «наблюдаемой» случайной величины X или выборки $\boldsymbol{X} = (X_1, \ldots, X_n)$. По сложившейся здесь традиции говорят о *нулевой* и *альтернативной гипотезах*. В простейшем случае нам необходимо выбрать одну из двух возможностей: п. р. в./д. ф. р. f величины X равна f_0 или f_1. Мы говорим, что утверждение $f = f_0$ представляет собой *простую нулевую гипотезу*, а $f = f_1$ — *простая альтернатива*. При таком подходе изначально имеется некоторое неравноправие между f_0 и f_1, которое будет проявляться также и в последующем.

Предположим, что наблюдаемое значение с. в. X равно x. «Научный» подход заключается в том, чтобы разбить множество значений величины X (пусть это будет $\mathbb{R}$) на две непересекающиеся области C и $\bar{C} = \mathbb{R} \setminus C$ и отвергать гипотезу H_0 (т. е. принимать гипотезу H_1), когда $x \in C$, а принимать гипотезу H_0 в том случае, когда $x \in \bar{C}$. Следовательно, критерий проверки гипотезы определяется областью C (которая называется *критической областью*). Другими словами, для проверки гипотез мы хотим применить решающую функцию d с двумя значениями (индикатор множества C), и при $d(x) = 1$ будем отклонять нулевую гипотезу, а при $d(x) = 0$ будем принимать эту гипотезу.

Предположим, что беспокоиться нам нужно прежде всего о том, чтобы не отвергнуть гипотезу H_0 (т. е. чтобы не принять H_1) в случае, когда H_0 верна: мы полагаем, что это представляет собой главную опасность. В этом случае (когда гипотеза H_0 верна, но отклоняется) мы говорим, что совершена ошибка I *рода*, и мы хотели бы, чтобы ошибки такого рода были довольно редки. Менее опасной была бы ошибка II *рода*: принять гипотезу H_0 в случае, когда она ложна (т. е. справедлива гипотеза H_1): мы хотим, чтобы такие ошибки появлялись в разумных пределах, если уже есть гарантии того, что ошибка I рода непременно будет совершаться редко (что является нашей главной целью). Формально мы фиксируем некоторое пороговое значение для *вероятности ошибки* I *рода* (в. о. р. I), $\alpha \in (0, 1)$, и затем пытаемся минимизировать *вероятность ошибки* II *рода* (в. о. р. II).

В такой ситуации часто говорят, что H_0 — это *консервативная* гипотеза, которая отвергается лишь при наличии явных доказательств против нее.

Например, если вы врач и изучаете гистологические данные своего пациента, то вы рассматриваете гипотезу о том, что у этого пациента имеется опухоль, при этом альтернативой является гипотеза о том, что опухоли нет. При выполнении гипотезы H_1 (опухоли нет) данные должны быть сгруппированы вблизи «нормальных» средних значений, тогда как при гипотезе H_0 (опухоль есть) данные смещаются к некоторым «анормальным» значениям. Принимая во внимание то, что при тестировании точность измерений не достигает 100%, данные следует считать случайными. Вы хотите разработать научный метод диагностики заболевания. И ваша первостепенная задача — не пропустить пациента с опухолью, поскольку это может иметь серьезные последствия. С другой стороны, если вы совершите ошибку II рода (поднимете ложную тревогу по поводу здорового пациента), то это может привести лишь к некоторым относительно мелким неприятностям (повторное тестирование, возможная короткая госпитализация).

Дальнейший вопрос заключается в том, как же выбрать *критическую область* C. При заданной области C в. о. р. I равна

$$\mathrm{P}_0(C) = \int f_0(x) I(x \in C)\, dx \quad \text{или} \quad \sum_x f_0(x) I(x \in C) \tag{2.1.1}$$

и называется *размером* критической области C (а также размером критерия). Требование, чтобы вероятность $\mathrm{P}_0(C)$ не превосходила заданного значения α (называемого *уровнем значимости*), не определяет область C единственным образом: мы можем найти много таких областей. На интуитивном уровне понятно, что область C должна содержать исходы x, для которых значения $f_0(x)$ малы, независимо от местоположения x. Но нам хотелось бы достичь большей точности (определенности). Например, если п. р. в. f_0, соответствующая гипотезе H_0, является нормальной $\mathrm{N}(\mu_0, \sigma_0^2)$-плотностью, то может ли C содержать точки, лежащие вблизи среднего значения μ_0, в которых п. р. в. f_0 достигает относительно высоких значений? Или же область C должна быть полубесконечным интервалом $(\mu_0 + c, \infty)$, или $(-\infty, \mu_0 - c)$, или, возможно, их объединением? Например, мы можем выбрать c таким, чтобы интегралы

$$\frac{1}{\sqrt{2\pi}\sigma_0}\int\limits_{\mu_0+c}^{\infty} e^{-(x-\mu_0)^2/2\sigma_0^2}\,dx \quad \text{или} \quad \frac{1}{\sqrt{2\pi}\sigma_0}\int\limits_{-\infty}^{\mu_0-c} e^{-(x-\mu_0)^2/2\sigma_0^2}\,dx$$

или их сумма не превышали α. На самом деле, оказывается, именно альтернативная п. р. в./д. ф. р. f_1 позволяет сузить возможности для выбора C (и определить эту область по сути однозначно), поскольку мы стремимся минимизировать в. о. р. II

$$\mathrm{P}_1(\bar{C}) = \int f_1(x) I(x \notin C)\,dx \quad \text{или} \quad \sum_x f_1(x) I(x \notin C) \tag{2.1.2}$$

при заданном уровне значимости. Вероятность дополнения

$$\mathrm{P}_1(C) = \int f_1(x) I(x \in C)\,dx \quad \text{или} \quad \sum_x f_1(x) I(x \in C) \tag{2.1.3}$$

называется *мощностью* критерия; ее необходимо максимизировать.

Критерий, имеющий максимальную мощность среди всех критериев размера не выше α, называют *наиболее мощным* (н. м.) критерием для заданного уровня значимости α. В разговорной речи такой критерий называют н. м. критерием размера α (поскольку, как мы покажем далее, его размер и в самом деле равен α).

§ 2.2. Критерии отношения правдоподобия. Лемма Неймана—Пирсона и комментарии к ней

> Статистики нуждаются в верных выводах, поскольку во времена их молодости многие из их гипотез были отвергнуты.
>
> (Из серии «Почему их не понимают».)

Одна естественная (и элегантная) идея проверки гипотез состоит в рассмотрении *отношения правдоподобия* (о. п.):

$$\Lambda_{H_1,H_0}(x) = \frac{f_1(x)}{f_0(x)}. \tag{2.2.1}$$

Если для заданного x о. п. велико, то мы склоняемся к тому, чтобы считать гипотезу H_0 маловероятной, т. е. к тому, чтобы отклонить H_0; в противном случае мы не отклоняем H_0. Следующая идея, которая приходит на ум, — это идея, что следует рассматривать области вида

$$\{x\colon \Lambda_{H_1,H_0}(x) > k\} \tag{2.2.2}$$

и выбирать k так, чтобы удовлетворить такому требованию: размер критерия равен α. Скаляр $x \in \mathbb{R}$ здесь может быть заменен на вектор

$\mathbf{x}=(x_1, \ldots, x_n)\in\mathbb{R}^n$ с п. р. в. $f_0(\mathbf{x})=\prod f_0(x_i)$ и $f_1(\mathbf{x})=\prod f_1(x_i)$. Критическая область C в этом случае превращается в подмножество $\mathbb{R}^n$.

Такая идея работает в основном неплохо, что показывает известная лемма Неймана—Пирсона (лемма Н—П), приводимая ниже. Эта лемма названа в честь Ю. Неймана и Е. С. Пирсона (1895—1980), выдающегося британского статистика. Пирсон был сыном К. Пирсона (1857—1936), которого принято считать создателем современного статистического мышления. Оба они, и отец, и сын, оказали огромное влияние на статистическую литературу того времени; их имена неоднократно будут появляться в этой части книги.

Интересно сравнить жизненные пути авторов леммы Н—П. Пирсон провел всю свою активную жизнь в Лондонском университетском колледже, где был последователем своего отца. Нейман, напротив, жил в период гражданских волнений, революции и войны на территории Российской империи (см. [Rei]). В 1920 г. на Украине он был брошен в тюрьму как поляк и ожидал сурового приговора (в то время шла война с Польшей, и его подозревали в шпионаже). Он был спасен в результате длительных переговоров его жены с одним большевистским чиновником, который в прошлом был студентом медицины под руководством тестя Неймана; в конце концов ему предоставили возможность бежать в Польшу. В 1925 г. Нейман прибыл в Лондон и начал работать с Е. С. Пирсоном. Одним из их совместных результатов была лемма Н—П.

О своем сотрудничестве с Пирсоном Нейман писал: «Инициатором совместных исследований был Пирсон. К тому же, по крайней мере на ранних этапах, он был лидером. Наша совместная работа велась посредством обширной переписки и во время нечастых встреч, некоторые из которых были в Англии, некоторые — во Франции, еще некоторые — в Польше. Это сотрудничество продолжалось в течение десяти лет с 1928 по 1938 год».

Постановка задачи в лемме Н—П состоит в следующем: предположим, что необходимо проверить гипотезу H_0: $f=f_0$ против альтернативы H_1: $f=f_1$, где f_1 и f_0 — это две различные п. р. в./д. ф. р. Лемма Н—П утверждает следующее.

Для любого $k>0$ критерий с критической областью $C=\{\mathbf{x}\colon f_1(\mathbf{x})>kf_0(\mathbf{x})\}$ имеет наибольшую мощность $\mathrm{P}_1(C)$ среди всех критериев (т. е. критических областей C^*) размера $\mathrm{P}_0(C^*)\leqslant\alpha=\mathrm{P}_0(C)$.

Другими словами, критерий с областью $C=\{f_1(\mathbf{x})>kf_0(\mathbf{x})\}$ является н. м. критерием среди всех критериев с уровнем значимости $\alpha=\mathrm{P}_0(C)$. Здесь α оказывается функцией от k: $\alpha=\alpha(k)$, $k>0$.

Доказательство. Обозначая через I_C и I_{C^*} индикаторные функции для C и C^*, запишем

$$0\leqslant(I_C(\mathbf{x})-I_{C^*}(\mathbf{x}))(f_1(\mathbf{x})-kf_0(\mathbf{x})),$$

поскольку выражения в скобках никогда не будут иметь противоположных знаков (при $\mathbf{x}\in C$ выполняются неравенства $f_1(\mathbf{x})>kf_0(\mathbf{x})$ и $I_C(\mathbf{x})\geqslant I_{C^*}(\mathbf{x})$, тогда как для $\mathbf{x}\notin C$ мы имеем $f_1(\mathbf{x})\leqslant kf_0(\mathbf{x})$ и $I_C(\mathbf{x})\leqslant I_{C^*}(\mathbf{x})$). Тогда

$$0\leqslant\int(I_C(\mathbf{x})-I_{C^*}(\mathbf{x}))(f_1(\mathbf{x})-kf_0(\mathbf{x}))\,d\mathbf{x}$$

или

$$0\leqslant\sum_{\mathbf{x}}(I_C(\mathbf{x})-I_{C^*}(\mathbf{x}))(f_1(\mathbf{x})-kf_0(\mathbf{x})).$$

Но правая часть равна

$$\mathrm{P}_1(C) - \mathrm{P}_1(C^*) - k(\mathrm{P}_0(C) - \mathrm{P}_0(C^*)).$$

Таким образом,

$$\mathrm{P}_1(C) - \mathrm{P}_1(C^*) \geqslant k(\mathrm{P}_0(C) - \mathrm{P}_0(C^*)).$$

Следовательно, если $\mathrm{P}_0(C^*) \leqslant \mathrm{P}_0(C)$, то $\mathrm{P}_1(C^*) \leqslant \mathrm{P}_1(C)$. □

Критерий с критической областью $C = \{x\colon f_1 x) > kf_0(x)\}$ часто называют *критерием отношения правдоподобия* (*к. о. п.*) или *критерием Неймана—Пирсона* (Н—П).

Замечания. 1. Утверждение леммы Н—П остается в силе, если неравенство $f_1(x) > kf_0(x)$ заменить неравенством $f_1(x) \geqslant kf_0(x)$.

2. На практике нам необходимо решать «обратную» задачу: для заданного $\alpha \in (0, 1)$ мы хотим найти н. м. критерий размера не больше α. Это означает, что мы хотим построить k как функцию от α, а не как-либо иначе. Во всех (тщательно подобранных) примерах этой главы это сделать несложно, поскольку функция $k \mapsto \alpha(k)$ является взаимно однозначной и допускает прямое обращение $k = k(\alpha)$. Значит, для заданного α всегда можно найти такое $k > 0$, что $\mathrm{P}(C) = \alpha$ для $C = \{f_1(x) > kf_0(x)\}$ (и отыскание такого k является частью задачи).

Однако во многих примерах (особенно относящихся к дискретному случаю) критерий Н—П может не существовать для всех значений (размера) α. Чтобы обойти эту трудность, мы должны рассмотреть более общие *рандомизированные* критерии, у которых решающая функция d может принимать не только значения 0, 1, но также и промежуточные значения из интервала (0, 1). При этом если наблюдается значение x, то мы отклоняем гипотезу H_0 с вероятностью $d(x)$. (Фактически «рандомизированная» лемма Н—П гарантирует, что всегда будет существовать критерий Н—П, у которого d принимает не более чем три значения: 0, 1 и, возможно, еще одно значение между этими двумя.)

Действуя формально, мы хотим вначале расширить понятия в. о. р. I и в. о. р. II, что можно сделать прямым обобщением уравнений (2.1.1) и (2.1.2):

$$\mathrm{P}_0(d) = \int d(x)f_0(x)\,dx \quad \text{или} \quad \sum_x d(x)f_0(x) \tag{2.2.3}$$

и

$$\mathrm{P}_1(\bar{d}) = \int (1 - d(x))f_1(x)\,dx \quad \text{или} \quad \sum_x (1 - d(x))f_1(x). \tag{2.2.4}$$

Мощность критерия с решающей функцией d равна

$$\mathrm{P}_1(d) = \int d(x)f_1(x)\,dx \quad \text{или} \quad \sum_x d(x)f_1(x). \tag{2.2.5}$$

Затем, как и ранее, фиксируем $\alpha \in (0, 1)$ и, рассматривая все рандомизированные критерии d размера не больше α, ищем среди них тот, который максимизирует $\mathrm{P}_1(d)$.

Рандомизированный вариант леммы Н—П таков.

Для любой пары п. р. в./д. ф. р. f_0 и f_1 и любого $\alpha \in (0, 1)$ существуют единственное такое $k > 0$ и такое $\rho \in [0, 1]$, что для критерия вида

$$d_{\mathrm{NP}}(x) = \begin{cases} 1, & f_1(x) > kf_0(x), \\ \rho, & f_1(x) = kf_0(x), \\ 0, & f_1(x) < kf_0(x), \end{cases} \tag{2.2.6}$$

выполняется равенство $\mathrm{P}_0(d_{\mathrm{NP}}) = \alpha$. Этот критерий имеет наибольшую мощность среди всех рандомизированных критериев размера α:

$$\mathrm{P}_1(d_{\mathrm{NP}}) = \max[\mathrm{P}_1(d^*)\colon \mathrm{P}_0(d^*) \leqslant \alpha]. \tag{2.2.7}$$

Таким образом, d_{NP} является н. м. рандомизированным критерием размера не больше α для проверки гипотезы $H_0\colon f = f_0$ против альтернативы $H_1\colon f = f_1$.

Как и ранее, критерий d_{NP}, описанный в формуле (2.2.6), называют (рандомизированным) критерием Н—П. Мы хотим еще раз подчеркнуть, что константа ρ может совпадать (и часто совпадает) с 0 или 1 и в этом случае критерий d_{NP} становится нерандомизированным.

Доказательство рандомизированной леммы Н—П несколько длиннее, чем в нерандомизированном случае, но по-прежнему сохраняет свою элегантность. Предположим вначале, что мы имеем дело с непрерывным случаем и п. р. в. f_0 и f_1 таковы, что для любого $k > 0$ выполняется равенство

$$\int f_0(\boldsymbol{x}) I(f_1(\boldsymbol{x}) = kf_0(\boldsymbol{x}))\, d\boldsymbol{x} = 0. \tag{2.2.8}$$

В этом случае значение ρ будет равным 0.

Действительно, рассмотрим функцию

$$G\colon k \mapsto \int f_0(\boldsymbol{x}) I(f_1(\boldsymbol{x}) > kf_0(\boldsymbol{x}))\, d\boldsymbol{x}.$$

Такая функция $G(k)$ является монотонно невозрастающей по k, поскольку область интегрирования сужается с ростом k. Далее заметим, что функция G является непрерывной справа: $\lim\limits_{r \searrow k+} G(r) = G(k)$ для любого $k > 0$. В самом деле, при $r \searrow k+$ мы имеем

$$I(f_1(\boldsymbol{x}) > rf_0(\boldsymbol{x})) \nearrow I(f_1(\boldsymbol{x}) > kf_0(\boldsymbol{x})),$$

так как каждая точка $\boldsymbol{x}$, для которой $f_1(\boldsymbol{x}) > kf_0(\boldsymbol{x})$, попадет в конце концов в семейство (расширяющихся) областей $\{f_1(\boldsymbol{x}) > rf_0(\boldsymbol{x})\}$. Тогда сходимость интегралов является следствием стандартных теорем анализа. Более того, аналогичным образом можно доказать, что G имеет левосторонние пределы, т. е.

$$G(k-) = \lim_{l \nearrow k-} G(l)$$

существуют для любого $k > 0$ и равны

$$\int f_0(\boldsymbol{x}) I(f_1(\boldsymbol{x}) \geqslant k f_0(\boldsymbol{x}))\, d\boldsymbol{x}.$$

Если выполнено предположение (2.2.8), то разность

$$G(k-) - G(k) = \int f_0(\boldsymbol{x}) I(f_1(\boldsymbol{x}) = k f_0(\boldsymbol{x}))\, d\boldsymbol{x}$$

стремится к нулю и функция G является непрерывной. Наконец, мы замечаем, что

$$G(0+) = \lim_{k \to 0+} G(k) = 1, \qquad G(\infty) = \lim_{k \to \infty} G(k) = 0. \tag{2.2.9}$$

Следовательно, график функции G пересекает каждый уровень $\alpha \in (0, 1)$ в некоторой точке $k = k(\alpha)$ (возможно, не единственной). Тогда (нерандомизированный) критерий, у которого

$$d_{\mathrm{NP}}(\boldsymbol{x}) = \begin{cases} 1, & f_1(\boldsymbol{x}) > k f_0(\boldsymbol{x}), \\ 0, & f_1(\boldsymbol{x}) \leqslant k f_0(\boldsymbol{x}), \end{cases}$$

имеет размер $\mathrm{P}_0(d) = G(k) = \alpha$ и соответствует формуле (2.2.6) при $\rho = 0$.

Этот критерий является н. м. критерием с уровнем значимости α. Действительно, пусть d^* — это любой другой (возможно, рандомизированный) критерий размера $\mathrm{P}_0(d^*) \leqslant \alpha$. Мы вновь получаем, что для любого $\boldsymbol{x}$ выполняется неравенство

$$0 \leqslant (d_{\mathrm{NP}}(\boldsymbol{x}) - d^*(\boldsymbol{x}))(f_1(\boldsymbol{x}) - k f_0(\boldsymbol{x})), \tag{2.2.10}$$

поскольку если $d_{\mathrm{NP}}(\boldsymbol{x}) = 1$, то обе скобки неотрицательны, а если $d_{\mathrm{NP}}(\boldsymbol{x}) = 0$, то обе они положительны. Значит,

$$0 \leqslant \int (d_{\mathrm{NP}}(\boldsymbol{x}) - d^*(\boldsymbol{x}))(f_1(\boldsymbol{x}) - k(\alpha) f_0(\boldsymbol{x}))\, d\boldsymbol{x},$$

но здесь правая часть опять равна $\mathrm{P}_1(d) - \mathrm{P}_1(d^*) - k(\mathrm{P}_0(d) - \mathrm{P}_0(d^*))$. Отсюда следует, что

$$\mathrm{P}_1(d) - \mathrm{P}_1(d^*) \geqslant k(\mathrm{P}_0(d) - \mathrm{P}_0(d^*)), \quad \text{т. е.} \quad \mathrm{P}_1(d) \geqslant \mathrm{P}_1(d^*),$$

если $\mathrm{P}_0(d) \geqslant \mathrm{P}_0(d^*)$.

Мы теперь готовы перейти к общему случаю, т. е. отказаться от предположения (2.2.8). Опять рассмотрим функцию

$$G\colon k \mapsto \int f_0(\boldsymbol{x}) I(f_1(\boldsymbol{x}) > k f_0(\boldsymbol{x}))\, d\boldsymbol{x} \quad \text{или} \quad \sum_{\boldsymbol{x}} f_0(\boldsymbol{x}) I(f_1(\boldsymbol{x}) > k f_0(\boldsymbol{x})). \tag{2.2.11}$$

Она по-прежнему монотонно не возрастает по k, непрерывна справа и имеет левосторонние пределы. (Существует удобный термин французского происхождения «cadlag» (continue à droite, limits à gauche).) Также и уравнение (2.2.9) остается верным. Однако разность $G(k-) - G(k)$ может быть положительной, так как она равна интегралу или сумме

$$\int f_0(x) I(f_1(x) = k f_0(x))\, dx \quad \text{или} \quad \sum_x f_0(x) I(f_1(x) = k f_0(x)),$$

которые не обязательно стремятся к нулю. Значит, при заданном $\alpha \in (0, 1)$ можно только гарантировать существование такого $k = k(\alpha) > 0$, что

$$G(k-) \geqslant \alpha \geqslant G(k).$$

Если $G(k-) = G(k)$ (т. е. функция G непрерывна в точке k), то применимы предыдущие рассуждения. В противном случае положим

$$\rho = \frac{\alpha - G(k)}{G(k-) - G(k)}. \tag{2.2.12}$$

Тогда $\rho \in [0, 1]$ и мы можем задать критерий d_{NP} формулой (2.2.6). (Если $\rho = 0$ или 1, критерий d_{NP} является нерандомизированным.) См. рис. 2.1.

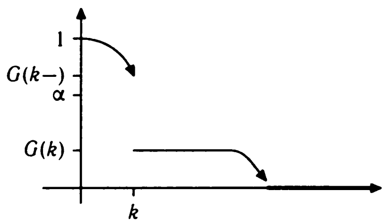


Рис. 2.1

Размер критерия

$$\mathrm{P}_0(d_{\mathrm{NP}}) = \int f_0(x) d_{\mathrm{NP}}(x)\, dx$$

равен

$$\int f_0(x) I(f_1(x) > k f_0(x))\, dx + \rho \int f_0(x) I(f_1(x) = k f_0(x))\, dx =$$
$$= G(k) + \frac{\alpha - G(k)}{G(k-) - G(k)} (G(k-) - G(k)) = \alpha.$$

Остается проверить, что d_{NP} является н. м. критерием размера не больше α. Это осуществляется так же, как и выше, поскольку неравенство (2.2.10) по-прежнему выполняется.

Имеет место полезное следствие рандомизированной леммы Н—П.

Лемма 1. *Если $d(\boldsymbol{x})$ является н. м. критерием размера $\alpha = \mathsf{P}_0(d)$, то его мощность $\beta = \mathsf{P}_1(d)$ не может быть менее α.*

Доказательство. Действительно, рассмотрим (рандомизированный) критерий, у которого $d^*(\boldsymbol{x}) \equiv \alpha$. Для него $\mathsf{P}_0(d^*) = \alpha = \mathsf{P}_1(d^*)$. Поскольку d является н. м. критерием, мы получаем $\beta \geqslant \mathsf{P}_1(d^*) = \alpha$. □

Критерий Н—П применим и для некоторых задач со *сложными* (т. е. не простыми) нулевыми гипотезами и альтернативами. Типичный пример сложной гипотезы: пусть $\Theta = \mathbb{R}$, H_0: $\theta \leqslant \theta_0$ и H_1: $\theta > \theta_0$ для некоторого заданного θ_0. Довольно часто одна из гипотез H_0 или H_1 является простой (например, H_0: $\theta = \theta_0$), а другая сложной ($\theta > \theta_0$, или $\theta < \theta_0$, или $\theta \neq \theta_0$).

В общем случае пара сложных гипотез задается как гипотеза H_0: $\theta \in \Theta_0$ против альтернативы H_1: $\theta \in \Theta_1$, где Θ_0 и Θ_1 являются непересекающимися подмножествами параметрического множества Θ. Для построения критерия нам необходимо опять выбрать такую критическую область $C \subset \mathbb{R}^n$, что гипотеза H_0 отклоняется, когда $\boldsymbol{x} \in C$. Как и в случае простых гипотез, вероятность

$$\mathsf{P}_\theta(C) = \mathsf{P}_\theta(\text{отклонить } H_0), \quad \theta \in \Theta_0, \tag{2.2.13}$$

трактуется как в. о. р. I. Аналогично вероятность

$$\mathsf{P}_\theta(\bar{C}) = \mathsf{P}_\theta(\text{принять } H_0), \quad \theta \in \Theta_1, \tag{2.2.14}$$

трактуется как в. о. р. II, а

$$\mathsf{P}_\theta(C) = \mathsf{P}_\theta(\text{отклонить } H_0), \quad \theta \in \Theta_1, \tag{2.2.15}$$

трактуется как мощность критерия с критической областью C. Здесь все эти величины являются функциями от параметра θ, принимающего значения либо в Θ_0, либо в Θ_1. (Уравнения (2.2.13) и (2.2.15) определяют фактически одну и ту же функцию $\theta \mapsto \mathsf{P}_\theta(C)$, рассматриваемую на паре взаимодополняющих друг друга множеств Θ_0 и Θ_1.)

Для постановки задачи мы используем ту же идею, что и ранее. А именно, фиксируем уровень значимости $\alpha \in (0, 1)$ и ищем такой критерий (т. е. критическую область), что

а) справедливо неравенство

$$\mathsf{P}_\theta(C) \leqslant \alpha, \quad \theta \in \Theta_0, \tag{2.2.16}$$

б) для любого критерия с критической областью C^*, удовлетворяющей условию (2.2.16), выполняется неравенство

$$\mathsf{P}_\theta(C) \geqslant \mathsf{P}_\theta(C^*), \quad \theta \in \Theta_1. \tag{2.2.17}$$

Такой критерий называется *равномерно наиболее мощным* (р. н. м.) критерием уровня α (для проверки гипотезы $H_0\colon \theta\in\Theta_0$ против альтернативы $H_1\colon \theta\in\Theta_1$).

Важным понятием, относящимся к данной теме, является понятие семейства п. р. в./д. ф. р. $f(\cdot;\theta)$, $\theta\in\Theta\subseteq\mathbb{R}$, имеющих *монотонное отношение правдоподобия* (м. о. п.). Для такого семейства мы требуем, чтобы для любой пары $\theta_1, \theta_2\in\Theta$, $\theta_1<\theta_2$, выполнялось соотношение

$$\Lambda_{\theta_2,\theta_1}(\boldsymbol{x})=\frac{f(\boldsymbol{x};\theta_2)}{f(\boldsymbol{x};\theta_1)}=g_{\theta_2,\theta_1}(T(\boldsymbol{x})), \tag{2.2.18}$$

где T — это действительнозначная статистика (т. е. скалярная функция, зависящая только от $\boldsymbol{x}$) и $g_{\theta_2,\theta_1}(y)$ — монотонно неубывающая функция ($g_{\theta_2,\theta_1}(y)\leqslant g_{\theta_2,\theta_1}(y')$ при $y\leqslant y'$). В этом определении мы также можем требовать, чтобы $g_{\theta_2,\theta_1}(y)$ являлась монотонно невозрастающей функцией от y; здесь важно только то, что вид монотонности должен быть одним и тем же для любой пары $\theta_1<\theta_2$.

Пример 2.2.1. *Гипергеометрическое распределение* $\mathrm{Hyp}(N, D, n)$. Предположим, вы рассматриваете множество из N изделий и выбираете из них для проверки n изделий случайным образом без возвращения, $n<N$. Если это множество содержит $D\leqslant N$ бракованных изделий, то число бракованных изделий в отобранной выборке имеет д. ф. р.

$$f_D(x)=\frac{C_D^xC_{N-D}^{n-x}}{C_N^n},\qquad x=\max[0, n+D-N], \ldots, \min[D, n]. \tag{2.2.19}$$

Здесь $\theta=D$. Отношение

$$\frac{f_{D+1}(x)}{f_D(x)}=\frac{D+1}{N-D}\,\frac{N-D-n+x}{D+1-x}$$

является монотонно возрастающим по x; следовательно, это семейство является семейством с м. о. п. и $T(x)=x$.

Гипергеометрическое распределение обладает рядом интересных свойств и находит применение в нескольких областях теоретической и прикладной теории вероятностей и статистике. В данной книге это распределение появляется только в этом примере. Однако мы приведем здесь полезное уравнение для производящей функции распределения $\mathsf{E}s^X$ с. в. $X\sim\mathrm{Hyp}(N, D, n)$:

$$\mathsf{E}s^X=\sum_{k=0}^{\min[n,D]}\frac{(-D)_k(-n)_k}{(-N)_k}\,\frac{(1-s)^k}{k!}.$$

В этой формуле $0\leqslant\max[n, D]\leqslant N$ и $(a)_k$ — это так называемый символ Покхаммера: $(a)_k=a(a+1)\ldots(a+k-1)$. Ряд в правой части может быть записан как ${}_2F_1(-D, -n; -N; 1-s)$ или ${}_2F_1\begin{pmatrix}-D, & -n \\ -N; & 1-s\end{pmatrix}$, где ${}_2F_1$ — гипергеометрическая функция Гаусса. В [GriS2] дается другой (элегантный) способ вычисления производящей функции распределения $\mathrm{Hyp}(N, D, n)$.

Как это ни удивительно, но во многих книгах область, на которой задается гипергеометрическое распределение (т. е. множество значений x, для которых $f_D(x) > 0$), представлена довольно противоречиво (что, возможно, забавно). В частности, левый конец $(n + D - N)_+ = \max[0, n + D - N]$ часто даже и не упоминается (или, что еще хуже, заменяется нулем). В других случаях правый и левый концы интервала значений x появляются в конце концов как результат рассуждений, и создается впечатление, что они получаются в процессе вычислений (что по сути неверно). □

Пример 2.2.2. *Биномиальное распределение* $\mathrm{Bin}(n, \theta)$. В этом примере изделие после проверки возвращается обратно. Тогда

$$f(x; \theta) = C_n^x \theta^x (1-\theta)^{n-x}, \qquad x = 0, 1, \ldots, n,$$

где $\theta = D/N \in [0, 1]$. Это семейство имеет м. о. п., и опять $T(x) = x$. □

Пример 2.2.3. В общем случае любое семейство п. р. в./д. ф. р. вида

$$f(x; \theta) = C(\theta)H(x)\exp[Q(\theta)R(x)],$$

где Q является строго возрастающей или строго убывающей функцией от θ, имеет м. о. п. Тогда

$$\frac{f(\boldsymbol{x}; \theta_2)}{f(\boldsymbol{x}; \theta_1)} = \frac{C(\theta_2)^n}{C(\theta_1)^n}\exp\left[(Q(\theta_2) - Q(\theta_1))\sum_i R(x_i)\right]$$

и правая часть монотонна по $T(\boldsymbol{x}) = \sum\limits_i R(x_i)$. В частности, семейство нормальных п. р. в. с фиксированной дисперсией имеет м. о. п. (относительно $\theta = \mu \in \mathbb{R}$), так же как и семейство нормальных п. р. в. с фиксированным средним (относительно $\theta = \sigma^2 > 0$). Еще одним примером является семейство показательных п. р. в. $f(x; \theta) = \theta e^{-\theta x} I(x \geqslant 0)$, $\theta > 0$. □

Пример 2.2.4. Пусть X — это с. в. и рассматривается нулевая гипотеза H_0: $X \sim \mathrm{N}(0, 1)$ против альтернативы H_1: $X \sim f(x) = \frac{1}{4}e^{-|x|/2}$ (двойное экспоненциальное распределение). Нас интересует н. м. критерий размера α, $\alpha < 0{,}3$.

В этом случае находим, что

$$f(x \mid H_1)/f(x \mid H_0) = C\exp\left(\frac{1}{2}(x^2 - |x|)\right).$$

Следовательно, $f(x \mid H_1)/f(x \mid H_0) > K$ тогда и только тогда, когда $x^2 - |x| > K'$, что, в свою очередь, выполняется тогда и только тогда, когда $|x| > t$ или $|x| < 1 - t$ для некоторого $t > 1/2$.

Мы хотим, чтобы значение $\alpha = \mathsf{P}_{H_0}(|X| > t) + \mathsf{P}_{H_0}(|X| < 1 - t)$ было меньше 0,3. Если $t \leqslant 1$, то

$$\mathsf{P}_{H_0}(|X| > t) \geqslant \mathsf{P}_{H_0}(|X| > 1) = 0{,}3174 > \alpha.$$

Таким образом, чтобы получить

$$\alpha = \mathsf{P}_{H_0}(|X| > t) = \Phi(-t) + 1 - \Phi(t),$$

нужно взять неравенство $t > 1$. Следовательно, $t = \Phi^{-1}(1 - \alpha/2)$, и критерий отклоняет гипотезу H_0, если $|X| > t$. Мощность критерия равна

$$\mathsf{P}_{H_1}(|X| > t) = 1 - \frac{1}{4}\int\limits_{-t}^{t} e^{-|x|/2}\,dx = e^{-t/2}. \qquad \square$$

Следующая теорема распространяет лемму Н—П на семейства с м. о. п. для случая *односторонней* нулевой гипотезы $H_0\colon \theta \leqslant \theta_0$ и односторонней альтернативы $H_1\colon \theta > \theta_0$.

Пусть $f(\boldsymbol{x}; \theta)$, $\theta \in \mathbb{R}$, — это семейство п. р. в./д. ф. р. с м. о. п. Зафиксируем $\theta_0 \in \mathbb{R}$, $k > 0$, и выберем любое $\theta_1 > \theta_0$. Тогда критерий проверки гипотезы H_0 против альтернативы H_1 с критической областью

$$C = \{\boldsymbol{x}\colon f(\boldsymbol{x}; \theta_1) > kf(\boldsymbol{x}; \theta_0)\} \tag{2.2.20}$$

является р. н. м. критерием с уровнем значимости $\alpha = \mathsf{P}_{\theta_0}(C)$.

Это утверждение выглядит несколько неожиданным, поскольку роль параметра θ_1 не ясна. На самом деле, как мы увидим далее, значение θ_1 требуется только для определения критической области и размера α. Более точно, в силу м. о. п. множество C будет представимо в виде

$$\{f(\boldsymbol{x}; \theta') > k(\theta')f(\boldsymbol{x}; \theta_0)\}$$

для любого $\theta' > \theta_0$ при подобранном надлежащим образом $k(\theta')$. Этот факт будет играть решающую роль при доказательстве теоремы.

Доказательство. Не теряя общности, предположим, что все функции g_{θ_2,θ_1} в определении м. о. п. являются монотонно неубывающими. Отметим, во-первых, что в силу леммы Н—П критерий (2.2.20) является критерием Н—П размера не больше $\alpha = \mathsf{P}_{\theta_0}(C)$ для проверки простой нулевой гипотезы $f = f(\cdot\,; \theta_0)$ против простой альтернативы $f = f(\cdot\,; \theta_1)$. Это означает, что

$$\mathsf{P}_{\theta_1}(C) \geqslant \mathsf{P}_{\theta_1}(C^*) \ \text{ для любого } C^* \quad \text{и} \quad \mathsf{P}_{\theta_0}(C^*) \leqslant \alpha. \tag{2.2.21}$$

В силу свойства м. о. п. критерий (2.2.20) эквивалентен критерию

$$C = \{\boldsymbol{x}\colon T(\boldsymbol{x}) > c\}$$

для некоторого значения c. Но тогда, опять в силу свойства м. о. п., для любого $\theta' > \theta_0$ критерий (2.2.20) эквивалентен критерию

$$C = \{f(\boldsymbol{x};\ \theta') > k(\theta')f(\boldsymbol{x};\ \theta_0)\} \tag{2.2.22}$$

для некоторого значения $k(\theta')$. И вновь в силу леммы Н—П критерий (2.2.22) (а следовательно, и критерий (2.2.20)) является н. м. критерием размера не больше $\mathsf{P}_{\theta_0}(C) = \alpha$ для проверки простой нулевой гипотезы $f = f(\cdot\,;\ \theta_0)$ против простой альтернативы $f = f(\cdot\,;\ \theta')$. Таким образом,

$$\mathsf{P}_{\theta'}(C) \geqslant \mathsf{P}_{\theta'}(C^*) \quad \text{для любых таких областей } C^*\text{, что } \mathsf{P}_{\theta_0}(C^*) \leqslant \alpha.$$

Другими словами, мы установили, что критерий (2.2.20) является р. н. м. критерием с уровнем значимости α для проверки простой нулевой гипотезы $f = f(\cdot\,;\ \theta_0)$ против односторонней альтернативы $f = f(\cdot\,;\ \theta')$ для некоторого $\theta' > \theta_0$. Формально говоря,

$$\mathsf{P}_{\theta'}(C) \geqslant \mathsf{P}_{\theta'}(C^*) \ \text{ при всех } \ \theta' > \theta_0 \quad \text{и} \quad \mathsf{P}_{\theta_0}(C^*) \leqslant \alpha.$$

Но тогда это же неравенство будет выполняться и при дополнительном ограничении (относительно C^*), что $\mathsf{P}_{\theta}(C^*) \leqslant \alpha$ для всех $\theta \leqslant \theta_0$. Этот факт и означает в точности, что критерий (2.2.22) (а, следовательно, и критерий (2.2.20)) представляет собой р. н. м. критерий с уровнем значимости α для проверки гипотезы H_0: $\theta \leqslant \theta_0$ против альтернативы $H_1 : \theta > \theta_0$. □

Анализируя доказательство, можно увидеть, что это же утверждение выполняется и в случае, когда в гипотезе H_0 предполагается, что $\theta < \theta_0$, а $H_1 : \theta \geqslant \theta_0$. Как и в случае простых гипотез, обратная задача (найти константы $k(\theta')$, $\theta' > \theta_0$, для заданного $\alpha \in (0,\ 1)$) требует рандомизации критерия. Имеет место соответствующее утверждение, гарантирующее существование рандомизированного р. н. м. критерия с решающей функцией, принимающей не более чем три значения.

§ 2.3. Критерии согласия. Проверка гипотез для нормальных распределений. Однородные выборки

Fit, Man, Test, Woman[1]

(Из серии «Фильмы, которые не вышли на большой экран».)

Лемма Н—П и ее модификации являются скорее примерами тех исключительных ситуаций, когда задача проверки гипотез может быть решена эффективно. Другие (практически важные) примеры — это те, когда с. в. являются нормально распределенными. Для таких случаев проверка гипотез тоже проводится эффективно (хотя и в несколько неполной по-

[1] Ср. название фильма «Eat, Drink, Man, Woman».

становке). Эта проверка основывается на теореме Фишера о том, что если $X_1, \ldots, X_n$ — это выборка н. о. р. с. в. $X_i \sim \mathrm{N}(\mu, \sigma^2)$, то

$$\frac{\sqrt{n}}{\sigma}(\bar{X} - \mu) \sim \mathrm{N}(0, 1), \quad \frac{1}{\sigma^2} S_{XX} \sim \chi^2_{n-1} \quad \text{и эти с. в. независимы.}$$

Здесь $\bar{X} = \sum\limits_i X_i/n$ и $S_{XX} = \sum\limits_i (X_i - \bar{X})^2$, см. § 1.5.

Выборка

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}, \quad \text{где } X_i \sim \mathrm{N}(\mu, \sigma^2),$$

называется однородной нормальной; неоднородные нормальные выборки — это такие выборки, что $X_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$, т. е. параметры распределения с. в. X_i различны при разных i.

Проверка гипотезы о среднем значении при неизвестной дисперсии. Рассмотрим однородную нормальную выборку $\boldsymbol{X} = (X_1, \ldots, X_n)$ и нулевую гипотезу $H_0\colon \mu = \mu_0$ с альтернативой $H_1\colon \mu \neq \mu_0$. Здесь μ_0 — заданное число. Наш критерий будет основан на t-*статистике Стьюдента*

$$T(\boldsymbol{X}) = \frac{\sqrt{n}(\bar{X} - \mu_0)/\sigma}{\sqrt{S_{XX}/(n-1)\sigma^2}} = \frac{\sqrt{n}(\bar{X} - \mu_0)}{s_{XX}}, \tag{2.3.1}$$

где s_{XX} — выборочное стандартное отклонение (см. уравнение (1.4.2)). Согласно определению t-распределения в примере 1.1.3 (см. равенство (1.1.6)) $T(\boldsymbol{X}) \sim \mathrm{t}_{n-1}$ при H_0. Здесь замечательно то, что для вычисления $T(\boldsymbol{x})$ не требуется знать σ^2. Следовательно, этот критерий будет работать независимо от того, знаем мы σ^2 или нет.

Следовательно, вполне естественно сделать заключение, что если при выполнении гипотезы H_0 абсолютное значение $|T(\boldsymbol{x})|$ t-статистики $T(\boldsymbol{x})$ велико, то гипотезу H_0 следует отклонить. Более точно, для заданного $\gamma \in (0, 1)$ обозначим через $\mathrm{t}_{n-1}(\gamma)$ верхнюю γ-точку (квантиль) t_{n-1}-распределения, определяемую как значение a, для которого

$$\int\limits_a^\infty f_{\mathrm{t}_{n-1}}(x)\,dx = \gamma$$

(нижней γ-точкой будет, конечно, $-\mathrm{t}_{n-1}(\gamma)$). Тогда мы отклоняем гипотезу H_0, если при $\mu = \mu_0$ получаем, что $|T(\boldsymbol{x})| > \mathrm{t}_{n-1}(\alpha/2)$.

Описанная процедура называется критерием *Стьюдента* или t-*критерием* (t-распределение также часто называют распределением Стьюдента). Критерий был предложен В. С. Госсеттом (1876—1937), британским статистиком, который работал в Ирландии и Англии и писал под псевдонимом «Стьюдент» (англ. «Student»).

Госсетт работал на пивоварне Гиннесса, где отвечал за экспериментальное пивоварение (по образованию он был химиком). Находясь на этой должности, он провел учебный год в биометрической лаборатории К. Пирсона в Лондоне, где изучал статистику. Госсетт был известен как мягкий и застенчивый человек; ходила шутка, что он был единственным человеком, которому удавалось поддерживать дружеские отношения и с К. Пирсоном, и с Фишером одновременно. (Фишер был знаменит не только благодаря своим научным результатам, но также прославился как очень раздражительный и вспыльчивый человек. Когда ему предъявлялись претензии (или даже просто вопросы в мягкой форме) по поводу противоречий или неясностей в изложенном или сказанном им, он зачастую приходил в ярость и покидал аудиторию. Однажды Тьюки (который и сам был знаменит своими «дерзостями») пришел в офис Фишера и затеял дискуссию относительно некоторых его суждений. Спустя пять минут, когда разговор стал накаляться, Фишер сказал: «Хорошо. А теперь Вы можете покинуть мой кабинет.» Тьюки ответил: «Нет, я этого не сделаю, потому что слишком Вас уважаю». «В таком случае», — отвечал Фишер, — «уйду я».)

Вернемся к t-критерию: возникает вопрос, что же делать, когда $|t| < \mathrm{t}_{n-1}(\alpha/2)$. Ответ довольно дипломатичный: в этом случае гипотеза H_0 не отклоняется при уровне значимости α (поскольку эта гипотеза будет все еще считаться консервативной).

Можно использовать t-статистику также и для построения доверительного интервала (д. и.) для μ (опять не имеет значения, известно σ^2 или нет). Действительно, в уравнении

$$\mathrm{P}\left(-\mathrm{t}_{n-1}(\alpha/2) < \frac{\sqrt{n}(\overline{X}-\mu)}{s_{XX}} < \mathrm{t}_{n-1}(\alpha/2)\right) = 1-\alpha$$

неравенства можно переписать относительно μ:

$$\mathrm{P}\left(\overline{X} - \frac{1}{\sqrt{n}} s_{XX}\mathrm{t}_{n-1}(\alpha/2) < \mu < \overline{X} + \frac{1}{\sqrt{n}} s_{XX}\mathrm{t}_{n-1}(\alpha/2)\right) = 1-\alpha.$$

Здесь P обозначает на самом деле $\mathrm{P}_{\mu,\sigma^2}$, распределение н. о. р. выборки с $X_i \sim \mathrm{N}(\mu, \sigma^2)$. Это означает, что симметричный $100(1-\alpha)\%$-д. и. для μ имеет вид

$$\left(\overline{X} - \frac{1}{\sqrt{n}} s_{XX}\mathrm{t}_{n-1}(\alpha/2),\ \overline{X} + \frac{1}{\sqrt{n}} s_{XX}\mathrm{t}_{n-1}(\alpha/2)\right). \tag{2.3.2}$$

Some statisticians don't drink because they are t-test totallers.

Некоторые статистики не пьют крепкого только потому, что они зациклены на критерии Стьюдента[1].

(Из серии «Почему их не понимают».)

Пример 2.3.1. Пусть необходимо проверить долговечность двух материалов a и b, используемых для изготовления подошвы в женских туфлях. Каждой из 10 добровольцев была предложена смешанная пара туфель, где

[1]Игра слов: tea-totaller — абсолютный трезвенник (пьет только чай), t-test totaller — абсолютный приверженец критерия Стьюдента.

одна подошва была изготовлена из материала a, а другая — из материала b. Износ (в соответствующих единицах) показан в таблице 2.1.

Таблица 2.1

Доброволец	1	2	3	4	5	6	7	8	9	10
a	14,7	9,7	11,3	14,9	11,9	7,2	9,6	11,7	9,7	14,0
b	14,4	9,2	10,8	14,6	12,1	6,1	9,7	11,4	9,1	13,2
Разность	0,3	0,5	0,5	0,3	−0,2	1,1	−0,1	0,3	0,6	0,8

Предположив, что разности $X_1, \ldots, X_{10}$ являются н. о. р. $\mathrm{N}(\mu, \sigma^2)$-с. в. с неизвестными μ и σ^2, проверяют гипотезу $H_0\colon \mu = 0$ против альтернативы $H_1\colon \mu \neq 0$. Здесь $\bar{x} = 0{,}41$, $S_{xx} = \sum_i x_i^2 - n\bar{x}^2 = 1{,}349$ и t-статистика равна

$$t = \sqrt{10} \cdot 0{,}41 / \sqrt{1{,}349/9} = 3{,}35.$$

В случае критерия размера 0,05 гипотеза H_0 отклоняется, так как $\mathrm{t}_9(0{,}025) = 2{,}262 < 3{,}35$, и приходят к заключению, что средний износ материалов a и b различен. Это t-критерий для парных сравнений, и 95%-доверительный интервал для среднего значения разности имеет вид

$$\left(0{,}41 - \frac{\sqrt{1{,}349/9} \cdot 2{,}262}{\sqrt{10}},\ 0{,}41 + \frac{\sqrt{1{,}349/9} \cdot 2{,}262}{\sqrt{10}}\right) = (0{,}133,\ 0{,}687). \quad \square$$

Исторически изобретение t-критерия явилось важной вехой в развитии предмета статистики. В сущности, проверка аналогичной гипотезы о дисперсии нормального распределения является более простой задачей, так как мы можем использовать только статистику S_{XX}.

Проверка гипотезы о дисперсии при неизвестном среднем. Возьмем опять однородную нормальную выборку $\boldsymbol{X} = (X_1, \ldots, X_n)$ и рассмотрим нулевую гипотезу $H_0\colon \sigma^2 = \sigma_0^2$ против альтернативы $H_1\colon \sigma^2 \neq \sigma_0^2$, где σ_0^2 — это заданное значение. Как было сказано выше, критерий основан на статистике

$$\frac{1}{\sigma_0^2} S_{XX} = \frac{1}{\sigma_0^2} \sum_i (X_i - \bar{X})^2 \sim \chi_{n-1}^2 \quad \text{при } H_0. \tag{2.3.3}$$

Следовательно, при заданном $\alpha \in (0, 1)$ мы отклоняем гипотезу H_0 в симметричном двустороннем критерии с уровнем α, когда значение S_{xx}/σ_0^2 либо меньше чем $h_{n-1}^{-}(\alpha/2)$ (что будет свидетельствовать в пользу того, что $\sigma^2 < \sigma_0^2$), либо больше чем $h_{n-1}^{+}(\alpha/2)$ (в этом случае более вероятно, что $\sigma^2 > \sigma_0^2$). Здесь и в дальнейшем $h_m^{+}(\gamma)$ обозначает верхнюю γ-точку (квантиль) распределения χ_m^2, т. е. такое значение a, что

$$\int_a^{\infty} f_{\chi_m^2}(x)\, dx = \gamma.$$

Аналогично $h_m^-(\gamma)$ — это нижняя γ-точка (квантиль), т. е. значение a, для которого

$$\int_0^a f_{\chi^2_m}(x)\,dx = \gamma;$$

как было отмечено выше, $h_m^-(\gamma) = h_m^+(1-\gamma)$, $0 < \gamma < 1$. Здесь и далее мы обозначаем через $f_{\chi^2_m}(x)$ п. р. в. χ^2-распределения с m степенями свободы.

Этот критерий называют нормальным критерием χ^2. Его применение никак не связано со значением μ, которое может быть известным или неизвестным.

Нормальный критерий χ^2 позволяет построить доверительный интервал для σ^2 независимо от того, знаем мы μ или нет. А именно, перепишем равенство

$$\mathrm{P}\left(h_{n-1}^-(\alpha/2) < \frac{1}{\sigma^2}S_{XX} < h_{n-1}^+(\alpha/2)\right) = 1-\alpha$$

в виде

$$\mathrm{P}\left(\frac{S_{XX}}{h_{n-1}^+(\alpha/2)} < \sigma^2 < \frac{S_{XX}}{h_{n-1}^-(\alpha/2)}\right) = 1-\alpha$$

(здесь под P подразумевается $\mathrm{P}_{\mu,\sigma^2}$, выборочное распределение для $X_i \sim \mathrm{N}(\mu, \sigma^2)$). Тогда очевидно, что симметричный $100(1-\alpha)\%$-д. и. для σ^2 имеет вид

$$\left(\frac{S_{XX}}{h_{n-1}^+(\alpha/2)}, \frac{S_{XX}}{h_{n-1}^-(\alpha/2)}\right). \tag{2.3.4}$$

Пример 2.3.2. Менеджер нового телефонного центра желает удостовериться, что время ожидания клиентами ответов на свои звонки не является слишком длительным. Выборка, состоящая из 30 звонков в дневное бизнес-время, приводит к значению среднего времени ожидания, равному 8 секундам (что считается приемлемым). В то же время, выборочное значение $S_{xx}/(n-1)$ равно 16, что значительно превосходит значение 9, полученное для других телефонных центров. Менеджер проверяет гипотезу H_0: $\sigma^2 = 9$ против альтернативы H_1: $\sigma^2 > 9$ в предположении, что времена ожидания ответа на звонок являются н. о. р. $\mathrm{N}(\mu, \sigma^2)$-с. в. (идеализированная модель).

Воспользуемся критерием χ^2. Мы имеем $S_{xx}/\sigma^2 = 29 \cdot 16/9 = 51{,}65$. Для $\alpha = 0{,}05$ и $n = 30$ верхняя α-точка $h_{n-1}^+(\alpha)$ равна 42,56. Следовательно, при уровне 5% мы отклоняем гипотезу H_0 и заключаем, что дисперсия в новом телефонном центре превосходит 9.

Доверительный интервал для σ^2 имеет вид

$$\left(\frac{16\cdot 29}{45{,}72}, \frac{16\cdot 29}{16{,}05}\right) = (10{,}15,\ 28{,}91),$$

так как $h_{29}^+(0{,}025) = 45{,}72$ и $h_{29}^+(0{,}975) = 16{,}05$. □

Оба критерия, t-критерий и нормальный критерий χ^2, являются примерами так называемых *критериев согласия*. Здесь нулевая гипотеза H_0 соответствует «узкому» подмножеству Θ_0 параметрического множества Θ. В примерах 2.3.1 и 2.3.2 это подмножество представляло собой полупрямую $\{\mu = \mu_0, \sigma^2 > 0 \text{ произвольное}\}$ или всю прямую $\{\mu \in \mathbb{R}, \sigma^2 = \sigma_0^2\}$, лежащую на полуплоскости $\Theta = \{\mu \in \mathbb{R}, \sigma^2 > 0\}$. Нужно найти статистику критерия (в этих примерах — T или S_{XX}), имеющую «стандартное» распределение при H_0. Тогда мы отклоняем гипотезу H_0 при заданном уровне значимости α, если значение статистики не попадает в область высоковероятных значений, определенную для этого α (см. рис. 2.2). Такая область была интервалом

$$(-\mathrm{t}_{n-1}(\alpha/2), \mathrm{t}_{n-1}(\alpha/2))$$

для t-критерия и интервалом

$$(h_{n-1}^{-}(\alpha/2), h_{n-1}^{+}(\alpha/2))$$

для χ^2-критерия. В этом случае говорят, что данные значимы на уровне α. В противном случае данные не считаются значимыми и гипотеза H_0 не отклоняется.

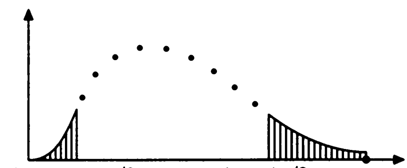


Рис. 2.2

§ 2.4. Критерий Пирсона χ^2. Теорема Пирсона

Statisticians do it with significance.

Статистики делают это со значением.

(Из серии «Как они делают это».)

Исторически идея построения критериев согласия для проверки гипотез восходит к Пирсону и датируется 1900 г. Свое дальнейшее развитие эта идея получила в 1920-е гг. в виде так называемого *критерия Пирсона хи-квадрат*, или *критерия Пирсона*, или *χ^2-критерия*, который основывается на *хи-квадрат статистике Пирсона*, или *статистике Пирсона*, или *χ^2-статистике*. Характерной чертой критерия Пирсона является его «универсальность» в том смысле, что он может быть использован для проверки гипотезы о принадлежности случайной выборки $\boldsymbol{X}$ вероятностному распределению с любой заданной п. р. в./д. ф. р. Гипотеза отклоняется, если значение статистики Пирсона не попадает в интервал высоковероятных значений. Универсальность проявляется в том, что формула для статистики Пирсона и распределение этой статистики не зависят от вида проверяемой д. ф. р./п. р. в. Однако мы платим за это: критерий является только асимптотически точным, когда объем выборки n стремится к ∞.

Предположим, что мы проверяем гипотезу о том, что н. о. р. выборка $\boldsymbol{X} = (X_1, \ldots, X_n)$ принадлежит вероятностному закону с п. р. в./д. ф. р. f^0. Разобьем пространство $\mathbb{R}$ значений с. в. X_i на k непересекающихся множеств (например, интервалов) $D_1, \ldots, D_k$ и вычислим вероятности

$$p_l^0 = \int_{D_l} f^0(x)\,dx \quad \text{или} \quad \sum_{D_l} f^0(x). \tag{2.4.1}$$

Нулевая гипотеза H_0 состоит в том, что для любого ℓ вероятность $\mathsf{P}(X_i \in D_l)$ равна значению p_ℓ^0, заданному формулой (2.4.1). Альтернативой здесь является утверждение, что эти вероятности произвольные (но такие, что $p_l \geqslant 0$ и $p_1 + \ldots + p_k = 1$). Для $\boldsymbol{x} = (x_1, \ldots, x_n)$ χ^2-статистика Пирсона определяется как

$$P(\boldsymbol{x}) = \sum_{l=1}^{k} \frac{(n_l - e_l)^2}{e_l}, \tag{2.4.2}$$

где n_l $(= n_l(\boldsymbol{x}))$ — это число значений x_i, попавших в множество D_l, и $n_1 + \ldots + n_k = n$, а $e_l = np_l^0$ — это среднее значение при гипотезе H_0. Символ P используется в честь Пирсона. Пусть задано $\alpha \in (0, 1)$, тогда мы отклоняем нулевую гипотезу при уровне значимости α, если значение π статистики P превосходит $h_{k-1}^{+}(\alpha)$, верхнюю α-квантиль χ_{k-1}^2-распределения.

Обоснование этого метода — следующая *теорема Пирсона.*

Предположим, что $X_1, X_2, \ldots$ *— это последовательность н. о. р. с. в. Пусть* $D_1, \ldots, D_k$ *— это разбиение пространства* $\mathbb{R}$ *на попарно непересекающиеся множества. Положим* $q_l = \mathsf{P}(X_i \in D_l)$, $l = 1, \ldots, k$, *где* $q_1 + \ldots + q_k = 1$. *Затем для любого* $l = 1, \ldots, k$ *и* $n \geqslant 1$ *определим*

$$N_{l,n} = \textit{число с. в. } X_i \textit{ среди таких } X_1, \ldots, X_n, \textit{ что } X_i \in D_l,$$

где $N_{1,n} + \ldots + N_{k,n} = n$, *и*

$$P_n = \sum_{l=1}^{k} \frac{(N_{l,n} - nq_l)^2}{nq_l}. \tag{2.4.3}$$

Тогда для любого $\alpha > 0$ выполняется равенство

$$\lim_{n\to\infty} \mathsf{P}(P_n > \alpha) = \int_{\alpha}^{\infty} f_{\chi^2_{k-1}}(x)\,dx. \tag{2.4.4}$$

Доказательство. Воспользуемся тем фактом, что соотношение (2.4.4) эквивалентно сходимости характеристических функций $\psi_{P_n}(t) = \mathsf{E}e^{itP_n}$:

$$\lim_{n\to\infty} \psi_{P_n}(t) = \int_0^{\infty} f_{\chi^2_{k-1}}(x) e^{itx}\,dx, \qquad t \in \mathbb{R}. \tag{2.4.5}$$

Положим

$$Y_{l,n} = \frac{N_{l,n} - nq_l}{\sqrt{nq_l}}, \qquad \text{так что} \quad P_n = \sum_{l=1}^{k} Y_{l,n}^2 \tag{2.4.6}$$

и

$$\sum_{l=1}^{k} Y_{l,n}\sqrt{q_l} = \frac{1}{\sqrt{n}} \sum_{l} (N_{l,n} - nq_l) = \frac{1}{\sqrt{n}} \left(\sum_{l} N_{l,n} - n \sum_{l} q_l \right) = 0. \tag{2.4.7}$$

Наша цель — определить предельное распределение случайного вектора

$$\boldsymbol{Y}_n = \begin{pmatrix} Y_{1,n} \\ \vdots \\ Y_{k,n} \end{pmatrix}.$$

Возьмем единичный вектор

$$\boldsymbol{\varkappa} = \begin{pmatrix} \sqrt{q_1} \\ \vdots \\ \sqrt{q_k} \end{pmatrix}$$

и рассмотрим действительную ортогональную матрицу A размера $k \times k$ с k-м столбцом $\boldsymbol{\varkappa}$. Такая матрица всегда существует: нужно просто дополнить вектор $\boldsymbol{\varkappa}$ до ортогонального базиса в $\mathbb{R}^k$ и использовать этот базис в качестве столбцов матрицы A. Рассмотрим случайный вектор

$$\boldsymbol{Z}_n = \begin{pmatrix} Z_{1,n} \\ \vdots \\ Z_{k,n} \end{pmatrix},$$

определяемый равенством

$$\boldsymbol{Z}_n = A^T \boldsymbol{Y}_n.$$

Его последний элемент обращается в нуль:

$$Z_{k,n} = \sum_{l=1}^{k} Y_{l,n} A_{l,k} = \sum_{l=1}^{k} Y_{l,n}\sqrt{q_l} = 0.$$

В то же время, в силу ортогональности матрицы A мы имеем

$$P_n = \sum_{l=1}^{k} Z_{l,n}^2 = \sum_{l=1}^{k-1} Z_{l,n}^2. \tag{2.4.8}$$

Формула (2.4.7) демонстрирует структуру с. в. P_n. Мы видим, что для доказательства предельного соотношения (2.4.4) достаточно проверить, что с. в. $Z_{1,n}, \ldots, Z_{k-1,n}$ будут асимптотически независимыми $\mathrm{N}(0, 1)$-с. в. Вновь используя х. ф., достаточно проверить, что совместная х. ф. сходится к произведению:

$$\lim_{n\to\infty} \mathrm{E} e^{it_1 Z_{1,n}+\ldots+it_{k-1}Z_{k-1,n}} = \prod_{l=1}^{k-1} e^{-t_l^2/2}. \tag{2.4.9}$$

Чтобы доказать соотношение (2.4.9), вернемся к с. в. $N_{l,n}$. Запишем

$$\mathrm{P}(N_{1,n} = n_1, \ldots, N_{k,n} = n_k) = \frac{n!}{n_1!\ldots n_k!} q_1^{n_1} \ldots q_k^{n_k}$$

для любых таких неотрицательных целых чисел $n_1, \ldots, n_k$, что $n_1 + \ldots$ $\ldots + n_k = n$. Тогда совместная х. ф. имеет вид

$$\mathrm{E} e^{i\sum\limits_l t_l N_{l,n}} = \sum_{n_1,\ldots,n_k:\ \sum\limits_l n_l = n} \frac{n!}{n_1!\ldots n_k!} q_1^{n_1} \ldots q_k^{n_k} e^{i\sum\limits_l t_l n_l} = \left(\sum_{l=1}^{k} q_l e^{it_l}\right)^n.$$

Переходя к величинам $Y_{1,n}, \ldots, Y_{k,n}$, находим

$$\mathrm{E} e^{i\sum\limits_l t_l Y_{l,n}} = e^{-i\sqrt{n}\sum\limits_l t_l\sqrt{q_l}} \left(\sum_{l=1}^{k} q_l e^{it_l/\sqrt{nq_l}}\right)^n$$

и

$$\ln \mathrm{E} e^{i\sum\limits_l t_l Y_{l,n}} = n\ln\left(\sum_{l=1}^{k} q_l e^{it_l\sqrt{nq_l}}\right) - i\sqrt{n}\sum_l t_l\sqrt{q_l} =$$

$$= n\ln\left(1 + \frac{i}{\sqrt{n}}\sum_{l=1}^{k} t_l\sqrt{q_l} - \frac{1}{2n}\sum_{l=1}^{k} t_l^2 + O(n^{-3/2})\right) - i\sqrt{n}\sum_l t_l\sqrt{q_l} =$$

$$= -\frac{1}{2}\sum_{l=1}^{k} t_l^2 + \frac{1}{2}\left(\sum_{l=1}^{k} t_l\sqrt{q_l}\right)^2 + O(n^{-1/2})$$

в силу разложения Тейлора.

При $n \to \infty$ полученное выражение сходится к

$$-\frac{1}{2}\|t\|^2 + \frac{1}{2}(A^T t)_k^2 = -\frac{1}{2}\|A^T t\|^2 + \frac{1}{2}(A^T t)_k^2 = -\frac{1}{2}\sum_{l=1}^{k-1}(A^T t)_l^2, \qquad t = \begin{pmatrix} t_1 \\ \vdots \\ t_k \end{pmatrix} \in \mathbb{R}^n.$$

Здесь A — это введенная выше ортогональная матрица размера $k \times k$, для которой $A^T = A^{-1}$, и $\|\cdot\|^2$ обозначает квадрат нормы (или длины) вектора из $\mathbb{R}^k$ (так что $\|t\|^2 = \sum\limits_l t_l^2$ и $\|A^T t\|^2 = \sum\limits_l (A^T t)_l^2$).

Следовательно, поскольку $\langle t, Y_n\rangle = \sum\limits_{l=1}^{k} t_l Y_{l,n}$, мы получаем

$$\lim_{n\to\infty} \mathsf{E} e^{i\langle t, Y_n\rangle} = \prod_{l=1}^{k-1} e^{-(A^T t)_l^2/2}. \tag{2.4.10}$$

Тогда для с. в. $Z_{l,n}$ в тех же обозначениях имеем

$$\mathsf{E} e^{i\sum\limits_l t_l Z_{l,n}} = \mathsf{E} e^{i\langle t, Z_n\rangle} = \mathsf{E} e^{i\langle t, A^T Y_n\rangle} = \mathsf{E} e^{i\langle At, Y_n\rangle}.$$

В силу равенства (2.4.9) последнее выражение должно сходиться при $n \to \infty$ к

$$\prod_{l=1}^{k-1} e^{-(A^T A t)_l^2/2} = \prod_{l=1}^{k-1} e^{-t_l^2/2}.$$

Подчеркнем еще раз, что χ^2-критерий Пирсона становится точным только в пределе при $n \to \infty$. Однако, поскольку скорость сходимости достаточно высока, этот критерий используют и при умеренных значениях n, обосновывая это ссылкой на теорему Пирсона.

Интересно отметить, что К. Пирсон был также крупным ученым в области философии. В его книге «Грамматика науки» (1891 г.) дано яркое изложение научной философии венской школы конца XIX в. Критический обзор этой книги был дан в работах Ленина (которые оба автора данного задачника должны были изучать в студенческие годы). Однако Ленин проводил четкое различие между Пирсоном и Махом, являвшимся основным представителем этого философского течения (и выдающимся физиком того времени), и определенно считал Пирсона выше Маха.

Пример 2.4.1. В своих знаменитых экспериментах Мендель скрестил 556 мужских растений круглого желтого гороха с женскими растениями морщинистого зеленого гороха. В полученном потомстве мы обозначим

N_1 = число круглых желтых горошин,

N_2 = число круглых зеленых горошин,

N_3 = число морщинистых желтых горошин,

N_4 = число морщинистых зеленых горошин и рассмотрим следующую нулевую гипотезу о значениях пропорций:

$$H_0\colon (p_1, p_2, p_3, p_4) = \left(\frac{9}{16}, \frac{3}{16}, \frac{3}{16}, \frac{1}{16}\right),$$

как это предсказывает теория Менделя. Альтернатива состоит в том, что p_i любые, но $p_i \geqslant 0$ и $\sum\limits_{i=1}^{4} p_i = 1$.

На основании наблюдений было получено, что $(n_1, n_2, n_3, n_4) = (301, 121, 102, 32)$ и

$$e = (312{,}75;\ 104{,}25;\ 104{,}25;\ 34{,}75), \qquad P = 3{,}39888,$$

а верхняя 0,25-точка равна $h_3^+(0{,}25) = 4{,}10834$. Следовательно, у нас нет оснований отвергнуть предположение Менделя, даже при уровне 25%.

Отметим, что рассмотренная выше нулевая гипотеза о том, что $f = f^0$, полностью задает вероятности p_l^0. Во многих случаях нужно следовать менее точной нулевой гипотезе о том, что p_l^0 принадлежит некоторому заданному семейству $\{p_l^0(\theta), \theta \in \Theta\}$. В этой ситуации применяем описанную процедуру, после того как оценим значения параметра по той же самой выборке. Это означает, что χ^2-статистика вычисляется в виде

$$P(\boldsymbol{x}) = \sum_{l=1}^{k} \frac{(n_l - \hat{e}_l)^2}{\hat{e}_l}, \tag{2.4.11}$$

где $\hat{e}_l = np_l^0(\hat{\theta})$ и $\hat{\theta}$ $(= \hat{\theta}(\boldsymbol{x}))$ — это оценка параметра θ.

Обычно в качестве $\hat{\theta}$ берут о. м. п. Тогда значение $\pi = \sum\limits_{i=1}^{k} (n_l - \hat{e})^2/\hat{e}$ статистики P сравнивают не с $h_{k-1}^+(\alpha)$, а с $h_{k-1-|\Theta|}^+(\alpha)$, где $|\Theta|$ — это размерность множества Θ. Таким образом, при уровне значимости $\alpha \in (0, 1)$ мы отклоняем гипотезу H_0: p_l^0 принадлежит семейству $\{p_l^0(\theta), \theta \in \Theta\}$, если $\pi > h_{k-1-|\Theta|}^+(\alpha)$. (Обычно значение $h_{k_1}^+(\alpha)$ выше, чем $h_{k_2}^+(\alpha)$ при $k_1 > k_2$.) Такая процедура основывается на модифицированной теореме Пирсона, которая аналогична приведенной выше. □

§ 2.5. Критерии обобщенного отношения правдоподобия. Теорема Уилкса

Silence of the Lambdas[1]

(Из серии «Фильмы, которые не вышли на большой экран».)

Идея об использовании о. м. п. для вычисления статистики критерия получает свое дальнейшее развитие при рассмотрении так называемого

[1] Ср. название фильма «The Silence of the Lambs».

критерия обобщенного отношения правдоподобия. В этом случае рассматривают нулевую гипотезу H_0 о том, что $\theta \in \Theta_0$, где $\Theta_0 \subset \Theta$, и общую альтернативу $H_1 : \theta \in \Theta$. Частным примером является случай H_0: $f = f^{(0)}$, когда точно заданы вероятности $p_l^0 = \mathrm{P}(D_l)$; в этом случае Θ — множество всех возможных д. ф. р. $(p_1, \ldots, p_k)$, а Θ_0 сводится к одной точке $(p_1^{(0)}, \ldots, p_k^{(0)})$. Более общим является случай, когда p_l^0 зависят от параметра θ. Будем действовать аналогично предыдущему: рассмотрим максимумы

$$\max[f(\boldsymbol{x}; \theta) : \theta \in \Theta_0]$$

и

$$\max[f(\boldsymbol{x}; \theta) : \theta \in \Theta]$$

и введем их отношение (которое будет не меньше 1):

$$\Lambda_{H_1;H_0}(\boldsymbol{x}) = \frac{\max[f(\boldsymbol{x}; \theta) : \theta \in \Theta]}{\max[f(\boldsymbol{x}; \theta) : \theta \in \Theta_0]}. \tag{2.5.1}$$

Отношение называют *обобщенным отношением правдоподобия* (о. о. п.) для H_0 и H_1; иногда знаменатель называют правдоподобием гипотезы H_0, а числитель — правдоподобием гипотезы H_1.

В некоторых случаях распределение о. о. п. $\Lambda_{H_1;H_0}$ можно найти. В общем случае вводят величину

$$R = 2 \ln \Lambda_{H_1;H_0}(\boldsymbol{X}). \tag{2.5.2}$$

Тогда если при заданном $\alpha \in (0, 1)$ значение r величины R в формуле (2.5.2) превосходит верхнюю точку $h_p(\alpha)$, то при уровне α гипотезу H_0 отклоняют. Здесь число степеней свободы p распределения χ^2 равно разности $|\Theta| - |\Theta_0|$ размерностей множеств Θ и Θ_0.

Описанная процедура называется *критерием обобщенного отношения правдоподобия* (к. о. о. п.) и основывается на *теореме Уилкса*, которая является обобщением теоремы Пирсона об асимптотических свойствах χ^2-статистики. Неформально это теорема состоит в следующем.

Пусть $\boldsymbol{X} = (X_1, \ldots, X_n)$ — случайная выборка с н. о. р. компонентами X_i и с п. р. в./д. ф. р. $f(\cdot\,; \theta)$, зависящей от параметра $\theta \in \Theta$, где Θ — это открытая область размерности $|\Theta|$ в евклидовом пространстве. Предположим, что о. м. п. $\hat{\theta}(\boldsymbol{X})$ является асимптотически нормальной с. в. при $n \to \infty$. Зафиксируем подмножество $\Theta_0 \subset \Theta$ так, чтобы Θ_0 было открытой областью в евклидовом пространстве меньшей размерности $|\Theta_0|$. Пусть нулевая гипотеза H_0 состоит в том, что $\theta \in \Theta_0$. Тогда (как в теореме Пирсона) при выполнении гипотезы H_0 с. в. R, заданная формулой (2.5.2), *имеет асимптотически*

χ^2_p-распределение с $p = |\Theta| - |\Theta_0|$ степенями свободы. Более точно, для любых $\theta \in \Theta_0$ и $h > 0$ справедливо равенство

$$\lim_{n\to\infty} \mathrm{P}_\theta(R > h) = \int\limits_h^\infty f_{\chi^2_p}(x)\,dx.$$

Существует также вариант теоремы Уилкса для независимых, но неодинаково распределенных с. в. $X_1, X_2, \ldots$ Эта теорема была названа в честь С. С. Уилкса (1906—1964), американского статистика, который некоторое время работал вместе с Фишером. Мы не будем доказывать теорему Уилкса, но проиллюстрируем ее роль в нескольких примерах.

Пример 2.5.1. Один простой пример — это случай, когда $X_i \sim \mathrm{N}(\mu, \sigma^2)$, где дисперсия σ^2 известна или неизвестна. Предположим сначала, что дисперсия известна. Как и в предыдущем параграфе, фиксируем значение μ_0 и проверим гипотезу $H_0\colon \mu = \mu_0$ против альтернативы $H_1\colon \mu \neq \mu_0$. Тогда при выполнении гипотезы H_1 значение $\bar{x}$ является о. м. п. для μ и

$$\max[f(\boldsymbol{x};\mu)\colon \mu \in \mathbb{R}] = \frac{1}{(\sqrt{2\pi}\sigma)^n}\exp\left(-\frac{S_{xx}}{2\sigma^2}\right),$$

где

$$S_{xx} = \sum_i (x_i - \bar{x})^2,$$

в то время как при выполнении гипотезы H_0 мы имеем

$$f(\boldsymbol{x};\mu_0) = \frac{1}{(\sqrt{2\pi}\,\sigma)^n}\exp\left(-\frac{\Sigma^2_{\mu_0}}{2\sigma^2}\right),$$

где

$$\Sigma^2_{\mu_0} = \sum_i (x_i - \mu_0)^2.$$

Следовательно,

$$\Lambda_{H_1;H_0} = \exp\left[\frac{1}{2\sigma^2}\left(\Sigma^2_{\mu_0} - S_{xx}\right)\right]$$

и

$$2\ln\Lambda_{H_1;H_0} = \frac{1}{\sigma^2}(\Sigma^2_{\mu_0} - S_{xx}) = \frac{n}{\sigma^2}(\bar{x} - \mu)^2 \sim \chi^2_1.$$

В этом примере величина $2\ln\Lambda_{H_1;H_0}(\boldsymbol{X})$ имеет в точности χ^2_1-распределение (т. е. утверждение теоремы Уилкса выполняется точно, а не асимптотически). Согласно к. о. о. п. мы отклоняем гипотезу H_0 при уровне α, если $2\ln\Lambda_{H_1;H_0}(\boldsymbol{x})$ превосходит $h_1^+(\alpha)$, верхнюю α-квантиль χ^2_1-распределения. Это эквивалентно тому, чтобы отвергать гипотезу H_0, когда $|\bar{x} - \mu|\sqrt{n}/\sigma$ превосходит $z^+(\alpha/2)$, верхнюю $\alpha/2$-квантиль нормального $\mathrm{N}(0,\ 1)$-распределения.

Если дисперсия σ^2 неизвестна, нужно использовать о. м. п. $\hat{\sigma}^2$, которая равна $\Sigma^2_{\mu_0}/n$ при H_0 и S_{XX}/n при H_1. В этой ситуации σ^2 рассматривается как *мешающий* параметр: он не входит в нулевую гипотезу, однако все же должен приниматься во внимание. Тогда при H_1 имеем

$$\max[f(\boldsymbol{x};\ \mu,\ \sigma^2)\colon \mu \in \mathbb{R},\ \sigma^2 > 0] = \frac{1}{(2\pi S_{xx}/n)^{n/2}} e^{-n/2},$$

а при H_0 имеем

$$\max[f(\boldsymbol{x};\ \mu_0,\ \sigma^2)\colon \sigma^2 > 0] = \frac{1}{(2\pi\Sigma^2_{\mu_0}/n)^{n/2}} e^{-n/2}.$$

Следовательно,

$$\Lambda_{H_1;H_0} = \left(\frac{\Sigma^2_{\mu_0}}{S_{xx}}\right)^{n/2} = \left(1 + \frac{n(\bar{x} - \mu_0)^2}{S_{xx}}\right)^{n/2},$$

и мы отклоняем гипотезу H_0, когда $n(\bar{x} - \mu_0)^2/S_{xx}$ велико. Но это вновь в точности совпадает с t-критерием, и у нас нет необходимости привлекать теорему Уилкса.

Мы видим, что t-критерий можно рассматривать как (важный) пример к. о. о. п. □

Пример 2.5.2. Пусть теперь μ известно и мы проверяем гипотезу $H_0\colon \sigma^2 = \sigma_0^2$ против альтернативы $\sigma^2 \neq \sigma_0^2$. Оценка максимального правдоподобия для σ^2 равна Σ^2/n, где $\Sigma^2 = \sum\limits_i (x_i - \mu)^2$. Следовательно,

$$\text{при } H_1 \quad \text{имеем} \quad \max[f(\boldsymbol{x};\ \mu,\ \sigma^2)\colon \sigma^2 > 0] = \left(\frac{\sqrt{n}}{\sqrt{2\pi\Sigma^2}}\right)^n e^{-n/2},$$

а

$$\text{при } H_0 \quad \text{имеем} \quad f(\boldsymbol{x};\ \mu,\ \sigma_0^2) = \frac{1}{(2\pi\sigma_0^2)^{n/2}} \exp\left[-\frac{1}{2\sigma_0^2}\Sigma^2\right],$$

тогда

$$\Lambda_{H_1;H_0} = \left(\frac{n\sigma_0^2}{\Sigma^2}\right)^{n/2} \exp\left[-\frac{n}{2} + \frac{\Sigma^2}{2\sigma_0^2}\right]$$

и

$$2\ln\Lambda_{H_1;H_0} = n\ln\frac{\sigma_0^2 n}{\Sigma^2} - n + \frac{\Sigma^2}{\sigma_0^2} = -n\ln\frac{\Sigma^2}{n\sigma_0^2} + \frac{\Sigma^2 - n\sigma_0^2}{\sigma_0^2} =$$
$$= -n\ln\left(1 + \frac{\Sigma^2 - n\sigma_0^2}{n\sigma_0^2}\right) + \frac{\Sigma^2 - n\sigma_0^2}{\sigma_0^2}.$$

Мы знаем, что при H_0 $\Sigma^2/\sigma^2 \sim \chi^2_n$. В силу ЗБЧ отношение Σ^2/n при H_0 сходится к $\mathbb{E}(X_1 - \mu)^2 = \operatorname{Var} X_1 = \sigma_0^2$. Следовательно, при больших n отношение $(\Sigma^2 - n\sigma_0^2)/n\sigma_0^2$ близко к 0. Воспользуемся разложением Тейлора для логарифма и найдем

$$2\ln \Lambda_{H_1;H_0} \approx -n\left[\frac{\Sigma^2 - n\sigma_0^2}{n\sigma_0^2} - \frac{1}{2}\left(\frac{\Sigma^2 - n\sigma_0^2}{n\sigma_0^2}\right)^2\right] + \frac{\Sigma^2 - n\sigma_0^2}{\sigma_0^2} =$$

$$= \frac{n}{2}\left(\frac{\Sigma^2 - n\sigma_0^2}{n\sigma_0^2}\right)^2 = \left(\frac{\Sigma^2 - n\sigma_0^2}{\sqrt{2n\sigma_0^4}}\right)^2.$$

Следующий факт состоит в том, что в силу ЦПТ при $n \to \infty$ выполняется условие

$$\frac{\Sigma^2 - n\sigma_0^2}{\sqrt{2n\sigma_0^4}} \sim \mathrm{N}(0,\ 1),$$

поскольку с. в. $\Sigma^2(\boldsymbol{X})$ представляет собой сумму н. о. р. с. в. $(X_i - \mu)^2$ с

$$\mathsf{E}(X_i - \mu)^2 = \sigma^2 \quad \text{и} \quad \mathrm{Var}(X_i - \mu)^2 = 2\sigma_0^4.$$

Следовательно, при H_0 и $n \to \infty$ имеем

$$\left(\frac{\Sigma^2 - n\sigma_0^2}{\sqrt{2n\sigma_0^4}}\right)^2 \sim \chi_1^2, \quad \text{т. е.} \quad 2\ln \Lambda_{H_1;H_0} \sim \chi_1^2,$$

что согласуется с теоремой Уилкса.

Мы видим, что при больших n к. о. о. п. состоит в том, чтобы отвергнуть H_0 при уровне α, когда

$$\frac{(\Sigma^2 - n\sigma_0^2)^2}{2n\sigma_0^4} > h_1^+(\alpha).$$

Конечно, в этой ситуации есть лучший способ действий: можно использовать тот факт, что

$$\frac{\Sigma^2}{\sigma_0^2} \sim \chi_n^2.$$

Следовательно, мы отклоняем гипотезу H_0, когда $\Sigma^2/\sigma_0^2 > h_n^+(\alpha/2)$ или $\Sigma^2/\sigma_0^2 < h_n^-(\alpha/2)$. Здесь используется та же статистика, что и в к. о. о. п., но с другой критической областью.

С другой стороны, можно рассмотреть

$$\frac{S_{XX}}{\sigma_0^2} \sim \chi_{n-1}^2,$$

где $S_{XX} = \sum\limits_{i=1}^{n}(X_i - \overline{X})^2$. В этом примере нормальный критерий χ^2, который предписывает

$$\text{отвергнуть } H_0\text{, если } \frac{1}{\sigma_0^2}S_{xx} > h_{n-1}^+(\alpha/2) \text{ или } \frac{1}{\sigma_0^2}S_{xx} < h_{n-1}^-(\alpha/2),$$

опять является более точным, чем к. о. о. п.

Аналогичным образом исследуется и случай, когда H_0: $\sigma^2 = \sigma_0^2$ и H_1: $\sigma^2 \neq \sigma_0^2$, а μ неизвестно (т. е. является мешающим параметром). В этом случае $\hat{\mu} = \bar{x}$, $\hat{\sigma}^2 = S_{XX}/n$,

$$\text{при } H_1 \text{ имеем} \quad \max[f(\boldsymbol{x};\, \mu,\, \sigma^2) \colon \mu \in \mathbb{R},\, \sigma^2 > 0] = \frac{1}{(2\pi S_{xx}^2/n)^{n/2}} e^{-n/2},$$

$$\text{при } H_0 \text{ имеем} \quad \max[f(\boldsymbol{x};\, \mu,\, \sigma_0^2) \colon \mu \in \mathbb{R}] = \frac{1}{(2\pi\sigma_0^2)^{n/2}} \exp\left(-\frac{1}{2\sigma_0^2} S_{xx}\right),$$

и статистика о. о. п. имеет вид

$$\Lambda_{H_1;H_0} = (ne)^{-n/2} \left(\frac{\sigma_0^2}{S_{xx}}\right)^{n/2} \exp\left(\frac{1}{2}\frac{S_{xx}}{\sigma_0^2}\right).$$

Мы видим, что $\Lambda_{H_1;H_0}$ велико, когда S_{xx}/σ_0^2 велико или близко к 0. Следовательно, согласно к. о. о. п. гипотеза H_0 отклоняется, если S_{XX}/σ_0^2 либо велико, либо мало. Но $S_{XX}/\sigma_0^2 \sim \chi^2_{n-1}$. В этом примере стандартный критерий χ^2 и к. о. о. п. приводят к использованию одних и тех же таблиц (однако работают с разными критическими областями). □

Еще одним аспектом к. о. о. п. является его связь с критерием Пирсона χ^2. Рассмотрим нулевую гипотезу H_0: $p_i = p_i(\theta)$, $1 \leqslant i \leqslant k$, для некоторого параметра $\theta \in \Theta_0$, где множество Θ_0 имеет размерность (число независимых координат) $|\Theta_0| = k_0 < k - 1$. Альтернативная гипотеза H_1 состоит в том, что вероятности p_i являются произвольными (лишь с тем ограничением, что $p_i \geqslant 0$ и $\sum\limits_i p_i = 1$).

Пример 2.5.3. Предположим, что имеются с. в. $X_1, \ldots, X_m \sim \mathrm{N}(\mu_X, \sigma_X^2)$ и $Y_1, \ldots, Y_n \sim \mathrm{N}(\mu_Y, \sigma_Y^2)$, где все параметры μ_X, σ_X^2, μ_Y и σ_Y^2 неизвестны. Пусть гипотеза H_0 состоит в том, что $\sigma_X^2 = \sigma_Y^2$, а гипотеза H_1 состоит в том, что $\sigma_X^2 > \sigma_Y^2$. Необходимо построить к. о. о. п. и определить распределение соответствующей статистики.

В этом случае

$$L_{xy}(H_0) = \max[f_X(x;\, \mu_X,\, \sigma^2) f_Y(y;\, \mu_Y,\, \sigma^2) \colon \mu_X \in \mathbb{R},\, \mu_Y \in \mathbb{R},\, \sigma > 0] = \\ = \max\left[\frac{1}{(2\pi\sigma^2)^{(n+m)/2}} \exp\left(-\frac{S_{xx} + S_{yy}}{2\sigma^2}\right) \colon \sigma > 0\right].$$

Отметим, что для $g(x) = x^a e^{-bx}$, где a, $b > 0$, выполняется равенство

$$\max_{x>0} g(x) = g\left(\frac{a}{b}\right) = \left(\frac{a}{b}\right)^a e^{-a}.$$

Следовательно,

$$L_{xy}(H_0) = \frac{1}{(2\pi\hat{\sigma}_0^2)^{(m+n)/2}} e^{-(m+n)/2}, \qquad \hat{\sigma}_0^2 = \frac{S_{xx} + S_{yy}}{m+n}.$$

Аналогично при H_1 имеем

$$L_{xy}(H_1) = \max[f_X(x;\ \mu_X,\ \sigma_X^2) f_Y(y;\ \mu_Y,\ \sigma_Y^2) \colon \mu_X,\ \mu_Y \in \mathbb{R},\ \sigma_X,\ \sigma_Y > 0] = \\ = \frac{1}{(2\pi\hat{\sigma}_X^2)^{m/2}(2\pi\hat{\sigma}_Y^2)^{n/2}} e^{-(m+n)/2},$$

где $\hat{\sigma}_X^2 = S_{xx}/m$, $\hat{\sigma}_Y^2 = S_{yy}/n$ (при условии, что $\hat{\sigma}_X^2 > \hat{\sigma}_Y^2$). В результате получаем, что

$$\text{если } \frac{S_{xx}}{m} > \frac{S_{yy}}{n}, \text{ то } \Lambda = \left(\frac{S_{xx}+S_{yy}}{m+n}\right)^{(m+n)/2} \left(\frac{S_{xx}}{m}\right)^{-m/2} \left(\frac{S_{yy}}{n}\right)^{-n/2},$$
$$\text{а если } \frac{S_{xx}}{m} \leqslant \frac{S_{yy}}{n}, \text{ то } \Lambda = 1.$$

Далее,

$$\text{если } \frac{S_{xx}}{S_{yy}} > \frac{m}{n}, \text{ то } 2\ln\Lambda = c + f\left(\frac{S_{xx}}{S_{yy}}\right).$$

Здесь

$$f(u) = (m+n)\ln(1+u) - m\ln u,$$

а c — константа. Находим далее, что

$$f'(u) = \frac{m+n}{1+u} - \frac{m}{u} = \frac{nu - m}{u(1+u)},$$

т. е. f возрастает, когда растет $u > m/n$. Следовательно, мы отклоняем гипотезу H_0, если S_{xx}/S_{yy} велико. При гипотезе H_0 имеем

$$\sigma_X^2 = \sigma_Y^2 = \sigma^2 \text{ и } \frac{S_{xx}}{\sigma^2} \sim \chi^2_{m-1},\ \frac{S_{yy}}{\sigma^2} \sim \chi^2_{n-1}, \text{ и эти величины независимы.}$$

Следовательно,

$$\frac{S_{xx}/(m-1)}{S_{yy}/(n-1)} \sim \mathrm{F}_{m-1,n-1}. \qquad \square$$

Пример 2.5.4. Пусть заданы $k+1$ вероятностей $p_0, \ldots, p_k$. Нулевая гипотеза H_0 состоит в том, что эти вероятности имеют вид

$$p_i(\theta) = C_k^i \theta^i (1-\theta)^{k-i}, \qquad 0 \leqslant i \leqslant k,$$

где $\theta \in (0,\ 1)$. (Здесь $|\Theta_0| = 1$.) Альтернатива H_1 состоит в том, что эти вероятности образуют k-мерное многообразие.

Оценка максимального правдоподобия при H_0 максимизирует сумму $\sum\limits_{i=0}^{k} n_i \ln p_i(\theta)$, где n_i — это число появлений значения i, $i = 0, \ldots, k$. Пусть $\hat{\theta}$ — точка максимума. Простые вычисления показывают, что при выполнении гипотезы H_0 мы имеем $\hat\theta = \left(\sum\limits_{i=1}^{k} in_i\right) \Big/ (kn)$. При выполнении гипотезы H_1 о. м. п. — это $\hat p_i = n_i/n$, где $n = n_0 + \ldots + n_k$. Получаем следующее выражение для логарифма о. о. п.:

$$2\ln\Lambda = 2\ln\frac{\prod\limits_{i=0}^{k}\hat p_i^{n_i}}{\prod\limits_{i=0}^{k}(p_i(\hat\theta))^{n_i}} = 2\sum_i n_i \ln\frac{n_i}{np_i(\hat\theta)}.$$

Число степеней свободы равно $k+1-1-|\Theta_0| = k-1$. Мы отклоняем гипотезу H_0, когда $2\ln\Lambda > h^+_{k-1}(\alpha)$.

Запишем $np_i(\hat\theta) = e_i$, $\delta_i = n_i - e_i$, $0 \leqslant i \leqslant k$, где $\sum\limits_i \delta_i = 0$. С помощью простых вычислений получаем

$$2\ln\Lambda = 2\sum_i n_i\ln\frac{n_i}{e_i} = 2\sum_i (e_i+\delta_i)\ln\left(1+\frac{\delta_i}{e_i}\right) \approx$$

$$\approx 2\sum_i\left(\delta_i + \frac{\delta_i^2}{e_i} - \frac{\delta_i^2}{2e_i}\right) = \sum_i\frac{\delta_i^2}{e_i} = \sum_i\frac{(n_i-e_i)^2}{e_i}.$$

Это в точности χ^2-статистика Пирсона. □

Следует сказать, что, вообще говоря, к. о. о. п. остается универсальным и мощным средством для проверки согласия. Мы рассмотрим два примера, где этот критерий используется для проверки гипотезы однородности для неоднородных выборок.

Пример 2.5.5. Пусть $X_1, \ldots, X_n$ — независимые случайные величины, $X_i \sim \mathrm{Po}(\lambda_i)$ с неизвестным средним λ_i, $i = 1, \ldots, n$. Требуется найти статистику обобщенного отношения правдоподобия для проверки гипотезы H_0: $\lambda_1 = \ldots = \lambda_n$ и показать, что это выражение можно аппроксимировать величиной $Z = \sum\limits_{i=1}^{n}(X_i - \overline{X})^2/\overline{X}$, где $\overline{X} = n^{-1}\sum\limits_{i=1}^{n} X_i$. Принимается ли гипотеза H_0, если для $n = 7$ получено значение этой статистики, равное 26,9?

Решение. Если $X_i \sim \mathrm{Po}(\lambda_i)$, то

$$f_{\boldsymbol{X}}(\boldsymbol{x};\lambda) = \prod_{i=1}^{n}\frac{\lambda_i^{x_i}}{x_i!}e^{-\lambda_i}.$$

При H_0 о. м. п. $\hat\lambda$ параметра λ — это $\bar x = \sum\limits_i x_i/n$, а при H_1 мы имеем $\hat\lambda_i = x_i$. Тогда о. о. п. имеет вид

$$\Lambda_{H_1;H_0}(\boldsymbol{x}) = \frac{f(\boldsymbol{x};\hat\lambda_1,\ldots,\hat\lambda_n)}{f(\boldsymbol{x};\hat\lambda)} = \frac{\prod\limits_i x_i^{x_i}}{\prod\limits_i \bar x^{x_i}}$$

и

$$2\ln\Lambda_{H_1;H_0}(\boldsymbol{x}) = 2\sum_i x_i \ln\frac{x_i}{\bar{x}} = 2\sum_i [\bar{x} + (x_i - \bar{x})]\ln\left(1 + \frac{x_i - \bar{x}}{\bar{x}}\right) \approx \sum_i \frac{(x_i - \bar{x})^2}{\bar{x}},$$

поскольку $\ln(1+x) \approx x - x^2/2$.

Наконец, нам известно, что для $Z \sim \chi^2_6$ выполняются равенства $\mathsf{E}Z = 6$ и $\operatorname{Var} Z = 12$. Значение 26,9 оказывается слишком большим. Таким образом, нам следует отклонить гипотезу H_0. □

Пример 2.5.6. Предположим, что задан набор np независимых случайных величин, представленных в виде n выборок, каждая длины p:

$$\begin{aligned} X^{(1)} &= (X_{11}, \ldots, X_{1p}), \\ X^{(2)} &= (X_{21}, \ldots, X_{2p}), \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ X^{(n)} &= (X_{n1}, \ldots, X_{np}). \end{aligned}$$

Случайная величина X_{ij} имеет распределение Пуассона с неизвестным параметром λ_j, $1 \leqslant j \leqslant p$. Требуется проверить гипотезу о том, что $\lambda_1 = \ldots = \lambda_p$, против альтернативы, что хотя бы два значения λ_j различны, выписать выражение для статистики критерия отношения правдоподобия, показать, что ее можно аппроксимировать величиной

$$\frac{n}{\bar{\bar{X}}}\sum_{j=1}^{p}(\bar{X}_j - \bar{\bar{X}})^2,$$

где

$$\bar{X}_j = \frac{1}{n}\sum_{i=1}^{n} X_{ij}, \qquad \bar{\bar{X}} = \frac{1}{np}\sum_{i=1}^{n}\sum_{j=1}^{p} X_{ij},$$

и объяснить, как проверяется гипотеза однородности при больших n.

В этом примере правдоподобие

$$\prod_{i=1}^{n}\prod_{j=1}^{p}\frac{e^{-\lambda_j}\lambda_j^{x_{ij}}}{x_{ij}!}$$

достигает максимума при $\hat{\lambda} = \bar{\bar{x}}$, если гипотеза H_0 справедлива, и это максимальное значение пропорционально величине $e^{-np\bar{\bar{x}}}\bar{\bar{x}}^{np\bar{\bar{x}}}$. При справедливости гипотезы H_1 мы имеем $\hat{\lambda}_j = \bar{x}_j$ и максимальное значение пропорционально произведению $e^{-np\bar{\bar{x}}}\prod\limits_{j=1}^{p}\bar{x}_j^{n\bar{x}_j}$. Следовательно,

$$\Lambda_{H_1;H_0} = \frac{1}{\bar{\bar{x}}^{np\bar{\bar{x}}}}\prod_{j=1}^{p}\bar{x}_j^{n\bar{x}_j}.$$

Мы отклоняем гипотезу H_0, когда $2\ln\Lambda_{H_1;H_0}$ принимает большие значения. Далее находим, что $\ln\Lambda_{H_1,H_0}$ имеет вид

$$\left(\sum_{j=1}^{p} n\bar{x}_j\ln\bar{x}_j - np\bar{x}\ln\bar{x}\right) = n\left[\sum_j(\bar{x}_j\ln\bar{x}_j - \bar{x}_j\ln\bar{x})\right] =$$

$$= n\left[\sum_j(\bar{x}+\bar{x}_j-\bar{x})\ln\left(\frac{\bar{x}+\bar{x}_j-\bar{x}}{\bar{x}}\right)\right] = n\left[\sum_j(\bar{x}+\bar{x}_j-\bar{x})\ln\left(1+\frac{\bar{x}_j-\bar{x}}{\bar{x}}\right)\right].$$

При выполнении гипотезы H_0 мы имеем $n\bar{X}_j \sim \mathrm{Po}(n\lambda)$ и $np\bar{X} \sim \mathrm{Po}(np\lambda)$. В силу ЗБЧ обе величины $\bar{X}_j$ и $\bar{X}$ сходятся к константе λ, следовательно, $(\bar{X}_j - \bar{X})/\bar{X}$ сходится к 0. Это означает, что для достаточно больших n можно взять в качестве хорошей аппроксимации логарифма его разложение до первого порядка включительно. Следовательно, если n велико, то $2\ln\Lambda_{H_1,H_0}$ приближенно равно

$$2n\sum_{j=1}^{p}(\bar{x}+\bar{x}_j-\bar{x})\left[\frac{\bar{x}_j-\bar{x}}{\bar{x}} - \frac{1}{2}\left(\frac{\bar{x}_j-\bar{x}}{\bar{x}}\right)^2\right] \approx n\sum_{i=1}^{p}\frac{(\bar{x}_j-\bar{x})^2}{\bar{x}}.$$

Заметим, что отношения $Y_{l,n} = (\bar{X}_l - \bar{X})/\sqrt{\bar{X}}$ напоминают отношения $Y_{l,n}$ из определения (2.4.6). На самом деле можно проверить, что при $n \to \infty$ мы имеем

$$\sqrt{n}\frac{\bar{X}_l - \bar{X}}{\sqrt{\bar{X}}} \sim \mathrm{N}(0,\ 1).$$

(Это составляет основную часть теоремы Уилкса.) Как и в уравнении (2.4.7), величины $Y_{l,n}$ удовлетворяют линейному соотношению

$$\sum_{l=1}^{p} Y_{l,n}\sqrt{\bar{X}} = \sum_{l=1}^{p}(\bar{X}_l - \bar{X}) = 0.$$

Аналогично доказательству теоремы Пирсона при $n \to \infty$ получаем

$$2\ln\Lambda_{H_1,H_0} \sim \chi^2_{p-1}. \qquad \square$$

Завершая этот параграф, мы приведем пример типичного вопроса в стиле кембриджских «Математических треножников».

Пример 2.5.7. Что понимают под критерием обобщенного отношения правдоподобия? Подробно объясните, как применять такой критерий.

Решение. Критерий обобщенного отношения правдоподобия предназначен для проверки гипотезы H_0: $\theta \in \Theta_0$ против общей альтернативы

$H_1: \theta \in \Theta$, где $\Theta_0 \subset \Theta$. При этом используют о. о. п.

$$\Lambda_{H_1;H_0}(\boldsymbol{x}) = \frac{\max[f(\boldsymbol{x};\,\theta) : \theta \in \Theta]}{\max[f(\boldsymbol{x};\,\theta) : \theta \in \Theta_0]}, \quad \text{где } \boldsymbol{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Критерий обобщенного отношения правдоподобия отклоняет гипотезу H_0 при больших значениях $\Lambda_{H_1;H_0}$.

Если случайная выборка $\boldsymbol{X}$ имеет н. о. р. компоненты $X_1, \ldots, X_n$ и n велико, то при H_0 мы имеем $2\ln\Lambda_{H_1;H_0}(\boldsymbol{X}) \sim \chi^2_p$, где число степеней свободы $p = |\Theta| - |\Theta_0|$ (теорема Уилкса). Следовательно, для заданного $\alpha \in (0, 1)$ мы отклоняем гипотезу H_0 с помощью критерия размера α, если

$$2\ln\Lambda_{H_1;H_0}(\boldsymbol{x}) > h^+_p(\alpha). \qquad \square$$

Интересно заметить, что к. о. о. п. был предложен Ю. Нейманом и Э. Пирсоном в 1928 г. Однако лемма Н—П была доказана лишь в 1933 г.

§ 2.6. Таблицы сопряженности признаков

Статистики делают это при помощи таблиц, если могут.

(Из серии «Как они делают это».)

Распространенным примером применения к. о. о. п. является случай, когда пытаются проверить независимость различных «категорий» или «признаков», приписываемых нескольким типам индивидуумов или вещей (объектов).

Пример 2.6.1. Проводится IQ-тест (IQ — intelligence quotient — коэффициент умственного развития) с целью приблизительно классифицировать людей по трем категориям согласно уровням: превосходный (A), хороший (B), посредственный (C). Тестирование проводилось в трех выбранных районах в Андлии, Гельсе и Грогландии, и число людей, получивших степени A, B и C соответственно, представлено в табл. 2.2.

Таблица 2.2

Район	A	B	C	Всего n_{i+}
Андлия	3009	2832	3008	8849
Гельс	3047	3051	2997	9095
Грогландия	2974	3038	3018	9030
Всего n_{+j}	9030	8921	9023	26974

Требуется построить к. о. о. п. для проверки независимости этой классификации.

Здесь значения $e_{ij} = n_{i+}n_{+j}/n_{++}$ равны

2962,35	2926,59	2960,06
3044,7	3007,95	3042,34
3022,94	2986,45	3020,6

Значение статистики Пирсона (которая определена ниже формулой (2.6.3)) равно

$$4{,}56828 + 1{,}29357 + 1{,}68437 = 7{,}54622,$$

и число степеней свободы равно $(3-1)\cdot(3-1) = 4$. При уровне 0,05 находим, что $h_4^+(0{,}05) = 9{,}488$. Следовательно, при таком уровне (а тем более при уровне 0,01) нет оснований отклонить гипотезу о том, что все группы однородны. □

В общем случае задана *таблица сопряженности признаков* с r строками и c столбцами. Пусть p_{ij} — это вероятность того, что объект i-го типа попадает в категорию j, или имеет признак j; независимость означает, что p_{ij} имеет вид $\alpha_i\beta_j$, где

$$\alpha_i,\ \beta_j \geqslant 0 \quad \text{и} \quad \sum_{i=1}^{r}\alpha_i = \sum_{j=1}^{c}\beta_j = 1.$$

Это составляет нашу нулевую гипотезу H_0. Альтернативе H_1 соответствуют общие ограничения $p_{ij} \geqslant 0$ и $\sum\limits_{i=1}^{r}\sum\limits_{j=1}^{c} p_{ij} = 1$. Положим

$$p_{i+} = \sum_j p_{ij}, \quad p_{+j} = \sum_i p_{ij} \quad \text{и} \quad p_{++} = \sum_i p_{i+} = \sum_j p_{+j} = 1$$

(эти обозначения сложились исторически).

В основе примера 2.6.1 лежит следующая модель: имеется n объектов или индивидуумов (26974 в этом примере), и n_{ij} из них попадают в ячейку (i, j) таблицы, так что $\sum\limits_{i,j} n_{ij} = n$. (Сравнивая с примером 2.2.1, видим, что мы имеем дело с моделью выбора с возвращением, которая обобщает пример 2.2.2.) Положим

$$n_{i+} = \sum_j n_{ij},\ n_{+j} = \sum_i n_{ij} \text{ и } n_{++} = \sum_i n_{i+} = \sum_j n_{+j} = n.$$

Случайные величины N_{ij} имеют совместное *мультиномиальное распределение* с параметрами $(n;\ (p_{ij}))$:

$$\mathrm{P}(N_{ij} = n_{ij}\ \forall i, j) = n! \prod \frac{1}{n_{ij}!} p_{ij}^{n_{ij}} \qquad (2.6.1)$$

(p_{ij} часто называют «вероятностями ячеек»). Нетрудно распознать здесь основания для применения к. о. о. п., где $|\Theta_1| = rc - 1$ (так как $p_{++} = 1$) и $|\Theta_0| = r - 1 + c - 1$, следовательно,

$$|\Theta_1| - |\Theta_0| = (r-1)(c-1).$$

При выполнении гипотезы H_1 о. м. п. для p_{ij} равна $\hat{p}_{ij} = n_{ij}/n$. Действительно, для любых таких n_{ij}, что $n_{++} = n$, д. ф. р. (т. е. правдоподобие) равна

$$f_{(N_{ij})}((n_{ij});\ (p_{ij})) = n! \prod_{i,j} \frac{p_{ij}^{n_{ij}}}{n_{ij}!}$$

и логарифм правдоподобия равен

$$\ell((n_{ij});\ (p_{ij})) = \sum_{i,j} n_{ij} \ln p_{ij} + \ln(n!) - \sum_{i,j} \ln(n_{ij}!).$$

Необходимо максимизировать ℓ по p_{ij} при ограничениях

$$p_{ij} \geqslant 0,\ \sum_{i,j} p_{ij} = 1.$$

Пренебрегая выражением

$$\ln(n!) - \sum_{i,j} \ln(n_{ij}!),$$

выпишем функцию Лагранжа

$$\mathcal{L} = \sum_{i,j} n_{ij} \ln p_{ij} - \lambda\left(\sum_{i,j} p_{ij} - 1\right).$$

Ее максимум достигается, когда для любых i, j выполняется равенство

$$\frac{\partial}{\partial p_{ij}} \mathcal{L} = \frac{n_{ij}}{p_{ij}} - \lambda = 0,$$

т. е. когда $p_{ij} = n_{ij}/\lambda$. Принимая во внимание ограничение $p_{++} = 1$, находим $\lambda = n$.

При выполнении гипотезы H_0 о. м. п. для α_i равна $\hat{\alpha}_i = n_{i+}/n$; о. м. п. для β_j — это $\hat{\beta}_j = n_{+j}/n$. Действительно, в этом случае правдоподобие равно

$$f_{(N_{ij})}((n_{ij});\ (\alpha_i),\ (\beta_j)) = n! \prod_{i,j} \frac{(\alpha_i\beta_j)^{n_{ij}}}{n_{ij}!} = n! \prod_i \alpha_i^{n_{i+}} \prod_j \beta_j^{n_{+j}} \Big/ \prod_{i,j} n_{ij}!$$

и логарифм правдоподобия равен

$$\ell((n_{ij});\ (\alpha_i),\ (\beta_j)) = \sum_i n_{i+} \ln \alpha_i + \sum_j n_{+j} \ln \beta_j + \ln(n!) - \sum_{i,j} \ln(n_{ij}!).$$

Необходимо максимизировать это выражение по α_i и β_j при ограничениях

$$\alpha_i,\ \beta_j \geqslant 0, \quad \sum_i \alpha_i = \sum_j \beta_j = 1.$$

Лагранжиан теперь равен

$$\mathcal{L} = \sum_i n_{i+} \ln \alpha_i + \sum_j n_{+j} \ln \beta_j - \lambda\Big(\sum_i \alpha_i - 1\Big) - \mu\Big(\sum_j \beta_j - 1\Big)$$

(здесь мы опять пренебрегаем выражением $[\ln(n!) - \sum \ln(n_{ij}!)]$). Условие для стационарных точек

$$\frac{\partial}{\partial \alpha_i}\mathcal{L} = \frac{\partial}{\partial \beta_j}\mathcal{L} = 0, \quad 1 \leqslant i \leqslant r,\ 1 \leqslant j \leqslant c,$$

приводит к оценкам $\hat{\alpha}_i = n_{i+}/\lambda$ и $\hat{\beta}_j = n_{+j}/\mu$. Применение ограничений приводит к равенствам $\lambda = \mu = n$.

Тогда о. о. п. имеет вид

$$\Lambda_{H_1;H_0} = \prod_{i,j}\Big(\frac{n_{ij}}{n}\Big)^{n_{ij}} \Big/ \prod_i \Big(\frac{n_{i+}}{n}\Big)^{n_{i+}} \prod_j \Big(\frac{n_{+j}}{n}\Big)^{n_{+j}}$$

и статистика $2\ln\Lambda_{H_1;H_0}$ равна

$$2\sum_{i,j} n_{ij} \ln\frac{n_{ij}}{n} - 2\sum_i n_{i+} \ln\frac{n_{i+}}{n} - 2\sum_j n_{+j}\ln\frac{n_{+j}}{n} = 2\sum_{i,j} n_{ij}\ln\frac{n_{ij}}{e_{ij}}, \qquad (2.6.2)$$

где $e_{ij} = (n_{i+}n_{+j})/n$. Представив n_{ij} в виде $n_{ij} = e_{ij} + \delta_{ij}$ и записав разложение для логарифма, находим, что правая часть равна

$$2\sum_{i,j}(e_{ij} + \delta_{ij})\ln\Big(1 + \frac{\delta_{ij}}{e_{ij}}\Big) = 2\sum_{i,j}(e_{ij} + \delta_{ij})\Big(\frac{\delta_{ij}}{e_{ij}} - \frac{1}{2}\frac{\delta_{ij}^2}{e_{ij}^2} + \ldots\Big),$$

что приблизительно равно

$$\sum_{i,j}\frac{(\delta_{ij})^2}{e_{ij}} = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(n_{ij} - e_{ij})^2}{e_{ij}}$$

в силу той же аппроксимации, что и ранее.

Следовательно, мы можем повторить обычную процедуру проверки, основанную на к. о. о. п.: составляем статистику

$$2\ln\Lambda = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(n_{ij} - e_{ij})^2}{e_{ij}} \qquad (2.6.3)$$

и отклоняем гипотезу H_0 при уровне α, когда значение статистики превосходит $h^+_{(r-1)(c-1)}(\alpha)$ — верхнюю α-точку $\chi^2_{(r-1)(c-1)}$-распределения.

Таблицы сопряженности приводят также к другой модели, когда фиксируют не только общее число наблюдений n_{++}, но также и некоторые компоненты, например все суммы элементов по строкам n_{i+} (или часть таких сумм). Тогда случайные величины N_{ij} в i-й строке с фиксированным n_{i+} будут иметь мультиномиальное распределение с параметрами $(n_{i+};\ p_{i1}, \ldots, p_{ic})$ и будут независимыми от остальных частот. В этом случае нулевая гипотеза состоит в том, что $p_{ij} = p_j$, т. е. p_{ij} не зависит от номера строки i для любых $j = 1, \ldots, c$. Альтернатива состоит в том, что p_{ij} произвольные, но $p_{i+} = 1$, $i = 1, \ldots, r$. Эта ситуация возникает в связи с проверкой гипотезы об однородности.

Пример 2.6.2. В эксперименте 150 пациентов разбили на три группы по 45, 45 и 60 человек. В двух группах пациенты принимали новое лекарство в разной дозировке, а пациенты третьей группы получали плацебо (безвредное, не лекарственное вещество, буквально «пустышка» (лат.)). Результаты приведены в табл. 2.3.

Таблица 2.3

	Улучшение	Без изменений	Ухудшение
Плацебо	16	20	9
Половина дозы	17	18	10
Полная доза	26	20	14

Введем гипотезу H_0:

вероятности $p_{\text{улучшение}}$, $p_{\text{без изменений}}$ и $p_{\text{ухудшение}}$

одинаковы для всех трех групп пациентов

и гипотезу H_1:

эти вероятности могут быть разными для разных групп.

Тогда при выполнении гипотезы H_1 правдоподобие равно

$$f_{(N_{ij})}((n_{ij}) \mid (p_{ij})) = \prod_{i=1}^{r} \frac{n_{i+}!}{n_{i1}! \ldots n_{ic}!} p_{i1}^{n_{i1}} \ldots p_{ic}^{n_{ic}}$$

и л. п. равен

$$\ell((n_{ij});\ (p_{ij})) = \sum_{i,j} n_{ij} \ln p_{ij} + (\text{члены, не зависящие от } p_{ij}).$$

Здесь о. м. п. $\hat{p}_{ij}$ для p_{ij} равна n_{ij}/n_{i+}. Аналогично при выполнении гипотезы H_0 л. п. равен

$$\ell((n_{ij});\ (p_{ij})) = \sum_{j} n_{+j} \ln p_j + (\text{члены, не зависящие от } p_j)$$

и о. м. п. $\hat{p}_j$ для p_j равна n_{+j}/n_{++}. Тогда, как и ранее, о. о. п. равно

$$2\ln \Lambda_{H_1;H_0} = 2\sum_{i,j} n_{ij} \ln \frac{\hat{p}_{ij}}{\hat{p}_j} = 2\sum_{i,j} n_{ij} \ln \frac{n_{ij}}{e_{ij}} \approx \sum_{i,j} \frac{(n_{ij}-e_{ij})^2}{e_{ij}},$$

где $e_{ij} = (n_{i+}n_{+j})/n_{++}$. Число степеней свободы равно $(r-1)(c-1)$, так как $|\Theta_1| = r(c-1)$ (имеется $c-1$ независимая переменная p_{ij} в каждой из r строк) и $|\Theta_0| = c-1$ (переменные в столбцах остаются постоянными).

В рассматриваемом примере $r=3$, $c=3$, $(r-1)(c-1)=4$, $h_4^+(0{,}05) = 9{,}488$. Массив данных приведен в табл. 2.4, а полученные средние значения показаны в табл. 2.5.

Таблица 2.4

n_{ij}	Улучшение	Без изменений	Ухудшение	
Плацебо	16	20	9	45
Половина дозы	17	18	10	45
Целая доза	26	20	14	60
	59	58	33	150

Таблица 2.5

e_{ij}	Улучшение	Без изменений	Ухудшение
Плацебо	17,7	17,4	9,9
Половина дозы	17,7	17,4	9,9
Целая доза	23,6	23,2	13,2

Итак, $\Lambda_{H_1;H_0}$ вычисляется следующим образом:

$$\frac{(2{,}4)^2}{23{,}6} + \frac{(1{,}7)^2}{17{,}7} + \frac{(0{,}7)^2}{17{,}7} + \frac{(2{,}6)^2}{17{,}4} + \frac{(0{,}6)^2}{17{,}4} + \\ + \frac{(3{,}2)^2}{23{,}2} + \frac{(0{,}9)^2}{9{,}9} + \frac{(0{,}1)^2}{9{,}9} + \frac{(0{,}8)^2}{13{,}2} = 1{,}41692.$$

Это значение меньше 9,488, следовательно, оно незначимо при уровне значимости 5%, и гипотезу H_1 нужно отвергнуть. □

Мы завершим этот параграф кратким обсуждением часто наблюдаемого *парадокса Симпсона*, который демонстрирует, что таблицы сопряженности признаков — это деликатный объект и они могут быть искажены, когда данные объединяются. Однако никакой мистики в этом нет. Рассмотрим следующий пример. Появилась возможность применять новое сложное лечение для потенциально смертельной болезни. Неудивительно, что врачи решили применять его главным образом в наиболее серьезных случаях. Вследствие этого новые данные охватывают намного больше именно таких случаев, в противоположность предыдущим [старым] данным, которые равномерно охватывали широкий круг пациентов. Это может приводить к обманчивой картине (см., например, табл. 2.6).

Таблица 2.6

	Не выздоровевшие	Выздоровевшие	Выздоровление, %
Предыдущее лечение	7500	5870	43,9
Новое лечение	12520	1450	10,38

Можно подумать, что новое лечение в четыре раза хуже, чем старое. Однако все дело в том, что новое лечение применялось значительно чаще, чем старое, в больницах, имеющих дело с наиболее серьезными случаями. С другой стороны, в тех клиниках, где обычно имеют дело с менее серьезными случаями, новое лечение применялось реже. Соответствующие данные приведены в табл. 2.7 и 2.8.

Таблица 2.7

Больницы	Не выздоровевшие	Выздоровевшие	Выздоровление, %
Предыдущее лечение	1100	70	5,98
Новое лечение	12500	1200	8,76

Таблица 2.8

Клиники	Не выздоровевшие	Выздоровевшие	Выздоровление, %
Предыдущее лечение	6400	5800	47,54
Новое лечение	20	250	92,60

Теперь очевидно, что новое лечение дает лучшие результаты для обеих категорий пациентов.

Вообще говоря, метод, основанный на таблицах сопряженности, не лишен логических трудностей. Предположим, что рассматриваются данные табл. 2.9, относящиеся к 100 семьям с двумя детьми.

Таблица 2.9

		1-й ребенок Мальчик	1-й ребенок Девочка	
2-й ребенок	Мальчик	30	20	50
	Девочка	20	30	50
		50	50	100

Рассмотрим две нулевые гипотезы:

H_0^1: вероятность того, что в семье есть мальчик, равна 1/2, и пол одного ребенка не зависит от пола другого ребенка;

H_0^2: пол одного ребенка не зависит от пола другого ребенка.

Каждая из этих гипотез проверяется против альтернативы: эти вероятности произвольны (точнее, вероятности $p_{\text{мм}}$, $p_{\text{мд}}$, $p_{\text{дм}}$, $p_{\text{дд}}$ подчинены только ограничениям $p_{..} \geqslant 0$ и $p_{\text{мм}} + p_{\text{мд}} + p_{\text{дм}} + p_{\text{дд}} = 1$; при этом случай возможной зависимости тоже охватывается). Значение статистики Пирсона равно

$$\frac{(30-25)^2}{25} + \frac{(20-25)^2}{25} + \frac{(20-25)^2}{25} + \frac{(30-25)^2}{25} = 4.$$

Число степеней свободы в случае гипотезы H_0^1 равно 3, а 5%-квантиль равна 7,815. В случае гипотезы H_0^2 число степеней свободы равно 1; та же квантиль равна 3,841.

Мы видим, что при уровне значимости 5% нет оснований, чтобы отвергнуть H_0^1, однако есть серьезные основания, чтобы отвергнуть H_0^2, хотя логически H_0^1 влечет за собой H_0^2. Мы благодарим А. Хокса за этот пример (см. [Haw]).

§ 2.7. Проверка гипотез для нормальных распределений. Неоднородные выборки

Variance is what any two statisticians are at.

Разногласие — это то, что существует между любыми двумя статистиками.

(Из серии «Почему их не понимают».)

Рассмотрим типичную ситуацию: имеется несколько (в простейшем случае две) выборок из нормального распределения, которые могут отличаться своими параметрами. Задача состоит в проверке гипотезы о том, что заданный параметр имеет одно и то же значение для всех выборок.

1. Проверка гипотез для двух нормальных выборок. Рассмотрим случай двух независимых выборок

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_m \end{pmatrix} \quad \text{и} \quad \boldsymbol{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix},$$

где X_i — н. о. р. $\mathrm{N}(\mu_1, \sigma_1^2)$-с. в., а Y_j — н. о. р. $\mathrm{N}(\mu_2, \sigma_2^2)$-с. в.

1а. Проверка гипотезы о равенстве средних при известной общей дисперсии. В этой модели предполагают, что дисперсия $\sigma_1^2 = \sigma_2^2 = \sigma^2$ известна, и проверяют гипотезу H_0: $\mu_1 = \mu_2$ против альтернативы $H_1: \mu_1 \neq \mu_2$. Используют для этого к. о. о. п. (который работает достаточно хорошо). При выполнении гипотезы H_1 функция правдоподобия

$f_{\boldsymbol{X}}(\boldsymbol{x};\,\mu_1,\,\sigma^2)f_{\boldsymbol{Y}}(\boldsymbol{y};\,\mu_2,\,\sigma^2)$ равна

$$\frac{1}{(\sqrt{2\pi}\sigma)^{m+n}}\exp\left[-\frac{1}{2\sigma^2}\sum_i (x_i-\mu_1)^2-\frac{1}{2\sigma^2}\sum_j (y_j-\mu_2)^2\right]=$$
$$=\frac{1}{(\sqrt{2\pi}\sigma)^{m+n}}\exp\left[-\frac{1}{2\sigma^2}(S_{xx}+m(\bar{x}-\mu_1)^2+S_{yy}+n(\bar{y}-\mu_2)^2)\right]$$

и достигает максимума при

$$\hat{\mu}_1=\bar{x} \quad \text{и} \quad \hat{\mu}_2=\bar{y}, \qquad \text{где} \quad \bar{x}=\frac{1}{m}\sum_i x_i, \quad \bar{y}=\frac{1}{n}\sum_j y_j.$$

При выполнении гипотезы H_0 функция правдоподобия $f_{\boldsymbol{X}}(\boldsymbol{x};\,\mu,\,\sigma^2)f_{\boldsymbol{Y}}(\boldsymbol{y};\,\mu,\,\sigma^2)$ равна

$$\frac{1}{(\sqrt{2\pi}\sigma)^{m+n}}\exp\left[-\frac{1}{2\sigma^2}(S_{xx}+m(\bar{x}-\mu)^2+S_{yy}+n(\bar{y}-\mu)^2)\right]$$

и достигает максимума при

$$\hat{\mu}=\frac{m\bar{x}+n\bar{y}}{m+n}.$$

Тогда о. о. п. имеет вид

$$\Lambda_{H_1;H_0}=\exp\left(\frac{1}{2\sigma^2}\frac{mn(\bar{x}-\bar{y})^2}{m+n}\right),$$

и мы отклоняем гипотезу H_0, если $|\bar{x}-\bar{y}|$ велико.

При выполнении гипотезы H_0 мы имеем

$$\bar{X}-\bar{Y}\sim \mathrm{N}\left(0,\,\sigma^2\left(\frac{1}{m}+\frac{1}{n}\right)\right),$$

следовательно,

$$Z=\frac{\bar{X}-\bar{Y}}{\sigma\sqrt{1/m+1/n}}\sim \mathrm{N}(0,\,1),$$

и мы отклоняем гипотезу H_0 при уровне значимости α, если $|Z|>z_+(\alpha/2)=$ $=\Phi^{-1}(1-\alpha/2)$. Этот критерий называют нормальным z-критерием. Отметим, что в этом случае распределение в теореме Уилкса является не асимптотическим, а точным.

В дальнейшем $z_+(\gamma)$ будет обозначать верхнюю γ-точку (квантиль) нормального $\mathrm{N}(0,\,1)$-распределения, т. е. значение, для которого

$$\frac{1}{\sqrt{2\pi}}\int\limits_{z_+(\gamma)}^{\infty} e^{-x^2/2}\,dx=\gamma.$$

Как было отмечено, это значение равно $\Phi^{-1}(1-\gamma)$, $0<\gamma<1$.

1б. Проверка гипотезы о равенстве средних при неизвестной общей дисперсии. Предположим теперь, что дисперсия $\sigma_1^2 = \sigma_2^2 = \sigma^2$ неизвестна и, следовательно, является мешающим параметром. Здесь опять рассматривают гипотезы H_0: $\mu_1 = \mu_2$ и H_1: $\mu_1 \neq \mu_2$. Как мы увидим, в результате будет получен t-критерий.

При выполнении гипотезы H_1 нам необходимо максимизировать

$$\frac{1}{(\sqrt{2\pi}\sigma)^{m+n}} \exp\left[-\frac{1}{2\sigma^2}(S_{xx} + m(\bar{x} - \mu_1)^2 + S_{yy} + n(\bar{y} - \mu_2)^2)\right]$$

по μ_1, μ_2 и σ^2. Как и ранее, $\hat{\mu}_1 = \bar{x}$, $\hat{\mu}_2 = \bar{y}$. Оценка максимального правдоподобия для σ^2 равна

$$\frac{S_{xx} + S_{yy}}{m+n}, \quad \text{где} \quad S_{xx} = \sum_i (x_i - \bar{x})^2, \quad S_{yy} = \sum_j (y_j - \bar{y})^2.$$

(Соответствующие вычисления аналогичны тем, которые привели к оценке $\hat{\sigma}^2 = S_{xx}/n$ в случае одной выборки.) Следовательно, при выполнении гипотезы H_1 мы имеем

$$\max[f(\boldsymbol{x}; \mu_1, \sigma^2) f(\boldsymbol{y}; \mu_2, \sigma^2) \colon \mu_1, \mu_2 \in \mathbb{R}, \sigma^2 > 0] = \\ = \left(2\pi \frac{S_{xx} + S_{yy}}{m+n}\right)^{-(m+n)/2} e^{-(m+n)/2}.$$

При выполнении гипотезы H_0 о. м. п. для μ будет, как и ранее, иметь вид

$$\hat{\mu} = \frac{m\bar{x} + n\bar{y}}{m+n},$$

а о. м. п. для σ^2 является величина

$$\hat{\sigma}^2 = \frac{1}{m+n}\left(\sum_i (x_i - \hat{\mu})^2 + \sum_j (y_j - \hat{\mu})^2\right) = \frac{1}{m+n}\left(S_{xx} + S_{yy} + \frac{mn}{m+n}(\bar{x} - \bar{y})^2\right).$$

Следовательно, при выполнении гипотезы H_0 мы имеем

$$\max[f(\boldsymbol{x}; \mu, \sigma^2) f(\boldsymbol{y}; \mu, \sigma^2) \colon \mu \in \mathbb{R}, \sigma^2 > 0] = \\ = \left(\frac{2\pi}{m+n}\left(S_{xx} + S_{yy} + \frac{mn}{m+n}(\bar{x} - \bar{y})^2\right)\right)^{-(m+n)/2} e^{-(m+n)/2}.$$

Это приводит к следующему выражению для о. о. п.:

$$\Lambda_{H_1;H_0} = \left(\frac{(m+n)(S_{xx} + S_{yy}) + mn(\bar{x} - \bar{y})^2}{(m+n)(S_{xx} + S_{yy})}\right)^{(m+n)/2} = \\ = \left(1 + \frac{mn(\bar{x} - \bar{y})^2}{(m+n)(S_{xx} + S_{yy})}\right)^{(m+n)/2}.$$

Таким образом, применяя к. о. о. п., мы отклоняем гипотезу H_0, если

$$\frac{|\bar{x}-\bar{y}|}{\sqrt{(S_{xx}+S_{yy})(1/m+1/n)}}$$

велико. Удобно умножить последнее выражение на $\sqrt{n+m-2}$. Это приводит к следующей статистике:

$$T=\frac{\bar{X}-\bar{Y}}{\sqrt{\dfrac{S_{XX}+S_{YY}}{m+n-2}\left(\dfrac{1}{m}+\dfrac{1}{n}\right)}}\sim \mathrm{t}_{n+m-2}.$$

Действительно, при выполнении гипотезы H_0 мы имеем

$$\frac{\bar{X}-\bar{Y}}{\sigma\sqrt{1/m+1/n}}\sim \mathrm{N}(0,\,1),\qquad \frac{1}{\sigma^2}S_{XX}\sim\chi^2_{m-1}\quad\text{и}\quad\frac{1}{\sigma^2}S_{YY}\sim\chi^2_{n-1},$$

и эти величины независимы. Поэтому

$$\frac{S_{XX}+S_{YY}}{\sigma^2}\sim\chi^2_{m+n-2}.$$

Утверждение теоремы Уилкса, конечно, выполняется асимптотически при $m,\ n\to\infty$, но этого нам здесь и не потребовалось.

Пример 2.7.1. Семена некоторого вида растений случайным образом высеивали либо в обогащенную почву (опытная группа), либо в стандартную (контрольная группа). В определенный срок все растения были собраны, высушены и взвешены, полученный вес в граммах указан в табл. 2.10.

Таблица 2.10

Контрольная группа	4,17	5,58	5,18	6,11	4,50	4,61	5,17	4,53	5,33	5,14
Опытная группа	4,81	4,17	4,41	3,59	5,87	3,83	6,03	4,89	4,32	4,69

Здесь контрольные наблюдения $X_1,\ \ldots,\ X_m$ являются н. о. р. $\mathrm{N}(\mu_X,\ \sigma^2)$-с. в., а опытные наблюдения $Y_1,\ \ldots,\ Y_n$ — н. о. р. $\mathrm{N}(\mu_Y,\ \sigma^2)$-с. в. Проверяется гипотеза

$$H_0\colon \mu_X=\mu_Y$$

против альтернативы

$$H_1\colon \mu_X\neq\mu_Y.$$

Находим

$$m=n=10,\quad \bar{x}=5{,}032,\quad S_{xx}=3{,}060,\quad \bar{y}=4{,}661,\quad S_{yy}=5{,}669.$$

Тогда

$$\hat{\sigma}^2=\frac{S_{xx}+S_{yy}}{m+n-2}=0{,}485$$

и

$$|t| = \frac{|\bar{x} - \bar{y}|}{\sqrt{\hat{\sigma}^2(1/m + 1/n)}} = 1{,}19.$$

Получив значение $t_{18}(0{,}025) = 2{,}101$, мы не отклоняем гипотезу H_0 при уровне 95%, утверждая, следовательно, что различные условия выращивания растений не приводят к различным значениям среднего веса.

Предположим теперь, что значение дисперсии σ^2 известно и равно 0,480 и не оценивается по выборке. Вычисляя величину

$$z = |\bar{x} - \bar{y}| \Big/ \sigma\sqrt{\frac{1}{m} + \frac{1}{n}},$$

находим значение $z = 1{,}1974$. Поскольку $\Phi(z) = 0{,}8847$, что меньше чем 0,975, у нас все еще нет оснований отклонить H_0. □

1в. Проверка гипотезы о равенстве дисперсий. Вначале рассмотрим случай, когда μ_1 и μ_2 известны (но не обязательно равны между собой). Рассмотрим нулевую гипотезу H_0: $\sigma_1^2 = \sigma_2^2$ и альтернативу H_1 : $\sigma_1^2 \neq \sigma_2^2$. При выполнении гипотезы H_0 функция правдоподобия $f(\boldsymbol{x};\, \mu_1,\, \sigma^2)f(\boldsymbol{y};\, \mu_2,\, \sigma^2)$ достигает максимума при $\hat{\sigma}^2 = (\Sigma_{xx}^2 + \Sigma_{yy}^2)/(m + n)$, и это максимальное значение равно

$$\left[\frac{2\pi(\Sigma_{xx}^2 + \Sigma_{yy}^2)}{m + n}\right]^{-(m+n)/2} e^{-(m+n)/2}.$$

Здесь $\Sigma_{xx}^2 = \sum\limits_i (x_i - \mu_1)^2$ и $\Sigma_{yy}^2 = \sum\limits_j (y_j - \mu_2)^2$.

При выполнении альтернативной гипотезы H_1 функция правдоподобия $f(\boldsymbol{x};\, \mu_1,\, \sigma^2)f(\boldsymbol{y};\, \mu_2,\, \sigma^2)$ достигает максимума при $\hat{\sigma}_1^2 = \Sigma_{xx}^2/m$, $\hat{\sigma}_2^2 = \Sigma_{yy}^2/n$ и максимальное значение равно

$$\left(\frac{2\pi\Sigma_{xx}^2}{m}\right)^{-m/2} \left(\frac{2\pi\Sigma_{yy}^2}{n}\right)^{-n/2} e^{-(m+n)/2}.$$

Тогда о. о. п. имеет вид

$$\Lambda_{H_1;H_0} = \frac{m^{m/2} n^{n/2}}{(m + n)^{(m+n)/2}} \left(\frac{\Sigma_{xx}^2 + \Sigma_{yy}^2}{\Sigma_{xx}^2}\right)^{m/2} \left(\frac{\Sigma_{xx}^2 + \Sigma_{yy}^2}{\Sigma_{yy}^2}\right)^{n/2} \propto$$

$$\propto \left(1 + \frac{\Sigma_{yy}^2}{\Sigma_{xx}^2}\right)^{m/2} \left(1 + \frac{\Sigma_{xx}^2}{\Sigma_{yy}^2}\right)^{n/2}.$$

Последнее выражение велико, когда $\Sigma_{yy}^2/\Sigma_{xx}^2$ велико или мало. А при выполнении гипотезы H_0 мы имеем

$$\frac{1}{\sigma^2}\Sigma_{XX}^2 \sim \chi_m^2, \qquad \frac{1}{\sigma^2}\Sigma_{YY}^2 \sim \chi_n^2, \qquad \text{и эти величины независимы.}$$

Таким образом, при выполнении гипотезы H_0 получаем $\Sigma^2_{YY}/\Sigma^2_{XX} \sim \mathrm{F}_{n,m}$. Следовательно, мы отклоняем гипотезу H_0 при уровне значимости α, если значение $\Sigma^2_{yy}/\Sigma^2_{xx}$ либо больше чем $\varphi^+_{n,m}(\alpha/2)$, либо меньше чем $\varphi^-_{n,m}(\alpha/2)$. Здесь и далее $\varphi^+_{n,m}(\gamma)$ обозначает верхнюю, а $\varphi^-_{n,m}(\gamma)$ — нижнюю γ-точку (квантиль) распределения Фишера $\mathrm{F}_{n,m}$, т. е. значение a, для которого

$$\int_a^\infty f_{\mathrm{F}_{n,m}}(x)\,dx = \gamma \quad \text{или} \quad \int_0^a f_{\mathrm{F}_{n,m}}(x)\,dx = \gamma,$$

и $\varphi^-_{n,m}(\gamma) = \varphi^+_{n,m}(1-\gamma)$, $0 < \gamma < 1$.

И вновь строгое применение к. о. о. п. приводит к несколько иной критической области (двустороннему несимметричному критерию).

Предположим теперь, что μ_1 и μ_2 — неизвестные мешающие параметры, и проверим гипотезу H_0: $\sigma_1^2 = \sigma_2^2$ против альтернативы H_1 : $\sigma_1^2 \neq \sigma_2^2$. При выполнении гипотезы H_0 функция правдоподобия $f(\boldsymbol{x};\, \mu_1,\, \sigma^2) f(\boldsymbol{y};\, \mu_2,\, \sigma^2)$ достигает максимума в точке

$$\hat{\mu}_1 = \bar{x}, \quad \hat{\mu}_2 = \bar{y}, \quad \hat{\sigma}^2 = \frac{1}{m+n}(S_{xx} + S_{yy})$$

и максимальное значение равно

$$\left(2\pi\frac{S_{xx}+S_{yy}}{m+n}\right)^{-(m+n)/2}.$$

При выполнении гипотезы H_1 функция правдоподобия равна

$$f(\boldsymbol{x};\, \mu_1,\, \sigma_1^2) f(\boldsymbol{y};\, \mu_2,\, \sigma_2^2).$$

Ее максимальное значение достигается при

$$\hat{\mu}_1 = \bar{x}, \quad \hat{\mu}_2 = \bar{y}, \quad \hat{\sigma}_1^2 = \frac{1}{m}S_{xx}, \quad \hat{\sigma}_2^2 = \frac{1}{n}S_{yy}$$

и равно

$$\left(\frac{1}{2\pi S_{xx}/m}\right)^{m/2}\left(\frac{1}{2\pi S_{yy}/n}\right)^{n/2} e^{-(m+n)/2}.$$

Тогда

$$\Lambda_{H_1;H_0} = \left(\frac{S_{xx}+S_{yy}}{S_{xx}}\right)^{m/2}\left(\frac{S_{xx}+S_{yy}}{S_{yy}}\right)^{n/2} \frac{m^{m/2}n^{n/2}}{(m+n)^{(m+n)/2}},$$

и мы отклоняем гипотезу H_0, если

$$\left(1+\frac{S_{xx}}{S_{yy}}\right)^{n/2}\left(1+\frac{S_{yy}}{S_{xx}}\right)^{m/2} \quad \text{велико.}$$

Однако, как следует из примера 1.1.4 (см. уравнение (1.1.7)),

$$\frac{S_{XX}/(m-1)}{S_{YY}/(n-1)} \sim \mathrm{F}_{m-1,n-1}.$$

Таким образом, при уровне значимости α мы отклоняем гипотезу H_0, используя «верхний» односторонний критерий, когда

$$\frac{(n-1)S_{xx}}{(m-1)S_{yy}} > \varphi^+_{m-1,n-1}(\alpha),$$

а при использовании «нижнего» одностороннего критерия — когда

$$\frac{(m-1)S_{yy}}{(n-1)S_{xx}} > \varphi^+_{n-1,m-1}(\alpha).$$

Эти критерии определяются верхними α-точками соответствующих F-распределений. Мы также можем использовать «двусторонний» критерий, при котором гипотеза H_0 отклоняется, если, скажем,

$$\frac{(n-1)S_{xx}}{(m-1)S_{yy}} > \varphi^+_{m-1,n-1}(\alpha/2) \quad \text{или} \quad \frac{(m-1)S_{yy}}{(n-1)S_{xx}} > \varphi^+_{n-1,m-1}(\alpha/2).$$

В каждом конкретном случае выбор критической области мотивируется видом графика п. р. в. $f_{\mathrm{F}_{m,n}}$ для заданных значений m и n (см. рис. 1.3 на с. 250).

F-критерий для проверки равенства дисперсий был предложен Фишером.

Замечание. Можно повторить эту процедуру проверки гипотез в случае, когда альтернатива H_1 состоит не в том, что $\sigma_1^2 \neq \sigma_2^2$, а в том, что $\sigma_1^2 > \sigma_2^2$ (в этом случае говорят о *сравнении дисперсий*). Тогда о. о. п. $\Lambda_{H_1;H_0}$ имеет вид

$$\begin{cases} \left(\dfrac{S_{xx}+S_{yy}}{S_{xx}}\right)^{m/2}\left(\dfrac{S_{xx}+S_{yy}}{S_{yy}}\right)^{n/2}\dfrac{m^{m/2}n^{n/2}}{(m+n)^{(m+n)/2}}, & \text{если } \dfrac{1}{m}S_{xx} > \dfrac{1}{n}S_{yy}, \\ 1, & \text{если } \dfrac{1}{m}S_{xx} < \dfrac{1}{n}S_{yy}, \end{cases}$$

и при уровне значимости α мы отклоняем H_0, когда

$$\frac{(n-1)S_{xx}}{(m-1)S_{yy}} > \varphi^+_{m-1,n-1}(\alpha).$$

Аналогичная модификация существует и для других случаев, рассмотренных выше.

F-критерий полезно применять в случае, когда $\boldsymbol{X}$ берется из распределения $\mathrm{Exp}(\lambda)$, а $\boldsymbol{Y}$ — из распределения $\mathrm{Exp}(\mu)$ и проверяется гипотеза $H_0\colon \lambda = \mu$.

В контексте к. о. о. п. F-критерий появляется также и в случае, когда мы проверяем гипотезу $H_0\colon \mu_1 = \ldots = \mu_n$, где $X_i \sim \mathrm{N}(\mu_i, \sigma^2)$ — нормальные с. в. с одной и той же дисперсией σ^2.

Пример 2.7.2. Для определения концентрации никеля в растворе используют спирт (спиртовой метод) или воду (водный метод). Необходимо проверить, приводит ли использование спирта к большей дисперсии, чем использование воды. Полученные значения концентрации никеля (в десятых долях процента) показаны в табл. 2.11.

Таблица 2.11

Спиртовой метод	4,28	4,32	4,25	4,29	4,31	4,35	4,32	4,33	4,28	4,27	4,38	4,28
Водный метод	4,27	4,32	4,29	4,30	4,31	4,30	4,30	4,32	4,28	4,32		

Здесь рассматривается следующая модель: величины $X_1, \ldots, X_{12}$, полученные при применении спиртового метода, и $Y_1, \ldots, Y_{10}$, полученные при применении водного метода, независимы, и $X_i \sim \mathrm{N}(\mu_X, \sigma_X^2)$, $Y_j \sim \mathrm{N}(\mu_Y, \sigma_Y^2)$. Проверяется нулевая гипотеза $H_0\colon \sigma_X^2 = \sigma_Y^2$ против альтернативы $H_1\colon \sigma_X^2 > \sigma_Y^2$.

При этом

$$m = 12, \quad \bar{x} = 4{,}311, \quad S_{xx} = 0{,}01189$$

и

$$n = 10, \quad \bar{y} = 4{,}301, \quad S_{yy} = 0{,}00269.$$

Находим $(S_{xx}/11)/(S_{yy}/9) = 3{,}617$. Из процентных таблиц для χ^2-распределения $\varphi^+_{11,9}(0{,}05) = 3{,}10$ и $\varphi^+_{11,9}(0{,}01) = 5{,}18$. Таким образом, мы отклоняем гипотезу H_0 при уровне значимости 5%, но принимаем ее при уровне значимости 1%. Это дает некоторые основания утверждать, что $\sigma_X^2 > \sigma_Y^2$, но чтобы достичь большей степени определенности, необходимы дальнейшие исследования. Например, так как $\varphi^+_{11,9}(0{,}025) \approx 3{,}92$, гипотезу H_0 следует принять при уровне 2,5%. □

2. Неоднородные нормальные выборки. Пусть выборка $\boldsymbol{X} = (X_1, \ldots, X_n)$ имеет компоненты $X_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$ с параметрами, которые меняются от одной величины к другой.

2а. Проверка гипотезы о равенстве средних значений при известных дисперсиях. Предположим, что $\sigma_1^2, \ldots, \sigma_n^2$ известны (не обязательно равны между собой) и мы проверяем гипотезу

$$H_0\colon \mu_1 = \ldots = \mu_n$$

против

$$H_1\colon \mu_i \in \mathbb{R} \text{ произвольные.}$$

Воспользуемся опять к. о. о. п. При выполнении гипотезы H_1 функция правдоподобия равна

$$f(\boldsymbol{x}; \mu_1, \ldots, \mu_n; \sigma_1^2, \ldots, \sigma_n^2) = \left(\frac{1}{2\pi}\right)^{n/2} \frac{1}{\prod_i \sigma_i} \exp\left[-\frac{1}{2}\sum_i (x_i - \mu_i)^2/\sigma_i^2\right].$$

Она достигает максимума при $\hat{\mu}_i = x_i$, и ее максимальное значение равно $(2\pi)^{-n/2} \Big/ \Big(\prod_i \sigma_i\Big)$.

При выполнении гипотезы H_0 функция правдоподобия равна

$$f(\boldsymbol{x};\, \mu, \sigma_1^2, \ldots, \sigma_n^2) = \left(\frac{1}{2\pi}\right)^{n/2} \frac{1}{\prod_i \sigma_i} \exp\left[-\frac{1}{2}\sum_i (x_i - \mu)^2/\sigma_i^2\right]$$

и достигает максимального значения

$$\left(\frac{1}{2\pi}\right)^{n/2} \prod_i \frac{1}{\sigma_i} \exp\left[-\frac{1}{2}\sum_i (x_i - \hat{\mu})^2/\sigma_i^2\right]$$

при значении взвешенного среднего

$$\hat{\mu} = \hat{\mu}(\boldsymbol{x}) = \sum_i x_i/\sigma_i^2 \Big/ \sum_i 1/\sigma_i^2.$$

Тогда о. о. п. равно

$$\Lambda_{H_1;H_0} = \exp\left[\frac{1}{2}\sum_i (x_i - \hat{\mu})^2/\sigma_i^2\right].$$

Мы отклоняем гипотезу H_0, если сумма $\sum_i (x_i - \hat{\mu})^2/\sigma_i^2$ принимает большие значения. Более точно,

$$2\ln \Lambda_{H_1;H_0} = \sum_i \frac{(X_i - \hat{\mu}(\boldsymbol{X}))^2}{\sigma_i^2} \sim \chi_{n-1}^2$$

(еще один случай, когда распределение в теореме Уилкса является точным, а не приближенным). Таким образом, при уровне значимости α гипотеза H_0 отклоняется, когда $\sum_i (x_i - \hat{\mu}(\boldsymbol{x}))^2/\sigma_i^2$ превышает $h_{n-1}^{+}(\alpha)$, верхнюю α-точку распределения χ_{n-1}^2.

2б. Проверка гипотезы о равенстве средних значений при неизвестных дисперсиях.

Отношение статистика к дисперсии напоминает отношение евангелиста к греху; он видит его всюду, в большей или меньшей степени.

(Из серии «Почему их не понимают».)

Рассмотрим те же нулевую гипотезу

$$H_0\colon \mu_1 = \ldots = \mu_n$$

и альтернативу

$$H_1 : \mu_i \text{ произвольные}$$

в случае неизвестных дисперсий (которые могут быть равны или не равны между собой). Тогда при H_1 мы должны максимизировать функцию правдоподобия и по σ_i^2 тоже. Это означает, что в дополнение к условию $\hat{\mu}_i = x_i$ мы должны положить $\hat{\sigma}_i = 0$, что нереалистично. Это пример, когда процедура, основанная на о. о. п., неприменима.

Предположим, однако, что нормальная выборка $\boldsymbol{X}$ была разбита на несколько групп так, что хотя бы в одной из них имеется более чем одна с. в., и что известно, что в пределах группы средние значения одинаковы. Вдобавок пусть дисперсия будет одинакова для всех с. в. X_i. Тогда можно использовать процедуру, называемую ANOVA (analysis of variance — *анализ дисперсии*), для проверки нулевой гипотезы о том, что все средние значения одинаковы. По сути, это является обобщением критерия, рассмотренного выше в п. 2а.

Итак, предположим, что имеется k групп, по n_i наблюдений в группе i, и $n = n_1 + \ldots + n_k$. Положим

$$X_{ij} = \mu_i + \varepsilon_{ij}, \quad j = 1, \ldots, n_i, \quad i = 1, \ldots k.$$

Предположим, что $\mu_1, \ldots, \mu_k$ — это фиксированные неизвестные константы и величины $\varepsilon_{ij} \sim \mathrm{N}(0, \sigma^2)$ независимы. Дисперсия σ^2 неизвестна. Рассматривается нулевая гипотеза

$$H_0\colon \mu_1 = \ldots = \mu_k = \mu$$

и альтернатива $H_1 : \mu_i$ произвольны.

При выполнении гипотезы H_1 функция правдоподобия

$$\frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \mu_i)^2\right]$$

достигает максимума при

$$\hat{\mu}_i = \bar{x}_{i+}, \ \hat{\sigma}^2 = \frac{1}{n} \sum_i \sum_j (x_{ij} - \bar{x}_{i+})^2, \quad \text{где} \ \ \bar{x}_{i+} = \frac{1}{n_i} \sum_j x_{ij}.$$

Максимальное значение равно

$$\left[2\pi\frac{1}{n} \sum_i \sum_j (x_{ij} - \bar{x}_{i+})^2\right]^{-n/2} e^{-n/2}.$$

Удобно обозначить

$$s_1 = \sum_i \sum_j (x_{ij} - \bar{x}_{i+})^2;$$

эту сумму называют *внутригрупповой суммой квадратов*. Если предположить, что хотя бы одно из чисел $n_1, \ldots, n_k$, скажем n_i, больше 1, то соответствующая сумма $\sum\limits_j (X_{ij} - \bar{X}_{i+})^2$ будет положительной с вероятностью 1.

Тогда случайная величина $S_1 = \sum_{i,j}(X_{ij} - \bar{X}_{i+})^2$, соответствующая внутригрупповой сумме квадратов, будет также положительной с вероятностью 1.

При выполнении гипотезы H_0 функция правдоподобия

$$\frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \mu)^2\right)$$

достигает максимума при

$$\hat{\mu} = \bar{x}_{++}, \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_i\sum_j(x_{ij} - \bar{x}_{++})^2,$$

и это максимальное значение равно

$$\left(2\pi\frac{1}{n}\sum_i\sum_j(x_{ij} - \bar{x}_{++})^2\right)^{-n/2} e^{-n/2},$$

где

$$\bar{x}_{++} = \frac{1}{n}\sum_{i,j} x_{ij}.$$

Обозначим

$$s_0 = \sum_{i,j}(x_{ij} - \bar{x}_{++})^2;$$

это *общая сумма квадратов.*

Обобщенное отношение правдоподобия равно

$$\Lambda_{H_1;H_0} = \left(\frac{s_0}{s_1}\right)^{n/2}.$$

Запишем s_0 в виде

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{i+} + \bar{x}_{i+} - \bar{x}_{++})^2 = \sum_{i,j}(x_{ij} - \bar{x}_{i+})^2 + \sum_i n_i(\bar{x}_{i+} - \bar{x}_{++})^2$$

(поскольку $\sum(x_{ij} - \bar{x}_{i+}) = 0$ для всех $i = 1, \ldots, k$, сумма попарных произведений обращается в нуль). Затем обозначим

$$s_2 = \sum_i n_i(\bar{x}_{i+} - \bar{x}_{++})^2;$$

эту сумму называют *межгрупповой суммой квадратов.*

Следовательно, $s_0 = s_1 + s_2$, и

$$\Lambda_{H_1;H_0} = \left(1 + \frac{s_2}{s_1}\right)^{n/2}.$$

Таким образом, мы отклоняем гипотезу H_0, когда s_2/s_1 принимает большие значения, или, что, эквивалентно, когда величина

$$\frac{s_2/(k-1)}{s_1/(n-k)}$$

принимает большие значения.

Теперь анализ распределений величин X_{ij}, $\bar{X}_{i+}$ и $\bar{X}_{++}$ распадается на три этапа.

Во-первых, для любого i при справедливости обеих гипотез H_0 и H_1 мы имеем

$$\frac{1}{\sigma^2}\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_{i+})^2 \sim \chi^2_{n_i-1}$$

согласно теореме Фишера (см. § 1.5). Затем, суммируя независимые с. в., имеющие χ^2-распределение, находим

$$\frac{S_1}{\sigma^2}=\frac{1}{\sigma^2}\sum_i\sum_j(X_{ij}-\bar{X}_{i+})^2 \sim \chi^2_{n-k}.$$

Далее, для любого i и при H_0, и при H_1 мы вновь получаем, что

$$\sum_{j=1}^{n_i}(X_{ij}-\bar{X}_{i+})^2 \quad \text{и} \quad \bar{X}_{i+} \quad \text{независимы,}$$

на основании теоремы Фишера. Тогда

$$S_1=\sum_{i,j}(X_{ij}-\bar{X}_{i+})^2 \quad \text{и} \quad S_2=\sum_i(X_{i+}-\bar{X}_{++})^2 \quad \text{независимы.}$$

Наконец, если выполняется гипотеза H_0, то X_{ij} являются н. о. р. $\mathrm{N}(\mu, \sigma^2)$-с. в. Тогда

$$\bar{X}_{++} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right) \quad \text{и} \quad \frac{n}{\sigma^2}(\bar{X}_{++}-\mu)^2 \sim \chi^2_1.$$

Далее, $\bar{X}_{i+}$ — это независимые $\mathrm{N}(\mu, \sigma^2/n_i)$-с. в. и $n_i(\bar{X}_{i+}-\mu)^2/\sigma^2 \sim \chi^2_1$. Следовательно,

$$\sum_i \frac{n_i}{\sigma^2}(\bar{X}_{i+}-\mu)^2 \sim \chi^2_k.$$

Более того, величины $\bar{X}_{++}$ и $\bar{X}_{i+}-\bar{X}_{++}$ независимы, поскольку они являются совместно нормальными с. в. и $\mathrm{Cov}(\bar{X}_{++}, \bar{X}_{i+}-\bar{X}_{++})=0$. Записав

$$\sum_i \frac{n_i}{\sigma^2}(\bar{X}_{i+}-\mu)^2=\frac{1}{\sigma^2}\sum_i n_i(\bar{X}_{i+}-\bar{X}_{++})^2+\frac{n}{\sigma^2}(\bar{X}_{++}-\mu)^2,$$

приходим к заключению, что в правой части

$$\frac{1}{\sigma^2}\sum_i n_i(\bar{X}_{i+}-\bar{X}_{++})^2=\frac{S_2}{\sigma^2}\sim\chi^2_{k-1} \quad \text{и} \quad \frac{n}{\sigma^2}(\bar{X}_{++}-\mu)^2\sim\chi^2_1$$

и эти величины независимы. С другой стороны, запишем S_2 в виде

$$S_2 = \sum_i n_i((\overline{X}_{i+} - \mu_i) + (\mu_i - \bar{\mu}) + (\bar{\mu} - \overline{X}_{++}))^2,$$

где $\bar{\mu} = \sum\limits_i \mu_i n_i / n$. Тогда прямые вычисления показывают, что

$$\mathsf{E} S_2 = (k-1)\sigma^2 + \sum_i n_i(\mu_i - \bar{\mu})^2.$$

Приходим к заключению, что при справедливости гипотезы H_1 величина S_2 имеет тенденцию к увеличению.

В итоге видим, что статистика

$$Q = \frac{S_2/(k-1)}{S_1/(n-k)}$$

распределена при H_0 как $\mathrm{F}_{k-1,n-k}$ и будет принимать бóльшие значения при H_1. Следовательно, при уровне значимости α мы отклоняем H_0, когда значение Q больше величины $\varphi^+_{k-1,n-k}(\alpha)$, верхней α-точки $\mathrm{F}_{k-1,n-k}$-распределения. Итог представлен в табл. 2.12.

Таблица 2.12

	Степени свободы	Сумма квадратов	Среднее квадратичное
Между группами	$k-1$	s_2	$s_2/(k-1)$
Внутри групп	$n-k$	s_1	$s_1/(n-k)$
Общее значение	$n-1$	s_0	

Пример 2.7.3. При проведении психологического исследования методов обучения математике в школе 45 учеников случайным образом были разбиты на 5 групп по 9 человек. Группы A и B обучались в отдельных классах обычным методом, а группы C, D и E обучались вместе. Ежедневно каждого ученика из группы C публично хвалили, каждого ученика из группы D публично порицали, а учеников из группы E вообще игнорировали. В конце эксперимента все ученики прошли тестирование и их результаты, выраженные в процентном отношении к высшему баллу, показаны в табл. 2.13.

Таблица 2.13

A (контроль)	34	28	48	40	48	46	32	30	48
B (контроль)	42	46	26	38	26	38	40	42	32
C (похвала)	56	60	58	48	54	60	56	56	46
D (порицание)	38	56	52	52	38	48	48	46	44
E (игнорирование)	42	28	26	38	30	30	20	36	40

Психологов интересует вопрос, есть ли существенная разница между группами. Величины X_{ij} предполагаются независимыми, и $X_{ij} \sim \mathrm{N}(\mu_i, \sigma^2)$, $i = 1, \ldots, 5$, $j = 1, \ldots, 9$. Рассматривается нулевая гипотеза H_0: $\mu_i \equiv \mu$ и альтернатива H_1: μ_i произвольные.

На основании результатов тестирования составим табл. 2.14 и 2.15.

Таблица 2.14

	Всего	$\bar{x}$	$\sum_j (x_{ij} - \bar{x}_{i+})^2$
A	354	39,3	568
B	330	36,7	408
C	494	54,9	192,9
D	422	46,9	304,9
E	290	32,2	419,6

Таблица 2.15

	Степени свободы	Сумма квадратов	Среднее квадратичное	F-отношение
Между группами	4	2891,48	722,869	722,869/47,33 = 15,273
Внутри групп	40	1893,3	47,33	
Общее значение	44	4784,78		

Наблюдаемое значение статистики Q равно 15,3. Так как $\varphi^+_{4,40}(0{,}001) = 5{,}7$, это значение попадает в глубину области отклонения гипотезы при уровне $\alpha = 0{,}001$. Следовательно, гипотеза H_0 отвергается, и мы приходим к заключению, что психологическое воздействие на детей оказывает сильное влияние на их успехи в математике. □

Statisticians are fond of curvilinear shapes and often own as a pet a large South American snake called ANOCOVA.

Статистики увлекаются искривленными формами и поэтому часто держат дома большую южноамериканскую змею под названием АНОКОВА.

(Из серии «Почему их не понимают».)

Nowadays in any Grand Slam you will be tested by an -Ova.

В наше время, чтобы выйти в финал, нужно пройти проверку на -ова.

(Разговор о русских спортсменках в кулуарах Уимблдонского женского турнира.)

2в. Проверка гипотезы о равенстве дисперсий при известном среднем значении. Предположим теперь, что $\mu_1 = \ldots = \mu_n = \mu$ известно. Попытаемся проверить гипотезу H_0: $\sigma_1^2 = \ldots = \sigma_n^2$ против альтернативы

$H_1: \sigma_i^2 > 0$ не связаны ограничениями. При выполнении гипотезы H_1 функция правдоподобия равна

$$f(\boldsymbol{x}; \mu, \sigma_1^2, \ldots, \sigma_n^2) = \left(\frac{1}{2\pi}\right)^{n/2} \frac{1}{\prod\limits_i \sigma_i} \exp\left[-\frac{1}{2}\sum_i (x_i - \mu)^2/\sigma_i^2\right]$$

и достигает максимума

$$\frac{1}{(2\pi)^{n/2}\left(\prod\limits_i (x_i - \mu)^2\right)^{1/2}} e^{-n/2}$$

при $\hat{\sigma}_i^2 = (x_i - \mu)^2$. При выполнении гипотезы H_0 функция правдоподобия достигает максимума при $\hat{\sigma}^2 = \sum\limits_i (x_i - \mu)^2/n$ и максимальное значение равно

$$\frac{1}{(2\pi)^{n/2}\left(\sum\limits_i (x_i - \mu)^2/n\right)^{n/2}} e^{-n/2}.$$

Это приводит к о. о. п.

$$\Lambda_{H_1;H_0} = \left(\frac{\prod\limits_i (x_i - \mu)^2}{\left(\sum\limits_i (x_i - \mu)^2/n\right)^n}\right)^{-1/2}.$$

Трудность здесь заключается в отыскании распределения о. о. п. при справедливости гипотезы H_0. Вследствие этого не существует эффективного критерия проверки нулевой гипотезы. Однако эта задача является очень важной во многих приложениях, в частности, в современной финансовой математике. В финансовой математике дисперсия приобретает большое значение (и рассматривается как существенно мешающий фактор).

§ 2.8. Линейная регрессия. Оценки метода наименьших квадратов

Ordinary Least Squares People[1]

(Из серии «Фильмы, которые не вышли на большой экран».)

Нашей следующей темой будет *линейная регрессия*. В рассматриваемой модели интересующая нас величина Y имеет вид $Y = g(x) + \varepsilon$, где а) ε — с. в. с нулевым средним (например, $\varepsilon \sim \mathrm{N}(0, \sigma^2)$ с известной или неизвестной дисперсией), б) $g(x) = g(x, \theta)$ — функция заданного вида (например, линейная функция $\gamma + \beta x$ или экспоненциальная функция

[1] Ср. название фильма «Ordinary People».

$ke^{\alpha x}$ (которая превращается в линейную после логарифмирования), при этом все или некоторые компоненты параметра θ неизвестны), и в) x — заданное значение (или, иногда, случайная величина). Главным образом будем рассматривать многомерную постановку задачи, при которой Y и ε заменяются случайными векторами

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \quad \text{и} \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix},$$

а функция $g(x)$ заменяется векторной функцией $g(\boldsymbol{x})$ векторного аргумента $\boldsymbol{x}$. Наша цель — определить функцию $g(x)$ или $g(\boldsymbol{x})$, т. е. оценить значения неизвестных параметров.

Интересный пример того, как можно использовать линейную регрессию, связан с таким фактом из астрономии, как закон Хаббла о линейном расширении. В 1929 г., Е. П. Хаббл (1889—1953), американский астроном, опубликовал важную статью, в которой утверждалось, что Вселенная расширяется (т. е. что галактики удаляются от Земли и чем более отдаленной от Земли является галактика, тем с большей скоростью она удаляется). Коэффициент пропорциональности, который появляется в этой линейной зависимости, был назван постоянной Хаббла, и его вычисление стало одной из центральных задач в астрономии: знание его позволило бы нам оценить возраст Вселенной.

Для решения этой задачи используют линейную регрессию, поскольку имеющиеся данные весьма скудны; соответствующие измерения относительно галактик являются длительной процедурой и имеют свои ограничения. Начиная с 1929 г. несколько раз проводились тщательные вычисления, и значение постоянной Хаббла несколько раз переоценивалось. Каждый новый цикл вычислений давал в результате все более ранний возраст возникновения Вселенной; было бы интересно увидеть, сохранится ли такой тренд и в дальнейшем.

Сосредоточимся главным образом на *простой линейной регрессии*, когда каждая компонента Y_i вектора $\boldsymbol{Y}$ определяется равенством

$$Y_i = \gamma + \beta x_i + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{2.8.1}$$

и $\varepsilon_1, \ldots, \varepsilon_n$ — это н. о. р. с. в. с $\mathsf{E}\varepsilon_i = 0$, $\operatorname{Var}\varepsilon_i = \sigma^2$. Здесь γ и β неизвестны, а $\boldsymbol{x} = (x_1, \ldots, x_n)$ — заданный вектор. Случайные величины ε_i интерпретируются как «шум», и их часто называют ошибками (наблюдений или измерений). Тогда, конечно, $Y_1, \ldots, Y_n$ независимы, $\mathsf{E}Y_i = \gamma + \beta x_i$ и $\operatorname{Var}\varepsilon_i = \sigma^2$. Удобно ввести новую параметризацию в уравнении (2.8.1) следующим образом:

$$Y_i = \alpha + \beta(x_i - \bar{x}) + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{2.8.2}$$

где

$$\bar{x} = \frac{1}{n}\sum_i x_i, \quad \alpha = \gamma + \beta\bar{x} \quad \text{и} \quad \sum_i (x_i - \bar{x}) = 0.$$

В такой модели получаем данные $(x_1, y_1), \ldots, (x_n, y_n)$, причем значения x_i известны, а значения y_i — это реализации с. в. Y_i. Определить значения γ и β (или, что эквивалентно, значения α и β) означает нарисовать *линию регрессии*

$$y = \gamma + \beta x \quad (\text{или } y = \alpha + \beta(x - \bar{x})),$$

что является наилучшей аппроксимацией имеющихся данных. В случае, когда $\gamma = 0$, имеют дело с линией регрессии, проходящей через начало координат.

Вполне естественно возникает идея рассмотреть пару *оценок метода наименьших квадратов* (МНК-оценок) $\hat{\alpha}$ и $\hat{\beta}$ для α и β, т. е. значения, минимизирующие сумму

$$\sum (y_i - \alpha - \beta(x_i - \bar{x}))^2.$$

Эта сумма является мерой отклонения данных $(x_1, y_1), \ldots, (x_n, y_n)$ от теоретической прямой $y = (\alpha - \beta\bar{x}) + \beta x$. Решая уравнения для стационарных точек

$$\frac{\partial}{\partial\alpha}\sum (y_i - \alpha - \beta(x_i - \bar{x}))^2 = 0, \quad \frac{\partial}{\partial\beta}\sum (y_i - \alpha - \beta(x_i - \bar{x}))^2 = 0,$$

находим, что МНК-оценки имеют вид

$$\hat{\alpha} = \bar{y}, \quad \hat{\beta} = \frac{S_{xy}}{S_{xx}}. \tag{2.8.3}$$

Здесь

$$\begin{gathered} \bar{y} = \frac{1}{n}\sum_i y_i, \quad S_{xx} = \sum_i (x_i - \bar{x})^2, \\ S_{xy} = \sum_i (x_i - \bar{x}) y_i = \sum_i (x_i - \bar{x})(y_i - \bar{y}) \end{gathered} \tag{2.8.4}$$

и $\hat{\gamma} = \hat{\alpha} - \hat{\beta}\bar{x} = \bar{y} - S_{xy}\bar{x}/S_{xx}$. Очевидно, нужно предположить, что $S_{xx} \neq 0$, т. е. не все значения x_i равны между собой. Отметим, что

$$\bar{y} = \hat{\gamma} + \hat{\beta}\bar{x}, \quad \text{т. е.} \quad (\bar{x}, \bar{y}) \text{ лежит на линии регрессии.}$$

Кроме того, $\hat{\alpha}$ и $\hat{\beta}$ являются линейными функциями от $y_1, \ldots, y_n$.

Из последнего замечания непосредственно следует ряд свойств МНК-оценок, которые перечислены в нижеследующих утверждениях:

а) $\mathsf{E}\hat{\alpha} = \alpha$, $\mathsf{E}\hat{\beta} = \beta$;

б) $\operatorname{Var}\hat{\alpha} = \dfrac{\sigma^2}{n}$, $\operatorname{Var}\hat{\beta} = \dfrac{\sigma^2}{S_{xx}}$;

в) $\operatorname{Cov}(\hat{\alpha}, \hat{\beta}) = 0$;

г) *пара* МНК-*оценок* $(\hat{\alpha}, \hat{\beta})$ *является* наилучшей линейной несмещенной оценкой (используется также аббревиатура BLUE — от англ. «best linear unbiased estimator») *для* (α, β), *т. е. несмещенной оценкой с наименьшей дисперсией среди всех несмещенных оценок, линейных относительно* $Y_1, \ldots, Y_n$.

Действительно, утверждение а) следует из соотношений

$$\mathsf{E}\hat{\alpha} = \frac{1}{n}\sum(\alpha + \beta(x_i - \bar{x})) = \alpha$$

и

$$\mathsf{E}\hat{\beta} = \frac{1}{S_{xx}}\sum_i(x_i - \bar{x})\,\mathsf{E}Y_i = \frac{1}{S_{xx}}\sum_i(x_i - \bar{x})(\alpha + \beta(x_i - \bar{x})) = \beta.$$

Далее, утверждение б) получим так:

$$\operatorname{Var}\hat{\alpha} = \frac{1}{n^2}\sum \operatorname{Var} Y_i = \frac{\sigma^2}{n}$$

и

$$\operatorname{Var}\hat{\beta} = \frac{1}{S_{xx}^2}\sum_i(x_i - \bar{x})^2 \operatorname{Var} Y_i = \frac{\sigma^2}{S_{xx}}.$$

Утверждение в) есть следствие равенств

$$\operatorname{Cov}(\hat{\alpha}, \hat{\beta}) = \frac{1}{nS_{xx}}\sum \operatorname{Cov}(Y_i, (x_i - \bar{x})Y_i) = \frac{1}{nS_{xx}}\sigma^2\sum_i(x_i - \bar{x}) = 0.$$

Наконец, докажем утверждение г): пусть $\tilde{A} = \sum\limits_i c_iY_i$ и $\tilde{B} = \sum\limits_i d_iY_i$ — это линейные несмещенные оценки для α и β соответственно,

$$\sum_i c_i(\alpha + \beta(x_i - \bar{x})) \equiv \alpha \quad \text{и} \quad \sum_i d_i(\alpha + \beta(x_i - \bar{x})) \equiv \beta.$$

Тогда должны выполняться равенства

$$\sum_i c_i = 1, \quad \sum_i c_i(x_i - \bar{x}) = 0, \quad \sum_i d_i = 0 \quad \text{и} \quad \sum_i d_i(x_i - \bar{x}) = 1.$$

Минимизация $\operatorname{Var}\tilde{A} = \sigma^2 \sum\limits_i c_i^2$ и $\operatorname{Var}\tilde{B} = \sigma^2 \sum\limits_i d_i^2$ при указанных выше ограничениях сводится к минимизации функций Лагранжа

$$\mathcal{L}_1(c_1, \ldots, c_n; \alpha) = \sigma^2 \sum_i c_i^2 - \lambda\left(\sum_i c_i - 1\right) - \mu \sum_i c_i(x_i - \bar{x})$$

и

$$\mathcal{L}_2(d_1, \ldots, d_n; \beta) = \sigma^2 \sum_i d_i^2 - \lambda \sum_i d_i - \mu\left(\sum_i d_i(x_i - \bar{x}) - 1\right).$$

Значения, при которых достигается минимум обеих функций $\mathcal{L}_1$ и $\mathcal{L}_2$, таковы:

$$c_i = d_i = \frac{1}{2\sigma^2}(\lambda + \mu(x_i - \bar{x})).$$

Принимая во внимание соответствующие ограничения, получим, что

$$c_i = \frac{1}{n} \quad \text{и} \quad d_i = \frac{x_i - \bar{x}}{S_{xx}}, \qquad \text{т. е.} \quad \tilde{A} = \hat{\alpha} \quad \text{и} \quad \tilde{B} = \hat{\beta}.$$

§ 2.9. Линейная регрессия для нормальных распределений

Statisticians must stay away from children's toys because they regress easily.

Статистикам следует держаться подальше от детских игрушек, потому что они легко впадают в детство.

(Из серии «Почему их не понимают».)

Отметим, что до сих пор не введено никаких предположений о виде совместного распределения ошибок $\varepsilon_1, \ldots, \varepsilon_n$. Например, если $\varepsilon_i \sim \mathrm{N}(0, \sigma^2)$, то пара МНК-оценок $(\hat{\alpha}, \hat{\beta})$ совпадает с о. м. п. Действительно, в этом случае

$$Y_i \sim \mathrm{N}(\alpha + \beta(x_i - \bar{x}), \sigma^2)$$

и минимизация суммы

$$\sum_i (y_i - \alpha - \beta(x_i - \bar{x}))^2$$

эквивалентна максимизации логарифма правдоподобия

$$\ell(\boldsymbol{x}, \boldsymbol{y}; \mu, \sigma^2) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_i (y_i - \alpha - \beta(x_i - \bar{x}))^2.$$

Этот факт объясняет те аналогии с приведенными ранее утверждениями (например, теоремой Фишера), которые появляются при более детальном рассмотрении нормальной линейной регрессии.

В самом деле, предположим, что $\varepsilon_i \sim \mathrm{N}(0, \sigma^2)$ и величины ε_i независимы. Рассмотрим минимальное значение суммы $\sum\limits_i (Y_i - \alpha - \beta(x_i - \bar{x}))^2$:

$$R = \sum (Y_i - \hat{\alpha} - \hat{\beta}(x_i - \bar{x}))^2. \tag{2.9.1}$$

Его называют *остаточной суммой квадратов* (о. с. к.). Имеет место следующая теорема:

а) $\hat{\alpha} \sim \mathrm{N}\left(\alpha, \dfrac{\sigma^2}{n}\right)$;

б) $\hat{\beta} \sim \mathrm{N}\left(\beta, \dfrac{\sigma^2}{S_{xx}}\right)$;

в) $\dfrac{R}{\sigma^2} \sim \chi^2_{n-2}$;

г) *величины* $\hat{\alpha}$, $\hat{\beta}$ *и* R *являются независимыми*;

д) $\dfrac{R}{n-2}$ *является несмещенной оценкой для* σ^2.

Для доказательства утверждений а) и б) заметим, что $\hat{\alpha}$ и $\hat{\beta}$ являются линейными комбинациями независимых нормальных с. в. Y_i. Для доказательства утверждений в) и г) мы придерживаемся той же стратегии, что и при доказательстве теорем Фишера и Пирсона.

Пусть A — действительная ортогональная матрица размера $n \times n$, у которой первый и второй столбцы имеют вид

$$\begin{pmatrix} 1/\sqrt{n} \\ \vdots \\ 1/\sqrt{n} \end{pmatrix}, \qquad \begin{pmatrix} (x_1 - \bar{x})/\sqrt{S_{xx}} \\ \vdots \\ (x_n - \bar{x})/\sqrt{S_{xx}} \end{pmatrix},$$

а остальные столбцы произвольные (выбранные так, чтобы матрица была ортогональной). Такая матрица всегда существует: первый и второй столбцы являются ортогональными в силу равенства $\sum\limits_i (x_i - \bar{x}) = 0$, и всегда можно добавить к ним еще $n-2$ вектора так, чтобы получить ортогональный базис в $\mathbb{R}^n$. Рассмотрим теперь случайный вектор $\boldsymbol{Z} = A^T \boldsymbol{Y}$ с первыми двумя компонентами

$$Z_1 = \sqrt{n}\,\bar{Y} = \sqrt{n}\,\hat{\alpha} \sim \mathrm{N}(\sqrt{n}\,\alpha, \sigma^2)$$

и

$$Z_2 = \frac{1}{\sqrt{S_{xx}}} \sum_i (x_i - \bar{x}) Y_i = \sqrt{S_{xx}}\,\hat{\beta} \sim \mathrm{N}(\sqrt{S_{xx}}\,\beta, \sigma^2).$$

В силу ортогональности матрицы A все компоненты $Z_1, \ldots, Z_n$ являются независимыми нормальными с. в. с одинаковой дисперсией. Более того,

$$\sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n} Y_i^2.$$

С другой стороны,

$$\sum_{i=3}^{n} Z_i^2 = \sum_{i=1}^{n} Z_i^2 - Z_1^2 - Z_2^2 = \sum_{i=1}^{n} Y_i^2 - n\hat{\alpha}^2 - S_{xx}\hat{\beta}^2 =$$

$$= \sum_{i=1}^{n}(Y_i - \overline{Y})^2 + \hat{\beta}^2 S_{xx} - 2\hat{\beta} S_{xY} \sum_{i=1}^{n}(Y_i - \hat{\alpha} - \hat{\beta}(x_i - \bar{x}))^2 = R.$$

Следовательно, величины $\hat{\alpha}$, $\hat{\beta}$ и R являются независимыми.

Пусть $A = (A_{ij})$, тогда, во-первых, для любого $j \geqslant 2$ выполняется равенство $\sum\limits_i A_{ij} = 0$, так как столбцы с номерами $2, \ldots, n$ являются ортогональными к первому столбцу, и, во-вторых, для любого $j \geqslant 3$ выполняется равенство $\sum\limits_i A_{ij}(x_i - \bar{x}) = 0$, так как столбцы с номерами $3, \ldots, n$ являются ортогональными ко второму столбцу. Значит, для любого $j \geqslant 3$ мы имеем

$$\mathsf{E} Z_j = \sum_i \mathsf{E} Y_i A_{ij} = \sum_i (\alpha + \beta(x_i - \bar{x}))A_{ij} = \alpha \sum_i A_{ij} + \beta \sum_i (x_i - \bar{x}) A_{ij} = 0$$

и

$$\operatorname{Var} Z_j = \sigma^2 \sum_i a_{ij}^2 = \sigma^2.$$

Следовательно,

$$Z_3, \ldots, Z_n \sim \mathrm{N}(0, \sigma^2) \ \text{ и } \ Z_i,\ i = 3, \ldots, n, \text{ независимы.}$$

Но тогда $R/\sigma^2 \sim \chi^2_{n-2}$. Отсюда немедленно следует утверждение д): $\mathsf{E}R = (n-2)\sigma^2$.

Полезно помнить, что

$$R = \sum_{i=1}^{n} Y_i^2 - n\hat{\alpha}^2 - S_{xx}\hat{\beta}^2. \tag{2.9.2}$$

Теперь можно проверить нулевую гипотезу $H_0\colon \beta = \beta_0$ против альтернативы $H_1\colon \beta \neq \beta_0$. Действительно, получаем, что при выполнении гипотезы H_0 справедливы утверждения:

$$\hat{\beta} \sim \mathrm{N}\left(\beta_0, \frac{\sigma^2}{S_{xx}}\right) \quad \text{и} \quad \frac{1}{\sigma^2} R \sim \chi^2_{n-2}$$

независимы.

Тогда

$$T = \frac{\hat{\beta} - \beta_0}{\sqrt{R}/\sqrt{(n-2)S_{xx}}} \sim \mathrm{t}_{n-2}. \tag{2.9.3}$$

Следовательно, критерий размера α отклоняет гипотезу H_0 в случае, когда значение t величины $|T|$ превосходит $\mathrm{t}_{n-2}(\alpha/2)$, верхнюю $\alpha/2$-точку (квантиль) t_{n-2}-распределения. Часто $\beta_0 = 0$, а тогда мы проверяем, нужен ли вообще член $\beta(x_i - \bar{x})$.

Аналогично можно использовать статистику

$$T = \frac{\hat{\alpha} - \alpha_0}{\sqrt{R}/\sqrt{(n-2)n}} \sim \mathrm{t}_{n-2} \tag{2.9.4}$$

для проверки гипотезы H_0: $\alpha = \alpha_0$ против альтернативы $H_1 : \alpha \neq \alpha_0$.

Изложенные выше критерии приводят к $100(1-\gamma)\%$-доверительным интервалам для α и β:

$$\left(\hat{\alpha} - \frac{\sqrt{R}}{\sqrt{(n-2)n}}\mathrm{t}_{n-2}(\gamma/2),\ \hat{\alpha} + \frac{\sqrt{R}}{\sqrt{(n-2)n}}\mathrm{t}_{n-2}(\gamma/2)\right), \tag{2.9.5}$$

$$\left(\hat{\beta} - \frac{\sqrt{R}}{\sqrt{(n-2)S_{xx}}}\mathrm{t}_{n-2}(\gamma/2),\ \hat{\beta} + \frac{\sqrt{R}}{\sqrt{(n-2)S_{xx}}}\mathrm{t}_{n-2}(\gamma/2)\right). \tag{2.9.6}$$

Аналогичная конструкция применима и к сумме $\alpha + \beta$. В этом случае

$$\hat{\alpha} + \hat{\beta} \sim \mathrm{N}\left(\alpha + \beta,\ \sigma^2\left(\frac{1}{n} + \frac{1}{S_{xx}}\right)\right) \quad \text{и} \quad \hat{\alpha} + \hat{\beta} \text{ не зависит от } R.$$

Тогда

$$T = \frac{\hat{\alpha} + \hat{\beta} - (\alpha + \beta)}{\left(\left(\frac{1}{n} + \frac{1}{S_{xx}}\right)R/(n-2)\right)^{1/2}} \sim \mathrm{t}_{n-2}.$$

Следовательно,

$$\left(\hat{\alpha} + \hat{\beta} - \mathrm{t}_{n-2}(\gamma/2)\left(\left(\frac{1}{n} + \frac{1}{S_{xx}}\right)R/(n-2)\right)^{1/2},\right.$$
$$\left.\hat{\alpha} + \hat{\beta} + \mathrm{t}_{n-2}(\gamma/2)\left(\left(\frac{1}{n} + \frac{1}{S_{xx}}\right)R/(n-2)\right)^{1/2}\right) \tag{2.9.7}$$

— $100(1-\gamma)\%$-д. и. для $\alpha + \beta$.

Далее, зафиксируем значение x и построим так называемый *интервал предсказания* для случайной величины $Y \sim \mathrm{N}(\alpha + \beta(x - \bar{x}), \sigma^2)$, которая не зависит от $Y_1, \ldots, Y_n$. Здесь $\bar{x} = \left(\sum\limits_{i=1}^{n} x_i\right) \Big/ n$, $x_1, \ldots, x_n$ — это точки, в которых были получены наблюдения $Y_1, \ldots, Y_n$, а x рассматривается как новая точка (в которой ранее не было наблюдений). Вполне естественно рассмотреть значение

$$\hat{Y} = \hat{\alpha} + \hat{\beta}(x - \bar{x})$$

как прогноз для Y в нашей модели, при этом

$$\mathrm{E}\hat{Y} = \alpha + \beta(x - \bar{x}) = \mathrm{E}Y \tag{2.9.8}$$

и

$$\operatorname{Var}(\hat{Y} - Y) = \operatorname{Var}\hat{Y} + \operatorname{Var}Y = \operatorname{Var}\hat{\alpha} + (x - \bar{x})^2 \operatorname{Var}\hat{\beta} + \sigma^2 = \\ = \sigma^2\left(1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}}\right). \tag{2.9.9}$$

Следовательно,

$$\hat{Y} - Y \sim \mathrm{N}\left(0, \sigma^2\left(1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}}\right)\right),$$

и мы получаем, что $100(1-\gamma)\%$-интервал предсказания для Y имеет центр в точке $\hat{Y}$ и длину

$$2\hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}}}\,\mathrm{t}_{n-2}\left(\frac{\gamma}{2}\right),$$

где $\hat{\sigma} = \sqrt{R/(n-2)}$.

Наконец, построим д. и. для линейной комбинации $\alpha + l\beta$. В этой ситуации, применяя аналогичные рассуждения, получим, что

$$\left(\hat{\alpha} + l\hat{\beta} - \frac{\sqrt{R}}{\sqrt{n-2}}\sqrt{\frac{1}{n} + \frac{l^2}{S_{xx}}}\mathrm{t}_{n-2}(\gamma/2),\ \hat{\alpha} + l\hat{\beta} + \frac{\sqrt{R}}{\sqrt{n-2}}\sqrt{\frac{1}{n} + \frac{l^2}{S_{xx}}}\mathrm{t}_{n-2}(\gamma/2)\right) \tag{2.9.10}$$

является $100(1-\gamma)\%$-д. и. для $\alpha + l\beta$.

В частности, если выбрать $l = x - \bar{x}$, то получим д. и. для среднего значения $\mathrm{E}Y$ в заданной точке наблюдения x, т. е.

$$\alpha + \beta(x - \bar{x}), \qquad \bar{x} = \frac{1}{n}\sum_i x_i.$$

Возникает естественный вопрос: насколько точно построенная линия регрессии приближает данные? Плохое приближение может возникнуть из-за больших значений σ^2, но значение σ^2 нам не известно. Чтобы оценить точность приближения, выполняют несколько наблюдений $Y_{i1}, \ldots, Y_{im_i}$ для каждого значения x_i:

$$Y_{ij} = \alpha + \beta(x_i - \bar{x}) + \varepsilon_{ij}, \quad j = 1, \ldots, m_i,\ i = 1, \ldots, n. \tag{2.9.11}$$

Тогда нулевая гипотеза H_0 о том, что среднее значение $\mathsf{E}Y_{..}$ зависит линейно от x, проверяется следующим образом. Усреднение

$$\overline{Y}_{i+} = \frac{1}{m_i}\sum_{j=1}^{m_i} Y_{ij}$$

равно

$$\alpha + \beta(x_i - \overline{x}) + \overline{\varepsilon}_{i+}, \qquad \text{где} \quad \overline{\varepsilon}_{i+} \sim \mathrm{N}\left(0,\ \frac{\sigma^2}{m_i}\right).$$

При выполнении гипотезы H_0 измеряем «отклонение от линейности» величиной о. с. к. $\sum\limits_i m_i(\overline{Y}_{i+} - \hat{\alpha} - \hat{\beta}(x_i - \overline{x}))^2$, которая удовлетворяет соотношению

$$\frac{1}{\sigma^2}\sum_i m_i(\overline{Y}_{i+} - \hat{\alpha} - \hat{\beta}(x_i - \overline{x}))^2 \sim \chi^2_{n-2}. \tag{2.9.12}$$

Далее, независимо от того, верна гипотеза H_0 или ложна, для любого $i = 1, \ldots, n$ получаем, что

$$\frac{1}{\sigma^2}\sum_j (Y_{ij} - \overline{Y}_{i+})^2 \sim \chi^2_{m_i - 1} \tag{2.9.13}$$

и эта сумма не зависит от $\overline{Y}_{1+}, \ldots, \overline{Y}_{n+}$. Следовательно,

$$\frac{1}{\sigma^2}\sum_{i,j} (Y_{ij} - \overline{Y}_{i+})^2 \sim \chi^2_{m_1 + \ldots + m_n - n}, \tag{2.9.14}$$

и эта сумма не зависит от $\overline{Y}_{1+}, \ldots, \overline{Y}_{n+}$. При выполнении гипотезы H_0 получаем, что

$$\frac{\sum\limits_i m_i(\overline{Y}_{i+} - \hat{\alpha} - \hat{\beta}(x_i - \overline{x}))^2/(n-2)}{\sum\limits_{i,j}(Y_{ij} - \overline{Y}_{i+})^2 \Big/ \left(\sum\limits_i m_i - n\right)} \sim \mathrm{F}_{n-2, m_1 + \ldots + m_n - n}. \tag{2.9.15}$$

Тогда при заданном $\alpha \in (0, 1)$ гипотезу H_0 отклоняют, когда значение статистики (2.9.15) будет больше чем $\varphi^+_{n-2, m_1 + \ldots + m_n - n}(\alpha)$.

Глава 3

Задачи кембриджских «Математических треножников» к курсу «Статистика»

Статистики, вероятно, сделают это.

(Из серии «Как они делают это».)

Манипулирование статистическими формулами не заменяет понимания того, что делается.

Г. М. Блалок (1926—), американский социолог

Задачи и решения, представленные ниже, следуют в обратном хронологическом порядке (однако порядок в пределах каждого года сохраняется).

Задача 1. Используя критерий обобщенного отношения правдоподобия, получите t-критерий Стьюдента для проверки гипотезы о равенстве средних значений двух популяций. Вам следует тщательно объяснить предположения, на которых основан этот критерий.

Решение. См. §2.7, п. 1в. □

Задача 2. а) Сформулируйте и докажите теорему Рао—Блекуэлла.

б) Предположим, что $X_1, X_2, \ldots, X_n$ — н. о. р. с. в., имеющие распределение

$$\mathrm{P}(X_1 = r) = p^{r-1}(1-p), \quad r = 1, 2, \ldots,$$

где p, $0 < p < 1$, является неизвестным параметром. Найдите одномерную достаточную статистику T для p.

в) Найдите несмещенную оценку для p, которая является функцией от T, отыскав для этого сперва простую несмещенную оценку для p (или иным способом).

Решение. б) Функция правдоподобия равна

$$f_{\boldsymbol{X}}(\boldsymbol{x}; p) = (1-p)^n \prod_{i=1}^{n} p^{x_i - 1} = \frac{(1-p)^n}{p^n} p^{x_1 + \ldots + x_n},$$
$$\boldsymbol{x} = (x_1, \ldots, x_n), \quad x_i = 1, 2, \ldots$$

Следовательно, $T_n(\boldsymbol{x}) = \sum\limits_{i=1}^{n} x_i$ является достаточной статистикой.

в) Простая несмещенная оценка параметра p имеет вид

$$\hat{p} = 1 - \frac{1}{n}\sum_{i=1}^{n} I(X_i = 1),$$

для нее

$$\mathsf{E}\hat{p} = 1 - \frac{1}{n}\sum_{i=1}^{n}\mathsf{P}(X_i = 1) = 1 - \frac{1}{n}\sum_{i=1}^{n}(1-p) = 1 - 1 + p = p.$$

Тогда для n, $t = 2, 3, \ldots, n \leqslant t$, мы имеем

$$\hat{p}^* = \mathsf{E}(\hat{p} \mid T_n = t) = \mathsf{E}\left(1 - \frac{1}{n}\sum_{i=1}^{n} I(X_i = 1) \;\middle|\; T_n = t\right) =$$

$$= 1 - \frac{1}{n}\sum_{i=1}^{n}\mathsf{P}(X_i = 1 \mid T_n = t) = 1 - C_{t-2}^{n-2}/C_{t-1}^{n-1} = \frac{t-n}{t-1}$$

(в предположении, что $C_{t-2}^{n-2} = 1$ для $t \geqslant n = 2$). Таким образом, статистика Рао—Блекуэлла имеет вид

$$\hat{p}^* = \mathsf{E}(I(X_1 > 1 \mid T_n) = \frac{T_n - n}{T_n - 1}.$$

В самом деле,

$$\mathsf{P}(X_i = 1 \mid T_n = t) = \frac{\mathsf{P}(X_i = 1,\, T_{n-1} = t-1)}{\mathsf{P}(T_n = t)} = \frac{(1-p)\,\mathsf{P}(T_{n-1} = t-1)}{\mathsf{P}(T_n = t)},$$

$$\mathsf{P}(T_{n-1} = t-1) = \frac{(1-p)^{n-1}}{p^{n-1}} \sum_{\substack{r_1,\ldots,r_{n-1}=1,2,\ldots:\\ r_1+\ldots+r_{n-1}=t-1}} p^{t-1}$$

и

$$\mathsf{P}(T_n = t) = \frac{(1-p)^n}{p^n} \sum_{\substack{r_1,\ldots,r_n=1,2,\ldots:\\ r_1+\ldots+r_n=t}} p^t.$$

Наконец, для $t \geqslant n \geqslant 2$ мы имеем

$$\sum_{\substack{r_1,\ldots,r_{n-1}=1,2,\ldots:\\ r_1+\ldots+r_{n-1}=t-1}} 1 = C_{t-2}^{n-2}, \qquad \sum_{\substack{r_1,\ldots,r_n=1,2,\ldots:\\ r_1+\ldots+r_n=t}} 1 = C_{t-1}^{n-1},$$

поскольку представление натурального числа m в виде суммы s положительных слагаемых соответствует следующей задаче: расположить $s-1$ «голубя» в $m-1$ «гнезде» так, чтобы в каждом гнезде находилось не более одного голубя (см. рис. 3.1).

Для $n = 1$ оценка $\hat{p}$ уже является функцией от T_n. □

Рис. 3.1

Задача 3. При исследовании 60 мужчин и 90 женщин все они были классифицированы в зависимости от цвета глаз, что привело к следующим цифрам.

Таблица 3.1

	Голубой	Карий	Серый
Мужчины	20	20	20
Женщины	20	50	20

Объясните, как бы вы проанализировали эти результаты. Точно укажите все предположения, на которых вы будете основывать свои суждения.

При дальнейших исследованиях 150 человек были классифицированы по двум признакам: по цвету глаз и по тому, является ли человек левшой или правшой. Была получена следующая таблица.

Таблица 3.2

	Голубой	Карий	Серый
Левша	20	20	20
Правша	20	50	20

Как изменится ваш анализ в этом случае? Опять укажите точно все предположения, на которых вы основываетесь. Возможно, вы пожелаете воспользоваться приведенными ниже процентилями χ^2-распределения.

Таблица 3.3

	χ_1^2	χ_2^2	χ_3^2	χ_4^2	χ_5^2	χ_6^2
95%-квантиль	3,84	5,99	7,81	9,49	11,07	12,59
99%-квантиль	6,64	9,21	11,34	13,28	15,09	16,81

Решение. Мы опустим стандартные неравенства, задающие параметры как неотрицательные. При первом исследовании (мужчины/женщины) таблица сопряженности признаков имеет следующий вид.

Таблица 3.4

h_1	h_2	h_3
q_1	q_2	q_3

Выбирается нулевая гипотеза H_0: $h_i = q_i$, $i = 1, 2, 3$ (строки одинаковы), и альтернатива H_1: $h_1 + h_2 + h_3 = 1$, $q_1 + q_2 + q_3 = 1$. При выполнении гипотезы H_0 мы имеем

$$\hat{h}_1 = \frac{4}{15}, \quad \hat{h}_2 = \frac{7}{15}, \quad \hat{h}_3 = \frac{4}{15}.$$

Таблица средних значений имеет следующий вид.

Таблица 3.5

16	28	16
24	42	24

Таким образом, значение статистики Пирсона равно

$$\frac{4^2}{16} + \frac{8^2}{28} + \frac{4^2}{16} + \frac{4^2}{24} + \frac{8^2}{42} + \frac{4^2}{24} \approx 7{,}33.$$

Число степеней свободы равно 2. Сравнивая полученное значение с верхней 5%-квантилью распределения χ^2_2, мы отвергаем гипотезу H_0 на уровне значимости 95%, а сравнивая с верхней 1%-квантилью, видим, что нет оснований отклонить гипотезу H_0 на уровне 99%.

При втором исследовании (левши/правши) рассматривается нулевая гипотеза H_0: $h_{ij} = h_i q_j$, $h_1 + h_2 + h_3 = 1$, $q_1 + q_2 = 1$, и альтернатива H_1: h_{ij} произвольно с ограничением $\sum\limits_{i,j} h_{ij} = 1$. При выполнении гипотезы H_0 имеем

$$\hat{q}_1 = \frac{6}{15}, \quad \hat{q}_2 = \frac{9}{15}$$

и $\hat{h}_i$, $i = 1, 2, 3$, такие же, как и раньше. Средние значения и статистика критерия те же, что и раньше, и, следовательно, выводы те же, что и при первом исследовании.

Некоторые студенты рассматривают другие нулевые гипотезы и альтернативы. Например, при гипотезе

H_0: цвет глаз не зависит от пола и имеет равномерное распределение

число степеней свободы равно пяти. Эта гипотеза отклоняется при уровне 99% (а тем более при уровне 95%).

Замечание. В этом примере две процедуры проверки гипотез основаны на разных моделях данных. При первом исследовании число мужчин и женщин фиксировано, а при втором исследовании 150 человек выбраны случайным образом. Кроме того, проверяются различные нулевые гипотезы и альтернативы. Однако средние значения и применяемые статистики одинаковы в обоих случаях, а следовательно, одинаковы и выводы. □

Задача 4. а) Точно определив необходимые понятия, сформулируйте и докажите лемму Неймана—Пирсона.

б) Пусть X — это одно наблюдение из распределения с плотностью

$$f(x;\,\theta)=\frac{1}{2}e^{-|x-\theta|},\qquad -\infty<x<\infty,$$

с неизвестным параметром θ. Найдите наилучший критерий размера α, $0<\alpha<1$, для проверки гипотезы $H_0\colon \theta=\theta_0$ против альтернативы $H_1\colon \theta=\theta_1$, где $\theta_1>\theta_0$.

Если $\alpha=0{,}05$, то для каких значений θ_0 и θ_1 мощность наилучшего критерия будет не менее чем 0,95?

Решение. б) Функция правдоподобия

$$\Lambda_{H_1,H_0}=\frac{f(x;\,\theta_1)}{f(x;\,\theta_0)}=\frac{e^{-|x-\theta_1|}}{e^{-|x-\theta_0|}}$$

принимает большие значения, когда величина

$$-|x-\theta_1|+|x-\theta_0|$$

велика. Логарифм правдоподобия имеет вид

$$\ln\Lambda_{H_1,H_0}(x)=\begin{cases}\theta_1-\theta_0, & x\geqslant\theta_1,\\ 2x-(\theta_0+\theta_1), & \theta_0<x<\theta_1,\\ \theta_0-\theta_1, & x\leqslant\theta_0;\end{cases}$$

см. рис. 3.2.

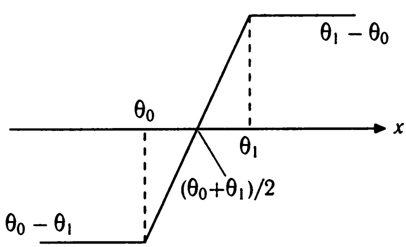


Рис. 3.2

Мы отвергаем нулевую гипотезу при больших значениях $\ln\Lambda_{H_1,H_0}$. Рассмотрим функцию, которая задает в. о. р. I:

$$b\in\mathbb{R}\mapsto B(b)=\frac{1}{2}\int e^{-|x-\theta_0|}I(\ln\Lambda_{H_1,H_0}(x)>b)\,dx;$$

она имеет вид

$$B(b) = \begin{cases} 0, & b \geqslant \theta_1 - \theta_0, \\ e^{(\theta_0 - \theta_1 - b)/2}/2, & \theta_0 - \theta_1 \leqslant b < \theta_1 - \theta_0, \\ 1, & b < \theta_0 - \theta_1. \end{cases}$$

Следовательно,

1) при $0 \leqslant \alpha \leqslant e^{\theta_0 - \theta_1}/2$ мы отклоняем гипотезу H_0, если

$$x \geqslant \theta_0 + \ln \frac{1}{2\alpha} \geqslant \theta_1;$$

ясно, что в этом случае мощность будет не больше $1/2$;

2) при $e^{\theta_0 - \theta_1}/2 < \alpha \leqslant 1/2$ мы отклоняем гипотезу H_0, если

$$x \geqslant \theta_0 + \ln \frac{1}{2\alpha};$$

действительно, решая уравнение $B(b) = \alpha$, мы получаем для этого случая, что $b = \theta_0 - \theta_1 + \ln 1/(2\alpha)$;

3) при $\alpha \geqslant 1/2$ мы отклоняем гипотезу H_0, если

$$x \geqslant \theta_0 - \ln \frac{1}{2(1-\alpha)}.$$

Следовательно, единственный случай, представляющий для нас интерес, — это случай 2. Мощность наилучшего критерия уровня $\alpha = 0{,}05$ будет не менее чем 0,95, когда

$$\frac{1}{2} + \frac{1}{2}(1 - e^{c - \theta_1}) \geqslant 0{,}95,$$

где

$$c = \theta_0 + \ln \frac{1}{2\alpha}.$$

Это эквивалентно неравенству

$$\exp\left(-\theta_0 - \ln \frac{1}{2\alpha} + \theta_1\right) \geqslant 10,$$

или

$$\theta_1 - \theta_0 - \ln \frac{1}{2\alpha} \geqslant \ln 10,$$

т. е.

$$\theta_1 - \theta_0 > \ln 100. \qquad \square$$

Задача 5. Предположим, что $Y_1, \ldots, Y_n$ — это такие независимые случайные величины, что Y_i имеет нормальное распределение со средним βx_i и дисперсией σ^2, где β, σ^2 неизвестны, а $x_1, \ldots, x_n$ — известные константы.

Найдите МНК-оценку для β.

Подробно объясните, как проверить гипотезу H_0: $\beta = 0$ против альтернативы H_1: $\beta \neq 0$.

Решение. См. § 2.9. □

Задача 6. Требуется оценить неизвестный параметр θ по единственному наблюдению с. в. X с плотностью распределения вероятностей $f(x;\ \theta)$; параметр θ имеет априорное распределение с плотностью $\pi(\theta)$, и функция потерь равна $L(\theta,\ a)$. Покажите, что оптимальная байесовская точечная оценка минимизирует апостериорные средние потери.

Предположим далее, что $f(x;\ \theta) = \theta e^{-\theta x}$, $x > 0$, и $\pi(\theta) = \mu e^{-\mu\theta}$, $\theta > 0$, где $\mu > 0$ неизвестно. Определите апостериорное распределение θ при заданном X.

Найдите оптимальную байесовскую точечную оценку параметра θ в случае, когда

а) $L(\theta,\ a) = (\theta - a)^2$,

б) $L(\theta,\ a) = |(\theta - a)/\theta|$.

Решение. Что касается первой части, см. § 1.9.

В данном примере апостериорная п. р. в. $\pi(\theta\,|\,x)$ равна

$$f(x;\ \theta)\pi(\theta) \propto \theta e^{-\theta x} e^{-\mu\theta} I(\theta > 0).$$

Это означает, что

$$\pi(\theta\,|\,x) = (x+\mu)^2 \theta e^{-(x+\mu)\theta} I(\theta > 0),$$

где мы воспользовались равенством

$$\int_0^\infty \theta e^{-(x+\mu)\theta}\, d\theta = \frac{1}{(x+\mu)^2}.$$

Наконец, в случае а), когда $L(\theta,\ a) = (\theta - a)^2$, оптимальная байесовская точечная оценка параметра θ равна

$$d^*(x) = \int_0^\infty \theta\pi(\theta\,|\,x)\, d\theta = \frac{2}{x+\mu}$$

(т. е. среднему значению п. р. в. $(\mu + x)^2 \theta e^{-(x+\mu)\theta} I(\theta > 0)$), а в случае б), когда $L(\theta,\ a) = |\theta - a|/\theta$, она имеет вид

$$d^*(x) = \arg\min_a \int |\theta - a| e^{-(\mu+x)\theta}\, d\theta = \frac{\ln 2}{\mu + x}$$

(т. е. равна медиане п. р. в. $(\mu + x)e^{-(\mu+x)\theta} I(\theta > 0)$). □

Задача 7. а) Найдите МНК-оценки $\hat{\alpha}$ и $\hat{\beta}$ для коэффициентов модели простой линейной регрессии

$$Y_i = \alpha + \beta(x_i - \bar{x}) + \varepsilon_i, \quad i = 1,\ \ldots,\ n,$$

где $x_1, \ldots, x_n$ — заданные константы, $\bar{x} = n^{-1} \sum_{i=1}^{n} x_i$, $i = 1, \ldots, n$, а величины ε_i независимы,

$$\mathrm{E}\varepsilon_i = 0, \quad \mathrm{Var}\,\varepsilon_i = \sigma^2, \quad i = 1, \ldots, n.$$

б) Производитель оптического оборудования располагает следующими данными о стоимости (в фунтах стерлингов за каждую единицу) некоторых изготовленных на заказ линз и количестве единиц в каждом заказе:

Кол-во единиц x_i	1	3	5	10	12
Стоимость единицы y_i	58	55	40	37	22

Предположим, что выполнены условия, при которых применим простой линейный регрессионный анализ. Оцените коэффициенты регрессии и воспользуйтесь полученным уравнением регрессии, чтобы спрогнозировать стоимость одной единицы в заказе, состоящем из 8 таких линз.

Указание. Для вышеприведенных данных $S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x}) y_i = -257{,}4$.

Решение. а) МНК-оценки $\hat{\alpha}$ и $\hat{\beta}$ минимизируют сумму

$$S = \sum_{i=1}^{n} (y_i - \alpha - \beta(x_i - \bar{x}))^2$$

и, следовательно, являются решениями уравнений

$$\sum_i y_i - n\alpha = 0,$$

$$\beta \sum_i (x_i - \bar{x})^2 = \sum_i (x_i - \bar{x}) y_i = \sum_i (x_i - \bar{x})(y_i - \bar{y}).$$

Другими словами, $\hat{\alpha} = \bar{y}$, $\hat{\beta} = S_{xy}/S_{xx}$ (очевидно, что эта стационарная точка является точкой минимума).

б) Проведем вычисления:

$$\bar{y} = \frac{1}{5}(58 + 55 + 40 + 37 + 22) = \frac{212}{5} = 42{,}4,$$
$$\bar{x} = \frac{1}{5}(1 + 3 + 5 + 10 + 12) = \frac{31}{5} = 6{,}2,$$
$$S_{xx} = 5{,}2^2 + 3{,}2^2 + 1{,}2^2 + 3{,}8^2 + 5{,}8^2 = 86{,}8.$$

В результате получаем $\hat{\alpha} = \bar{y} = 42{,}4$, $\hat{\beta} = \frac{S_{xy}}{S_{xx}} = -\frac{257{,}4}{86{,}8} \approx -2{,}965$ и находим, что $y(x) = 42{,}4 - 2{,}965(x - 6{,}2)$. При $x = 8$ мы получаем $y(8) \approx 37{,}063$. □

Задача 8. Предположим, что 6 наблюдений $X_1, \ldots, X_6$ извлечены случайным образом из нормального распределения, у которого оба параметра, среднее μ_X и дисперсия σ_X^2, неизвестны, но найдено, что $S_{xx} = \sum_{i=1}^{6}(x_i - \bar{x})^2 = 30$, где $\bar{x} = \frac{1}{6}\sum_{i=1}^{6} x_i$. Предположим также, что 21 наблюдение $Y_1, \ldots, Y_{21}$ извлечено случайным образом из другого распределения, у которого и среднее μ_Y, и дисперсия σ_Y^2 неизвестны и найдено, что $S_{yy} = 40$. Получите критерий отношения правдоподобия для проверки гипотезы H_0: $\sigma_X^2 = \sigma_Y^2$ против альтернативы H_1: $\sigma_X^2 > \sigma_Y^2$ и примените его к указанным данным при уровне значимости 0,05.

Указание.

Распределение	χ_5^2	χ_6^2	χ_{20}^2	χ_{21}^2	$F_{5,20}$	$F_{6,21}$
95%-квантиль	11,07	12,59	31,41	32,68	2,71	2,57

Решение. Мы имеем $X_1, \ldots, X_m \sim N(\mu_X, \sigma_X^2)$, $Y_1, \ldots, Y_n \sim N(\mu_Y, \sigma_Y^2)$. При выполнении гипотезы H_0 получаем $\sigma_X^2 = \sigma_Y^2$ и, следовательно,

$$L_{xy}(H_0) = \sup_{\substack{\sigma_X=\sigma_Y,\\ \mu_X,\mu_Y}} f_X(x; \mu_X, \sigma^2) f_Y(y; \mu_Y, \sigma^2) = \sup_{\sigma} \frac{1}{(2\pi\sigma^2)^{(n+m)/2}} \exp\left(-\frac{S_{xx}+S_{yy}}{2\sigma^2}\right).$$

Заметим теперь, что для $g(x) = x^a e^{-bx}$ $(a, b > 0)$ выполняется соотношение

$$\sup_{x>0} g(x) = g\left(\frac{a}{b}\right) = \left(\frac{a}{b}\right)^a e^{-a}.$$

Следовательно,

$$L_{xy}(H_0) = \frac{1}{(2\pi\hat{\sigma}_0^2)^{(m+n)/2}} e^{-(m+n)/2}, \quad \hat{\sigma}_0^2 = \frac{S_{xx} + S_{yy}}{m+n}.$$

Аналогично при выполнении гипотезы H_1 имеем

$$L_{xy}(H_1) = \sup_{\substack{\sigma_X,\sigma_Y,\\ \mu_X,\mu_Y}} f_X(x; \mu_X, \sigma_X^2) f_Y(y; \mu_Y, \sigma_Y^2) = \frac{1}{(2\pi\hat{\sigma}_X^2)^{m/2}(2\pi\hat{\sigma}_Y^2)^{n/2}} e^{-\frac{m+n}{2}},$$

где $\hat{\sigma}_X^2 = \frac{S_{xx}}{m}$, $\hat{\sigma}_Y^2 = \frac{S_{yy}}{n}$ (при условии, что $\hat{\sigma}_X^2 > \hat{\sigma}_Y^2$). В результате получаем

$$\Lambda = \begin{cases} \left(\dfrac{S_{xx} + S_{yy}}{m+n}\right)^{(m+n)/2} \left(\dfrac{S_{xx}}{m}\right)^{-m/2} \left(\dfrac{S_{yy}}{n}\right)^{-n/2}, & \text{если } \dfrac{S_{xx}}{m} > \dfrac{S_{yy}}{n}, \\ 1 & \text{в противном случае.} \end{cases}$$

Если $\frac{S_{xx}}{S_{yy}} > \frac{m}{n}$, то

$$2\ln\Lambda = \mathrm{Const} + g\left(\frac{S_{xx}}{S_{yy}}\right),$$

где

$$g(z) = (m+n)\ln(1+z) - m\ln z.$$

Тогда $g'(z) = \frac{m+n}{1+z} - \frac{m}{z} = \frac{nz-m}{z(1+z)}$, т. е. если функция $g(z)$ возрастает при $z > m/n$, то гипотеза H_0 отвергается при условии, что $\frac{S_{xx}}{S_{yy}}$ велико. При выполнении гипотезы H_0 мы имеем $\sigma_X^2 = \sigma_Y^2 = \sigma^2$, $\frac{S_{xx}}{\sigma^2} \sim \chi^2_{m-1}$, $\frac{S_{yy}}{\sigma^2} \sim \chi^2_{n-1}$ и эти величины независимы; следовательно,

$$\frac{S_{xx}/(m-1)}{S_{yy}/(n-1)} \sim \mathrm{F}_{m-1,n-1}.$$

Для рассматриваемых данных правая часть будет равна

$$\frac{30/5}{40/20} = 3 > 2{,}71$$

(здесь используется распределение $\mathrm{F}_{5,20}$), и, таким образом, гипотезу H_0 следует отклонить при заданном уровне. □

Задача 9. Пусть $X_1, \ldots, X_n$ — случайная выборка из $\mathrm{N}(\theta, \sigma^2)$-распределения, и предположим, что априорным распределением для θ является $\mathrm{N}(\mu, \tau^2)$, где σ^2, μ, τ^2 известны. Определите апостериорное распределение для θ при заданных $X_1, \ldots, X_n$ и наилучшую байесовскую оценку параметра θ при квадратической функции потерь и функции потерь, равной абсолютной ошибке.

Решение. По поводу первой части см. задачи 2.3.8, 2.3.9 части А. Вторая часть задачи: как было показано в задаче 2.3.8 части А, апостериорное распределение имеет вид

$$\pi(\theta \mid \boldsymbol{x}) = \frac{\pi(\theta) f(\boldsymbol{x};\theta)}{\int \pi(\theta') f(\boldsymbol{x};\theta')\,d\theta'} = \frac{1}{\sqrt{2\pi}\,\tau_n} \exp\left[-\frac{(\theta-\mu_n)^2}{2\tau_n^2}\right],$$

где

$$\frac{1}{\tau_n^2} = \frac{1}{\tau^2} + \frac{n}{\sigma^2}, \qquad \mu_n = \frac{\dfrac{\mu}{\tau^2} + \dfrac{n\bar{x}}{\sigma^2}}{\dfrac{1}{\tau^2} + \dfrac{n}{\sigma^2}}.$$

Так как нормальное распределение является симметричным относительно своего среднего значения, наилучшей оценкой для обеих функций потерь будет $\mathsf{E}(\theta \mid \boldsymbol{x}) = \mu_n$. □

Задача 10. Проведено исследование уровня знаний старшеклассников, при котором в каждом из двух городов было рассмотрено по 500 учеников и уровень знаний каждого классифицировался согласно шкале низкий-средний-высокий. Результаты приведены в таблице:

	Низкий	Средний	Высокий
Город A	103	145	252
Город B	140	136	224

Получите критерий однородности и проверьте гипотезу о том, что распределения уровней знаний учеников одинаковы для обоих городов.

Указание.

Распределение	χ_1^2	χ_2^2	χ_3^2	χ_5^2	χ_6^2
99%-квантиль	6,63	9,21	11,34	15,09	16,81
95%-квантиль	3,84	5,99	7,81	11,07	12,59

Решение. Составим таблицы

$$\begin{array}{cccc} n_{11} & n_{12} & n_{13} & n_{1+} \\ n_{21} & n_{22} & n_{23} & n_{2+} \\ n_{+1} & n_{+2} & n_{+3} & n_{++} \end{array}$$

$$\begin{array}{ccc} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \end{array}$$

При выполнении гипотезы H_0 мы имеем $(N_{i1}, N_{i2}, N_{i3}) \sim \mathrm{Mult}(n_{i+}, p_1, p_2, p_3)$, где $p_j = p_{1j} = p_{2j}$, $j = 1, 2, 3$, $p_+ = 1$ и $|\Theta_0| = 2$. Далее,

$$\ln \Lambda = \mathrm{Const} + \sum_j n_{+j} \ln p_j.$$

Следовательно,

$$\bar{p}_j = \frac{n_{+j}}{n_{++}}, \qquad |\Theta_0| = 3 - 1 = 2.$$

При выполнении гипотезы H_1 мы имеем $(N_{11}, N_{12}, N_{13}) \sim \mathrm{Mult}(n_{1+}, p_{11}, p_{12}, p_{13})$ (мультиномиальное распределение с параметрами $(n_{1+}, p_{11}, p_{12}, p_{13})$) и $(N_{21}, N_{22}, N_{23}) \sim \mathrm{Mult}(n_{2+}, p_{21}, p_{22}, p_{23})$, $p_{11}+p_{12}+p_{13}=p_{21}+p_{22}+p_{23}=1$. Таким образом, $|\Theta_1| = 4$. Далее, при $\hat{p}_{ij} = n_{ij}/n_{i+}$ и $e_{ij} = n_{i+}n_{+j}/n_{++}$ получаем

$$2\ln\Lambda = 2\sum n_{ij}\ln(\hat{p}_{ij}/\bar{p}_j) = 2\sum n_{ij}\ln\left(\frac{n_{ij}}{n_{i+}n_{+j}/n_{++}}\right) =$$

$$= 2\sum n_{ij}\ln\frac{n_{ij}}{e_{ij}} = 2\sum(e_{ij} + n_{ij} - e_{ij})\ln\left(1 + \frac{n_{ij} - e_{ij}}{e_{ij}}\right) \approx \sum\frac{(n_{ij} - e_{ij})^2}{e_{ij}};$$

так как $|\Theta_1| - |\Theta_0| = 2$, мы сравниваем $2\ln\Lambda$ с χ_2^2.

В данном примере имеем

	L	M	H	n_{i+}
A	103	145	252	500
B	140	136	224	500
n_{+j}	243	281	476	1000
e_{ij}	121,5	140,5	238	

$n_{ij} - e_{ij}$ для A	−18,5	4,5	14
$n_{ij} - e_{ij}$ для B	18,5	−4,5	−14

Следовательно,

$$2\ln\Lambda = 2\left(\frac{(18{,}5)^2}{121{,}5} + \frac{(4{,}5)^2}{140{,}5} + \frac{(14)^2}{238}\right) \approx 7{,}569.$$

Значение значимо на уровне 5% и не значимо на уровне 1%:

$$2\ln\Lambda = 7{,}569 < \chi^2_2(0{,}99) = 9{,}21 \quad \text{и} \quad 2\ln\Lambda > \chi^2_2(0{,}95) = 5{,}99. \qquad \square$$

Задача 11. Следующая таблица содержит распределение, полученное в результате 320 подбрасываний 6 монет, и соответствующие ожидаемые (теоретические) частоты, вычисленные по формуле биномиального распределения с параметрами $p = 0{,}5$ и $n = 6$.

Число гербов	0	1	2	3	4	5	6
Наблюдаемые частоты	3	21	85	110	62	32	7
Теоретические частоты	5	30	75	100	75	30	5

Примените критерий согласия с уровнем 0,05 для проверки нулевой гипотезы о том, что все монеты правильные.

Указание.

Распределение	χ^2_5	χ^2_6	χ^2_7
95%-квантиль	11,07	12,59	14,07

Решение. Необходимые данные для применения критерия согласия запишем в таблице:

Число гербов	n	e	$n - e$
0	3	5	−2
1	21	30	−9
2	85	75	10
3	110	100	10
4	62	75	−13
5	32	30	2
6	7	5	2

Вычислим статистику $T = \sum\limits_{i=1}^{6}(n_i - e_i)^2/e_i$:

$$T = \frac{4}{5} + \frac{81}{30} + \frac{100}{75} + \frac{100}{100} + \frac{169}{75} + \frac{4}{30} + \frac{4}{5} = \frac{1}{150}(240 + 425 + 538 + 150) = 9{,}02.$$

Находим $|\Theta_1| = 6$, $|\Theta_0| = 0$, $\chi^2_6 = 12{,}59$ при уровне 0,05; таким образом, $T < \chi^2_6$, и нет оснований отвергать гипотезу H_0. $\square$

Задача 12. а) Сформулируйте и докажите теорему Рао—Блекуэлла.

б) Пусть $X_1, \ldots, X_n$ — независимые случайные величины, равномерно распределенные на $(\theta, 3\theta)$. Найдите двумерную достаточную статистику

$T(X)$ для θ. Покажите, что $\hat{\theta} = X_1/2$ является несмещенной оценкой параметра θ.

Найдите несмещенную оценку параметра θ, которая является функцией от $T(X)$ и среднеквадратическая ошибка которой не больше, чем среднеквадратическая ошибка оценки $\hat{\theta}$.

Решение. а) Теорема Р—Б: пусть T — это достаточная статистика для θ, а $\tilde{\theta}$ — несмещенная оценка для θ, у которой $\mathsf{E}\tilde{\theta}^2 < \infty$ при всех θ. Определим

$$\hat{\theta} = \mathsf{E}[\tilde{\theta} \mid T].$$

Тогда для всех θ выполняется неравенство

$$\mathsf{E}[(\hat{\theta} - \theta)^2] \leqslant \mathsf{E}[(\tilde{\theta} - \theta)^2].$$

Неравенство будет строгим, если $\tilde{\theta}$ не является функцией от T.

По формуле условного математического ожидания

$$\mathsf{E}[\hat{\theta}] = \mathsf{E}[\mathsf{E}(\tilde{\theta} \mid T)] = \mathsf{E}\tilde{\theta},$$

следовательно, $\hat{\theta}$ и $\tilde{\theta}$ имеют одинаковое смещение. По формуле условной дисперсии

$$\mathrm{Var}(\tilde{\theta}) = \mathsf{E}[\mathrm{Var}(\tilde{\theta} \mid T)] + \mathrm{Var}[\mathsf{E}(\tilde{\theta} \mid T)] = \mathsf{E}[\mathrm{Var}(\tilde{\theta} \mid T)] + \mathrm{Var}(\hat{\theta}).$$

Следовательно, $\mathrm{Var}(\tilde{\theta}) \geqslant \mathrm{Var}(\hat{\theta})$, и равенство выполняется, только если $\mathrm{Var}(\tilde{\theta} \mid T) = 0$.

б) Относительно второй части задачи: ср. ее с примером 1.6.2. □

Задача 13. а) Сформулируйте критерий факторизации для достаточной статистики и докажите его в дискретном случае.

б) Пусть $X_1, \ldots, X_n$ — это случайная выборка из распределения Пуассона и значение параметра θ неизвестно. Найдите одномерную достаточную статистику для θ.

Решение. б) Если $X_1, \ldots, X_n$ — независимые $\mathrm{Po}(\theta)$-с. в., то

$$f(\boldsymbol{x}; \theta) = \prod_{i=1}^{n} \frac{e^{-\theta}\theta^{x_i}}{x_i!} = \left(\prod_i \frac{1}{x_i!}\right) e^{-n\theta}\theta^{\sum x_i}.$$

Следовательно, $T(\boldsymbol{X}) = \sum\limits_i X_i$ является достаточной статистикой. □

Задача 14. Предположим, что мы опросили 50 мужчин и 150 женщин, являются они «жаворонками», «совами» или же не принадлежат ни к одному из этих типов (неопределенный тип). Полученные данные содержатся в следующей таблице.

	Жаворонки	Совы	Неопределенный тип	Всего
Мужчины	17	22	11	50
Женщины	43	78	29	150
Всего	60	100	40	200

Получите *критерий независимости классификации* в терминах (обобщенного) отношения правдоподобия. Каким будет результат применения этого критерия при уровне 0,01?

Указание.

Распределение	χ^2_1	χ^2_2	χ^2_3	χ^2_5	χ^2_6
99%-квантиль	6,63	9,21	11,34	15,09	16,81

Решение. Мы проверяем гипотезу

$$H_0\colon p_{ij} = \alpha_i\beta_j, \quad \alpha_i, \beta_j \geqslant 0, \quad i = 1, 2; \quad j = 1, 2, 3; \quad \sum \alpha_i = \sum \beta_j = 1,$$

против альтернативы

$$H_1\colon p_{ij} \geqslant 0, \quad \sum p_{ij} = 1.$$

При выполнении гипотезы H_0 о. м. п. имеют вид

$$\hat{\alpha}_i = \frac{n_{i+}}{n}, \quad \hat{\beta}_j = \frac{n_{+j}}{n},$$

$n_{i+} = \sum_j n_{ij}$, $n_{+j} = \sum_i n_{ij}$ и оцениваемые элементы равны

$$e_{ij} = n\hat{p}_{i+}\hat{p}_{+j} = \frac{n_{i+}n_{+j}}{n}.$$

Таблица полученных оценок такова:

e_{ij}	Жаворонки	Совы	Неопределенный тип	Всего
Мужчины	15	25	10	50
Женщины	45	75	30	150
Всего	60	100	40	200

Далее,

$$2\ln\Lambda = 2\sum_{i=1}^{2}\sum_{j=1}^{3} n_{ij}\ln\frac{n_{ij}}{e_{ij}}.$$

Поскольку разности $\delta_{ij} = n_{ij} - e_{ij}$ относительно малы ($|\delta_{ij}/n_{ij}| \leqslant 2/15 \approx 0{,}133$) и $|\Theta_1| - |\Theta_0| = (2-1)\cdot(3-1) = 2$, можно записать

$$2\ln\Lambda \approx \sum_{i=1}^{2}\sum_{j=1}^{3}\frac{(n_{ij}-e_{ij})^2}{e_{ij}} \sim \chi^2_2.$$

Проведем вычисления:

$$\sum_{i=1}^{2}\sum_{j=1}^{3}\frac{(n_{ij}-e_{ij})^2}{e_{ij}}=\frac{4}{15}+\frac{4}{45}+\frac{9}{25}+\frac{9}{75}+\frac{1}{10}+\frac{1}{30}\approx 0{,}969.$$

Полученное число меньше чем $\chi_2^2(0{,}01)=9{,}21$. Таким образом, при заданном уровне мы не отвергаем гипотезу H_0. □

Задача 15. а) Объясните, что понимают под равномерно наиболее мощным критерием, его функцией мощности и размером.

б) Пусть $Y_1, \ldots, Y_n$ — независимые одинаково распределенные случайные величины с общей плотностью $\rho e^{-\rho y}$, $y \geqslant 0$. Укажите равномерно наиболее мощный критерий для проверки равенства $\rho=\rho_0$ против альтернативы $\rho<\rho_0$ и определите функцию мощности критерия.

Решение. а) Пусть $X_1, \ldots, X_n$ — н. о. р. с. в. со значениями в $\mathcal{V}$ и общей п. р. в./д. ф. р. f, зависящей от параметра $\theta\in\Theta$. Предположим, что нам необходимо проверить гипотезу H_0: $\theta\in\Theta_0$ против альтернативы H_1 : $\theta\in\Theta_1$, где $\Theta_0\cap\Theta_1=\varnothing$, $\Theta_0\cup\Theta_1=\Theta$. Определим критическую область $\mathcal{C}\subset\mathcal{V}^n$ так, что гипотеза H_0 отвергается, когда $\boldsymbol{x}\in\mathcal{C}$, и принимается, когда $\boldsymbol{x}\notin\mathcal{C}$. Тогда вероятность отклонить гипотезу H_0 при заданном $\theta\in\Theta$ равна

$$w(\theta)\ (=w(\theta,\mathcal{C}))=\mathsf{P}_\theta(\boldsymbol{X}\in\mathcal{C}).$$

Чтобы оценить размер критерия, возьмем

$$\alpha\ (=\alpha(\mathcal{C}))=\sup[w(\theta,\mathcal{C})\colon \theta\in\Theta_0].$$

Функцию мощности определим как сужение w на Θ_1. Разность $1-w(\theta)$ для $\theta\in\Theta_1$ равна в. о. р. II, т. е. вероятности принять гипотезу H_0, когда она ложна. Критерий с критической областью $\mathcal{C}$ называют р. н. м. критерием размера a (для проверки гипотезы H_0 против альтернативы H_1), если 1) $\alpha(\mathcal{C})\leqslant a$ и 2) для любого критерия с такой критической областью $\mathcal{C}^*$, что $\alpha(\mathcal{C}^*)\leqslant a$, выполняется следующее неравенство:

$$w(\theta,\mathcal{C})\geqslant w(\theta,\mathcal{C}^*),\qquad \theta\in\Theta_1.$$

б) Для рассматриваемых с. в. $Y_1, \ldots, Y_n$ мы воспользуемся монотонным отношением правдоподобия. Для начала рассмотрим гипотезу H_0 против простой альтернативы H_{ρ_1} : $\rho=\rho_1$, где ρ_1 — некоторое выбранное значение, меньшее ρ_0. Тогда о. п. равно

$$\Lambda_{H_{\rho_1};H_0}(\boldsymbol{y})=\frac{\rho_1^n\exp\left(-\rho_1\sum\limits_i y_i\right)}{\rho_0^n\exp\left(-\rho_0\sum\limits_i y_i\right)}=\left(\frac{\rho_1}{\rho_0}\right)^n\exp\left[(\rho_0-\rho_1)\sum_i y_i\right].$$

Критическая область имеет вид $\{\sum y_i > k\}$ и, следовательно, не зависит от выбора $\rho_1 < \rho_0$. Тогда для любого $k > 0$ критерий с такой критической областью является р. н. м. критерием размера $\alpha(k) = \mathrm{P}_{\rho_0}(\sum y_i > k)$.

Функцию мощности найдем так: поскольку $2\rho Y_i \sim \mathrm{Exp}(1/2)$, мы имеем $2\rho \sum Y_i \sim \mathrm{Gam}(n, 1/2) \sim \chi^2_{2n}$. Размер критерия равен

$$\alpha(k) = \mathrm{P}_{\rho_0}\Big(\sum y_i > k\Big) = \mathrm{P}_{\chi^2_{2n}}(X > 2\rho_0 k), \quad k > 0;$$

это соотношение позволяет записать обратную функцию $k(\alpha)$, $0 < \alpha < 1$. Следовательно, при заданном α функция мощности критерия имеет вид

$$\mathrm{P}_\rho\Big(\sum y_i > k(\alpha)\Big) = \mathrm{P}_{\chi^2_{2n}}(X > 2\rho k(\alpha)), \quad \rho < \rho_0. \qquad \square$$

Задача 16. Для десяти стальных слитков, полученных в результате некоторого производственного процесса, были произведены измерения степени твердости, и результаты таковы:

$$73{,}2,\ 74{,}3,\ 75{,}4,\ 73{,}8,\ 74{,}4,\ 76{,}7,\ 76{,}1,\ 73{,}0,\ 74{,}6,\ 74{,}1.$$

В предположении, что распределение степени твердости достаточно хорошо описывается нормальной функцией, получите оценку среднего значения твердости.

Производитель утверждает, что он поставляет сталь с показателем твердости 75. Получите критерий (обобщенного) отношения правдоподобия для проверки этого утверждения. Пусть известно, что для указанных данных

$$S_{XX} = \sum_{j=1}^{n}(X_i - \bar{X})^2 = 12{,}824.$$

Каким тогда будет результат, полученный при применении этого критерия с уровнем значимости 5%?

Указание.

Распределение	t_9	t_{10}
95%-квантиль	1,83	1,81
97,5%-квантиль	2,26	2,23

Решение. Удобно перенормировать данные: $z_i = 10(x_i - 73)$, что приводит к значениям

$$z_i = 2,\ 13,\ 24,\ 6,\ 14,\ 37,\ 31,\ 0,\ 16,\ 11, \qquad \sum z_i = 156, \qquad \sum z_i^2 = 3716$$

и

$$\bar{z} = \frac{1}{n}\sum z_i = 15{,}6, \qquad S_{zz} = \sum z_i^2 - \frac{1}{n}\Big(\sum z_i\Big)^2 = 3716 - \frac{24336}{10} = 1282{,}4.$$

Рассмотрим оценки $\bar{X}=\frac{1}{n}\sum_i X_i\sim\mathrm{N}(\mu,\,\sigma^2/n)$ и $\hat{\sigma}^2=S_{XX}/(n-1)$, где $\hat{\sigma}^2/\sigma^2\sim$ $\sim\chi^2_{n-1}$. Вычислим их значения: $\bar{x}=73+\bar{z}/10=74{,}56$, $\frac{S_{xx}}{9}=\frac{S_{zz}}{9\cdot 10^2}\approx 1{,}425$.

Мы хотим проверить гипотезу H_0: $\mu=\mu_0=75$ против альтернативы $H_1:\mu\neq 75$, полагая $X_i\sim\mathrm{N}(\mu,\,\sigma^2)$ с неизвестным σ^2. Совместная п. р. в. имеет вид

$$f(\boldsymbol{x};\,\mu,\,\sigma^2)=\frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left[-\frac{1}{2\sigma^2}\sum(x_i-\mu)^2\right].$$

Правдоподобие гипотезы H_0 равно

$$\sup[f(\boldsymbol{x};\,\mu_0,\,\sigma^2):\,\sigma^2>0]=\frac{e^{-n/2}}{(2\pi\hat{\sigma}_0^2)^{n/2}},$$

где $\hat{\sigma}_0^2=\frac{1}{n}\sum(x_i-\mu_0)^2=\frac{1}{n}S_{xx}+(\bar{x}-\mu_0)^2$. Аналогично правдоподобие гипотезы H_1 равно

$$\sup[f(\boldsymbol{x};\,\mu,\,\sigma^2):\,\mu\in\mathbb{R},\,\sigma^2>0]=\frac{e^{-n/2}}{(2\pi\hat{\sigma}_1^2)^{n/2}},$$

где $\hat{\sigma}_1^2=\frac{1}{n}S_{xx}$.

Тогда

$$\Lambda_{H_1;H_0}(\boldsymbol{x})=\left(1+\frac{n(\bar{x}-\mu_0)^2}{S_{xx}}\right)^{n/2}.$$

Полагая $t=\frac{\sqrt{n}(\bar{x}-\mu_0)}{s_{xx}}$, где $s_{xx}=\sqrt{S_{xx}/(n-1)}$, мы отклоняем гипотезу H_0 при больших значениях Λ_{H_1,H_0}, т. е. когда $|t|$ велико.

Замечание. Для решения задачи нам не потребовалось привлекать о. о. п. $\Lambda_{H_1;H_0}$. Действительно, при выполнении гипотезы H_0 мы имеем

$$T=\frac{\sqrt{n}(\bar{X}-\mu_0)}{s_{XX}}\sim\mathrm{t}_{n-1}.$$

Таким образом, применяя критерий размера α, мы отклоняем H_0, если $|t|>\mathrm{t}_{n-1}(\alpha/2)$. Вычислим значение

$$|t|=\frac{\sqrt{10}|\bar{x}-75|}{\sqrt{S_{xx}/9}}=\left(\frac{10}{1{,}425}\right)^{1/2}\cdot 0{,}44\approx 1{,}166,$$

это число меньше $\mathrm{t}_9(0{,}025)=2{,}26$. Следовательно, гипотеза H_0 не отвергается при уровне значимости $\alpha=0{,}05$. □

Задача 17. Из каждой сотни бетонных смесей выбирают шесть блоков-образцов, которые проходят тестирование на прочность, и число тех блоков из шести, которые не прошли тест, заносят в таблицу:

число x блоков, не прошедших тест	0	1	2	3	4	5	6
число смесей, в которых x блоков не прошло тест	53	32	12	2	1	0	0

В предположении, что вероятность не пройти тест одинакова для всех блоков, получите несмещенную оценку этой вероятности и объясните, как получить 95%-доверительный интервал для этой вероятности.

Решение. В этом примере рассматриваются $n = 100$ н. о. р. с. в. $X_i \sim \sim \mathrm{Bin}(6, p)$. Несмещенной оценкой для p является $\hat{p} = \frac{1}{6n}\sum_{k=0}^{6} kn_k$, и ее значение равно

$$\frac{1 \cdot 32 + 2 \cdot 12 + 3 \cdot 2 + 4 \cdot 1}{600} = \frac{66}{600} = 0{,}11.$$

В силу ЦПТ мы имеем $\hat{P} = \frac{1}{6n}\sum_{k=0}^{6} kN_k \sim \mathrm{N}(p,\, p(1-p)/(6n))$ при достаточно больших n, так что $(\hat{p} - p)\sqrt{\frac{6n}{p(1-p)}} \sim \mathrm{N}(0,\, 1)$. Поскольку $\hat{p} \approx p$, запишем

$$\frac{p(1-p)}{6n} \approx \frac{\hat{p}(1-\hat{p})}{6n} = \frac{0{,}11 \cdot 0{,}89}{600} \approx (0{,}0128)^2.$$

Для $\mathrm{N}(0,\, 1)$-распределения имеем $z_{0{,}975} \approx 1{,}96$, следовательно, приближенный доверительный интервал имеет вид

$$\left(\hat{p} - z_{0{,}975}\sqrt{\frac{\hat{p}(1-\hat{p})}{6n}},\, \hat{p} + z_{0{,}975}\sqrt{\frac{\hat{p}(1-\hat{p})}{6n}}\right) \approx$$
$$\approx (0{,}11 - 1{,}96 \cdot 0{,}0128,\ 0{,}11 + 1{,}96 \cdot 0{,}0128) \approx (0{,}085,\ 0{,}135). \qquad \square$$

Задача 18. а) Объясните, что понимают под априорным распределением, апостериорным распределением и байесовской оценкой. Как связана байесовская оценка с апостериорным распределением при выборе квадратической функции потерь и при функции потерь, равной абсолютной ошибке?

б) Пусть $X_1, \ldots, X_n$ — независимые одинаково распределенные случайные величины с равномерным распределением на интервале $(\theta - 1,\, \theta + 1)$, и пусть априорное распределение параметра θ является равномерным на интервале $(20,\, 50)$. Вычислите апостериорное распределение параметра θ при заданном $\boldsymbol{x} = (x_1, \ldots, x_n)$ и найдите оптимальную байесовскую оценку для θ при квадратической функции потерь и при функции потерь, равной абсолютной ошибке.

Решение. а) Байесовская оценка минимизирует средние апостериорные потери $h(a)$. Для квадратической ф. п. имеем

$$h(a) = \int (a - \theta)^2 \pi(\theta \,|\, \boldsymbol{x})\, d\theta.$$

Следовательно, минимум функции h достигается в такой точке a, что $h'(a) = 0$, т. е.

$$a \int \pi(\theta \mid \boldsymbol{x})\, d\theta = \int \theta\pi(\theta \mid \boldsymbol{x})\, d\theta.$$

Таким образом, $a(\boldsymbol{x}) = \int \theta\pi(\theta \mid \boldsymbol{x})\, d\theta$, т. е. это апостериорное среднее.

Для функции потерь, равной абсолютной ошибке,

$$h(a) = \int |a - \theta| \pi(\theta \mid \boldsymbol{x})\, d\theta,$$

и $h'(a) = 0$, когда

$$\int_{-\infty}^{a} \pi(\theta \mid \boldsymbol{x})\, d\theta = \int_{a}^{\infty} \pi(\theta \mid \boldsymbol{x})\, d\theta = \frac{1}{2},$$

т. е. в апостериорной медиане.

б) Если апостериорная плотность $\pi(\theta)$ равна $\frac{1}{30} I(20 < \theta < 50)$, то функция правдоподобия принимает вид

$$f(\boldsymbol{x}; \theta) = \left(\frac{1}{2}\right)^n \prod_{i=1}^{n} I(|x_i - \theta| < 1)$$

и апостериорная плотность такова:

$$\pi(\theta \mid \boldsymbol{x}) \propto I(20 < \theta < 50) I(\max x_i - 1 < \theta < \min x_i + 1) =$$
$$= I(20 \vee (\max x_i - 1) < \theta < 50 \wedge (\min x_i + 1)),$$

где $a \vee b = \max\{a, b\}$, $a \wedge b = \min\{a, b\}$. Таким образом, апостериорное распределение является равномерным на этом интервале. Следовательно, обе функции потерь, и квадратическая, и абсолютной ошибки, приводят к одной и той же байесовской оценке

$$\hat{\theta} = \frac{1}{2}[20 \vee (\max x_i - 1) + 50 \wedge (\min x_i + 1)]. \qquad \square$$

Задача 19. Пусть $X_1, \ldots, X_n$ — независимые одинаково распределенные $\mathrm{N}(\mu, \mu^2)$-с. в., $\mu > 0$.

Найдите двумерную достаточную статистику для μ, цитируя при этом все необходимые вам результаты.

Чему равна оценка максимального правдоподобия параметра μ?

Решение. Мы имеем

$$f(\boldsymbol{x}, \mu) = \prod_{i=1}^{n} f(x_i, \mu) \propto \exp\left(-\frac{1}{2\mu^2} \sum_{i=1}^{n} x_i^2 + \frac{1}{\mu} \sum_{i=1}^{n} x_i - n \ln \mu\right).$$

В силу критерия факторизации $(T_2, T_1) = \left(\sum\limits_{i=1}^{n} x_i^2, \sum\limits_{i=1}^{n} x_i\right)$ является достаточной статистикой для μ.

Оценка максимального правдоподобия максимизирует функцию $f(\boldsymbol{x};\,\mu)$ или

$$\ell(\boldsymbol{x};\,\mu) = -\frac{1}{2\mu^2}T_2 + \frac{1}{\mu}T_1 - n\ln\mu.$$

Решая уравнение $\dfrac{\partial}{\partial\mu}\ell = 0$, находим

$$\hat{\mu} = \frac{1}{2n}\left(-T_1 + \sqrt{T_1^2 + 4nT_2}\right). \qquad \square$$

Задача 20. а) Что называют простой гипотезой? Определите понятия размера и мощности для критерия проверки простой гипотезы против простой альтернативы. Сформулируйте (без доказательства) лемму Неймана—Пирсона.

б) Пусть X — случайная величина с распределением F. Рассмотрим нулевую гипотезу

H_0: F является стандартным нормальным распределением $\mathrm{N}(0,\,1)$

против альтернативы

H_1: F — двойное экспоненциальное распределение с плотностью $\dfrac{1}{4}e^{-|x|/2}$, $x\in\mathbb{R}$.

Найдите критерий размера α, $\alpha < 1/4$, имеющий наибольшую мощность, и покажите что его мощность равна $e^{-t/2}$, где $\Phi(t) = 1 - \dfrac{\alpha}{2}$ и Φ — функция нормального распределения $\mathrm{N}(0,\,1)$.

Указание. Используйте соотношение $\mathrm{P}_{H_0}(|X| > 1) = \Phi(-1) + 1 - \Phi(1) = 0{,}3174$.

Решение. См. пример 2.2.4. $\square$

Задача 21. Предположим, что одномерная случайная величина X имеет равномерное распределение на отрезке $[0,\,\theta]$ и требуется оценить θ при функции потерь

$$\ell(\theta,\,a) = c(\theta - a)^2,$$

где $c > 0$.

Найдите апостериорное распределение для θ и оптимальную байесовскую оценку относительно априорного распределения с плотностью $\pi(\theta) = \theta e^{-\theta}$, $\theta > 0$.

Решение. Мы хотим минимизировать по a средние потери

$$h(a) = \int \pi(\theta\,|\,x)\ell(\theta,\,a)\,d\theta.$$

Здесь апостериорная плотность равна

$$\pi(\theta\,|\,x) \propto \pi(\theta)f(x;\,\theta) = \theta e^{-\theta}\frac{1}{\theta}I(0 \leqslant x \leqslant \theta) = e^{-\theta}I(\theta \geqslant x \geqslant 0).$$

Значит,

$$h(a) = c \int_x^\infty e^{-\theta}(\theta - a)^2\, d\theta;$$

для отыскания минимума мы решаем уравнение $h'(a) = 0$, т. е.

$$\int_x^\infty e^{-\theta}\theta\, d\theta = a \int_x^\infty e^{-\theta}\, d\theta \quad \text{или} \quad (x+1)e^{-x} = ae^{-x}.$$

Следовательно, $a = x + 1$. □

Задача 22. а) Что такое критерий обобщенного отношения правдоподобия? Объясните, как применяют такой критерий.

б) Пусть $X_1, \ldots, X_n$ — независимые случайные величины и X_i имеет распределение Пуассона с неизвестным средним λ_i, $i = 1, \ldots, n$.

Найдите вид статистики обобщенного отношения правдоподобия для проверки гипотезы $H_0\colon \lambda_1 = \ldots = \lambda_n$ и покажите, что его можно аппроксимировать выражением

$$\frac{1}{\overline{X}} \sum_{i=1}^{n} (X_i - \overline{X})^2,$$

где $\overline{X} = n^{-1} \sum\limits_{i=1}^{n} X_i$.

Если для $n = 7$ вы определили, что значение этой статистики равно 27,3, то следует ли вам принять гипотезу H_0? Обоснуйте свой ответ.

Решение. См. примеры 2.5.5 и 2.5.6. □

Задача 23. Рассмотрим линейную модель регрессии

$$Y_i = \beta x_i + \varepsilon_i,$$

$i = 1, \ldots, n$, где $x_1, \ldots, x_n$ — заданные константы, а $\varepsilon_1, \ldots, \varepsilon_n$ — независимые одинаково распределенные $\mathrm{N}(0, \sigma^2)$-с. в. с неизвестной дисперсией σ^2.

Найдите МНК-оценку $\hat{\beta}$ для β. Укажите (без доказательства) распределение оценки $\hat{\beta}$ и опишите, как бы вы проверяли гипотезу $H_0\colon \beta = \beta_0$, где β_0 — заданное значение.

Решение. Заметим, что $Y_i \sim \mathrm{N}(\beta x_i, \sigma^2)$. МНК-оценка $\hat{\beta}$ минимизирует сумму $\sum\limits_i (Y_i - \mathsf{E} Y_i)^2 = \sum\limits_i (Y_i - \beta x_i)^2$. Продифференцировав по β, находим $\hat{\beta} = \sum\limits_i x_i Y_i \big/ \sum\limits_i x_i^2 \sim \mathrm{N}\left(\beta, \dfrac{\sigma^2}{\sum\limits_i x_i^2}\right)$.

Положим теперь

$$R = \sum_i (Y_i - \hat{\beta} x_i)^2 \sim \sigma^2 \chi^2_{n-1}.$$

Тогда R не зависит от $\hat{\beta}$. Чтобы проверить гипотезу H_0, вычислим значение

$$t = \frac{\hat{\beta} - \beta_0}{\sqrt{R/(n-1)}} \sqrt{\sum_i x_i^2}$$

и сравним его с $\mathrm{t}_{n-1}(\alpha/2)$, где $\mathrm{t}_{n-1}(\alpha/2)$ ищется из соотношения $\mathrm{P}_{\mathrm{t}_{n-1}}(X < \mathrm{t}_{n-1}(\alpha/2)) = 1 - \alpha/2$. Двусторонний критерий состоит в том, чтобы отвергнуть гипотезу H_0, если $t > \mathrm{t}_{n-1}(\alpha/2)$ или $t < -\mathrm{t}_{n-1}(\alpha/2)$. □

Задача 24. а) Пусть $X_1, \ldots, X_n$ — независимые одинаково распределенные $\mathrm{N}(\mu, \sigma^2)$-случайные величины, где μ и σ^2 неизвестны.

Найдите оценки максимального правдоподобия $\hat{\mu}$, $\hat{\sigma}^2$ параметров μ, σ^2, основываясь на $X_1, \ldots, X_n$. Покажите, что $\hat{\mu}$ и $\hat{\sigma}^2$ независимы, и найдите их распределения.

б) Предположим, что требуется построить интервал предсказания $\mathcal{I}(X_1, \ldots, X_n)$ для $\mathrm{N}(\mu, \sigma^2)$-случайной величины X_0, независимой от предыдущих. Потребуем, чтобы выполнялось ограничение

$$\mathrm{P}\{X_0 \in \mathcal{I}(X_1, \ldots, X_n)\} = 1 - \alpha$$

на совместное распределение $X_0, X_1, \ldots, X_n$.

Пусть

$$\mathcal{I}_\gamma(X_1, \ldots, X_n) = \left(\hat{\mu} - \gamma\hat{\sigma}\sqrt{1 + \frac{1}{n}},\ \hat{\mu} + \gamma\hat{\sigma}\sqrt{1 + \frac{1}{n}}\right).$$

Рассмотрев распределение с. в. $(X_0 - \hat{\mu}) \Big/ \left(\hat{\sigma}\sqrt{\dfrac{n+1}{n-1}}\right)$, найдите значение γ, для которого $\mathrm{P}\{X_0 \in \mathcal{I}_\gamma(X_1, \ldots, X_n)\} = 1 - \alpha$.

Решение. а) Запишем

$$f(\boldsymbol{x}; \mu, \sigma^2) = \prod_{i=1}^{n} f(x_i; \mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_i (x_i - \mu)^2\right].$$

Для того чтобы найти о. м. п., решаем уравнение

$$\frac{\partial}{\partial\mu} \ln f(\boldsymbol{x}; \mu, \sigma^2) = \frac{\partial}{\partial\sigma^2} \ln f(\boldsymbol{x}; \mu, \sigma^2) = 0$$

и получаем, что $\hat{\mu} = \bar{x}$, $\hat{\sigma}^2 = \dfrac{1}{n} S_{xx}$, где $S_{xx} = \sum_i (x_i - \bar{x})^2$.

б) Утверждается, что $\bar{X} \sim \mathrm{N}\left(\mu, \dfrac{\sigma^2}{n}\right)$, $\dfrac{1}{\sigma^2} S_{XX} \sim \chi^2_{n-1}$ и эти с. в. независимы. Действительно, запишем

$$\sum_i (X_i - \mu)^2 = \sum_i (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2.$$

Пусть A — такая действительная ортогональная матрица размера $n \times n$, что вектор $\boldsymbol{Y} = A^T(\boldsymbol{X} - \mu)$ имеет первую компоненту $Y_1 = \sqrt{n}(\overline{X} - \mu)$. Компоненты $Y_1, \ldots, Y_n$ вектора $\boldsymbol{Y}$ являются н. о. р. $\mathrm{N}(0, \sigma^2)$-с. в. Поэтому

$$\sum_{i=2}^{n} Y_i^2 \sim \sigma^2 \chi^2_{n-1} \text{ и эта сумма не зависит от } Y_1.$$

Более того, $\sum\limits_{i=1}^{n} Y_i^2 = \sum\limits_{i=1}^{n} (X_i - \mu)^2$. Следовательно,

$$\sum_{i=2}^{n} Y_i^2 = \sum_{i=1}^{n} Y_i^2 - Y_1^2 = \sum_{i=1}^{n} (X_i - \mu)^2 - n(\overline{X} - \mu)^2 = \sum_{i=1}^{n} (X_i - \overline{X})^2.$$

Далее, $\dfrac{X_0 - \hat{\mu}}{\sqrt{1 + \frac{1}{n}}\sigma} \sim \mathrm{N}(0, 1)$, и эта величина не зависит от $\dfrac{1}{\sigma^2} S_{XX} \sim \chi^2_{n-1}$.

Поскольку $s_{xx} = \hat{\sigma}\sqrt{n/(n-1)}$, видим, что $\dfrac{X_0 - \hat{\mu}}{\sqrt{\frac{n+1}{n-1}}\hat{\sigma}} \sim \mathrm{t}_{n-1}$. Тогда вероятность $\mathrm{P}(X_0 \in \mathcal{I}_\gamma)$ равна

$$\mathrm{P}\left(\frac{|X_0 - \hat{\mu}|}{\sqrt{\frac{n+1}{n-1}}\hat{\sigma}} \leqslant \gamma\sqrt{\frac{n-1}{n}}\right) = \mathrm{P}_{\mathrm{t}_{n-1}}\left(|X| \leqslant \gamma\sqrt{\frac{n-1}{n}}\right).$$

Отсюда получаем

$$\gamma = \sqrt{\frac{n}{n-1}}\mathrm{t}_{n-1}(\alpha/2). \qquad \square$$

Задача 25. Сформулируйте и докажите лемму Неймана—Пирсона. В своем ответе вам следует точно определить следующие понятия: размер, критическая область, вероятность ошибки II рода.

Решение. Пусть необходимо проверить простую нулевую гипотезу H_0: $f = f_0$ против альтернативы H_1: $f = f_1$, где f_0 и $f_1 > 0$ заданы в одной и той же области и непрерывны. Лемма Неймана—Пирсона гласит: *среди всех критериев размера не больше α критерий с наименьшей в. о. р. II (т. е. ошибкой принять гипотезу H_0, когда верна гипотеза H_1) задается критической областью $C = \{x\colon f_1(x)/f_0(x) > K\}$, где K определяется соотношением $\alpha = \mathrm{P}(X \in C \mid H_0) = \int\limits_C f_0(x)\,dx$.*

Доказательство леммы Н—П: см. § 2.2. $\square$

Задача 26. Пусть $X_1, \ldots, X_n$ — независимые одинаково распределенные $\mathrm{N}(\mu, \sigma^2)$-с. в., где дисперсия σ^2 неизвестна.

а) Укажите (без доказательства) вид совместного распределения величин $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$ и $S_{XX} = \sum_{i=1}^{n}(X_i - \bar{X})^2$.

б) Подробно опишите t-критерий Стьюдента для проверки нулевой гипотезы H_0: $\mu = \mu_0$ против альтернативной гипотезы H_1: $\mu \neq \mu_0$.

в) Что понимают под критерием обобщенного отношения правдоподобия? Подробно объясните, как применять этот критерий.

Постройте критерий обобщенного отношения правдоподобия для проверки гипотезы H_0: $\mu = \mu_0$ и покажите, что этот критерий является эквивалентным критерию из п. б).

Решение. а) Мы имеем $\bar{X} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$ и $\frac{1}{\sigma^2}S_{XX} \sim \chi^2_{n-1}$, и эти величины независимы.

б) По определению t-распределения находим, что

$$T(\boldsymbol{X}) = (\bar{X} - \mu) \Big/ \sqrt{\frac{\sigma^2}{n\sigma^2} s_{XX}} = \frac{\sqrt{n}(\bar{X} - \mu)}{s_{XX}} \sim \mathrm{t}_{n-1},$$

где $s_{XX} = \sqrt{S_{XX}/(n-1)}$. Следовательно, $T(\boldsymbol{X}) \sim \mathrm{t}_{n-1}$ при H_0. Таким образом, мы отвергаем гипотезу H_0 в критерии размера α, если наблюдаемое значение t статистики T либо больше чем верхняя квантиль $\mathrm{t}^+_{n-1}(\alpha/2)$, либо меньше чем $-\mathrm{t}^+_{n-1}(\alpha/2)$.

в) Предположим, что необходимо проверить гипотезу H_0: $\theta \in \Theta_0$ против общей альтернативы H_1: $\theta \in \Theta$, где $\Theta \supset \Theta_0$. Положим

$$\Lambda_{H_1;H_0}(\boldsymbol{x}) = \frac{\sup\left[f(\boldsymbol{x};\theta) : \theta \in \Theta\right]}{\sup\left[f(\boldsymbol{x};\theta) : \theta \in \Theta_0\right]},$$

при этом число степеней свободы $p = |\Theta| - |\Theta_0|$.

Критерий обобщенного отношения правдоподобия отклоняет гипотезу H_0 при больших значениях $\Lambda_{H_1;H_0}$. Так как выборка $\boldsymbol{X} = (X_1, \ldots X_n)$ имеет н. о. р. компоненты $X_1, \ldots, X_n$, при больших n мы имеем $2\ln\Lambda_{H_1;H_0}(\boldsymbol{X}) \sim \chi^2_p$ при H_0. Следовательно, мы отклоняем гипотезу H_0, применяя критерий размера (приблизительно) α, если $2\ln\Lambda_{H_1;H_0}(\boldsymbol{x}) > c$, где $\mathrm{P}(\chi^2_p > c) = \alpha$.

При выполнении гипотезы H_1 функция правдоподобия имеет вид

$$f(\boldsymbol{x};\mu,\sigma^2) = (2\pi\sigma^2)^{-n/2}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right].$$

Эта функция достигает максимума по μ, σ^2 в точке $\hat{\mu} = \bar{x}$ и $\hat{\sigma}^2 = \frac{1}{n}S_{xx}$. При выполнении гипотезы H_0 зафиксируем $\mu = \mu_0$ и найдем максимум по σ^2. Получим $\tilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_0)^2$.

Таким образом, при выполнении гипотезы H_1 имеем

$$\sup[f(\boldsymbol{x};\ \mu,\ \sigma^2)\colon \mu \in \mathbb{R},\ \sigma^2 > 0] = (2\pi S_{xx}/n)^{-n/2} e^{-n/2},$$

а если верна гипотеза H_0, то

$$\sup[f(\boldsymbol{x};\ \mu_0,\ \sigma^2)\colon \sigma^2 > 0] = \left(2\pi \sum_i (x_i - \mu_0)^2\right)^{-n/2} e^{-n/2}.$$

Следовательно,

$$\Lambda_{H_1:H_0}(\boldsymbol{x}) = \left[\frac{\sum (x_i - \mu_0)^2}{S_{xx}}\right]^{n/2} = \left[\frac{S_{xx} + n(\bar{x} - \mu_0)^2}{S_{xx}}\right]^{n/2} =$$
$$= \left[1 + \frac{n(\bar{x} - \mu_0)^2}{S_{xx}}\right]^{n/2} = \left(1 + \frac{t^2}{n-1}\right)^{n/2}.$$

Мы отвергаем гипотезу H_0 для больших значений $\Lambda_{H_1:H_0}$ (или $2\ln\Lambda_{H_1:H_0}$), т. е. когда t^2 (или $|t|$) велико, что совпадает с критерием из п. б). □

Задача 27. а) Объясните, что понимают под достаточной статистикой. Сформулируйте (без доказательства) критерий факторизации для достаточной статистики.

б) Пусть $X_1, \ldots, X_n$ — независимые одинаково распределенные с. в. с общей плотностью

$$f(x;\ \theta) = \theta e^{-\theta x + 1}, \qquad x \geqslant \frac{1}{\theta}.$$

Найдите двумерную достаточную статистику для θ.

Решение. а) Статистика $T = T(\boldsymbol{x})$ является достаточной для параметра θ, если условное распределение случайной выборки $\boldsymbol{X}$ при условии $T = t$ не зависит от θ при каждом t.

б) Совместная п. р. в. величин $X_1, \ldots, X_n$ равна

$$\prod_{i=1}^{n} f(x_i;\ \theta) = \prod_{i=1}^{n} \theta e^{-\theta x_i + 1} I\left(x_i > \frac{1}{\theta}\right) = \theta^n e^{-\theta \sum\limits_{i=1}^{n} x_i + n} I\left(\min x_i > \frac{1}{\theta}\right).$$

Таким образом, в силу критерия факторизации $T = \left(\min x_i,\ \sum\limits_{i=1}^{n} x_i\right)$ является достаточной статистикой для θ. □

Задача 28. Найдите выражения для оценок максимального правдоподобия α, β и σ^2 в линейной модели регрессии

$$Y_i = \alpha + \beta x_i + \varepsilon_i, \qquad 1 \leqslant i \leqslant n,$$

где $\varepsilon_1, \ldots, \varepsilon_n$ — независимые одинаково распределенные $\mathrm{N}(0, \sigma^2)$-с. в. и $\sum\limits_{i=1}^{n} x_i = 0$.

Найдите совместное распределение оценок максимального правдоподобия $\hat{\alpha}$, $\hat{\beta}$, $\hat{\sigma}^2$.

Подробно объясните, как

а) проверить гипотезу $H_0\colon \beta=\beta_0$ против альтернативы $H_1: \beta\neq\beta_0$,

б) построить 95%-доверительный интервал для $\alpha+\beta$.

Решение. Оценка максимального правдоподобия максимизирует по α, β и σ^2 функцию правдоподобия $f(\cdot;\, \alpha, \beta, \sigma^2)$ для величин $Y_1, \ldots, Y_n$, которая равна произведению

$$\prod_{i=1}^{n} f(y_i;\, \alpha, \beta, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\alpha-\beta x_i)^2\right].$$

Ясно, что о. м. п. для α и β минимизируют $\sum\limits_{i=1}^{n}(Y_i-\alpha-\beta x_i)^2$; они имеют вид $\hat{\alpha}=\overline{Y}$ и $\hat{\beta}=S_{xY}/S_{xx}$, где

$$\overline{Y}=n^{-1}\sum_i Y_i, \quad S_{xY}=\sum_i x_iY_i, \quad S_{xx}=\sum_i x_i^2.$$

Тогда $\hat{\sigma}^2=\dfrac{1}{n}R$, где

$$R=\sum_i (Y_i-\hat{\alpha}-\hat{\beta}x_i)^2.$$

Запишем $\boldsymbol{x}=\begin{pmatrix}x_1\\ \vdots\\ x_n\end{pmatrix}$ и отметим, что $R=\sum\limits_i Y_i^2-n\overline{Y}^2-(\boldsymbol{x}^T\boldsymbol{x})\hat{\beta}^2$. Чтобы найти совместное распределение, рассмотрим действительную ортогональную матрицу $A=(A_{ij})$ размера $n\times n$ вида

$$A=\begin{pmatrix} 1/\sqrt{n} & (\boldsymbol{x}^T\boldsymbol{x})^{-1/2}x_1 & \ldots & \ldots\\ 1/\sqrt{n} & (\boldsymbol{x}^T\boldsymbol{x})^{-1/2}x_2 & \ldots & \ldots\\ \ldots & \ldots & \ldots & \ldots\\ 1/\sqrt{n} & (\boldsymbol{x}^T\boldsymbol{x})^{-1/2}x_n & \ldots & \ldots \end{pmatrix},$$

где все столбцы начиная с третьего выбраны произвольным образом, но так, чтобы матрица A была ортогональной. Такая матрица существует, поскольку два указанных столбца

$$\begin{pmatrix}1/\sqrt{n}\\ \vdots\\ 1/\sqrt{n}\end{pmatrix} \quad \text{и} \quad \begin{pmatrix}(\boldsymbol{x}^T\boldsymbol{x})^{-1/2}x_1\\ \vdots\\ (\boldsymbol{x}^T\boldsymbol{x})^{-1/2}x_n\end{pmatrix}$$

являются ортонормальными (они ортогональны, так как $\sum\limits_{i=1}^{n} x_i=0$).

Положим $\boldsymbol{Z} = A^T\boldsymbol{Y}$, где $Z_j = \sum_i A_{ij}Y_i$. Тогда $\boldsymbol{Z} \sim \mathrm{N}(A^T(\alpha\mathbf{1} + \beta\boldsymbol{x}), \sigma^2\boldsymbol{I})$, где $\mathbf{1} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$, а $\boldsymbol{I}$ — единичная матрица. Таким образом, случайные величины $Z_1, \ldots, Z_n$ независимы, и $Z_i \sim \mathrm{N}((A^T(\alpha\mathbf{1} + \beta\boldsymbol{x}))_i, \sigma^2)$. Поскольку $Z_1 = \sqrt{n}\hat{\alpha} \sim \mathrm{N}(\sqrt{n}\alpha, \sigma^2)$, мы получаем $\hat{\alpha} \sim \mathrm{N}(\alpha, \sigma^2/n)$. Аналогично, поскольку $Z_2 = (\boldsymbol{x}^T\boldsymbol{x})^{1/2}\hat{\beta} \sim \mathrm{N}((\boldsymbol{x}^T\boldsymbol{x})^{1/2}\beta, \sigma^2)$, имеем $\hat{\beta} \sim \mathrm{N}(\beta, \sigma^2/(\boldsymbol{x}^T\boldsymbol{x}))$. Далее, $Z_3, \ldots, Z_n \sim \mathrm{N}(0, \sigma^2)$.

В силу ортогональности матрицы A выполняется равенство $\sum_i Z_i^2 = \sum_i Y_i^2$. При этом, с одной стороны, для $n \geqslant 3$ получаем

$$\sum_i Z_i^2 = n\hat{\alpha}^2 + (\boldsymbol{x}^T\boldsymbol{x})\hat{\beta}^2 + \sum_3^n Z_i^2,$$

а с другой стороны,

$$\sum_i Y_i^2 = R + n\overline{Y}^2 + (\boldsymbol{x}^T\boldsymbol{x})\hat{\beta}^2.$$

Следовательно,

$$\frac{1}{\sigma^2}n\hat{\sigma}^2 = \frac{1}{\sigma^2}R = \frac{1}{\sigma^2}\sum_3^n Z_i^2 \sim \chi^2_{n-2},$$

и эта величина не зависит от Z_1 и Z_2, т. е. от $\hat{\alpha}$ и $\hat{\beta}$.

Рассмотрим теперь п. а): если гипотеза H_0 верна, то

$$T = \frac{(\hat{\beta} - \beta_0)(\boldsymbol{x}^T\boldsymbol{x})^{1/2}}{(R/(n-2))^{1/2}} \sim \mathrm{t}_{n-2}.$$

Таким образом, мы отвергаем гипотезу H_0 в критерии размера α, если значение t статистики T будет больше $\mathrm{t}_{n-2}(\alpha/2)$, верхней $\alpha/2$-точки распределения t_{n-2}, или меньше $-\mathrm{t}_{n-2}(\alpha/2)$.

Далее рассмотрим п. б): запишем

$$\hat{\alpha} + \hat{\beta} \sim \mathrm{N}\left(\alpha + \beta, \sigma^2\left(\frac{1}{n} + \frac{1}{\boldsymbol{x}^T\boldsymbol{x}}\right)\right);$$

эта сумма не зависит от R. Таким образом,

$$\frac{\hat{\alpha} + \hat{\beta} - (\alpha + \beta)}{\left(\left(\frac{1}{n} + \frac{1}{\boldsymbol{x}^T\boldsymbol{x}}\right)R/(n-2)\right)^{1/2}} \sim \mathrm{t}_{n-2}.$$

Следовательно,

$$\mathrm{P}\left(-\mathrm{t}_{n-2}(\alpha/2) < \frac{\hat{\alpha} + \hat{\beta} - (\alpha + \beta)}{\left(\left(\frac{1}{n} + \frac{1}{\boldsymbol{x}^T\boldsymbol{x}}\right)R/(n-2)\right)^{1/2}} < \mathrm{t}_{n-2}(\alpha/2)\right) = 1 - \alpha,$$

т. е.

$$\mathrm{P}\Big(\hat\alpha+\hat\beta-\mathrm{t}_{n-2}(\alpha/2)\Big(\Big(\frac{1}{n}+\frac{1}{\boldsymbol{x}^T\boldsymbol{x}}\Big)R/(n-2)\Big)^{1/2}<\alpha+\beta<$$
$$<\hat\alpha+\hat\beta+\mathrm{t}_{n-2}(\alpha/2)\Big(\Big(\frac{1}{n}+\frac{1}{\boldsymbol{x}^T\boldsymbol{x}}\Big)R/(n-2)\Big)^{1/2}\Big)=1-\alpha.$$

Это означает, что

$$\Big(\hat\alpha+\hat\beta-\mathrm{t}_{n-2}(0{,}025)\Big(\Big(\frac{1}{n}+\frac{1}{\boldsymbol{x}^T\boldsymbol{x}}\Big)R/(n-2)\Big)^{1/2},$$
$$\hat\alpha+\hat\beta+\mathrm{t}_{n-2}(0{,}025)\Big(\Big(\frac{1}{n}+\frac{1}{\boldsymbol{x}^T\boldsymbol{x}}\Big)R/(n-2)\Big)^{1/2}\Big)$$

является 95%-доверительным интервалом для $\alpha+\beta$. □

Задача 29. а) Что понимают под 2×2-таблицей сопряженности признаков? Опишите две выборочные схемы, которые приводят к такой таблице.

б) Подробно объясните, как применяется критерий χ^2 для проверки независимости строк и столбцов в таблице сопряженности признаков размера 2×2. Кратко прокомментируйте, как влияет выборочная модель на проверку независимости.

Решение. а) Таблица сопряженности признаков размера 2×2 представляет собой 2×2-массив *частот*, полученных (например) при перекрестной классификации совокупности объектов (или индивидуумов) по *двум* признакам, где каждый из признаков имеет *две* ступени (разряда, уровня),

		Признак I	
		уровень 1	уровень 2
Признак II	уровень 1	a	b
	уровень 2	c	d

где a — число объектов, принадлежащих уровню 1 по признаку I и уровню 1 по признаку II, и т. д.

Рассматривают два случая: 1) фиксируется общая сумма $n=a+b+$ $+c+d$, и тогда совместное распределение величин (A, B, C, D) мультиномиальное с вероятностями ячеек $p_{11}, p_{12}, p_{21}, p_{22}$, или 2) фиксируют две частные суммы (например, суммы элементов по строкам $a+b$ и $c+d$), и тогда имеют дело с двумя независимыми биномиальными распределениями, например $A\sim \mathrm{Bin}(a+b, p_{11})$, $C\sim\mathrm{Bin}(c+d, p_{21})$.

б) Для проверки гипотезы о независимости в случае 1) выбирают нулевую гипотезу $p_{ij}=\alpha_i\beta_j$, $i, j=1, 2$, а в случае 2) нулевая гипотеза состоит в том, что $p_{11} = p_{21} = p$. Рассматривают статистику вида

$$Z^2 = \sum_{\text{ячейки}} \frac{(\text{наблюдаемое} - \text{ожидаемое})^2}{\text{ожидаемое}},$$

где ожидаемая частота ячейки (i, j) равна

(сумма элементов i-й строки) $\times$ (сумма элементов j-го столбца)$/n$.

При нулевой гипотезе $Z^2 \sim \chi_1^2$. Следовательно, (приблизительный) критерий размера α отклоняет нулевую гипотезу, когда Z^2 превосходит $h_1(\alpha)$, верхнюю α-точку распределения χ_1^2.

Выборочная модель не влияет на проверку гипотезы о независимости, поскольку статистика и ее асимптотическое распределение при нулевой гипотезе имеют один и тот же вид для обоих случаев. $\square$

Задача 30. а) Объясните термины «априорная» и «апостериорная» плотность, которые используются при байесовской методологии оценивания параметра θ.

При заданной функции потерь $\ell(\theta, d)$, которая определяет потери, понесенные при оценивании интересующего нас параметра θ величиной d, опишите процедуру вычисления оптимальной байесовской оценки параметра θ. Найдите общий вид точечной оценки, если $\ell(\theta, d) = (\theta - d)^2$.

б) Пусть $X_1, \ldots, X_n$ — независимые одинаково распределенные относительно θ с. в. и $\mathrm{P}(X_1 = 1 \mid \theta) = 1 - \mathrm{P}(X_1 = 0 \mid \theta) = \theta$. Предположим, что априорное распределение параметра θ является равномерным на интервале $(0, 1)$. Найдите оптимальную байесовскую оценку для функции потерь $\ell(\theta, d) = (\theta - d)^2$. Чему равна оценка максимального правдоподобия параметра θ?

Покажите, что оптимальная байесовская оценка θ является смещенной, а оценка максимального правдоподобия является несмещенной.

Указание. При целых a и b выполняется равенство

$$\int_0^1 x^{a-1}(1 - x)^{b-1}\, dx = (a - 1)!\,(b - 1)!/(a + b - 1)!$$

Решение. а) При байесовском оценивании параметр θ рассматривается как случайная величина. Априорное распределение выражает наши первоначальные предположения относительно θ в виде определенной п. р. в./д. ф. р. $\pi(\theta)$. Апостериорная плотность распределения имеет п. р. в./д. ф. р. $\pi(\theta \mid \boldsymbol{x})$; она уточняет наши предположения о θ в свете наблюдаемой выборки $\boldsymbol{x}$:

$$\pi(\theta \mid \boldsymbol{x}) \propto \pi(\theta) f(\boldsymbol{x}; \theta).$$

Оптимальная байесовская оценка d выбирается так, чтобы минимизировать (по d) средние апостериорные потери $\mathrm{E}_{\boldsymbol{x}}\ell(\theta, d)$, где $\mathrm{E}_{\boldsymbol{x}}$ — это математическое ожидание (по θ) относительно апостериорной плотности $\pi(\cdot\,|\,\boldsymbol{x})$:

$$\hat{d}\ (=\hat{d}(\boldsymbol{x})) = \arg\min \int \ell(\theta,\ d)\pi(\theta\,|\,\boldsymbol{x})\,d\theta \quad \text{или} \quad \sum_{\theta} \ell(\theta,\ d)\pi(\theta\,|\,\boldsymbol{x}).$$

Если $\ell(\theta,\ d) = (\theta - d)^2$, то минимум достигается при

$$\hat{d} = \int \theta\pi(\theta\,|\,\boldsymbol{x})\,d\theta \quad \text{или} \quad \sum_{\theta} \theta\pi(\theta\,|\,\boldsymbol{x}),$$

т. е. при апостериорном среднем.

б) В данном примере выборка $\boldsymbol{x} = (x_1,\ \ldots,\ x_n)$ имеет вид $x_i = \begin{cases} 0 \\ 1 \end{cases}$ и

$$f(\boldsymbol{x};\ \theta) = \prod_{i=1}^{n} f(x_i;\ \theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} = \theta^{\sum x_i}(1-\theta)^{n-\sum x_i}.$$

Тогда

$$\pi(\theta\,|\,\boldsymbol{x}) \propto \pi(\theta)f(\boldsymbol{x};\ \theta) \propto \theta^{\sum x_i}(1-\theta)^{n-\sum x_i},$$

т. е. апостериорным распределением является $\mathrm{Bet}(t+1,\ n-t+1)$-распределение:

$$\pi(\theta\,|\,\boldsymbol{x}) \propto \pi(\theta)f(\boldsymbol{x};\ \theta) = \frac{\theta^t(1-\theta)^{n-t}}{\int_0^1 \tilde{\theta}^t(1-\tilde{\theta})^{n-t}\,d\tilde{\theta}}, \qquad 0 < \theta < 1,$$

где $t := \sum x_i$.

Таким образом, оптимальная байесовская оценка при $\ell(\theta,\ d) = (\theta - d)^2$ задается формулой

$$\hat{d} = \frac{\int_0^1 \theta^{t+1}(1-\theta)^{n-t}\,d\theta}{\int_0^1 \theta^t(1-\theta)^{n-t}\,d\theta} = \frac{(t+1)!\,(n-t)!\,(n+1)!}{(n+2)!\,t!\,(n-t)!} = \frac{t+1}{n+2} = \frac{\sum x_i + 1}{n+2}$$

(здесь использовано указание).

С другой стороны, максимизируя $\ln f(\boldsymbol{x};\ \theta)$ по θ, находим $\hat{\theta} = \bar{x} = \sum x_i/n$. Оценка максимального правдоподобия имеет среднее $\mathsf{E}(\hat{\theta}) = \theta$, т. е. является несмещенной. А среднее $\mathsf{E}\hat{d} = \dfrac{1}{n+2}(n\theta + 1)$ отлично от θ, и, следовательно, $\hat{d}$ является смещенной оценкой. □

Задача 31. Пусть $X_1,\ \ldots,\ X_6$ — выборка из равномерного распределения на отрезке $[0,\ \theta]$, где $\theta \in [1,\ 2]$ — неизвестный параметр. Найдите несмещенную оценку параметра θ с дисперсией, меньшей чем $1/10$.

Решение. Положим

$$M = \max[X_1, \ldots, X_6].$$

Тогда M является достаточной статистикой для θ, и мы можем использовать ее для построения несмещенной оценки для θ. Находим

$$F_M(y;\, \theta) = \mathsf{P}_\theta(M < y) = \mathsf{P}_\theta(X_1 < y, \ldots, X_6 < y) = \\ = \prod_{i=1}^{6} \mathsf{P}_\theta(X_i < y) = (F_X(y;\, \theta))^6, \quad 0 \leqslant y \leqslant \theta.$$

Плотность распределения вероятностей случайной величины M имеет вид

$$f_M(y;\, \theta) = \frac{d}{dy} F_M(y;\, \theta) = \frac{6y^5}{\theta^6}, \quad 0 \leqslant y \leqslant \theta,$$

а среднее значение равно

$$\mathsf{E}M = \int_0^\theta y \frac{6y^5}{\theta^6}\, dy = \frac{6y^7}{7\theta^6}\Big|_0^\theta = \frac{6\theta}{7}.$$

Таким образом, несмещенной оценкой для θ будет величина $7M/6$. Далее,

$$\operatorname{Var} \frac{7M}{6} = \mathsf{E}\left(\frac{7M}{6}\right)^2 - \left(\mathsf{E}\frac{7M}{6}\right)^2 = \int_0^\theta \frac{7^2 \cdot 6y^7}{6^2\theta^6}\, dy - \theta^2 = \frac{7^2\theta^2}{8 \cdot 6} - \theta^2 = \frac{\theta^2}{48}.$$

Если $\theta \in [1,\, 2]$, то

$$\frac{\theta^2}{48} \in \left[\frac{1}{48},\, \frac{1}{12}\right],$$

т. е. $\theta^2/48 < 1/10$, как и требовалось. Следовательно, получен ответ:

$$\text{требуемая оценка} = \frac{7}{6} \max[X_1, \ldots, X_6]. \qquad \square$$

Задача 32. Пусть $X_1, \ldots, X_n$ — выборка из равномерного распределения на $[0,\, \theta]$, где $\theta \in (0,\, \infty)$ — неизвестный параметр.

а) Найдите одномерную достаточную статистику M для параметра θ и постройте 95%-доверительный интервал для θ, основываясь на M.

б) Предположим теперь, что θ — это случайная величина с априорной плотностью

$$\pi(\theta) \propto I(\theta \geqslant a)\theta^{-k},$$

где $a > 0$ и $k > 2$. Вычислите апостериорную плотность для θ и найдите оптимальную байесовскую оценку $\hat{\theta}$ при квадратической функции потерь $(\theta - \hat{\theta})^2$.

Решение. а) Положим

$$M = \max[X_1, \ldots, X_n].$$

Тогда M является достаточной статистикой для θ. Действительно, функция правдоподобия имеет вид

$$f(\boldsymbol{x}; \theta) = \begin{cases} \dfrac{1}{\theta^n}, & \theta > x_{\max}, \\ 0 & \text{в противном случае.} \end{cases}$$

Таким образом, имея θ и $x_{\max}$, можно вычислить $f(\cdot; \theta)$. Это означает, что M является достаточной статистикой. (Формально это следует из критерия факторизации.)

Далее, $\mathsf{P}_\theta(\theta \geqslant M) = 1$. Следовательно, мы можем построить доверительный интервал с левым концом M и таким правым концом $b(M)$, что $\mathsf{P}(\theta \leqslant b(M)) = 0{,}95$. Запишем

$$\mathsf{P}(\theta \leqslant b(M)) = 1 - \mathsf{P}(b(M) < \theta) = 1 - \mathsf{P}(M < b^{-1}(\theta)) = 1 - 0{,}05;$$

так как п. р. в. равна

$$f_M(t; \theta) = n\frac{t^{n-1}}{\theta^n} I(0 < x < \theta),$$

мы получаем

$$\frac{(b^{-1}(\theta))^n}{\theta^n} = 0{,}05, \quad \text{следовательно,} \quad b^{-1}(\theta) = \theta(0{,}05)^{1/n}.$$

Тогда из равенства $M = \theta(0{,}05)^{1/n}$ получаем $\theta = M/(0{,}05)^{1/n}$. Итак, 95%-д. и. для θ имеет вид $(M, M/(0{,}05)^{1/n})$.

б) Из формулы Байеса

$$\pi(\theta \mid \boldsymbol{x}) = \frac{f(\boldsymbol{x}; \theta)\pi(\theta)}{\int f(\boldsymbol{x} \mid \varphi)\pi(\varphi)\, d\varphi}$$

следует, что апостериорная п. р. в. имеет вид

$$\pi(\theta \mid \boldsymbol{x}) \propto f(\boldsymbol{x}; \theta)\pi(\theta) \propto \frac{1}{\theta^{n+k}} I(\theta \geqslant c),$$

где

$$c = c(\boldsymbol{x}) = \max[a, x_{\max}].$$

Таким образом,

$$\pi(\theta \mid \boldsymbol{x}) = (n + k - 1)c^{n+k-1}\theta^{-n-k} I(\theta \geqslant c).$$

Далее, при квадратической ф. п. нужно минимизировать

$$\int \pi(\theta \mid \boldsymbol{x})(\theta - \hat{\theta})^2\, d\theta.$$

Отсюда апостериорное среднее равно

$$\hat{\theta}^* = \arg\min_{\hat{\theta}} \int \pi(\theta \mid \boldsymbol{x})(\theta - \hat{\theta})^2 \, d\theta = \int \theta\pi(\theta \mid \boldsymbol{x}) \, d\theta.$$

Поскольку

$$\int_c^\infty \theta^{-n-k+1} \, d\theta = \frac{1}{-n-k+2} \theta^{-n-k+2} \Big|_c^\infty = \frac{1}{n+k-2} c^{2-n-k},$$

после нормировки получим

$$\hat{\theta}^* = \frac{k+n-1}{k+n-2} c = \frac{k+n-1}{k+n-2} \max[a, x_1, \ldots, x_n]. \qquad \square$$

Задача 33. Кратко опишите стандартную процедуру, которую статистики применяют для проверки гипотез. В своем изложении вам нужно объяснить, в частности, почему нулевую гипотезу рассматривают иначе, чем альтернативу, а также описать, что понимают под критерием отношения правдоподобия.

Решение. Пусть даны наблюдения

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

из п. р. в./д. ф. р. f. Мы формулируем две взаимоисключающие гипотезы относительно f: H_0 (нулевая гипотеза) и H_1 (альтернатива).

Эти гипотезы имеют разный статус. Гипотеза H_0 трактуется как консервативная гипотеза, которая отклоняется только при явных свидетельствах против нее. Например,

а) H_0: $f = f_0$ против альтернативы $H_1: f = f_1$; обе функции f_0 и f_1 заданы (этот случай охватывает лемму Неймана—Пирсона);

б) H_0: $f = f_0$ против альтернативы $H_1: f \neq f_0$; функция f_0 задана (этот случай включает в себя теорему Пирсона, которая приводит к критерию χ^2);

в) $f = f(\cdot\,; \theta)$ определяется значением параметра; H_0: $\theta \in \Theta_0$ против $H_1: \theta \in \Theta_1$, где $\Theta_0 \cap \Theta_1 = \varnothing$ (например, семейства с монотонным отношением правдоподобия);

г) H_0: $\theta \in \Theta_0$ против $H_1: \theta \in \Theta$, где $\Theta_0 \subset \Theta$ и Θ имеет большую размерность, чем Θ_0.

Критерий задается такой критической областью C, что если $\boldsymbol{x} \in C$, то гипотеза H_0 отклоняется, тогда как при $\boldsymbol{x} \notin C$ она не отклоняется (что отражает консервативную природу гипотезы H_0). Ошибку I рода совершают, когда гипотеза H_0 отклоняется, в то время как она верна. Далее, вероятность ошибки I рода определяется как $\mathrm{P}(C)$ при выполнении гипотезы H_0;

мы говорим, что критерий имеет размер a (или $\leqslant a$), если $\max\limits_{H_0} \mathrm{P}(C) \leqslant a$. Мы выбираем a по своему усмотрению (например, 0,1, 0,01 и т. д.), устанавливая тем самым допустимый шанс отвергнуть гипотезу H_0, когда она верна. Затем мы ищем критерий заданного размера a, который минимизирует вероятность ошибки II рода $1 - \mathrm{P}(C)$, т. е. максимизирует мощность $\mathrm{P}(C)$ при гипотезе H_1.

Чтобы определить соответствующую критическую область, рассматривают отношение правдоподобия

$$\frac{\max f(\boldsymbol{x} \mid H_1)}{\max f(\boldsymbol{x} \mid H_0)},$$

где максимумы берут от п. р. в./д. ф. р. для гипотез H_0 и H_1 соответственно. Критическую область определяют как множество выборочных значений $\boldsymbol{x}$, для которых отношение правдоподобия принимает большие значения, учитывая заданный размер a, как указано выше. □

Задача 34. а) Сформулируйте и докажите лемму Неймана—Пирсона. Объясните, что понимают под равномерно наиболее мощным критерием.

б) Пусть $X_1, \ldots, X_n$ — выборка из нормального распределения со средним θ и дисперсией 1, где $\theta \in \mathbb{R}$ — неизвестный параметр. Найдите р. н. м. критерий размера $1/100$ для проверки гипотезы

$$H_0\colon \theta \leqslant 0 \quad \text{против } H_1\colon \theta > 0,$$

выразите ответ в терминах соответствующей функции распределения. Подробно объясните, почему критерий действительно является равномерно наиболее мощным критерием размера $1/100$.

Решение. а) Лемма Н—П применима, когда обе гипотезы, и нулевая, и альтернатива, являются простыми, т. е. $H_0\colon f = f_0$, $H_1\colon f = f_1$, где f_1 и f_0 — две п. р. в./д. ф. р., заданные в одной и той же области. Лемма Н—П утверждает: *для любого $k > 0$ критерий с критической областью $C = \{\boldsymbol{x}\colon f_1(\boldsymbol{x}) > k f_0(\boldsymbol{x})\}$ имеет наибольшую мощность $\mathrm{P}_1(C)$ среди всех критериев (т. е. критических областей) размера $\mathrm{P}(C)$.*

Доказательство леммы Н—П см. в §2.2.

Равномерно наиболее мощный критерий размера a для проверки гипотезы $H_0\colon \theta \in \Theta_0$ против альтернативы $H_1\colon \theta \in \Theta_1$ имеет такую критическую область C, что 1) $\max[\mathrm{P}_\theta(C)\colon \theta \in \Theta_0] \leqslant a$ и 2) для любой такой области C^*, что $\max[\mathrm{P}_\theta(C^*)\colon \theta \in \Theta_0] \leqslant a$, мы имеем $\mathrm{P}_\theta(C) \geqslant \mathrm{P}_\theta(C^*)$ для всех $\theta \in \Theta_1$.

б) В этом примере, когда $X_i \sim \mathrm{N}(\theta, 1)$, $H_0\colon \theta \leqslant 0$ и $H_1\colon \theta > 0$, мы зафиксируем $\theta_1 > 0$ и рассмотрим простую нулевую гипотезу о том, что $\theta = 0$, против простой альтернативы: $\theta = \theta_1$. Логарифм правдоподобия

$$\ln \frac{f(\boldsymbol{x};\, \theta_1)}{f(\boldsymbol{x};\, 0)} = \theta_1 \sum_i x_i - \frac{n}{2}\theta_1^2$$

принимает большие значения, когда сумма $\sum\limits_i x_i$ принимает большие значения. Выберем $k_1 > 0$ так, что

$$\frac{1}{100} = \mathrm{P}_0\Bigl(\sum_i X_i > k_1\Bigr) = \mathrm{P}_0\left(\frac{\sum\limits_i X_i}{\sqrt{n}} > \frac{k_1}{\sqrt{n}}\right) = 1 - \Phi\Bigl(\frac{k_1}{\sqrt{n}}\Bigr),$$

т. е. $k_1/\sqrt{n} = z_+(0{,}01) = \Phi^{-1}(0{,}99)$. Тогда

$$\mathrm{P}_\theta\Bigl(\sum_i X_i > k_1\Bigr) < \frac{1}{100}$$

для всех $\theta < 0$. Таким образом, критерий с критической областью

$$C = \Bigl\{\boldsymbol{x}\colon \sum_i X_i > k_1\Bigr\}$$

имеет размер 0,01 для случая, когда $H_0\colon \theta \leqslant 0$.

Далее, для любого $\theta' > 0$ область C можно записать в виде

$$C = \Bigl\{\boldsymbol{x}\colon \frac{f(\boldsymbol{x};\theta')}{f(\boldsymbol{x};0)} > k'\Bigr\}$$

при некотором $k' = k'(\theta') > 0$. По лемме Н—П мы имеем $\mathrm{P}_{\theta'}(C^*) \leqslant \mathrm{P}_{\theta'}(C)$, для любых таких областей C^*, что $\mathrm{P}_0(C^*) \leqslant 0{,}01$. Аналогичным образом проверяется, что для любого $\theta' > 0$ выполняется неравенство

$$\mathrm{P}_{\theta'}(C^*) \leqslant \mathrm{P}_{\theta'}(C) \quad \text{для любых таких областей } C^*, \text{ что} \quad \mathrm{P}_\theta(C^*) \leqslant \frac{1}{100} \text{ для любых } \theta \leqslant 0.$$

Таким образом, $C = \Bigl\{\boldsymbol{x}\colon \sum\limits_i X_i > k_1\Bigr\}$ определяет р. н. м. критерий размера 0,01 для проверки гипотезы H_0 против H_1. □

Задача 35. Студенты-математики большого университета во время итогового экзамена за год получают оценки, выраженные в процентах к высшему баллу. В выборке из девяти студентов зафиксированы следующие оценки:

28 32 34 39 41 42 42 46 56.

Студенты-историки также получают оценки в процентах. В выборке из пяти студентов зафиксированы следующие оценки:

53 58 60 61 68.

Свидетельствуют ли эти данные в поддержку гипотезы о том, что оценки математиков варьируются больше, чем оценки по истории? Подтвердите свои заключения вычислениями. Прокомментируйте выбор модели.

Указание.

Распределение	N(0, 1)	$F_{9,5}$	$F_{8,4}$	χ^2_{14}	χ^2_{13}	χ^2_{12}
95%-квантиль	1,65	4,78	6,04	23,7	22,4	21,0

Решение. Рассмотрим независимые с. в.

$$X_i \ (= X_i^M) \sim \mathrm{N}(\mu_1, \sigma_1^2), \quad i = 1, \ldots, 9,$$

и

$$Y_j \ (= Y_j^H) \sim \mathrm{N}(\mu_2, \sigma_2^2), \quad j = 1, \ldots, 5.$$

Если $\sigma_1^2 = \sigma_2^2$, то

$$F = \frac{1}{8}\sum_{i=1}^{9}(X_i - \bar{X})^2 \Big/ \left(\frac{1}{4}\sum_{i=1}^{5}(Y_i - \bar{Y})^2\right) \sim \mathrm{F}_{8,4},$$

где

$$\bar{X} = \frac{1}{9}\sum_i X_i, \quad \bar{Y} = \frac{1}{5}\sum_j Y_j.$$

Находим $\bar{X} = 40$ и значения, показанные в табл. 3.6.

Таблица 3.6

X_i	28	32	34	39	41	42	42	46	56
$X_i - \bar{X}$	−12	−8	−6	−1	1	2	2	6	16
$(X_i - \bar{X})^2$	144	64	36	1	1	4	4	36	256

При этом $\sum_i (X_i - \bar{X})^2 = 546$.

Аналогично $\bar{Y} = 60$, и мы находим значения, показанные в табл. 3.7.

Таблица 3.7

Y_j	53	58	60	61	68
$Y_j - \bar{Y}$	−7	−2	0	1	8
$(Y_j - \bar{Y})^2$	49	4	0	1	64

При этом $\sum_j (Y_j - \bar{Y})^2 = 118$.
Тогда

$$F = \frac{\frac{1}{8}\cdot 546}{\frac{1}{4}\cdot 118} = \frac{273}{118} \approx 2{,}31.$$

Однако $\varphi^+_{8,4}(0{,}05) = 6{,}04$. Таким образом, нет причин отвергнуть гипотезу H_0: $\sigma_1^2 = \sigma_2^2$ при уровне 95%, т. е. мы не принимаем гипотезу о том, что $\sigma_1^2 > \sigma_2^2$. □

Задача 36. Рассмотрим линейную модель регрессии

$$Y_i = \alpha + \beta x_i + \varepsilon_i, \qquad \varepsilon_i \sim \mathrm{N}(0, \sigma^2), \qquad i = 1, \ldots, n,$$

где $x_1, \ldots, x_n$ известны и $\sum\limits_{i=1}^{n} x_i = 0$ и где $\alpha, \beta \in \mathbb{R}$ и $\sigma^2 \in (0, \infty)$ неизвестны.
Найдите оценки максимального правдоподобия $\hat{\alpha}$, $\hat{\beta}$, $\hat{\sigma}^2$ и выпишите их распределения.

Рассмотрите следующие данные:

x_i	-3	-2	-1	0	1	2	3
Y_i	-5	0	3	4	3	0	-5

Согласуйте с этими данными линейную модель регрессии и объясните, является ли она приемлемой.

Решение. Случайная величина Y_i имеет п. р. в.

$$f_{Y_i}(y;\, \alpha, \beta, \sigma^2) = (2\pi\sigma^2)^{-1/2} e^{-(y-\alpha-\beta x_i)^2/2\sigma^2},$$

и

$$\ln f_{\mathbf{Y}}(\mathbf{y};\, \alpha, \beta, \sigma^2) = -\frac{n}{2}\ln\sigma^2 - \sum_i \frac{(y_i - \alpha - \beta x_i)^2}{2\sigma^2}.$$

Чтобы найти о. м. п. $\hat{\alpha}$ и $\hat{\beta}$, рассмотрим стационарные точки:

$$\frac{\partial}{\partial\alpha}\sum_i (y_i - \alpha - \beta x_i)^2 = -2n(\bar{y} - \alpha),$$

откуда получаем $\hat{\alpha} = \overline{Y}$;

$$0 = \frac{\partial}{\partial\beta}\sum_i (y_i - \alpha - \beta x_i)^2 = -2\sum_i x_i(y_i - \alpha - \beta x_i),$$

откуда получаем $\hat{\beta} = \dfrac{S_{xY}}{S_{xx}}$, где $\overline{Y} = \sum\limits_i Y_i/n$, $S_{xY} = \sum\limits_i x_i Y_i$ и $S_{xx} = \sum\limits_i x_i^2$. Тот факт, что в этих точках достигается глобальный максимум, следует из единственности стационарной точки и из того, что $f_Y(\boldsymbol{y};\, \alpha, \beta, \sigma^2) \to 0$, если любой из параметров α, β или σ^2 стремится к ∞ или когда $\sigma^2 \to 0$.

Положим $R = \sum\limits_{i=1}^{n} (Y_i - \hat{\alpha} - \hat{\beta} x_i)^2$, тогда в точке $\hat{\sigma}^2$ имеем

$$0 = \frac{\partial}{\partial\sigma^2}\left(-\frac{n}{2}\ln\sigma^2 - \frac{R}{2\sigma^2}\right) = -\frac{n}{\sigma^2} + \frac{R}{\sigma^4}, \quad \text{откуда получаем} \quad \hat{\sigma}^2 = \frac{R}{n}.$$

Распределения оценок следующие:

$$\hat{\alpha} \sim \mathrm{N}\left(\alpha, \frac{\sigma^2}{n}\right), \quad \hat{\beta} \sim \mathrm{N}\left(\beta, \sigma^2\Big/\sum_i x_i^2\right) \quad \text{и} \quad R/\sigma^2 \sim \chi^2_{n-2}.$$

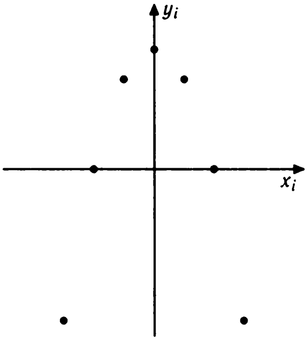


Рис. 3.3

В данном примере $\overline{Y} = 0$ и $S_{xY} = 0$, откуда следует, что $\hat{\alpha} = \hat{\beta} = 0$. Далее, $R = 84$, т. е. $\hat{\sigma}^2 = 14$.

Эта модель не особенно хороша, так как данные (x_i, Y_i) демонстрируют параболическую форму, а не линейную (см. рис. 3.3). □

Задача 37. Независимые наблюдения X_1, X_2 распределены как пуассоновские случайные величины с такими средними значениями μ_1, μ_2 соответственно, что

$$\ln \mu_1 = \alpha,$$
$$\ln \mu_2 = \alpha + \beta,$$

где α и β — неизвестные параметры. Выпишите функцию логарифма правдоподобия $\ell(\alpha, \beta)$ и затем найдите

а) $\dfrac{\partial^2 \ell}{\partial \alpha^2}$, $\dfrac{\partial^2 \ell}{\partial \alpha \, \partial \beta}$, $\dfrac{\partial^2 \ell}{\partial \beta^2}$,

б) оценку максимального правдоподобия $\hat{\beta}$ параметра β.

Решение. а) Для любых целых x_1, $x_2 \geqslant 0$ находим логарифм

$$\ell(x_1, x_2; \alpha, \beta) = \ln\left(e^{-\mu_1}\frac{\mu_1^{x_1}}{x_1!}e^{-\mu_2}\frac{\mu_2^{x_2}}{x_2!}\right) = -e^{\alpha} + x_1\alpha - e^{\alpha+\beta} + x_2(\alpha\beta) - \ln(x_1!\,x_2!).$$

Таким образом,

$$\frac{\partial^2}{\partial \alpha^2}\ell = -e^{\alpha}(1 + e^{\beta}), \qquad \frac{\partial^2}{\partial \alpha \, \partial \beta}\ell = -e^{\alpha+\beta}, \qquad \frac{\partial^2}{\partial \beta^2}\ell = -e^{\alpha+\beta}.$$

б) Рассмотрим уравнение для стационарной точки:

$$\frac{\partial}{\partial \alpha}\ell = 0 \;\Rightarrow\; x_1 + x_2 = e^{\hat{\alpha}} + e^{\hat{\alpha}+\hat{\beta}}, \qquad \frac{\partial}{\partial \beta}\ell = 0 \;\Rightarrow\; x_2 = e^{\hat{\alpha}+\hat{\beta}}.$$

Следовательно,

$$\hat{\alpha} = \ln x_1, \; \hat{\beta} = \ln \frac{x_2}{x_1}.$$

Стационарная точка доставляет глобальный максимум, так как она единственна и $\ell \to -\infty$ при $|\alpha|$, $|\beta| \to \infty$. □

Задача 38. Срок службы некоторого электронного устройства описывается показательной п. р. в.

$$f(t; \theta) = \frac{1}{\theta}\exp\left(-\frac{t}{\theta}\right) \qquad \text{для } t \geqslant 0,$$

где t — выборочное значение величины T.

Пусть $t_1, \ldots, t_n$ — случайная выборка из распределения с указанной плотностью. Основываясь на лемме Неймана—Пирсона, найдите наиболее мощный критерий размера 0,05 для проверки гипотезы

$$H_0\colon \theta = \theta_0 \quad \text{против} \quad H_1\colon \theta = \theta_1,$$

где θ_0 и θ_1 заданы и $\theta_0 < \theta_1$. Введем функцию

$$G_n(u) = \int\limits_0^u e^{-t} \frac{t^{n-1}}{(n-1)!}\, dt.$$

Покажите, что мощность этого критерия равна

$$1 - G_n\left(\frac{\theta_0}{\theta_1} G_n^{-1}(1-\alpha)\right),$$

где $\alpha = 0{,}05$.

Если для $n = 100$ наблюдается значение $\sum\limits_i t_i/n = 3{,}1$, то принимается ли гипотеза $H_0\colon \theta_0 = 2$? Обоснуйте свой ответ, используя асимптотическое распределение величины $(T_1 + \ldots + T_n)/n$.

Решение. Функция правдоподобия для выборочного вектора $\boldsymbol{t} \in \mathbb{R}^n$ имеет вид

$$f(\boldsymbol{t}; \theta) = \frac{1}{\theta^n} \exp\left(-\frac{1}{\theta} \sum_i t_i\right) I(\min t_i > 0), \qquad \boldsymbol{t} = \begin{pmatrix} t_1 \\ \vdots \\ t_n \end{pmatrix}.$$

По лемме Н—П н. м. критерию размера α будет соответствовать такая критическая область

$$C = \left\{\boldsymbol{t}\colon \frac{f_{\theta_1}(\boldsymbol{t})}{f_{\theta_0}(\boldsymbol{t})} > k\right\},$$

что

$$\int\limits_C f_{\theta_0}(\boldsymbol{t})\, d\boldsymbol{t} = 0{,}05.$$

Поскольку

$$\frac{f_{\theta_1}(\boldsymbol{t})}{f_{\theta_0}(\boldsymbol{t})} = \left(\frac{\theta_0}{\theta_1}\right)^n \exp\left[\left(\frac{1}{\theta_0} - \frac{1}{\theta_1}\right) \sum_i t_i\right] \quad \text{и} \quad \frac{1}{\theta_0} > \frac{1}{\theta_1},$$

область C имеет вид

$$\left\{\boldsymbol{t}\colon \sum_i t_i > c\right\}$$

для некоторого $c > 0$. При выполнении гипотезы H_0 имеем

$$X = \frac{1}{\theta_0}\sum_{i=1}^{n} T_i \sim \mathrm{Gam}(n, 1) \quad \text{и} \quad \mathsf{P}_{\theta_0}(X < u) = G_n(u).$$

Следовательно, чтобы получить н. м. критерий размера 0,05, выберем c так, чтобы выполнялось равенство

$$1 - G_n\left(\frac{c}{\theta_0}\right) = 0{,}05, \quad \text{т. е.} \quad \frac{c}{\theta_0} = G_n^{-1}(0{,}95).$$

Тогда мощность критерия равна

$$\int\limits_{C} f_{\theta_1}(t)\, dt = 1 - G_n\left(\frac{1}{\theta_1}c\right),$$

что равно

$$1 - G_n\left[\frac{\theta_0}{\theta_1} G_n^{-1}(0{,}95)\right],$$

как и требовалось.

Поскольку $\mathsf{E}T_i = \theta$ и $\mathrm{Var}\, T_i = n\theta^2$, для больших n имеем

$$\frac{\sum\limits_i T_i - n\theta}{\theta\sqrt{n}} \sim \mathrm{N}(0, 1)$$

в силу ЦПТ. При выполнении гипотезы H_0: $\theta_0 = 2$ получаем

$$\frac{\sum\limits_i T_i - n\theta_0}{\theta_0\sqrt{n}} = 5{,}5.$$

С другой стороны, $z_+(0{,}05) = 1{,}645$. Так как $5{,}5 > 1{,}645$, мы отвергаем гипотезу H_0. □

Задача 39. Рассматривается модель

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i, \qquad 1 \leqslant i \leqslant n,$$

где $x_1, \ldots, x_n$ — заданные значения, причем $\sum\limits_i x_i = 0$, и $\varepsilon_1, \ldots, \varepsilon_n$ — независимые нормальные ошибки, имеющие нулевое среднее и неизвестную дисперсию σ^2.

а) Получите уравнения для оценок максимального правдоподобия $(\hat\beta_0, \hat\beta_1, \hat\beta_2)$ для $(\beta_0, \beta_1, \beta_2)$. Не пытайтесь решать эти уравнения.

б) Получите выражение для оценки максимального правдоподобия $\hat\beta_1^*$ параметра β_1 в упрощенной модели

$$H_0\colon y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \qquad 1 \leqslant i \leqslant n,$$

где $\sum\limits_i x_i = 0$ и $\varepsilon_1, \ldots, \varepsilon_n$ распределены так же, как и ранее.

Решение. а) Так как с. в. $Y_i \sim \mathrm{N}(\beta_0 + \beta_1 x_i + \beta_2 x_i^2, \sigma^2)$ независимы, функция правдоподобия имеет вид

$$f(\boldsymbol{y} \mid \beta_0, \beta_1, \beta_2, \sigma^2) = \prod_{i=1}^{n} \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left[-\frac{1}{2\sigma^2}(y_i - \beta_0 - \beta_1 x_i - \beta_2 x_i^2)^2\right] =$$
$$= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i - \beta_2 x_i^2)^2\right].$$

Найдем уравнения для стационарной точки логарифма правдоподобия:

$$\frac{\partial}{\partial\beta_0}\ell = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i - \beta_2 x_i^2) = 0,$$
$$\frac{\partial}{\partial\beta_1}\ell = \frac{1}{\sigma^2}\sum_{i=1}^{n}x_i(y_i - \beta_0 - \beta_1 x_i - \beta_2 x_i^2) = 0,$$
$$\frac{\partial}{\partial\beta_2}\ell = \frac{1}{\sigma^2}\sum_{i=1}^{n}x_i^2(y_i - \beta_0 - \beta_1 x_i - \beta_2 x_i^2) = 0.$$

Замечание. В принципе σ^2 также должно фигурировать здесь как параметр. Полученная система уравнений все же содержит σ^2, но, к счастью, уравнение $\partial\ell/\partial\sigma^2 = 0$ можно опустить.

б) Эти же рассуждения справедливы и для упрощенной модели. Ответ имеет вид

$$\hat{\beta}_1 = \frac{S_{xY}}{S_{xx}},$$

где $S_{xY} = \sum\limits_i x_i Y_i$, $S_{xx} = \sum\limits_i x_i^2$, т. е. получены точные значения. И здесь опять σ^2 не появляется в выражении для $\hat{\beta}_1$. □

Задача 40. Пусть $(x_1, \ldots, x_n)$ — случайная выборка из распределения с нормальной п. р. в. со средним μ и дисперсией σ^2.

а) Выпишите функцию логарифма правдоподобия $\ell(\mu, \sigma^2)$.

б) Найдите пару достаточных статистик для неизвестных параметров (μ, σ^2), сославшись на соответствующую теорему.

в) Найдите оценку максимального правдоподобия $(\hat{\mu}, \hat{\sigma}^2)$ для (μ, σ^2). Используя стандартные распределения, постройте 95%-доверительный интервал для μ.

Решение. а) Логарифм правдоподобия имеет вид

$$\ell(\boldsymbol{x}; \mu, \sigma^2) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2.$$

б) В силу критерия факторизации $T(\boldsymbol{x})$ является достаточной статистикой для (μ, σ^2) тогда и только тогда, когда $\ell(\boldsymbol{x}; \mu, \sigma^2) = \ln g(T(\boldsymbol{x}), \mu, \sigma^2) + \ln h(\boldsymbol{x})$ для некоторых функций g и h. Запишем

$$\ell(\boldsymbol{x}; \mu, \sigma^2) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_i [(x_i - \bar{x}) + (\bar{x} - \mu)]^2.$$

Следующие вычисления касаются только суммы $\sum\limits_i$:

$$-\frac{1}{2\sigma^2}\sum_{i=1}^n [(x_i-\bar{x})^2+2(x_i-\bar{x})(\bar{x}-\mu)+(\bar{x}-\mu)^2] = -\frac{1}{2\sigma^2}\left[\sum_{i=1}^n (x_i-\bar{x})^2+n(\bar{x}-\mu)^2\right],$$

поскольку $\sum\limits_{i=1}^n (x_i - \bar{x}) = 0$.

Таким образом, выбор $T(\boldsymbol{X}) = \left(\bar{X}, \sum\limits_i (X_i - \bar{X})^2\right)$ удовлетворяет критерию факторизации (при $h \equiv 1$). Значит, $T(\boldsymbol{X}) = \left(\bar{X}, \sum\limits_i (X_i - \bar{X})^2\right)$ — достаточная статистика для (μ, σ^2).

в) Оценки максимального правдоподобия для (μ, σ^2) находятся из уравнений

$$\frac{\partial}{\partial\mu}\ell = \frac{\partial}{\partial\sigma^2}\ell = 0$$

и имеют вид

$$\hat{\mu} = \bar{x}, \quad \hat{\sigma}^2 = \frac{S_{xx}}{n}, \quad \text{где} \quad S_{xx} = \sum_i (x_i - \bar{x})^2.$$

Известно, что

$$\bar{X} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right) \quad \text{и} \quad \frac{1}{\sigma^2}S_{XX} \sim \chi^2_{n-1}.$$

Поэтому

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \Bigg/ \frac{1}{\sigma}\sqrt{\frac{S_{XX}}{n-1}} \sim \mathrm{t}_{n-1}.$$

Следовательно, если $\mathrm{t}_{n-1}(0{,}025)$ — это верхняя точка распределения t_{n-1} ($\mathrm{P}_{\mathrm{t}_{n-1}}(X > \mathrm{t}_{n-1}(0{,}025)) = 0{,}025$), то

$$\left(\bar{x} - \frac{1}{\sqrt{n}}s_{xx}\mathrm{t}_{n-1}(0{,}025),\ \bar{x} + \frac{1}{\sqrt{n}}s_{xx}\mathrm{t}_{n-1}(0{,}025)\right)$$

является 95%-д. и. для μ. Здесь $s_{xx} = \sqrt{S_{xx}/(n-1)}$. □

Задача 41. Предположим, что при заданном действительном параметре θ наблюдение X имеет нормальное распределение со средним θ и дисперсией v, где v известно. Пусть априорная плотность для θ равна

$$\pi(\theta) \propto \exp(-(\theta - \mu_0)^2/2v_0),$$

где μ_0 и v_0 заданы. Покажите, что апостериорная плотность для θ равна $\pi(\theta \mid x)$, где

$$\pi(\theta \mid x) \propto \exp(-(\theta - \mu_1)^2/2v_1),$$

а μ_1 и v_1 определяются равенствами

$$\mu_1 = \frac{\mu_0/v_0 + x/v}{1/v_0 + 1/v}, \qquad \frac{1}{v_1} = \frac{1}{v_0} + \frac{1}{v}.$$

Сделайте набросок типичных графиков кривых $\pi(\theta)$ и $\pi(\theta \mid x)$, отметив μ_0 и x на θ-оси.

Решение. Заметим, что

$$f(x;\, \theta) \propto \exp\left[-\frac{1}{2v}(x - \theta)^2\right], \quad -\infty < x < \infty.$$

Поэтому

$$\pi(\theta \mid x) \propto f(x;\, \theta)\pi(\theta) \propto \exp\left[-\frac{1}{2v}(x - \theta)^2 - \frac{1}{2v_0}(\theta - \mu_0)^2\right].$$

Запишем

$$\frac{(x-\theta)^2}{v} + \frac{(\theta - \mu_0)^2}{v_0} = \left(\frac{1}{v} + \frac{1}{v_0}\right)\theta^2 - 2\theta\left(\frac{x}{v} + \frac{\mu_0}{v_0}\right) + \frac{\mu_0^2}{v_0} + \frac{x^2}{v} =$$
$$= \left(\frac{1}{v} + \frac{1}{v_0}\right)\left(\theta - \frac{x/v + \mu_0/v_0}{1/v_0 + 1/v}\right)^2 - \left(\frac{x/v + \mu_0/v_0}{1/v_0 + 1/v}\right)^2 + \frac{\mu_0^2}{v_0} + \frac{x^2}{v} =$$
$$= \frac{1}{v_1}(\theta - \mu_1)^2 + \text{члены, не содержащие } \theta,$$

где μ_1, v_1 — искомые значения.

Таким образом,

$$\pi(\theta \mid x) \propto \exp\left[-\frac{1}{2v_1}(\theta - \mu_1)^2 - \frac{k}{2}\right] \propto \exp\left[-\frac{1}{2v_1}(\theta - \mu_1)^2\right].$$

Обе п. р. в. $\pi(\theta)$ и $\pi(\cdot \mid x)$ являются нормальными п. р. в.; как видно из рис. 3.4, дисперсия величины π больше, чем дисперсия величины $\pi(\cdot \mid x)$. □

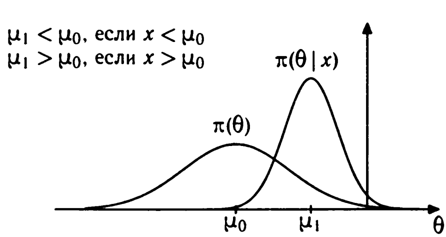


Рис. 3.4

Задача 42. а) Пусть (n_{ij}) — значения частот для таблицы сопряженности признаков размера $r \times c$, $n = \sum\limits_{i=1}^{r} \sum\limits_{j=1}^{c} n_{ij}$,

$$\mathsf{E}(n_{ij}) = np_{ij}, \quad 1 \leqslant i \leqslant r, \quad 1 \leqslant j \leqslant c,$$

и, таким образом, $\sum\limits_{i} \sum\limits_{j} p_{ij} = 1$.

При обычном предположении о том, что (n_{ij}) — это выборка из мультиномиального распределения, покажите, что статистика отношения правдоподобия для проверки гипотезы

$$H_0\colon p_{ij} = \alpha_i \beta_j$$

для всех (i, j) и для некоторых векторов $\alpha = (\alpha_1, \ldots, \alpha_r)$ и $\beta = (\beta_1, \ldots, \beta_c)$ имеет вид

$$D = 2 \sum_{i=1}^{r} \sum_{j=1}^{c} n_{ij} \ln(n_{ij}/e_{ij}),$$

причем (e_{ij}) необходимо определить. Покажите далее, что для малых значений $|n_{ij} - e_{ij}|$ статистику D можно аппроксимировать величиной

$$Z^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} (n_{ij} - e_{ij})^2 / e_{ij}.$$

б) В 1843 г. Уильям Гай собрал данные, приведенные в табл. 3.8, о 1659 бывших пациентах некоторого госпиталя, определяющие уровень их физических усилий во время работы и вид заболевания (туберкулез легких или другое заболевание).

Таблица 3.8

Уровень усилий во время работы	Вид заболевания	
	Туберкулез легких	Другое заболевание
Незначительное	125	385
Среднее (периодическое)	41	136
Постоянное	142	630
Значительное	33	167

Оказалось, что для этих данных значение Z^2 равно 9,84. Какой отсюда можно сделать вывод?

(Заметим, что на этот вопрос можно ответить и без применения калькулятора или статистических таблиц.)

Решение. а) Пусть n_{ij} обозначают частоты в таблице сопряженности. Положим

$$n = \sum_{i=1}^{r} \sum_{j=1}^{c} n_{ij}, \quad \mathsf{E}(n_{ij}) = np_{ij}, \quad 1 \leqslant i \leqslant r, \quad 1 \leqslant j \leqslant c,$$

с ограничением $\sum\limits_{i,j} p_{ij} = 1$. Далее,

$$H_0\colon p_{ij} = \alpha_i \beta_j;\ \sum_i \alpha_i = \sum_j \beta_j = 1; \quad H_1: p_{ij} \text{ произвольные}, \sum_{i,j} p_{ij} = 1.$$

При выполнении гипотезы H_1 и при мультиномиальном распределении вероятность наблюдать выборку (n_{ij}) равна $\prod\limits_{i,j} p_{ij}^{n_{ij}}/n_{ij}!$. Логарифм правдоподобия как функция аргументов p_{ij} равен

$$\ell(p_{ij}) = \sum_{i,j} n_{ij} \ln p_{ij} + A,$$

где $A = -\sum\limits_{i,j} \ln(n_{ij}!)$ не зависит от p_{ij}. Следовательно, при ограничениях $\sum\limits_{i,j} p_{ij} = 1$ функция ℓ достигает максимума в точке

$$\hat{p}_{ij} = \frac{n_{ij}}{n}.$$

При выполнении гипотезы H_0 логарифм правдоподобия равен

$$\ell(\alpha_i, \beta_j) = \sum_i n_{i+} \ln \alpha_i + \sum_j n_{+j} \ln \beta_j + B,$$

где B не зависит от α_i, β_j и

$$n_{i+} = \sum_j n_{ij}, \quad n_{+j} = \sum_i n_{ij}.$$

При ограничениях $\sum \alpha_i = \sum \beta_j = 1$ функция ℓ достигает максимума при

$$\hat{\alpha}_i = \frac{n_{i+}}{n}, \quad \hat{\beta}_j = \frac{n_{+j}}{n}.$$

Таким образом,

$$\tilde{p}_{ij} = \frac{e_{ij}}{n}, \quad \text{где} \quad e_{ij} = \frac{n_{i+}n_{+j}}{n}.$$

Тогда статистика о. п.

$$2 \ln \frac{\max\left[\ell(p_{ij})\colon \sum\limits_{i,j} p_{ij} = 1\right]}{\max[\ell(\alpha_i, \beta_j)\colon \sum \alpha_i = \sum \beta_j = 1]} = 2 \sum_{i,j} n_{ij} \ln(n_{ij}/e_{ij})$$

совпадает с D, что и требовалось показать.

Обозначив $n_{ij} - e_{ij} = \delta_{ij}$, запишем

$$D = 2\sum_{i,j}(e_{ij} + \delta_{ij})\ln\Big(1 + \frac{\delta_{ij}}{e_{ij}}\Big).$$

Опуская индексы, видим, что это выражение приближенно равно

$$2\sum(e+\delta)\Big(\frac{\delta}{e} - \frac{1}{2}\frac{\delta^2}{e^2} + \ldots\Big) = 2\sum\delta + \sum\frac{\delta^2}{e} + \ldots$$

Поскольку $\sum\delta = 0$, мы имеем

$$D \approx \sum\frac{\delta_{ij}^2}{e_{ij}} = Z^2.$$

б) Для указанных данных находим $Z^2 = 9{,}84$. Так как $r = 4$, $c = 2$, используем распределение χ_3^2. Значение 9,84 слишком велико для распределения χ_3^2, следовательно, мы отклоняем гипотезу H_0. Замечаем, что с увеличением уровня физической нагрузки число случаев заболевания туберкулезом уменьшается быстрее, чем другими болезнями. □

Задача 43. Рассмотрена большая группа молодых супружеских пар, и найдено, что стандартное отклонение для возраста мужа относительно среднего возраста в группе мужчин равно четырем годам, а для возраста жены — трем годам. Пусть D обозначает разность возрастов супругов в паре.

При каких условиях можно ожидать, что стандартное отклонение для величины D для рассматриваемой группы супружеских пар будет равно приблизительно 5 годам?

Предположим, что стандартное отклонение для D оказалось равным 2 годам. Одно возможное объяснение такого расхождения — случайная изменчивость. Найдите еще одно объяснение.

Решение. Записав $D = H - W$, видим, что если величины H и W независимы, то

$$\operatorname{Var} D = \operatorname{Var} H + \operatorname{Var} W = 4^2 + 3^2 = 5^2$$

и стандартное отклонение равно 5. В противном случае следует ожидать, что $\operatorname{Var} D \neq 5$, и, таким образом, значение 5 получают при независимости величин H и W.

Альтернативное объяснение возможности получить значение 2 состоит в том, что величины H и W являются коррелированными. Если $H = \alpha W + \varepsilon$, где W и ε независимы, причем $\operatorname{Var} H = 16$ и $\operatorname{Var} W = 9$, то

$$\operatorname{Var} H = 16 = \alpha^2 \operatorname{Var} W + \operatorname{Var}\varepsilon,$$

$$\operatorname{Var}(H - W) = 4 = (\alpha - 1)^2 \operatorname{Var} W + \operatorname{Var}\varepsilon.$$

Следовательно,

$$12 = (\alpha^2 - (\alpha - 1)^2)) \operatorname{Var} W,$$

и $\alpha = 7/6$. □

Задача 44. Предположим, что $X_1, \ldots, X_n$ и $Y_1, \ldots, Y_m$ представляют собой две независимые выборки, первая из них — из показательного распределения с параметром λ, вторая — из показательного распределения с параметром μ.

а) Постройте критерий отношения правдоподобия для проверки гипотезы $H_0 \colon \lambda = \mu$ против альтернативы $H_1 \colon \lambda \neq \mu$.

б) Покажите, что критерий для п. а) может основываться на статистике

$$T = \frac{\sum\limits_{i=1}^{n} X_i}{\sum\limits_{i=1}^{n} X_i + \sum\limits_{j=1}^{m} Y_j}.$$

в) Опишите, как квантили распределения статистики T при гипотезе H_0 можно определить при помощи квантилей F-распределения.

Решение. а) Поскольку $X_i \sim \operatorname{Exp}(\lambda)$, $Y_j \sim \operatorname{Exp}(\mu)$, функция $f(\boldsymbol{x}; \lambda) = \prod\limits_{i=1}^{n} (\lambda e^{-\lambda x_i}) I(\min x_i \geqslant 0)$ достигает максимума при $\hat{\lambda}^{-1} = \frac{1}{n} \sum\limits_{i=1}^{n} x_i$, при этом

$$\max[f(\boldsymbol{x}; \lambda) \colon \lambda > 0] = \hat{\lambda}^n e^{-n}.$$

Аналогично функция $f(\boldsymbol{y}; \mu) = \prod\limits_{j=1}^{m} (\mu e^{-\mu y_j}) I(\min y_j \geqslant 0)$ достигает максимума при $\hat{\mu}^{-1} = \frac{1}{m} \sum\limits_{j=1}^{m} y_j$, при этом

$$\max[f(\boldsymbol{y}; \mu) \colon \mu > 0] = \hat{\mu}^m e^{-m}.$$

При выполнении гипотезы H_0 функция правдоподобия равна

$$f(\boldsymbol{x}; \theta) f(\boldsymbol{y}; \theta);$$

она достигает максимума при $\hat{\theta}^{-1} = \frac{1}{n+m} \left(\sum\limits_{i=1}^{n} x_i + \sum\limits_{j=1}^{m} y_j \right)$, и

$$\max[f(\boldsymbol{x}; \theta) f(\boldsymbol{y}; \theta) \colon \theta > 0] = \hat{\theta}^{n+m} e^{-(n+m)}.$$

Тогда согласно критерию гипотеза H_0 отвергается, если отношение

$$\frac{\hat{\lambda}^n \hat{\mu}^m}{\hat{\theta}^{n+m}} = \left(\frac{\sum\limits_i x_i + \sum\limits_j y_j}{n+m} \right)^{n+m} \left(\frac{n}{\sum\limits_i x_i} \right)^n \left(\frac{m}{\sum\limits_j y_j} \right)^m$$

принимает большие значения.

б) Логарифм имеет вид

$$(n+m)\ln\Big(\sum_i x_i + \sum_j y_j\Big) - n\ln\sum_i x_i - m\ln\sum_j y_j$$

плюс члены, не зависящие от $\boldsymbol{x}$, $\boldsymbol{y}$. Основной частью является

$$-n\ln\frac{\sum\limits_i x_i}{\sum\limits_i x_i + \sum\limits_j y_j} - m\ln\frac{\sum\limits_j y_j}{\sum\limits_i x_i + \sum\limits_j y_j} = -n\ln T - m\ln(1-T).$$

Таким образом, критерий и в самом деле основан на статистике T. При этом гипотеза H_0 отвергается, если T близко к 0 или 1.

в) При гипотезе H_0 мы имеем $\lambda = \mu = \theta$ и

$$2\theta\sum_i X_i \sim \chi^2_{2n}, \quad 2\theta\sum_j Y_j \sim \chi^2_{2m}.$$

Следовательно,

$$T^{-1} = 1 + R\frac{m}{n}, \quad R = \frac{\sum\limits_j Y_j/(2m)}{\sum\limits_i X_i/(2n)} \sim \mathrm{F}_{2m,2n}.$$

Таким образом, симметричная критическая область равна объединению

$$C = \Big(0, \Big(1 + \frac{m}{n}\varphi^+_{2m,2n}(\alpha/2)\Big)^{-1}\Big) \cup \Big(\Big(1 + \frac{m}{n}\varphi^-_{2m,2n}(\alpha/2)\Big)^{-1}, 1\Big)$$

и определяется квантилями распределения $\mathrm{F}_{2m,2n}$. □

Задача 45. Объясните, что понимают под достаточной статистикой.

Рассмотрим независимые случайные величины $X_1, X_2, \ldots, X_n$, где $X_i \sim \mathrm{N}(\alpha + \beta c_i, \theta)$ с заданными константами c_i, $i = 1, 2, \ldots, n$, и неизвестными параметрами α, β и θ. Найдите три выборочных характеристики, которые, взятые в совокупности, образуют достаточную статистику.

Решение. Статистика $T = T(\boldsymbol{x})$ является достаточной статистикой для параметра θ, если $f(\boldsymbol{x}; \theta) = g(T, \theta)h(\boldsymbol{x})$.

Для выборочного вектора $\boldsymbol{x}$ с компонентами $x_1, \ldots, x_n$ находим

$$f(\boldsymbol{x}; \theta) = \prod_{i=1}^n f_{i,\theta}(x_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\theta}}\exp\Big[-\frac{1}{2\theta}(x_i - \alpha - \beta c_i)^2\Big] =$$

$$= \frac{1}{(2\pi\theta)^{n/2}}\exp\Big(-\frac{1}{2\theta}\Big[\sum_i x_i^2 - 2\sum_i x_i(\alpha + \beta c_i) + \sum_i(\alpha + \beta c_i)^2\Big]\Big) =$$

$$= \frac{1}{(2\pi\theta)^{n/2}}\exp\Big[-\frac{1}{2\theta}\sum_i x_i^2 + \frac{\alpha}{\theta}\sum_i x_i + \frac{\beta}{\theta}\sum_i x_i c_i - \frac{1}{2\theta}\sum_i(\alpha + \beta c_i)^2\Big].$$

Таким образом, вектор

$$T(\boldsymbol{x}) = \left(\sum_i x_i^2, \sum_i x_i, \sum_i c_i x_i\right)$$

является достаточной статистикой. □

Задача 46. Пусть $X_1, X_2, \ldots, X_n$ — случайная выборка из $\mathrm{N}(\theta, \sigma^2)$-распределения, и предположим, что априорное распределение для θ — это нормальное $\mathrm{N}(\mu, \tau^2)$-распределение, где σ^2, μ и τ^2 известны. Найдите апостериорное распределение для θ при заданных $X_1, X_2, \ldots, X_n$ и оптимальную оценку параметра θ при а) функции квадратичных потерь и б) функции потерь, равной абсолютной ошибке.

Решение. По поводу первой половины задачи см. задачи 2.3.8 и 2.3.9 части А. Решение для второй части задачи: как было показано в задаче 2.3.8 части А, апостериорная плотность распределения равна

$$\pi(\theta \mid \boldsymbol{x}) = \frac{\pi(\theta) f(\boldsymbol{x}; \theta)}{\int \pi(\theta') f(\boldsymbol{x}; \theta')\, d\theta'} = \frac{1}{\sqrt{2\pi}\tau_n} \exp\left[-\frac{(\theta - \mu_n)^2}{2\tau_n^2}\right],$$

где

$$\frac{1}{\tau_n^2} = \frac{1}{\tau^2} + \frac{n}{\sigma^2}, \qquad \mu_n = \frac{\mu/\tau^2 + n\bar{x}/\sigma^2}{1/\tau^2 + n/\sigma^2}.$$

Поскольку нормальное распределение является симметричным относительно своего среднего значения, наилучшей оценкой при каждой из функций потерь а) и б) будет $\mathsf{E}(\theta \mid \boldsymbol{x}) = \mu_n$. □

Задача 47. Пусть $X_1, X_2, \ldots, X_n$ — случайная выборка из равномерного распределения на интервале $(-\theta, 2\theta)$, где значение положительного параметра θ неизвестно. Найдите оценку максимального правдоподобия для параметра θ.

Решение. Функция правдоподобия $f(\boldsymbol{x}; \theta)$ равна

$$\frac{1}{(3\theta)^n} I(-\theta < x_1, \ldots, x_n < 2\theta) = \frac{1}{(3\theta)^n} I(-\theta < \min x_i) I(\max x_i < 2\theta).$$

Следовательно, о. м. п. имеет вид

$$\hat{\theta} = \max\left[-\min x_i, \frac{1}{2}\max x_i\right]. \qquad \square$$

Задача 48. а) Статистику χ^2 часто используют как меру расхождения между наблюдаемыми (эмпирическими) частотами и ожидаемыми (теоретическими) частотами при выполнении нулевой гипотезы. Опишите статистику χ^2 и критерий согласия χ^2.

Число звонков, поступающих ежедневно в справочное бюро, подсчитывают в течение K недель. Можно предположить, что число звонков в любой день имеет распределение Пуассона, что число звонков в один день не зависит от числа звонков в другой день и что ожидаемое число звонков зависит только от дня недели. Пусть n_i, $i = 1, 2, \ldots, 7$, — это общее число звонков, поступивших соответственно в понедельник, вторник, ..., воскресенье.

Постройте приближенный критерий проверки гипотезы о том, что звонки поступают с одинаковой интенсивностью во все дни недели, за исключением воскресенья.

б) Найдите также критерий для проверки второй гипотезы о том, что ожидаемое число принятых звонков одинаково для трех дней недели со вторника по четверг и что ожидаемое число принятых звонков одинаково для понедельника и пятницы.

Решение. а) Пусть мы наблюдаем n_i — число появлений состояния $i = 1, \ldots, n$ и e_i — ожидаемые (теоретические) частоты этих состояний. Статистика χ^2 задается формулой

$$P = \sum_{i=1}^{n} \frac{(n_i - e_i)^2}{e_i}.$$

Критерий согласия χ^2 применяют для проверки гипотезы H_0: $p_i = p_i(\theta)$, $\theta \in \Theta$, против альтернативы H_1: p_i произвольные, где p_i — вероятность появления состояния i.

В данном примере предполагаем, что доля звонков, поступивших за все дни недели, кроме воскресенья, является фиксированным числом (и вычисляется на основании данных). Такое предположение является вполне естественным при наличии больших объемов данных. Однако доля звонков, поступивших в заданный день с понедельника по пятницу, является разной для разных дней, и мы поступим следующим образом. Пусть $e^* = \frac{1}{6}(n_1 + \ldots + n_6)$, $e_1^* = \frac{1}{3}(n_2 + n_3 + n_4)$, $e_2^* = \frac{1}{2}(n_1 + n_5)$.

При гипотезе H_0: *с понедельника по субботу звонки поступают с одинаковой интенсивностью* выбирается статистика

$$\sum_{i=1}^{6} \frac{(n_i - e^*)^2}{e^*},$$

которая асимптотически имеет χ_5^2-распределение. (Здесь число степеней свободы равно пяти, поскольку подгоняется один параметр.)

б) Во втором варианте гипотеза H_0 состоит в том, что *во вторник, среду и четверг звонки поступают с одной интенсивностью, а в по-*

недельник и пятницу — с другой интенсивностью. В этом случае рассматривается статистика

$$\sum_{i=2}^{4} \frac{(n_i - e_1^*)^2}{e_1^*} + \sum_{j=1,5} \frac{(n_j - e_2^*)^2}{e_2^*},$$

которая асимптотически имеет χ_3^2-распределение. □

Задача 49. а) Произведены воздушные наблюдения x_1, x_2, x_3, x_4 значений внутренних углов θ_1, θ_2, θ_3, θ_4 некоторого четырехугольника, расположенного на земле. Предполагается, что эти измерения выполнены с незначительными независимыми ошибками, имеющими нулевые средние и общую дисперсию σ^2. Найдите МНК-оценки для θ_1, θ_2, θ_3, θ_4.

б) Найдите несмещенную оценку σ^2 в ситуации, описанной в части а).

в) Предположим теперь, что известно, что рассматриваемый четырехугольник — это параллелограмм, у которого $\theta_1 = \theta_3$ и $\theta_2 = \theta_4$. Какими теперь будут МНК-оценки величин углов? Найдите несмещенную оценку σ^2 в этом случае.

Решение. а) МНК-оценки должны минимизировать сумму $\sum\limits_{i=1}^{4}(\theta_i - x_i)^2$ при условии, что $\sum\limits_{i=1}^{4}\theta_i = 2\pi$. Для функции Лагранжа

$$L = \sum_i (\theta_i - x_i)^2 - \lambda\Big(\sum_i \theta_i - 2\pi\Big)$$

находим

$$\frac{\partial}{\partial\theta_i} L = 0 \quad \text{при} \quad 2(\theta_i - x_i) - \lambda = 0, \quad \text{т. е.} \quad \hat\theta_i = x_i + \frac{\lambda}{2}.$$

Воспользовавшись тем, что $\sum \hat\theta_i = 2\pi$, получаем $\lambda = \frac{1}{2}\Big(2\pi - \sum\limits_i x_i\Big)$ и

$$\hat\theta_i = x_i + \frac{1}{4}\Big(2\pi - \sum_i x_i\Big).$$

б) Применим метод наименьших квадратов. Случайные величины X_i, $i = 1, 2, 3, 4$, имеют средние значения θ_i, дисперсию σ^2, и они независимы. Запишем

$$\mathsf{E}(X_i - \hat\theta_i)^2 = \mathsf{E}\left[\frac{1}{4}(2\pi - \sum_i X_i)\right]^2 = \frac{1}{16}\mathsf{E}\left[\left(\Big(\sum_i X_i - \mathsf{E}\Big(\sum_i X_i\Big)\Big)\right)^2\right] =$$

$$= \frac{1}{16}\operatorname{Var}\Big(\sum_i X_i\Big) = \frac{1}{16}\cdot 4\sigma^2 = \frac{\sigma^2}{4}.$$

Следовательно,

$$\mathsf{E}\left(\sum_i (X_i - \hat{\theta}_i)^2\right) = \sigma^2,$$

и $\sum_i (x_i - \hat{\theta}_i)^2$ является несмещенной оценкой для σ^2.

в) Если $\theta_1 = \theta_3$ и $\theta_2 = \theta_4$, то ограничение принимает вид $2(\theta_1 + \theta_2) = 2\pi$, т. е. $\theta_1 + \theta_2 = \pi$. Тогда функция Лагранжа

$$L = (\theta_1 - x_1)^2 + (\theta_2 - x_2)^2 + (\theta_1 - x_3)^2 + (\theta_2 - x_4)^2 - 2\lambda(\theta_1 + \theta_2 - \pi)$$

имеет производные

$$\frac{\partial}{\partial\theta_1} L = 2(\theta_1 - x_1) + 2(\theta_1 - x_3) - 2\lambda,$$
$$\frac{\partial}{\partial\theta_2} L = 2(\theta_2 - x_2) + 2(\theta_2 - x_4) - 2\lambda$$

и достигает минимума при

$$\hat{\theta}_1 = \frac{1}{2}(x_1 + x_3 + \lambda), \qquad \hat{\theta}_2 = \frac{1}{2}(x_2 + x_4 + \lambda).$$

Ограничение $\hat{\theta}_1 + \hat{\theta}_2 = \pi$ затем приводит к равенству $\lambda = \left(\pi - \sum_i x_i/2\right)$, и

$$\hat{\theta}_1 = \frac{1}{2}(x_1 + x_3) + \frac{1}{4}\left(2\pi - \sum_i x_i\right) = \frac{1}{4}(x_1 + x_3 - x_2 - x_4) + \frac{\pi}{2}.$$

Аналогично

$$\hat{\theta}_2 = \frac{1}{4}(x_2 + x_4 - x_1 - x_3) + \frac{\pi}{2}.$$

Далее,

$$\mathsf{E}(X_1 - \hat{\theta}_1)^2 = \mathsf{E}\left(\frac{3X_1}{4} - \frac{X_3}{4} + \frac{X_2}{4} + \frac{X_4}{4} - \frac{\pi}{2}\right)^2 = \operatorname{Var}\left(\frac{3X_1}{4} - \frac{X_3}{4} + \frac{X_2}{4} + \frac{X_4}{4}\right) =$$
$$= \left[\left(\frac{3}{4}\right)^2 + \left(\frac{1}{4}\right)^2 + \left(\frac{1}{4}\right)^2 + \left(\frac{1}{4}\right)^2\right]\sigma^2 = \frac{3}{4}\sigma^2.$$

Эти же рассуждения справедливы для $i = 2, 3, 4$. Следовательно, $\sum (x_i - \hat{\theta}_i)^2/3$ — несмещенная оценка для σ^2. □

Задача 50. а) Пусть $X_1, X_2, \ldots, X_n$ — случайная выборка из распределения с плотностью

$$f(x; \theta) = \begin{cases} 2x/\theta^2, & 0 \leqslant x \leqslant \theta, \\ 0 & \text{в противном случае,} \end{cases}$$

где значение положительного параметра θ неизвестно. Найдите оценку максимального правдоподобия медианы этого распределения.

б) Среди руководства компании «Надежные двигатели» общепринятой является точка зрения, что число X дефектных автомобилей, производимых ежемесячно, имеет биномиальное распределение

$$\mathsf{P}(X=s)=C_n^s p^s(1-p)^{n-s} \quad (s=0,\ 1,\ \ldots,\ n;\ 0\leqslant p\leqslant 1).$$

Существуют, однако, некоторые разногласия относительно параметра p. Генеральный директор полагает, что априорное распределение параметра p является равномерным (т. е. имеет п. р. в. $f_p(x)=I(0\leqslant x\leqslant 1)$). Более пессимистично настроенный менеджер производства полагается на априорное распределение с плотностью $f_p(x)=2xI(0\leqslant x\leqslant 1)$. Обе п. р. в. сосредоточены на $(0,\ 1)$.

За один месяц произведено s дефектных автомобилей. Пусть функция потерь генерального директора имеет вид $(\hat{p}-p)^2$, где $\hat{p}$ — это оценка генерального директора, а p — истинное значение. Покажите, что тогда наилучшей оценкой для p, полученной генеральным директором, будет оценка

$$\hat{p}=\frac{s+1}{n+2}.$$

Менеджер производства несет несколько иную ответственность, чем генеральный директор, и его функция потерь имеет вид $(1-p)(\hat{p}-p)^2$. Найдите наилучшую оценку для p с точки зрения менеджера производства и покажите, что эта оценка превосходит ту, которая получена генеральным директором, если только для s не будет выполняться неравенство $s\geqslant n/2$.

Указание. Для неотрицательных целых α, β выполняется равенство

$$\int\limits_0^1 p^\alpha(1-p)^\beta\,dp=\frac{\alpha!\,\beta!}{(\alpha+\beta+1)!}.$$

Решение. а) Если m — медиана, то из уравнения

$$\int\limits_0^m \frac{2x}{\theta^2}\,dx=\left.\frac{x^2}{\theta^2}\right|_0^m=\frac{1}{2}$$

находим, что $m=\theta/\sqrt{2}$. Тогда функция

$$f(x;\ m\sqrt{2})=\frac{2x}{(\sqrt{2}m)^2}=\frac{x}{m^2},\qquad 0\leqslant x\leqslant m\sqrt{2},$$

достигает максимума при m, равном $x/\sqrt{2}$, и произведение $f(\boldsymbol{x};\ m\sqrt{2})=\prod f(x_i;\ m\sqrt{2})$ достигает максимума при

$$\widehat{m}=\frac{1}{\sqrt{2}}\max x_i.$$

б) Поскольку $\mathsf{P}_p(X = s) \propto p^s(1-p)^{n-s}$, $s = 0, 1, \ldots n$, апостериорная плотность для оценки генерального директора (ГД) равна

$$\pi^{\text{ГД}}(p \mid s) \propto p^s(1-p)^{n-s} I(0 < p < 1),$$

а апостериорная плотность для оценки менеджера производства (МП) равна

$$\pi^{\text{МП}}(p \mid s) \propto pp^s(1-p)^{n-s} I(0 < p < 1).$$

Тогда средние потери для ГД минимизируются в точке, равной апостериорному среднему:

$$\hat{p}^{\text{ГД}} = \int_0^1 pp^s(1-p)^{n-s}\,dp \Big/ \int_0^1 p^s(1-p)^{n-s}\,dp = \\ = \frac{(s+1)!\,(n-s)!}{(n-s+s+2)!}\frac{(n-s+s+1)!}{s!\,(n-s)!} = \frac{s+1}{n+2}.$$

Для МП средние потери

$$\int_0^1 (1-p)(p-a)^2 \pi^{\text{МП}}(p \mid s)\,dp$$

минимизируются при

$$a = \int_0^1 p(1-p)\pi^{\text{МП}}(p \mid s)\,dp \Big/ \int_0^1 (1-p)\pi^{\text{МП}}(p \mid s)\,dp,$$

откуда находим, что

$$\hat{p}^{\text{МП}} = \int_0^1 p(1-p)pp^s(1-p)^{n-s}\,dp \Big/ \int_0^1 p(1-p)pp^s(1-p)^{n-s}\,dp = \\ = \frac{(s+2)!\,(n-s+1)!}{(n-s+s+4)!}\frac{(n-s+s+3)!}{(s+1)!\,(n-s+1)!} = \frac{s+2}{n+4}.$$

Видим, что $(s+2)/(n+4) > (s+1)/(n+2)$ тогда и только тогда, когда $s < n/2$. □

Задача 51. а) Что такое простая гипотеза? Определите понятия размера и мощности критерия для проверки простой гипотезы против простой альтернативы.

Сформулируйте и докажите лемму Неймана—Пирсона.

б) Производится одно наблюдение над случайной величиной X, имеющей п. р. в. $f(x)$. Постройте наилучший критерий размера 0,05 для проверки нулевой гипотезы

$$H_0\colon f(x) = \frac{1}{2} \quad (-1 \leqslant x \leqslant 1)$$

против альтернативы

$$H_1\colon f(x)=\frac{3}{4}(1-x^2)\quad (-1\leqslant x\leqslant 1).$$

Вычислите мощность этого критерия.

Решение. а) Простая гипотеза относительно параметра θ — это гипотеза вида $H_0\colon \theta=\theta_0$. Размер критерия равен вероятности отклонить гипотезу H_0, когда она верна. Мощность равна вероятности отклонить гипотезу H_0, когда она неверна; для простой альтернативы $H_1:\theta=\theta_1$ мощность — это число (в общем случае это функция параметра, пробегающего параметрическое множество, соответствующее гипотезе H_1).

Утверждение леммы Н—П для проверки гипотезы $H_0\colon f=f_0$ против альтернативы $H_1: f=f_1$ заключается в следующем.

Среди всех критериев размера не больше α критерий с наибольшей мощностью задается критической областью $C=\{x\colon f_1(x)>kf_0(x)\}$, где k таково, что $\mathrm{P}(X\in C\mid H_0)=\int\limits_C f_0(x)\,dx=\alpha$.

Замечание. Для любого $k>0$ критерий, который состоит в том, чтобы отвергнуть гипотезу H_0, когда $f_1(x)>kf_0(x)$, имеет наибольшую мощность среди критериев с размером не больше $\alpha:=\mathrm{P}_0(f_1(X)>kf_0(X))$.

Доказательство леммы Н—П дано в § 2.2.

б) Составим отношение правдоподобия $f_1(x)/f_0(x)=3(1-x^2)/2$; мы отклоняем гипотезу H_0, если о. п. не меньше k, т. е. $|x|\leqslant(1-2k/3)^{1/2}$. Необходимо, чтобы выполнялось равенство

$$0{,}05=\mathrm{P}\left(|X|\leqslant\left(1-\frac{2k}{3}\right)^{1/2}\ \Big|\ H_0\right)=\left(1-\frac{2k}{3}\right)^{1/2}.$$

Это означает, что условие $|x|\leqslant(1-2k/3)^{1/2}$ равносильно условию $|x|\leqslant 0{,}05$. В силу леммы Н—П критерий, при котором гипотеза H_0 отклоняется, когда $|x|\leqslant 0{,}05$, является наиболее мощным критерием размера 0,05.

При этом мощность критерия равна

$$\mathrm{P}(|X|\leqslant 0{,}05\mid H_1)=\frac{3}{4}\int\limits_{-0{,}05}^{0{,}05}(1-x^2)\,dx=\frac{3}{2}\left(x-\frac{x^3}{3}\right)\Big|_0^{0{,}05}\approx 0{,}075.\qquad\square$$

Задача 52. а) Пусть $X_1,\ldots,X_m$ и $Y_1,\ldots,Y_n$ — независимые случайные выборки из распределений $\mathrm{N}(\mu_1,\sigma^2)$ и $\mathrm{N}(\mu_2,\sigma^2)$ соответственно. Здесь все параметры μ_1, μ_2 и σ^2 неизвестны. Опишите, как бы вы проверяли гипотезу $H_0\colon \mu_1=\mu_2$ против альтернативы $H_1:\mu_1\neq\mu_2$.

б) Пусть $X_1,\ldots,X_n$ — случайная выборка из распределения с п. р. в.

$$f(x;\theta)=e^{-(x-\theta)},\quad \theta<x<\infty,$$

где θ имеет в качестве априорного распределения стандартное нормальное распределение $\mathrm{N}(0, 1)$. Найдите апостериорное распределение параметра θ.

Предположим, что необходимо оценить θ при функции потерь, равной абсолютной ошибке, т. е. при $L(a, \theta) = |a - \theta|$. Найдите оптимальную байесовскую оценку и выразите ее в терминах функции $c_n(x)$, которая задается равенством

$$2\Phi(c_n(x) - n) = \Phi(x - n), \quad -\infty < x < \infty,$$

где

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-y^2/2}\, dy$$

— стандартная нормальная функция распределения.

Решение. а) Поскольку

$$\bar{X} = \frac{1}{m}\sum_i X_i \sim \mathrm{N}(\mu_1, \sigma^2/m) \quad \text{и} \quad \bar{Y} = \frac{1}{n}\sum_i Y_j \sim \mathrm{N}(\mu_2, \sigma^2/n),$$

при выполнении гипотезы H_0 мы имеем

$$\frac{1}{\sigma}(\bar{X} - \bar{Y})\left(\frac{1}{m} + \frac{1}{n}\right)^{-1/2} \sim \mathrm{N}(0, 1).$$

Положим

$$S_{XX} = \sum_{i=1}^{m}(X_i - \bar{X})^2 \sim \sigma^2\chi^2_{m-1}, \quad S_{YY} = \sum_{j=1}^{n}(Y_j - \bar{Y})^2 \sim \sigma^2\chi^2_{n-1}.$$

Тогда

$$\frac{1}{\sigma^2}(S_{XX} + S_{YY}) \sim \chi^2_{m+n-2}$$

и

$$t = (\bar{X} - \bar{Y})\left(\frac{1}{m} + \frac{1}{n}\right)^{-1/2} \Big/ \left(\frac{S_{XX} + S_{YY}}{m+n-2}\right)^{1/2} \sim \mathrm{t}_{m+n-2}.$$

При н. м. критерии размера α гипотеза H_0 отвергается, когда $|t|$ превышает $\mathrm{t}_{m+n-2}(\alpha/2)$, верхнюю $\alpha/2$ точку t_{m+n-2}-распределения.

б) По формуле Байеса находим, что

$$\pi(\theta \,|\, \boldsymbol{x}) \propto \pi(\theta) f(\boldsymbol{x}; \theta) \propto e^{-\theta^2/2 + n\theta - \sum\limits_i x_i} I(\theta < \min x_i),$$

где коэффициент пропорциональности имеет вид

$$\left(\int e^{-\theta^2/2 + n\theta - \sum\limits_i x_i} I(\theta < \min x_i)\, d\theta\right)^{-1} = \frac{1}{\sqrt{2\pi}} \frac{\exp\left(-n^2/2 + \sum\limits_i x_i\right)}{\Phi(\min x_i - n)}.$$

При функции потерь, равной абсолютной ошибке $L(a, \theta) = |a - \theta|$, оптимальная байесовская оценка равна апостериорной медиане, т. е. это то значение s, при котором

$$\int_{-\infty}^{s} e^{-(\theta-n)^2/2}\, d\theta = \frac{1}{2} \int_{-\infty}^{\min x_i} e^{-(\theta-n)^2/2}\, d\theta,$$

или, что эквивалентно, $2\Phi(s - n) = \Phi(\min x_i - n)$ и $s = c_n(\min x_i)$, что и требовалось показать. □

Задача 53. а) Пусть $X_1, \ldots, X_n$ — случайная выборка из распределения с п. р. в.

$$f(x;\, \theta) = \frac{2x}{\theta^2}, \quad 0 \leqslant x \leqslant \theta.$$

Найдите оценку максимального правдоподобия M параметра θ и покажите, что $\left(M,\, M/(1-\gamma)^{1/2n}\right)$ является $100\gamma\%$-доверительным интервалом для θ, где $0 < \gamma < 1$.

б) Пусть $X_1, \ldots, X_n$ — независимая случайная выборка из равномерного распределения на отрезке $[0, \theta_1]$, и пусть $Y_1, \ldots, Y_n$ — независимая случайная выборка из равномерного распределения на отрезке $[0, \theta_2]$. Выведите критерий отношения правдоподобия для проверки гипотезы $H_0\colon \theta_1 = \theta_2$ против альтернативы $H_1\colon \theta_1 \neq \theta_2$ и выразите этот критерий в терминах статистики

$$T = \frac{\max\left(M_X,\, M_Y\right)}{\min\left(M_X,\, M_Y\right)},$$

где $M_X = \max\limits_{1 \leqslant i \leqslant n} X_i$ и $M_Y = \max\limits_{1 \leqslant i \leqslant n} Y_i$.

Заметив, что при гипотезе H_0 распределение статистики T не зависит от $\theta = \theta_1 = \theta_2$, воспользуйтесь этим для того, чтобы точно указать критическую область для критерий размера α (или найдите эту критическую область иным способом).

Решение. а) Функция правдоподобия

$$f(\boldsymbol{x};\, \theta) = \frac{2^n}{\theta^{2n}} \left(\prod_{i=1}^{n} x_i\right) I(\max x_i \leqslant \theta)$$

имеет вид $g(M(\boldsymbol{x}, \theta))h(\boldsymbol{x})$, где $M(\boldsymbol{x}) = \max x_i$. Следовательно, $M = M(\boldsymbol{X}) = \max X_i$ является достаточной статистикой. Эта величина является также и о. м. п., при этом

$$\mathbf{P}(M \leqslant u) = (\mathbf{P}(X_1 \leqslant u))^n = \left(\frac{u}{\theta}\right)^{2n}, \quad 0 \leqslant u \leqslant \theta.$$

Тогда

$$\mathsf{P}\left(M \leqslant \theta \leqslant \frac{M}{(1-\gamma)^{1/(2n)}}\right) = \mathsf{P}(\theta(1-\gamma)^{1/(2n)} \leqslant M \leqslant \theta) =$$
$$= 1 - \mathsf{P}(M < \theta(1-\gamma)^{1/(2n)}) = 1 - \frac{(\theta(1-\gamma)^{1/(2n)})^{2n}}{\theta^{2n}} = 1 - (1-\gamma) = \gamma.$$

Следовательно, $(M, M/(1-\gamma)^{1/(2n)})$ является $100\gamma\%$-д. и. для θ.

б) При выполнении гипотезы H_0 функция

$$f_{X,Y} = \left(\frac{1}{\theta}\right)^{2n} I(0 \leqslant x_1, \ldots, x_n, y_1, \ldots, y_n \leqslant \theta),$$

достигает максимума в о. м. п. $\hat{\theta} = \max(M_X, M_Y)$. При выполнении гипотезы H_1 функция

$$f_{X,Y} = \left(\frac{1}{\theta_1}\right)^n \left(\frac{1}{\theta_2}\right)^n I(0 \leqslant x_1, \ldots, x_n \leqslant \theta_1) I(0 \leqslant x_1, \ldots, x_n \leqslant \theta_2)$$

достигает максимума в о. м. п. $(\hat{\theta}_1, \hat{\theta}_2) = (M_X, M_Y)$.

Далее, отношение $\Lambda_{H_1;H_0}(\boldsymbol{x}, \boldsymbol{y})$ равно

$$\frac{\left(\frac{1}{M_X}\right)^n \left(\frac{1}{M_Y}\right)^n}{\left[\frac{1}{\max(M_X, M_Y)}\right]^{2n}} = \frac{[\max(M_X, M_Y)]^{2n}}{(M_X M_Y)^n} = \left[\frac{\max(M_X, M_Y)}{\min(M_X, M_Y)}\right]^n = T^n.$$

Следовательно, мы отклоняем гипотезу H_0, если $T(\boldsymbol{x}, \boldsymbol{y}) \geqslant k$ при некотором $k \geqslant 1$.

Далее, при выполнении гипотезы H_0 мы имеем

$$\mathsf{P}(M_X < x) = \left(\frac{x}{\theta}\right)^n, \quad \text{т. е.} \quad f_{M_X}(x) = n\frac{1}{\theta}\left(\frac{x}{\theta}\right)^{n-1},$$

и аналогично для M_Y. Тогда для $0 < \alpha < 1$ и $k \geqslant 1$ мы требуем, чтобы α было равно следующей вероятности:

$$\mathsf{P}(T \geqslant k \mid H_0) = \frac{1}{\theta^2} \int_0^\theta \int_0^\theta n^2 \left(\frac{x}{\theta}\right)^{n-1} \left(\frac{y}{\theta}\right)^{n-1} I\left(\frac{\max[x, y]}{\min[x, y]} \geqslant k\right) dy\, dx =$$
$$= 2n^2 \int_0^1 \int_0^{x/k} x^{n-1} y^{n-1}\, dy\, dx = 2n \int_0^1 x^{n-1} \frac{x^n}{k^n}\, dx = \frac{1}{k^n}.$$

Таким образом, $k = \alpha^{-1/n}$, и критическая область для критерия размера α имеет вид

$$\mathbb{C} = \{\boldsymbol{x}, \boldsymbol{y} \colon T > \alpha^{-1/n}\}. \qquad \square$$

Задача 54. а) Сформулируйте и докажите лемму Неймана—Пирсона.

б) Пусть $X_1, \ldots, X_n$ — случайная выборка из $\mathrm{N}(\mu, \sigma^2)$-распределения. Докажите, что случайные величины $\overline{X}$ (выборочное среднее) и $\sum\limits_{i=1}^{n}(X_i - \overline{X})^2$ ($(n-1) \times$ (выборочная дисперсия)) являются независимыми, и найдите их распределения.

Предположим, что

$$\begin{array}{lll} X_{11}, & \ldots & X_{1n}, \\ X_{21}, & \ldots & X_{2n}, \\ \ldots & \ldots & \ldots \\ X_{m1}, & \ldots & X_{mn} \end{array}$$

являются независимыми случайными величинами и X_{ij} имеет $\mathrm{N}(\mu_i, \sigma^2)$-распределение для $1 \leqslant j \leqslant n$, где $\mu_1, \ldots, \mu_m, \sigma^2$ — неизвестные константы. Сославшись на предыдущий результат, объясните, как бы вы проверили гипотезу $H_0\colon \mu_1 = \ldots = \mu_m$.

Решение. (Только для части б).) Мы утверждаем, что

1) $\overline{X} = \sum\limits_i X_i/n \sim \mathrm{N}(\mu, \sigma^2/n)$;

2) величины $\overline{X}$ и $S_{XX} = \sum\limits_i (X_i - \overline{X})^2$ независимы;

3) $S_{XX} = \sum\limits_i (X_i - \overline{X})^2/\sigma^2 \sim \chi^2_{n-1}$.

Для доказательства утверждения 1 заметим, что линейная комбинация нормальных с. в. является нормальной с. в., и в силу независимости этих величин получаем, что $\overline{X} \sim \mathrm{N}(\mu, \sigma^2/n)$. Кроме того, $(\overline{X} - \mu)^2/\sigma^2 \sim \chi^2_1$.

Для доказательства утверждений 2 и 3 заметим, что

$$\sum_i (X_i - \mu)^2 = \sum_i [(X_i - \overline{X}) + (\overline{X} - \mu)]^2 =$$
$$= \sum_i [(X_i - \overline{X})^2 + 2(X_i - \overline{X})(\overline{X} - \mu) + (\overline{X} - \mu)^2] = S_{XX} + n(\overline{X} - \mu)^2.$$

Пусть A — ортогональная матрица размера $n \times n$. Положим

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}, \ \boldsymbol{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \quad \text{причем } \boldsymbol{Y} = A^T(\boldsymbol{X} - \mu\mathbf{1}).$$

Мы хотим выбрать A так, чтобы первая компонента имела вид $Y_1 = \sqrt{n}(\overline{X} - \mu)$. Тогда матрица A должна иметь вид

$$A = \begin{pmatrix} 1/\sqrt{n} & \ldots & \ldots & \ldots \\ 1/\sqrt{n} & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ 1/\sqrt{n} & \ldots & \ldots & \ldots \end{pmatrix},$$

где остальные столбцы выбираются так, чтобы образовать ортогональную матрицу.

Теперь $Y_1 = \sqrt{n}(\bar{X} - \mu) \sim \mathrm{N}(0, \sigma^2)$ и Y_1 не зависит от $Y_2, \ldots, Y_n$. Поскольку $\sum\limits_i Y_i^2 = \sum\limits_i (X_i - \mu)^2$, мы получаем

$$\sum_{i=2}^{n} Y_i^2 = \sum_{i=1}^{n} (X_i - \mu)^2 - n(\bar{X} - \mu)^2 = S_{XX}.$$

Следовательно, $S_{XX} = \sum\limits_{i=2}^{n} Y_i^2$, где $Y_2, \ldots, Y_n$ — н. о. р. $\mathrm{N}(0, \sigma^2)$-с. в. Поэтому $S_{XX}/\sigma^2 \sim \chi^2_{n-1}$.

Рассмотрим с. в. X_{ij}. Для проверки гипотезы H_0: $\mu_1 = \ldots = \mu_m = \mu$ против H_1: «гипотеза H_0 неверна» используем анализ дисперсий (ANOVA). Запишем $N = mn$ и $X_{ij} = \mu_i + \varepsilon_{ij}$, $j = 1, \ldots, n$, $i = 1, \ldots, m$, где ε_{ij} — н. о. р. с. в. $\mathrm{N}(0, \sigma^2)$.

Применим к. о. о. п.: само о. п. $\Lambda_{H_1:H_0}((x_{ij}))$ равно

$$\frac{\max\limits_{\mu_1, \ldots, \mu_m, \sigma^2} (2\pi\sigma^2)^{-N/2} \exp\left[-\sum\limits_{i,j} (x_{ij} - \mu_i)^2/(2\sigma^2)\right]}{\max\limits_{\mu, \sigma^2} (2\pi\sigma^2)^{-N/2} \exp\left[-\sum\limits_{i,j} (x_{ij} - \mu)^2/(2\sigma^2)\right]} = \left(\frac{S_0}{S_1}\right)^{N/2}.$$

Здесь

$$S_0 = \sum_{i,j} (x_{ij} - \bar{x}_{++})^2, \quad S_1 = \sum_{i,j} (x_{ij} - \bar{x}_{i+})^2$$

и $\bar{x}_{i+} = \sum\limits_{j=1}^{n} \dfrac{x_{ij}}{n}$ — среднее значение для i-й группы (и о. м. п. для μ_i при H_1), $\bar{x}_{++} = \sum\limits_{i,j} \dfrac{x_{ij}}{N} = \sum\limits_i \dfrac{n\bar{x}_{i+}}{N}$ — глобальное среднее (и о. м. п. для μ при H_0).

Критерий состоит в следующем: гипотеза H_0 отклоняется, когда $\dfrac{S_0}{S_1}$ принимает большие значения. Далее,

$$S_0 = \sum_{i=1}^{m} \sum_{j=1}^{n} (x_{ij} - \bar{x}_{i+} + \bar{x}_{i+} - \bar{x}_{++})^2 =$$
$$= \sum_{i=1}^{m} \sum_{j=1}^{n} [(x_{ij} - \bar{x}_{i+})^2 + 2(x_{ij} - \bar{x}_{i+})(\bar{x}_{i+} - \bar{x}_{++}) + (\bar{x}_{i+} - \bar{x}_{++})^2] =$$
$$= \sum_{i,j} (x_{ij} - \bar{x}_{i+})^2 + n \sum_i (\bar{x}_{i+} - \bar{x}_{++})^2 = S_1 + S_2,$$

где

$$S_2 = n \sum_i (\bar{x}_{i+} - \bar{x}_{++})^2.$$

Таким образом, величина S_0/S_1 принимает большие значения, когда S_2/S_1 принимает большие значения. Величина S_1 называется внутригрупповой, а S_2 — межгрупповой суммой квадратов.

Далее, независимо от того, верна гипотеза H_0 или нет, получаем

$$\sum_i (X_{ij} - \bar{X}_{i+})^2 \sim \sigma^2\chi^2_{n-1} \quad \forall i,$$

поскольку $\mathsf{E}X_{ij}$ зависит только от i. Следовательно,

$$S_1 \sim \sigma^2\chi^2_{N-m},$$

так как выборки при разных i независимы, а также для любого i

$$\text{сумма } \sum_j (X_{ij} - \bar{X}_{i+})^2 \text{ не зависит от } \bar{X}_{i+}.$$

Поэтому S_1 не зависит от S_2. Если гипотеза H_0 верна, то

$$S_2 \sim \sigma^2\chi^2_{m-1}.$$

Далее, если гипотеза H_0 неверна, то S_2 имеет среднее значение

$$\mathsf{E}S_2 = (m-1)\sigma^2 + n\sum_{i=1}^{m}(\mu_i - \bar{\mu})^2,$$

где $\bar{\mu} = \sum\limits_i \mu_i/m$.

Более того, при выполнении гипотезы H_1 значение S_2 увеличивается («раздувается»). Таким образом, если гипотеза H_0 верна, то

$$Q = \frac{S_2/(m-1)}{S_1/(N-m)} \sim \mathrm{F}_{m-1,N-m},$$

тогда как при H_1 значение Q увеличивается. Таким образом, при размере критерия α мы отвергаем H_0, когда значение статистики Q больше чем $\varphi^+_{m-1,N-m}(\alpha)$, где $\varphi^+_{m-1,N-m}(\alpha)$ — значение верхней α-квантили распределения $\mathrm{F}_{m-1,N-m}$. □

The Private Life of C. A. S. Anova

Частная жизнь К. А. З. Анова

(Из серии «Фильмы, которые не вышли на большой экран».)

Задача 55. а) В трех ресторанах на одной фешенебельной улице в один и тот же час было замечено 43, 41 и 30 посетителей соответственно. Описав предположения, которые вам необходимы, объясните, как проверить гипотезу о том, что все эти три ресторана пользуются одинаковой популярностью, против альтернативной гипотезы, что популярность этих ресторанов различна.

б) Объясните значение следующих терминов в контексте проверки гипотез:

1) простая гипотеза;

2) сложная гипотеза;

3) вероятности ошибок I и II рода;

4) размер критерия;

5) мощность критерия.

Пусть $X_1, \ldots, X_n$ — выборка из $\mathrm{N}(\mu, 1)$-распределения. Постройте наиболее мощный критерий размера α для проверки гипотезы $H_0\colon \mu = \mu_0$ против альтернативы $H_1\colon \mu = \mu_1$, где $\mu_1 > \mu_0$.

Найдите критерий, который минимизирует наибольшую из двух вероятностей ошибок. Обоснуйте ваш ответ.

Решение. а) В общей сложности замечено 114 посетителей. Предположив, что каждый из посетителей независимо от другого выбирает рестораны с вероятностями p_1, p_2, p_3, получим мультиномиальное распределение. Нулевая гипотеза $H_0\colon p_1 = p_2 = p_3 = 1/3$, при этом ожидаемое число посетителей 38. Значение χ^2-статистики Пирсона таково:

$$P = \sum \frac{(\text{наблюдаемое} - \text{ожидаемое})^2}{\text{ожидаемое}} = \frac{25 + 9 + 64}{38} = \frac{98}{38} = 2{,}579.$$

При заданном $\alpha \in (0, 1)$ мы сравниваем значение P с $h_2^+(\alpha)$, верхней α-квантилью χ_2^2-распределения. Критерий размера α отклоняет гипотезу H_0, когда $P > h_2^+(\alpha)$.

б) 1. Простая гипотеза H_0 состоит в том, что п. р. в./д. ф. р. имеет вид $f = f_0$, где f_0 — заданная функция, что дает возможность вычислить все необходимые вероятности.

2. Сложная гипотеза имеет вид $f \in$ множеству п. р. в./д. ф. р.

3. Вероятность ошибки I рода — это P(отклонить $H_0 \mid H_0$ верна), а вероятность ошибки II рода — это P(принять $H_0 \mid H_0$ ложна). В случае простой гипотезы H_0 в. о. р. I — это число, а в случае сложной гипотезы H_0 — это функция (аргумент которой изменяется в пределах параметрического множества, соответствующего гипотезе H_0). Вероятность ошибки II рода имеет аналогичный смысл.

4. Размер критерия равен максимальному значению, которое в. о. р. I достигает на параметрическом множестве, соответствующем гипотезе H_0. Если гипотеза H_0 простая, то размер критерия попросту равен в. о. р. I.

5. Аналогично мощность равна 1 минус в. о. р. II и рассматривается как функция на параметрическом множестве, соответствующем гипотезе H_1.

Для построения н. м. критерия используем лемму Н—П. Отношение правдоподобия

$$\frac{f(\boldsymbol{x};\, \mu_1)}{f(\boldsymbol{x};\, \mu_0)} = \exp\left(-\frac{1}{2}\sum_i [(x_i-\mu_1)^2-(x_i-\mu_0)^2]\right) = \exp\left[n(\mu_1-\mu_0)\bar{x}+\frac{n}{2}(\mu_0^2-\mu_1^2)\right]$$

принимает большие значения, когда $\bar{x}$ велико. Таким образом, критическая область н. м. критерия размера α имеет вид

$$\mathcal{C} = \{\boldsymbol{x}\colon \bar{x} > c\},$$

где

$$\alpha = \mathsf{P}_{\mu_0}(\bar{X} > c) = 1 - \Phi(\sqrt{n}(c-\mu_0)),$$

т. е.

$$c = \mu_0 + \frac{1}{\sqrt{n}} z_+(\alpha), \quad \text{где} \quad \frac{1}{\sqrt{2\pi}} \int\limits_{z_+(\alpha)}^{\infty} e^{-x^2/2}\, dx = \alpha.$$

Критерий, минимизирующий наибольшую из вероятностей ошибок, должен опять быть критерием Н—П; в противном случае существовал бы лучший критерий. Для размера α в. о. р. II как функция от α равна

$$\beta(\alpha) = \mathsf{P}_{\mu_1}(\bar{X} < c) = \Phi(z_+(\alpha) + \sqrt{n}(\mu_0-\mu_1)),$$

где $z_+(\alpha)$ — верхняя α-квантиль распределения $\mathrm{N}(0,\, 1)$. Ясно, что максимум $\max\{\alpha,\, \beta(\alpha)\}$ будет достигать минимального значения при $\alpha = \beta(\alpha)$, см. рис. 3.5.

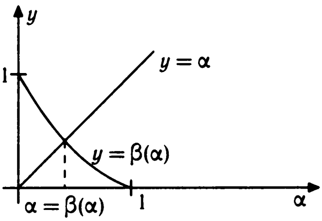


Рис. 3.5

Мы знаем, что когда α возрастает, $z_+(\alpha)$ убывает. Выберем α так, чтобы выполнялось равенство

$$\alpha = \Phi(z_+(\alpha) + \sqrt{n}(\mu_0-\mu_1)), \quad \text{т. е.} \quad z_+(\alpha) = -\frac{\sqrt{n}}{2}(\mu_0-\mu_1).$$

Отсюда находим

$$c = \mu_0 - \frac{1}{2}(\mu_0-\mu_1) = \frac{\mu_0+\mu_1}{2}. \qquad \square$$

Задача 56. а) Пусть $X_1, \ldots, X_m$ — это случайная выборка из $\mathrm{N}(\mu_1, \sigma_1^2)$-распределения и $Y_1, \ldots, Y_n$ — независимая выборка из $\mathrm{N}(\mu_2, \sigma_2^2)$-распределения. Параметры μ_1, μ_2, σ_1^2 и σ_2^2 неизвестны. Объясните, как проверить гипотезу $H_0\colon \sigma_1^2 = \sigma_2^2$ против альтернативы $H_1: \sigma_1^2 \neq \sigma_2^2$.

б) Пусть $Y_1, \ldots, Y_n$ — независимые случайные величины, где Y_i имеет $\mathrm{N}(\beta x_i, \sigma^2)$-распределение, $i = 1, \ldots, n$. При этом $x_1, \ldots, x_n$ известны, а β и σ^2 неизвестны. Требуется

1) найти оценки максимального правдоподобия $(\hat{\beta}, \hat{\sigma}^2)$ параметров (β, σ^2);

2) найти распределение оценки $\hat{\beta}$;

3) показав, что с. в. $Y_i - \hat{\beta}x_i$ и $\hat{\beta}$ независимы, найти совместное распределение величин $\hat{\beta}$ и $\hat{\sigma}^2$ (или найти это распределение иным способом);

4) объяснить, как проверить гипотезу $H_0\colon \beta = \beta_0$ против альтернативы $H_1: \beta \neq \beta_0$.

Решение. а) Положим

$$S_{xx} = \sum_{i=1}^{m} (x_i - \bar{x})^2, \quad \text{при этом} \quad \frac{1}{\sigma_1^2} S_{XX} \sim \chi^2_{m-1},$$

и

$$S_{yy} = \sum_{j=1}^{n} (y_j - \bar{y})^2, \quad \text{при этом} \quad \frac{1}{\sigma_2^2} S_{YY} \sim \chi^2_{n-1}.$$

Более того, величины S_{XX} и S_{YY} независимы.

Тогда при выполнении гипотезы H_0 мы имеем

$$\frac{1}{m-1} S_{XX} \Big/ \frac{1}{n-1} S_{YY} \sim \mathrm{F}_{m-1,n-1},$$

и в критерии размера α гипотеза H_0 отклоняется, если значение $(S_{xx}/(m-1))/(S_{yy}/(n-1))$ будет либо меньше $\varphi^-_{m-1,n-1}(\alpha/2)$, либо больше $\varphi^+_{m-1,n-1}(\alpha/2)$, где $\varphi^{\pm}_{m-1,n-1}(\alpha/2)$ — верхняя/нижняя квантиль распределения $\mathrm{F}_{m-1,n-1}$.

б) Функция правдоподобия имеет вид

$$f_{\beta,\sigma^2}(\boldsymbol{Y}) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left[-\frac{1}{2\sigma^2} \sum_i (Y_i - \beta x_i)^2\right].$$

1. Оценки максимального правдоподобия $(\hat{\beta}, \hat{\sigma}^2)$ для (β, σ^2) находят из уравнений

$$\frac{\partial}{\partial\beta} \ln f_{\beta,\sigma^2}(\boldsymbol{Y}) = \frac{1}{\sigma^2} \sum_i x_i(Y_i - \beta x_i) = 0,$$

$$\frac{\partial}{\partial\sigma^2} \ln f_{\beta,\sigma^2}(\boldsymbol{Y}) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_i (Y_i - \beta x_i)^2 = 0,$$

и эти оценки таковы:

$$\hat{\beta} = \frac{\sum_i x_i Y_i}{\sum_i x_i^2}, \ \hat{\sigma}^2 = \frac{1}{n}\sum_i (Y_i - \hat{\beta} x_i)^2.$$

На этих значениях достигается глобальный максимум, поскольку мы минимизируем выпуклую квадратичную функцию $-\ln f_{\beta,\sigma^2}(\boldsymbol{Y})$.

2. Мы имеем $\hat{\beta} \sim \mathrm{N}\left(\beta, \sigma^2 \Big/ \sum_i x_i^2\right)$, так как $\hat{\beta}$ является линейной комбинацией независимых нормальных с. в.

3. Поскольку величины $\hat{\beta}$ и $Y_i - \hat{\beta} x_i$ имеют совместное нормальное распределение, они являются независимыми тогда и только тогда, когда их ковариация равна нулю. Так как $\mathrm{E}(Y_i - \hat{\beta} x_i) = 0$, мы получаем

$$\mathrm{Cov}(Y_i - \hat{\beta} x_i, \hat{\beta}) = \mathrm{E}(Y_i - \hat{\beta} x_i)\hat{\beta} = \mathrm{E}(Y_i \hat{\beta}) - x_i\,\mathrm{E}(\hat{\beta}^2),$$

что равно нулю, поскольку

$$\mathrm{E}(Y_i\hat{\beta}) = \sum_{j\neq i} x_j \frac{\mathrm{E}Y_i\,\mathrm{E}Y_j \sum_k x_k^2}{+} x_i \frac{\mathrm{E}Y_i^2}{\sum_k x_k^2} = \sum_{j\neq i} \frac{\beta^2 x_i x_j^2}{\sum_k x_k^2} + x_i \frac{\sigma^2 + \beta^2 x_i^2}{\sum_k x_k^2} =$$

$$= x_i \left(\beta^2 + \frac{\sigma^2}{\sum_k x_k^2} \right) = x_i (\mathrm{Var}\,\hat{\beta} + (\mathrm{E}\hat{\beta})^2) = x_i\,\mathrm{E}(\hat{\beta}^2).$$

Аналогично можно проверить, что $Y_1 - \hat{\beta} x_1, \ldots, Y_n - \hat{\beta} x_n, \hat{\beta}$ являются независимыми и нормальными с. в. Тогда величины $\hat{\sigma}^2$ и $\hat{\beta}$ независимы.

Далее, сумма $\sum_i (Y_i - \beta x_i)^2$ равна

$$\sum_i [Y_i - \hat{\beta} x_i + (\hat{\beta} - \beta) x_i]^2 = \sum_i (Y_i - \hat{\beta} x_i)^2 + \left(\sum_i x_i^2\right)(\hat{\beta} - \beta)^2.$$

Поскольку

$$\frac{1}{\sigma^2}\sum_i (Y_i - \beta x_i)^2 \sim \chi_n^2 \quad \text{и} \quad \frac{1}{\sigma^2}\left(\sum_i x_i^2\right)(\hat{\beta} - \beta)^2 \sim \chi_1^2,$$

мы заключаем, что $\hat{\sigma}^2 n/\sigma^2 \sim \chi_{n-1}^2$.

4. При выполнении гипотезы H_0 мы имеем

$$T = \frac{(\hat{\beta} - \beta_0)\sqrt{\sum_i x_i^2}}{\sqrt{\frac{1}{n-1}\sum_i (Y_i - \hat{\beta} x_i)^2}} \sim \mathrm{t}_{n-1}.$$

Поэтому критерий состоит в следующем: гипотеза H_0 отвергается, если $|T| > \mathrm{t}_{n-1}(\alpha/2)$, где $\mathrm{t}_{n-1}(\alpha/2)$ — верхняя $\alpha/2$ точка t_{n-1}-распределения. □

Задача 57. а) Пусть X — случайная величина с п. р. в.

$$f(x;\,\theta) = e^{-(x-\theta)}, \qquad \theta < x < \infty,$$

где θ имеет априорное показательное распределение со средним значением 1. Найдите апостериорную плотность распределения θ.

Найдите оптимальную байесовскую оценку параметра θ для с. в. X при квадратической функции потерь.

б) Пусть $X_1, \ldots, X_n$ — выборка из распределения с п. р. в.

$$f(x;\,\lambda,\,\mu) = \begin{cases} \dfrac{1}{\lambda+\mu} e^{-x/\lambda}, & x \geqslant 0, \\ \dfrac{1}{\lambda+\mu} e^{x/\mu}, & x < 0, \end{cases}$$

где $\lambda > 0$ и $\mu > 0$ — неизвестные параметры. Найдите (простую) достаточную статистику для $(\lambda,\,\mu)$ и оценки максимального правдоподобия $(\hat{\lambda},\,\hat{\mu})$ параметров $(\lambda,\,\mu)$.

Предположим теперь, что $n = 1$. Будет ли $\hat{\lambda}$ несмещенной оценкой для λ? Обоснуйте свой ответ.

Решение. а) Запишем для апостериорной п. р. в.

$$\pi(\theta\,|\,\boldsymbol{x}) \propto e^{-\theta} e^{-(x-\theta)} I(x > \theta > 0) \propto I(0 < \theta < x).$$

Таким образом, апостериорным распределением является $\mathrm{U}(0,\,x)$-распределение.

При квадратической функции потерь оптимальной оценкой является апостериорное среднее, т. е. $x/2$.

б) Мы имеем

$$f(\boldsymbol{x};\,\lambda,\,\mu) = \frac{1}{(\lambda+\mu)^n} \exp\left[-\sum_i x_i I(x_i \geqslant 0)/\lambda + \sum_i x_i I(x_i < 0)/\mu\right].$$

Следовательно,

$$T(\boldsymbol{x}) = (S^+,\,S^-), \qquad S^+ = \sum_i x_i I(x_i \geqslant 0), \qquad S^- = \sum_j x_j I(x_j < 0),$$

является достаточной статистикой.

Для того чтобы найти о. м. п. $(\hat{\lambda},\,\hat{\mu})$, находим производные функции $\ell(\boldsymbol{x};\,\lambda,\,\mu) = \ln f(\boldsymbol{x};\,\lambda,\,\mu)$:

$$\frac{\partial}{\partial\lambda}\ell(\boldsymbol{x};\,\lambda,\,\mu) = -\frac{n}{\lambda+\mu} + \frac{S^+}{\lambda^2} = 0,$$

$$\frac{\partial}{\partial\mu}\ell(\boldsymbol{x};\,\lambda,\,\mu) = -\frac{n}{\lambda+\mu} - \frac{S^-}{\mu^2} = 0,$$

откуда получаем

$$\hat{\lambda} = \frac{1}{n}(S^+ + \sqrt{-S^-S^+}), \quad \hat{\mu} = \frac{1}{n}(-S^- + \sqrt{-S^-S^+}),$$

что является единственным решением. При этих значениях ℓ достигает максимума, т. е. они являются о. м. п. для (λ, μ), поскольку $\ell \to -\infty$, если λ или μ стремятся к 0 или ∞.

Эти рассуждения применимы в случае, когда S^+, $-S^- > 0$. Если одна из этих величин равна 0, то соответствующий параметр оценивают значением 0. Таким образом, вышеприведенная формула справедлива для любых выборок $\boldsymbol{x} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$.

При $n = 1$ оценка $\hat{\lambda}$ имеет вид $\hat{\lambda} = xI(x \geqslant 0)$. Так как значение

$$\mathsf{E}\hat{\lambda} = \frac{1}{\lambda + \mu} \int_0^\infty x e^{-x/\lambda}\, dx$$

зависит от μ, оно не может быть равным λ. Таким образом, оценка $\hat{\lambda}$ является смещенной.

Исключением является случай, когда $\mu = 0$. Тогда $\mathsf{E}\hat{\lambda} = \lambda$ и оценка является несмещенной. В общем случае $\mathsf{E}S^+ = n\lambda^2/(\lambda + \mu)$. Таким образом,

$$\mathsf{E}\hat{\lambda} = \frac{\lambda^2}{\lambda + \mu} + \frac{1}{n}\mathsf{E}\sqrt{-S^-S^+}.$$

Второе слагаемое неотрицательно при $n > 1$, если только не выполнены равенства $\mu = 0$ и $S^- = 0$. Таким образом, в исключительном случае, когда $\mu = 0$, $\mathsf{E}\hat{\lambda} = \lambda$ для всех $n \geqslant 1$. □

Задача 58. а) Пусть $x_1, \ldots, x_n$ — выборка из нормального распределения с неизвестным средним μ и известной дисперсией σ^2. Покажите, как построить наиболее мощный критерий заданного размера $\alpha \in (0, 1)$ для проверки нулевой гипотезы H_0: $\mu = \mu_0$ против альтернативы H_1: $\mu = \mu_1$ $(\mu_0 \neq \mu_1)$.

При каком значении α мощность этого критерия равна $1/2$?

б) Сформулируйте и докажите лемму Неймана—Пирсона. В случае простой нулевой гипотезы и простой альтернативы какой критерий вы бы предложили, чтобы минимизировать сумму вероятностей ошибок I и II рода? Обоснуйте свой ответ.

Решение. а) Поскольку обе гипотезы простые, н. м. критерий размера не больше α — это критерий о. п. с критической областью

$$C = \left\{\boldsymbol{x} : \frac{f(\boldsymbol{x} \mid H_1)}{f(\boldsymbol{x} \mid H_0)} > k\right\},$$

где k таково, что вероятность ошибки I рода $\mathrm{P}(C\,|\,H_0)=\alpha$. Находим

$$f(\boldsymbol{x}\,|\,H_0)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}e^{-(x_i-\mu_0)^2/(2\sigma^2)},\quad f(\boldsymbol{x}\,|\,H_1)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}e^{-(x_i-\mu_1)^2/(2\sigma^2)}$$

и

$$\Lambda_{H_1:H_0}(\boldsymbol{x})=\ln\frac{f(\boldsymbol{x}\,|\,H_1)}{f(\boldsymbol{x}\,|\,H_0)}=-\frac{1}{2\sigma^2}\sum_i[(x_i-\mu_1)^2-(x_i-\mu_0)^2]=$$

$$=\frac{n}{2\sigma^2}[2\bar{x}(\mu_1-\mu_0)-(\mu_1^2-\mu_0^2)].$$

Случай 1: $\mu_1>\mu_0$. Отклоняем гипотезу H_0, если $\bar{x}>k$, где $\alpha=\mathrm{P}(\bar{X}>k\,|\,H_0)$. При выполнении гипотезы H_0 имеет место соотношение $\bar{X}\sim\mathrm{N}(\mu_0,\sigma^2/n)$, так как $X_i\sim\mathrm{N}(\mu_0,\sigma^2)$ и X_i независимы. Тогда $\sqrt{n}(\bar{X}-\mu_0)/\sigma\sim\mathrm{N}(0,1)$. Таким образом, мы отклоняем гипотезу H_0, когда $\bar{x}>\mu_0+\sigma z_+(\alpha)/\sqrt{n}$, где $z_+(\alpha)=\Phi^{-1}(1-\alpha)$ — верхняя α-точка $\mathrm{N}(0,1)$-распределения.

Мощность критерия равна

$$\frac{1}{\sqrt{2\pi n}\sigma}\int\limits_{\sigma\sqrt{n}z_+(\alpha)+n\mu_0}^{\infty}e^{-(y-n\mu_1)^2/(2\sigma^2 n)}\,dy=\frac{1}{\sqrt{2\pi}}\int\limits_{z_+(\alpha)+\sqrt{n}(\mu_0-\mu_1)/\sigma}^{\infty}e^{-y^2/2}\,dy,$$

что равно $1/2$ при $z_+(\alpha)=(\mu_1-\mu_0)\sqrt{n}/\sigma$, т. е.

$$\alpha=\frac{1}{\sqrt{2\pi}}\int\limits_{\sqrt{n}(\mu_1-\mu_0)/\sigma}^{\infty}e^{-y^2/2}\,dy.$$

Случай 2: $\mu_1<\mu_0$. Отклоняем гипотезу H_0, если $\bar{x}<\bar{\mu}-\sigma z_+(\alpha)/\sqrt{n}$, в силу аналогичных аргументов. Мощность равна $1/2$, когда

$$\alpha=\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\sqrt{n}(\mu_1-\mu_0)/\sigma}e^{-y^2/2}\,dy.$$

Следовательно, мощность равна $1/2$, когда

$$\alpha=\frac{1}{\sqrt{2\pi}}\int\limits_{\sqrt{n}|\mu_0-\mu_1|/\sigma}^{\infty}e^{-y^2/2}\,dy.$$

б) (Отвечаем только на вторую часть вопроса.) Рассмотрим непрерывный случай: $f(\cdot\,|\,H_0)$ и $f(\cdot\,|\,H_1)$ — п. р. в. Рассмотрим критерий с критической областью $C=\{x\colon f(x\,|\,H_1)>f(x\,|\,H_0)\}$ и вероятностями ошибок

$\mathrm{P}(C \mid H_0)$ и $1 - \mathrm{P}(C \mid H_1)$. Тогда для любого критерия с критической областью C^* сумма вероятностей ошибок равна

$$\mathrm{P}(C^* \mid H_0) + 1 - \mathrm{P}(C^* \mid H_1) = 1 + \int\limits_{C^*} [f(x \mid H_0) - f(x \mid H_1)]\, dx =$$
$$= 1 + \int\limits_{C^* \cap C} [f(x \mid H_0) - f(x \mid H_1)]\, dx + \int\limits_{C^* \setminus C} [f(x \mid H_0) - f(x \mid H_1)]\, dx \geqslant$$
$$\geqslant 1 + \int\limits_{C^* \cap C} [f(x \mid H_0) - f(x \mid H_1)]\, dx,$$

поскольку значение интеграла по $C^* \setminus C$ будет неотрицательным.

Далее,

$$1 + \int\limits_{C^* \cap C} [f(x \mid H_0) - f(x \mid H_1)]\, dx =$$
$$= 1 + \int\limits_{C} [f(x \mid H_0) - f(x \mid H_1)]\, dx - \int\limits_{C \setminus C^*} [f(x \mid H_0) - f(x \mid H_1)]\, dx \geqslant$$
$$\geqslant 1 + \int\limits_{C} [f(x \mid H_0) - f(x \mid H_1)]\, dx = \mathrm{P}(C \mid H_0) + 1 - \mathrm{P}(C \mid H_1),$$

поскольку значение интеграла по $C \setminus C^*$ будет отрицательным. □

Следует сказать, что развитие статистики после леммы Неймана—Пирсона было омрачено затянувшимися на долгие годы спорами и дискуссиями среди статистиков, к которым примешивалась значительная доля личной вражды. Наиболее известным (и серьезным ввиду своих последствий) был, возможно, спор Фишера—Неймана, завязавшийся публично в 1935 г. и продолжавшийся даже после смерти Фишера в 1962 г. (одна из статей Неймана была озаглавлена «Серебряный юбилей моего спора с Фишером»). Две авторитетные книги по истории статистики, [FiB] и [Rei], приводят различные версии того, с чего все началось и как развивалось. Согласно [FiB, с. 263], это Нейман в 1934—1935 гг. «критиковал Фишера на своих лекциях и раздувал неугасимое пламя непонимания и разногласий между кафедрами Фишера и К. Пирсона в Лондонском университетском колледже с очевидным желанием посеять раздоры». С другой стороны, в книге [Rei] явно возлагается вина на Фишера, в обоснование этого цитируются мнения многих людей, таких как Р. Оппенгеймер (в будущем руководитель атомного проекта в Лос-Аламосе), который, как предполагают, сказал о Фишере в 1936 г.: «Я лишь раз взглянул на него и решил, что не хочу общаться с ним» (см. [Rei, с. 144]). В этой ситуации права была, возможно, Ф. Н. Дэвид (1909—1993), выдающийся британский статистик, которая знала все стороны, вовлеченные в конфликт, и говорила: «Они (Фишер, Нейман, К. Пирсон, Э. Пирсон) все завидовали друг другу и опасались, как бы кто-нибудь из них не опередил других». В частности, касаясь леммы Н—П, она утверждала: «Госсету не была свойственна зависть. Он поставил задачу. Э. Пирсон до некоторой степени перефразировал вопрос Госсета, выразив его на языке статистики. Нейман решил задачу математически» ([Rei, с. 133]). Согласно [FiB, с. 451], лемма Н—П «первоначально составляла часть работы Фишера», но вскоре «выделилась из нее. Она стала повсеместно признанной и всеми (широко) изучаемой, особенно в Соединенных Штатах. Это не вызвало одобрения у Фишера... ».

Дэвид была одной из внучатых племянниц Флоренс Найтингейл. Интересно отметить, что Дэвид была первой женщиной — профессором статистики в Британии, а Флоренс Найтингейл была первой женщиной — членом Королевского статистического общества. В конце своей карьеры Дэвид переехала в Калифорнию, но еще десятилетия содержала коттедж и садик на юго-востоке Англии.

Отметим, что продолжительные дискуссии и рассуждения по поводу леммы Н—П, теории и практики, произраставших из нее, сводятся, по сути, к следующему основному вопросу. Предположим, что наблюдается выборка $(x_1, \ldots, x_n)$. Что бы вы могли (аргументированно) сказать о предполагаемом случайном механизме, лежащем в основе этой выборки? Согласно сложившемуся мнению, методы Фишера будут занимать видное место в будущем развитии статистических наук (см., например, [E1]). Но даже признанные лидеры современной статистики не претендуют на то, что они располагают ясным видением этого вопроса (см. дискуссию, следующую за основным изложением в [E1]). Однако это не должно отвлекать нас от наших скромных целей.

Задача 59. а) Объясните, что означает построить доверительный интервал для неизвестного параметра θ по заданной выборке $x_1, \ldots, x_n$. Пусть задано следующее семейство плотностей распределения вероятностей $f(x;\, \theta)$, $-\infty < \theta < \infty$:

$$f(x;\, \theta) = \begin{cases} e^{-(x-\theta)}, & x \geqslant \theta, \\ 0, & x < \theta. \end{cases}$$

Пусть $n = 4$ и $x_1 = -1{,}0$, $x_2 = 1{,}5$, $x_3 = 0{,}5$, $x_4 = 1{,}0$. Постройте 95%-доверительный интервал для θ.

б) Пусть $f(x;\, \mu, \sigma^2)$ — семейство нормальных п. р. в. с неизвестным средним μ и неизвестной дисперсией $\sigma^2 > 0$. Объясните, как построить 95%-доверительный интервал для μ по выборке $x_1, \ldots, x_n$. Обоснуйте необходимые при этом требования относительно распределений.

Мой левый хвост[1]

(Из серии «Фильмы, которые не вышли на большой экран».)

Решение. а) Для построения $100(1-\gamma)\%$-д. и. необходимо найти две такие оценки $a = a(\boldsymbol{x})$, $b = b(\boldsymbol{x})$, что $\mathrm{P}(a(\boldsymbol{X}) \leqslant \theta \leqslant b(\boldsymbol{X})) \geqslant 1-\gamma$.

В данном примере

$$f(\boldsymbol{x};\, \theta) = e^{-\sum\limits_i x_i + n\theta} I(x_i \geqslant \theta\ \forall i).$$

Следовательно, $\min X_i$ является достаточной статистикой, и $\min X_i - \theta \sim \sim \mathrm{Exp}(n)$. Тогда можно взять $a = b - \delta$ и $b = \min x_i$, где

$$\int\limits_{\delta}^{\infty} n e^{-nx}\, dx = e^{-n\delta} = \gamma.$$

При $\gamma = 0{,}05$, $n = 4$ и $\min x_i = -1$ находим следующий д. и.:

$$\left[-1 - \frac{\ln 20}{4},\, -1\right].$$

[1]Ср. название фильма «My Left Foot».

Возможно и другое решение. Рассмотрим

$$\sum_{i=1}^{n} X_i - n\theta \sim \mathrm{Gam}(n,\ 1) \quad \text{или} \quad 2\sum_{i=1}^{n}(X_i - \theta) \sim \chi^2_{2n}.$$

Следовательно, можно взять

$$a = \left(\sum_i X_i - \frac{1}{2}h^+_{2n}(\gamma/2)\right) \Big/ n, \quad b = \left(\sum_i X_i - \frac{1}{2}h^-_{2n}(\gamma/2)\right) \Big/ n,$$

где $h^{\pm}_{2n}(\gamma/2)$ — верхняя/нижняя $\gamma/2$-квантиль χ^2_{2n}-распределения. При $\gamma = 0{,}05$, $n = 4$ и $\sum\limits_i x_i = 2$ находим

$$\left[\frac{1}{2} - \frac{h^+_8(0{,}025)}{8},\ \frac{1}{2} - \frac{h^-_8(0{,}025)}{8}\right].$$

Точное выражение для первого интервала: $[-1{,}749,\ -1]$, а для второго:

$$\left[\frac{1}{2} - \frac{17{,}530}{8},\ \frac{1}{2} - \frac{2{,}180}{8}\right] = [-1{,}6912,\ 0{,}2275].$$

Правая точка 0,2275 оказывается, конечно, бесполезной, поскольку мы знаем, что $\theta \leqslant x_1 = -1{,}0$. Заменив 0,2275 на -1, получим более короткий интервал, имеющий вид $[-1{,}6912,\ -1]$.

б) Положим

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Тогда 1) $\sqrt{n}(\hat{\mu} - \mu)/\sigma \sim \mathrm{N}(0,\ 1)$ и 2) $(n-1)\hat{\sigma}^2/\sigma^2 \sim \chi^2_{n-1}$, и эти величины независимы, откуда следует, что

$$\sqrt{n}\,\frac{\hat{\mu} - \mu}{\hat{\sigma}} \sim \mathrm{t}_{n-1}.$$

Следовательно, симметричный д. и. имеет вид

$$\left[\hat{\mu} - \frac{\hat{\sigma}}{\sqrt{n}}\mathrm{t}_{n-1}(0{,}025),\ \hat{\mu} + \frac{\hat{\sigma}}{\sqrt{n}}\mathrm{t}_{n-1}(0{,}025)\right],$$

где $\mathrm{t}_{n-1}(\alpha/2)$ — $\alpha/2$-квантиль критерия t_{n-1}, $\alpha = 0{,}05$. Обоснование утверждений 1, 2 было приведено выше. □

Задача 60. а) Сформулируйте и докажите критерий факторизации для достаточной статистики в случае дискретных случайных величин.

б) Величину, заданную линейной функцией $y = Ax + B$ с неизвестными коэффициентами A и B, многократно измеряют в различных точках $x_1, \ldots, x_k$: сначала n_1 раз в точке x_1, затем n_2 раз в точке x_2 и т. д.; и, наконец, n_k раз в точке x_k. Результатом i-й серии измерений является выборка $y_{i1}, \ldots, y_{in_i}$, $i = 1, \ldots, k$. Ошибки при всех измерениях являются независимыми нормальными величинами с нулевыми средними и дисперсией 1. Необходимо оценить A и B по всей выборке y_{ij}, $1 \leqslant j \leqslant n_i$, $1 \leqslant i \leqslant k$. Докажите, что оценки максимального правдоподобия и МНК-оценки пары (A, B) совпадают, и найдите их.

Обозначим через $\hat{A}$ оценку максимального правдоподобия параметра A и через $\hat{B}$ — оценку максимального правдоподобия параметра B. Найдите распределение пары $(\hat{A}, \hat{B})$.

Решение. б) Определим

$$\langle n\rangle = \sum_i n_i, \qquad \bar{x} = \frac{\sum\limits_i x_i n_i}{\langle n\rangle}$$

и

$$u_i = x_i - \bar{x}, \quad \text{причем} \quad \langle un\rangle = \sum_i u_i n_i = 0,\ \langle u^2 n\rangle = \sum_i u_i^2 n_i > 0.$$

Пусть $\alpha = A$, $\beta = B + A\bar{x}$, тогда $y_{ij} = \alpha u_i + \beta + \varepsilon_{ij}$ и $Y_{ij} \sim \mathrm{N}(\alpha u_i + \beta,\ 1)$, т. е.

$$f_{Y_{ij}}(y_{ij}) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y_{ij} - \alpha u_i - \beta)^2\right).$$

Для того чтобы найти о. м. п., необходимо минимизировать произведение $\prod\limits_{i,j} \exp\left(-\frac{1}{2}(y_{ij} - \alpha u_i - \beta)^2\right)$. Это эквивалентно отысканию минимума квадратичной функции $\sum\limits_{i,j}(y_{ij} - \alpha u_i - \beta)^2$. Последняя задача приводит в точности к МНК-оценкам. Следовательно, о. м. п. и МНК-оценки совпадают.

Для отыскания МНК-оценок решим уравнения

$$\frac{\partial}{\partial\beta} \sum_{i,j} (Y_{ij} - \beta - \alpha u_i)^2 = 0 \iff \hat{\beta} = \frac{1}{\langle n\rangle} \sum_{i,j} Y_{ij},$$

$$\frac{\partial}{\partial\alpha} \sum_{i,j} (Y_{ij} - \beta - \alpha u_i)^2 = 0 \iff \hat{\alpha} = \frac{1}{\langle u^2 n\rangle} \sum_{i,j} Y_{ij} u_i,$$

где

$$\hat{\beta} \sim \mathrm{N}\left(\beta,\ \frac{1}{\langle n\rangle}\right), \qquad \hat{\alpha} \sim \mathrm{N}\left(\alpha,\ \frac{1}{\langle u^2 n\rangle}\right).$$

Таким образом,

$$\hat{A} = \frac{1}{\langle u^2 n\rangle} \sum_{i,j} Y_{ij} u_i \sim \mathrm{N}\left(A,\ \frac{1}{\langle u^2 n\rangle}\right),$$

$$\hat{B} = \frac{1}{\langle n\rangle} \sum_{i,j} Y_{ij} - \frac{\bar{x}}{\langle u^2 n\rangle} \sum_{i,j} Y_{ij} u_i \sim \mathrm{N}\left(B,\ \frac{1}{\langle n\rangle} + \frac{\bar{x}^2}{\langle u^2 n\rangle}\right).$$

Заметим, что величины $(\hat{A}, \hat{B})$ совместно нормальны с ковариацией

$$\mathrm{Cov}(\hat{A}, \hat{B}) = \frac{\sum\limits_i n_i x_i}{\left(\sum\limits_i n_i x_i\right)^2 - \left(\sum\limits_i n_i\right)\left(\sum\limits_i n_i x_i^2\right)}. \qquad \square$$

Задача 61. а) Пусть $x_1, \ldots, x_n$ — случайная выборка из распределения с п. р. в. $f(x;\ \theta)$. Что имеют в виду, когда говорят, что $t(x_1, \ldots, x_n)$ является достаточной статистикой для θ?

Пусть

$$f(x;\ \theta) = \begin{cases} e^{-(x-\theta)}, & x > \theta, \\ 0, & x \leqslant \theta, \end{cases}$$

и предположим, что $n = 3$. Пусть $y_1 < y_2 < y_3$ — упорядоченные значения x_1, x_2, x_3. Покажите, что y_1 является достаточной статистикой для θ.

б) Покажите, что распределение $Y_1 - \theta$ является показательным распределением с параметром 3. Ваш заказчик предлагает рассматривать в качестве возможных оценок параметра θ следующие величины:

$$\bar{\theta}_1 = x_3 - 1,$$
$$\bar{\theta}_2 = y_1,$$
$$\bar{\theta}_3 = \frac{1}{3}(x_1 + x_2 + x_3) - 1.$$

Что бы вы ему посоветовали?

Указание. Следует сформулировать общие теоремы, которые вы используете, но доказывать их не обязательно.

Решение. а) Величина $T(\boldsymbol{x})$ является достаточной статистикой для параметра θ тогда и только тогда, когда условное распределение выборки $\boldsymbol{X}$ при заданном $T(\boldsymbol{X})$ не содержит параметра θ. Это означает, что $\mathsf{P}_\theta(\boldsymbol{X} \in B \mid T = t)$ не зависит от θ для любой области B, принадлежащей выборочному пространству. Критерий факторизации утверждает, что T является достаточной статистикой тогда и только тогда, когда выборочная п. р. в. имеет вид $f(\boldsymbol{x};\ \theta) = g(T(\boldsymbol{x}), \theta)h(\boldsymbol{x})$ для некоторых функций g и h.

Для указанной функции $f(x;\ \theta)$ при $n = 3$ и $\boldsymbol{x} = (x_1, x_2, x_3)$ имеем

$$f(\boldsymbol{x};\ \theta) = \prod_{i=1}^{3} e^{-(x_i - \theta)} I(x_i > \theta) = e^{3\theta} e^{-\sum\limits_i x_i} I(\min x_i > \theta).$$

Таким образом, $T = \min X_i := Y_1$ является достаточной статистикой: здесь $g(y, \theta) = e^{3\theta} I(y > \theta)$, $h(\boldsymbol{x}) = e^{-\sum\limits_i x_i}$.

б) Для любого $y > 0$ выполняется равенство

$$P_\theta(Y_1 - \theta > y) = P_\theta(X_1, X_2, X_3 > y + \theta) = \prod_{i=1}^{3} P_\theta(X_i > y + \theta) = e^{-3y}.$$

Следовательно, $Y_1 \sim \mathrm{Exp}(3)$.

Находим $\mathsf{E}(X_3 - 1) = \mathsf{E}(X_3 - \theta) + \theta - 1 = \theta$, $\operatorname{Var} X_3 = 1$.

Далее, $Y_1 = \min X_i$ является о. м. п., которая максимизирует $f(x;\, \theta)$ по θ; эта оценка является смещенной, так как

$$\mathsf{E} Y_1 = \mathsf{E}(Y_1 - \theta) + \theta = \frac{1}{3} + \theta.$$

Дисперсия $\operatorname{Var}\overline{\theta}_2$ равна $1/9$.

Наконец,

$$\mathsf{E}\left(\frac{1}{3}(X_1 + X_2 + X_3) - 1\right) = \theta, \quad \operatorname{Var}\left(\frac{1}{3}(X_1 + X_2 + X_3) - 1\right) = \frac{1}{3}.$$

Мы видим, что Y_1 имеет наименьшую дисперсию среди всех трех величин. Мы бы посоветовали заказчику использовать оценку $\overline{\theta}_2$, помня о смещении. Однако наилучшим выбором является $\overline{\theta}_2 - 1/3$. □

Замечание. Теорема Р—Б предлагает использовать оценку $\hat{\theta} = \mathsf{E}(\overline{\theta}_1 \,|\, \overline{\theta}_2 = t) = \mathsf{E}(\overline{\theta}_3 \,|\, \overline{\theta}_2 = t)$. Прямые вычисления показывают, что $\hat{\theta} = t - 1/3$. Следовательно, $\mathsf{E}\hat{\theta} = \theta$ и $\operatorname{Var}\hat{\theta} = 1/9$. Таким образом, процедура Р—Б дает возможность избежать смещения и уменьшает дисперсию до минимума.

Задача 62. а) Получите оценки максимального правдоподобия для параметров α, β и σ^2 в линейной модели

$$Y_i = \alpha + \beta x_i + \varepsilon_i,$$

$1 \leqslant i \leqslant n$, где $\varepsilon \sim \mathrm{N}(0,\, \sigma^2)$ и $\sum\limits_{i=1}^{n} x_i = 0$.

б) Определите совместное распределение оценок максимального правдоподобия $\hat{\alpha}$, $\hat{\beta}$ и $\hat{\sigma}^2$. Постройте 95%-доверительный интервал для

1) σ^2,

2) $\alpha + \beta$.

Решение. а) Очевидно, что $Y_i \sim \mathrm{N}(\alpha + \beta x_i,\, \sigma^2)$. Тогда

$$f(\boldsymbol{y};\, \alpha,\, \beta,\, \sigma^2) = \frac{1}{(\sqrt{2\pi\sigma^2})^n} \exp\left(-\frac{1}{2\sigma^2}\sum_i (y_i - \alpha - \beta x_i)^2\right)$$

и

$$\ell(\boldsymbol{y};\, \alpha,\, \beta,\, \sigma^2) = \ln f(\boldsymbol{y};\, \alpha,\, \beta,\, \sigma^2) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_i (y_i - \alpha - \beta x_i)^2.$$

Минимум достигается, когда

$$\frac{\partial}{\partial\alpha}\ell(\boldsymbol{y};\,\alpha,\,\beta,\,\sigma^2) = \frac{\partial}{\partial\beta}\ell(\boldsymbol{y};\,\alpha,\,\beta,\,\sigma^2) = \frac{\partial}{\partial\sigma^2}\ell(\boldsymbol{y};\,\alpha,\,\beta,\,\sigma^2) = 0,$$

т. е. при

$$\hat{\alpha} = \bar{y}, \quad \hat{\beta} = \frac{\boldsymbol{x}^T\boldsymbol{Y}}{(\boldsymbol{x}^T\boldsymbol{x})}, \quad \hat{\sigma}^2 = \frac{1}{n}\sum_i \hat{\varepsilon}_i^2, \quad \text{где } \hat{\varepsilon}_i = Y_i - \hat{\alpha} - \hat{\beta}x_i.$$

б) Далее,

$$\hat{\alpha} \sim \mathrm{N}\left(\alpha,\, \frac{\sigma^2}{n}\right), \quad \hat{\beta} \sim \mathrm{N}\left(\beta,\, \frac{\sigma^2}{\boldsymbol{x}^T\boldsymbol{x}}\right), \quad \left(\frac{n}{\sigma^2}\right)\hat{\sigma}^2 \sim \chi^2_{n-2}.$$

Кроме того,

$$\mathrm{Cov}(\hat{\alpha},\,\hat{\beta}) = \mathrm{Cov}\left(\frac{1}{n}\sum_i Y_i,\, \frac{1}{\boldsymbol{x}^T\boldsymbol{x}}\sum_i x_iY_i\right) =$$

$$= \frac{1}{n\boldsymbol{x}^T\boldsymbol{x}}\sum_i x_i\,\mathrm{Cov}(Y_i,\,Y_i) = \frac{\sigma^2}{n\boldsymbol{x}^T\boldsymbol{x}}\sum_i x_i = 0,$$

$$\mathrm{Cov}(\hat{\alpha},\,\hat{\varepsilon}_i) = \mathrm{Cov}(\hat{\alpha},\,Y_i - \hat{\alpha} - \hat{\beta}x_i) = \frac{1}{n}\sum_{j=1}^{n}\left(\delta_{ij} - \frac{1}{n}\right)\mathrm{Cov}(Y_j,\,Y_j) = 0$$

и

$$\mathrm{Cov}(\hat{\beta},\,\hat{\varepsilon}_i) = \mathrm{Cov}(\hat{\beta},\,Y_i - \hat{\alpha} - \hat{\beta}x_i) =$$

$$= \frac{1}{\boldsymbol{x}^T\boldsymbol{x}}\sum_{j=1}^{n} x_j\left(\delta_{ij} - \frac{1}{\boldsymbol{x}^T\boldsymbol{x}}x_jx_i\right)\mathrm{Cov}(Y_j,\,Y_j) = \frac{\sigma^2}{\boldsymbol{x}^T\boldsymbol{x}}(x_i - x_i) = 0.$$

Следовательно, величины $\hat{\alpha}$, $\hat{\beta}$ и $\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n$ независимы. А значит, $\hat{\alpha}$, $\hat{\beta}$ и $\hat{\sigma}^2$ также независимы.

1) Мы покажем, что 95%-д. и. для σ^2 имеет вид

$$\left(\frac{n\hat{\sigma}^2}{h^+},\, \frac{n\hat{\sigma}^2}{h^-}\right),$$

где h^- — нижняя, а h^+ — верхняя 0,025-точка распределения χ^2_{n-2}.

2) Мы имеем

$$\hat{\alpha} + \hat{\beta} \sim \mathrm{N}\left(\alpha + \beta,\, \sigma^2\left(\frac{1}{n} + \frac{1}{\boldsymbol{x}^T\boldsymbol{x}}\right)\right),$$

и $\hat{\alpha} + \hat{\beta}$ не зависит от $\hat{\sigma}^2$. Значит,

$$\frac{(\hat{\alpha} + \hat{\beta}) - (\alpha + \beta)}{\sqrt{\dfrac{1}{n} + \dfrac{1}{\boldsymbol{x}^T\boldsymbol{x}}}} \Big/ \hat{\sigma}\sqrt{\frac{n}{n-2}} \sim \mathrm{t}_{n-2}.$$

Следовательно, 95%-д. и. для $\alpha+\beta$ имеет вид

$$\left(\hat{\alpha}+\hat{\beta}-t\hat{\sigma}\sqrt{\frac{1}{n}+\frac{1}{\mathbf{x}^T\mathbf{x}}}\sqrt{\frac{n}{n-2}},\ \hat{\alpha}+\hat{\beta}+t\hat{\sigma}\sqrt{\frac{1}{n}+\frac{1}{\mathbf{x}^T\mathbf{x}}}\sqrt{\frac{n}{n-2}}\right),$$

где t — это верхняя 0,025-квантиль t_{n-2}-распределения. □

Задача 63. а) Кратко опишите процедуру отыскания байесовской точечной оценки в статистическом эксперименте. Включите в это описание определения следующих понятий: 1) априорное распределение; 2) апостериорное распределение.

б) Пусть $X_1, \ldots, X_n$ — независимые одинаково распределенные случайные величины, имеющие распределение $\mathrm{Gam}(k, \lambda)$. Предположим, что k известно, а *априорное распределение* параметра λ является показательным с параметром μ. Пусть потери, понесенные при оценивании параметра λ значением a, имеют вид $(a-\lambda)^2$. Вычислите апостериорное распределение параметра λ и найдите его оптимальную байесовскую оценку.

Решение. а) Пусть задана выборка из распределения с п. р. в./д. ф. р. $f(\boldsymbol{x}; \theta)$ и задана априорная п. р. в./д. ф. р. $\pi(\theta)$ параметра θ. Тогда мы рассматриваем апостериорное распределение

$$\pi(\theta\,|\,\boldsymbol{x}) \propto \pi(\theta) f(\boldsymbol{x}; \theta),$$

нормированное так, чтобы общая масса равнялась 1. Байесовская оценка определяется как точка, минимизирующая средние потери $\mathsf{E}_{\pi(\theta\,|\,\boldsymbol{x})}L(a, \theta)$, где $L(a, \theta)$ — заданная функция, определяющая потери, понесенные при оценивании параметра θ значением a.

б) При квадратичных потерях, когда $L(a, \theta) = (a-\theta)^2$, байесовская оценка равна апостериорному среднему. В данном примере апостериорным является гамма-распределение:

$$\pi(\lambda\,|\,\boldsymbol{x}) \propto e^{-\mu\lambda}\lambda^{kn}\prod_i e^{-\lambda x_i} \sim \mathrm{Gam}\left(kn+1,\ \mu+\sum_i x_i\right).$$

Следовательно, байесовская оценка задается как среднее значение этого распределения $(kn+1)\Big/\left(\mu+\sum\limits_i x_i\right)$. □

Задача 64. Пусть $X_1, X_2, \ldots, X_n$ — независимая выборка из нормального распределения с неизвестными средним значением μ и дисперсией σ^2. Покажите, что пара $(\bar{X}, \bar{S}^2)$, где

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad \bar{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2,$$

— это достаточная статистика для (μ, σ^2).

При заданном $\lambda > 0$ рассмотрим $\lambda\bar{S}^2$ в качестве оценки параметра σ^2. При каких значениях λ оценка $\lambda\bar{S}^2$ является

а) оценкой максимального правдоподобия,

б) несмещенной оценкой?

Какое значение λ минимизирует среднеквадратическую ошибку

$$\mathsf{E}(\lambda\bar{S}^2 - \sigma^2)^2?$$

Решение. Функция правдоподобия имеет вид

$$f(\boldsymbol{x};\, \mu,\, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right] =$$
$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\bar{x}+\bar{x}-\mu)^2\right] =$$
$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\bar{x})^2 - \frac{n}{2\sigma^2}(\bar{x}-\mu)^2\right].$$

Следовательно, в силу критерия факторизации $(\bar{X},\, \bar{S}^2)$ является достаточной статистикой.

Отыскание максимума по μ и σ^2 эквивалентно решению уравнений

$$\frac{\partial}{\partial\mu}\ln f(\boldsymbol{x};\, \mu,\, \sigma^2) = \frac{\partial}{\partial\sigma^2}\ln f(\boldsymbol{x};\, \mu,\, \sigma^2) = 0,$$

что приводит к значениям $\hat{\mu} = \bar{x}$, $\hat{\sigma}^2 = \sum\limits_i (x_i - \bar{x})^2/n$. Следовательно, а) $\lambda\bar{S}^2$ является о. м. п. при $\lambda = 1$.

Для п. б) находим

$$\mathsf{E}n\bar{S}^2 = \sum_{i=1}^{n}\mathsf{E}(X_i - \bar{X})^2 = \sum_{i=1}^{n}\mathsf{E}(X_i-\mu)^2 - n\,\mathsf{E}(\bar{X}-\mu)^2 =$$
$$= n\operatorname{Var}X - \frac{n}{n^2}n\operatorname{Var}X = (n-1)\sigma^2.$$

Следовательно, $\lambda\bar{S}^2$ является несмещенной оценкой при $\lambda = n/(n-1)$.

Наконец, положим $\varphi(\lambda) = \mathsf{E}(\lambda\bar{S}^2 - \sigma^2)^2$. После дифференцирования получим

$$\varphi'(\lambda) = 2\,\mathsf{E}(\lambda\bar{S}^2 - \sigma^2)\bar{S}^2.$$

Далее, $n\bar{S}^2 \sim (Y_1^2 + \ldots + Y_{n-1}^2)$, где $Y_i \sim \mathrm{N}(0,\, \sigma^2)$ и y_i независимы. Следовательно,

$$\mathsf{E}(\bar{S}^2)^2 = n^{-2}\,\mathsf{E}(Y_1^2 + \ldots + Y_{n-1}^2)^2 = n^{-2}[(n-1)\,\mathsf{E}Y_1^4 + (n-1)(n-2)(\mathsf{E}Y_1^2)^2] =$$
$$= n^{-2}[3(n-1)\sigma^4 + (n-1)(n-2)\sigma^4] = n^{-2}(n^2-1)\sigma^4.$$

Поскольку $\mathsf{E}\bar{S}^2 = (n-1)\sigma^2/n$, из уравнения $\varphi'(\lambda) = 0$ находим

$$\lambda = \frac{\sigma^2\,\mathsf{E}\bar{S}^2}{\mathsf{E}(\bar{S}^2)^2} = \frac{n}{n+1},$$

что, как легко видеть, является точкой минимума. □

The Mystery of Mean Mu and Squared Sigma[1]

(Из серии «Фильмы, которые не вышли на большой экран».)

Задача 65. Пусть задан набор из np независимых случайных величин, представленный в виде n выборок длины p каждая:

$$\begin{aligned}
X^{(1)} &= (X_{11}, \ldots, X_{1p}),\\
X^{(2)} &= (X_{21}, \ldots, X_{2p}),\\
&\ldots\ldots\ldots\ldots\ldots\ldots\\
X^{(n)} &= (X_{n1}, \ldots, X_{np}).
\end{aligned}$$

Случайная величина X_{ij} имеет распределение Пуассона с неизвестным параметром λ_j, $1 \leqslant j \leqslant p$. Требуется проверить гипотезу о равенстве $\lambda_1 = \ldots = \lambda_p$ против альтернативы, что хотя бы два значения λ_j различны. Найдите выражение для статистики критерия отношения правдоподобия. Покажите, что ее можно аппроксимировать величиной

$$\frac{n}{\bar{X}} \sum_{j=1}^{p} (\bar{X}_j - \bar{X})^2,$$

где

$$\bar{X}_j = \frac{1}{n} \sum_{i=1}^{n} X_{ij}, \qquad \bar{X} = \frac{1}{np} \sum_{i=1}^{n} \sum_{j=1}^{p} X_{ij}.$$

Объясните, как бы вы могли проверить гипотезу при больших значениях n.

Решение. См. пример 2.5.6. □

Задача 66. Пусть $X_1, X_2, \ldots, X_n$ — независимая выборка из нормального распределения с известным средним μ и неизвестной дисперсией σ^2, принимающей одно из двух значений σ_1^2 или σ_2^2. Объясните подробно, как построить наиболее мощный критерий размера α для проверки гипотезы $\sigma = \sigma_1$ против альтернативы $\sigma = \sigma_2$. Существует ли н. м. критерий размера α, имеющий мощность, строго меньшую чем α? Обоснуйте свой ответ.

Решение. В силу леммы Н—П, поскольку обе гипотезы простые, н. м. критерий, имеющий размер не больше α, — это критерий отношения правдоподобия, при котором отклоняется гипотеза $H_0\colon \sigma = \sigma_1$ и принимается

[1] Ср. названия фильмов: «The Mystery of the Crystal Orb» и «Hero Squared».

$H_1 : \sigma = \sigma_2$, если

$$\frac{(2\pi\sigma_2^2)^{-n/2} \exp\left[-\sum_i (x_i - \mu)^2/(2\sigma_2^2)\right]}{(2\pi\sigma_1^2)^{-n/2} \exp\left[-\sum_i (x_i - \mu)^2/(2\sigma_1^2)\right]} > k,$$

где $k > 0$ подобрано так, что в. о. р. I равна α. Такой критерий всегда существует, поскольку нормальная п. р. в. $f(x)$ монотонна по $|x|$ и непрерывна.

Переписав о. п. в виде

$$\left(\frac{\sigma_1}{\sigma_2}\right)^n \exp\left[\frac{1}{2}\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2}\right) \sum (x_i - \mu)^2\right],$$

легко видеть, что если $\sigma_2 > \sigma_1$, то гипотеза H_0 отклоняется в критической области

$$C^+ = \left\{\sum_i (x_i - \mu)^2 \geqslant k^+\right\},$$

а если $\sigma_1 > \sigma_2$, то критической областью является

$$C^- = \left\{\sum_i (x_i - \mu)^2 \leqslant k^-\right\}.$$

Более того,

$$\frac{1}{\sigma_1^2} \sum_i (X_i - \mu)^2 \sim \chi_n^2 \quad \text{при } H_0$$

и

$$\frac{1}{\sigma_2^2} \sum_i (X_i - \mu)^2 \sim \chi_n^2 \quad \text{при } H_1.$$

Плотность распределения вероятностей χ_n^2 имеет вид

$$f_{\chi_n^2}(x) = \frac{1}{\Gamma(n/2)} \frac{1}{2^{n/2}} x^{n/2-1} e^{-x/2} I(x > 0).$$

Тогда если $\sigma_2 > \sigma_1$, то выбираем k^+ так, чтобы для размера критерия выполнялось соотношение

$$\mathrm{P}(C^+ \mid H_0) = \int\limits_{k^+/\sigma_1^2}^{\infty} \frac{1}{\Gamma(n/2)} \frac{1}{2^{n/2}} x^{n/2-1} e^{-x/2}\, dx = \alpha,$$

а если $\sigma_1 > \sigma_2$, выбираем k^- так, что

$$\mathrm{P}(C^- \mid H_0) = \int\limits_0^{k^-/\sigma_1^2} \frac{1}{\Gamma(n/2)} \frac{1}{2^{n/2}} x^{n/2-1} e^{-x/2}\, dx = \alpha.$$

В случае, когда $\sigma_2 > \sigma_1$, мощность равна

$$\beta = \mathrm{P}(C^+ \mid H_1) = \int\limits_{k^+/\sigma_2^2}^{\infty} \frac{1}{\Gamma(n/2)} \frac{1}{2^{n/2}} x^{n/2-1} e^{-x/2}\, dx,$$

а в случае $\sigma_1 > \sigma_2$ мощность равна

$$\beta = \mathrm{P}(C^- \mid H_1) = \int\limits_0^{k^-/\sigma_2^2} \frac{1}{\Gamma(n/2)} \frac{1}{2^{n/2}} x^{n/2-1} e^{-x/2}\, dx.$$

Мы видим, что если $\sigma_2 > \sigma_1$, то $k^+/\sigma_2^2 < k^+/\sigma_1^2$ и $\beta > \alpha$. Аналогично если $\sigma_1 > \sigma_2$, то $k^-/\sigma_2^2 > k^-/\sigma_1^2$ и опять $\beta > \alpha$. Таким образом, неравенство $\beta < \alpha$ невозможно. □

Gamma and Her Sisters[1]

Гамма и ее сестры

(Из серии «Фильмы, которые не вышли на большой экран».)

Задача 67. Пусть $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ — независимые случайные величины, имеющие $\mathrm{N}(0, 1)$-распределение, а $x_1, x_2, \ldots, x_n$ — фиксированные действительные числа. Пусть случайные величины $Y_1, Y_2, \ldots, Y_n$ заданы в виде

$$Y_i = \alpha + \beta x_i + \sigma\varepsilon_i, \qquad 1 \leqslant i \leqslant n,$$

где $\alpha, \beta \in \mathbb{R}$ и $\sigma^2 \in (0, \infty)$ — неизвестные параметры. Найдите выражение для оценки наименьших квадратов (МНК-оценки) пары (α, β) и определите ее распределение. Объясните, как проверить гипотезу о том, что $\beta = 0$, против альтернативы $\beta \neq 0$ и как построить 95%-доверительный интервал для β.

(Сформулируйте (не приводя доказательства) те общие факты, которые используются при решении этой задачи.)

Решение. Положим

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i,\ a = \alpha + \beta\bar{x} \quad \text{и} \quad u_i = x_i - \bar{x},$$

так что

$$\sum_i u_i = 0 \quad \text{и} \quad Y_i = a + \beta u_i + \sigma\varepsilon_i.$$

Пара МНК-оценок для (α, β) минимизирует сумму квадратов $R = \sum\limits_{i=1}^{n}(Y_i - \mathrm{E}Y_i)^2 = \sum\limits_{i=1}^{n}(Y_i - a - \beta u_i)^2$, т. е. является решением уравнений

$$\frac{\partial}{\partial\alpha}R = \frac{\partial}{\partial\beta}R = 0.$$

Отсюда

$$\hat{a} = \overline{Y} \sim \mathrm{N}\left(\alpha + \beta\bar{x},\ \frac{\sigma^2}{n}\right), \qquad \hat{\alpha} \sim \mathrm{N}\left(\alpha,\ \frac{\sigma^2}{n}\right)$$

[1] Ср. название фильма Вуди Аллена «Hannah and Her Sisters».

и

$$\hat{\beta} = \frac{\boldsymbol{u}^T \boldsymbol{Y}}{\boldsymbol{u}^T \boldsymbol{u}} \sim \mathrm{N}\left(\beta, \frac{\sigma^2}{\boldsymbol{u}^T \boldsymbol{u}}\right),$$

и эти величины независимы в совокупности. Далее,

$$\hat{R} = \sum_i (Y_i - \hat{\alpha} - \hat{\beta} x_i)^2 = \sum_i Y_i^2 - n\overline{Y}^2 - \boldsymbol{u}^T \boldsymbol{u} \hat{\beta}^2 \sim \sigma^2 \chi^2_{n-2},$$

так как оценки для двух параметров уже найдены. Таким образом, $R/(n-2)$ — несмещенная оценка для σ^2.

Рассмотрим теперь гипотезы $H_0\colon \beta = \beta_0 = 0$, $H_1 : \beta \neq \beta_0$. Тогда

$$\hat{\beta} \sim \mathrm{N}\left(0, \frac{\sigma^2}{\boldsymbol{u}^T \boldsymbol{u}}\right) \quad \text{или} \quad \frac{\hat{\beta}\sqrt{\boldsymbol{u}^T \boldsymbol{u}}}{\sigma} \sim \mathrm{N}(0, 1), \qquad \text{если } \beta = 0,$$

т. е. при H_0. Таким образом,

$$T = \frac{\hat{\beta}\sqrt{\boldsymbol{u}^T \boldsymbol{u}}}{\sqrt{\hat{R}/(n-2)}} \sim \mathrm{t}_{n-2}.$$

Следовательно, при заданном α мы отвергаем H_0, когда значение $|T|$ превосходит $\mathrm{t}_{n-2}(\alpha/2)$, верхнюю $\alpha/2$-точку t_{n-2}-распределения.

Наконец, чтобы построить симметричный 95%-д. и. для β, возьмем $\mathrm{t}_{n-2}(0{,}025)$. Тогда из неравенств

$$-\mathrm{t}_{n-2}(0{,}025) < \frac{(\beta - \hat{\beta})\sqrt{\boldsymbol{u}^T \boldsymbol{u}}}{\sqrt{\hat{R}/(n-2)}} < \mathrm{t}_{n-2}(0{,}025)$$

находим, что искомый интервал имеет вид

$$\left(\hat{\beta} - \mathrm{t}_{n-2}(0{,}025)\sqrt{\frac{\hat{R}}{(n-2)\boldsymbol{u}^T \boldsymbol{u}}}, \quad \hat{\beta} + \mathrm{t}_{n-2}(0{,}025)\sqrt{\frac{\hat{R}}{(n-2)\boldsymbol{u}^T \boldsymbol{u}}}\right). \qquad \square$$

Каждый из нас занимается статистикой всю свою жизнь в том смысле, что каждый из нас со дня рождения то и дело извлекает выводы из эмпирических наблюдений.

В. Крускал (1919—), американский статистик

* * *

Мы завершаем этот том историей о Ф. Иэйтсе (1902—1994), выдающемся британском статистике и сподвижнике Фишера (история взята из [Wi, с. 204—205]). В студенческие годы, проведенные в колледже Святого Иоанна в Кембридже, Иэйтс увлекался одним местным видом спорта, передававшимся от поколения к поколению в течение длительного времени. Состоял этот вид спорта в умении взбираться на крыши и башни зданий колледжей в ночное время. (Ощущение было тем более захватывающим, что при этом нужно было не только преодолеть трудности такого восхождения, но и ускользнуть от бдительного ока смотрителей колледжа.) Массивная башня часовни в неоготическом стиле колледжа Святого Иоанна была, в частности, украшена статуями святых, и Иэйтсу было очевидно, что статуи только выиграют, если должным образом нарядить этих святых в стихари (церковную одежду). Однажды ночью он взобрался наверх и осуществил эту затею; на следующее утро результат привел всех в восторг. Однако администрация колледжа не оценила этого и начала искать пути, как бы освободить святых от новообретенных одежд. Это оказалось нелегко — до статуй нельзя было добраться с помощью обычных лестниц. Попытки сбросить стихари вниз, используя веревки с крюками на концах, не увенчались успехом, так как Иэйтс, предвидя такую возможность, закрепил стихари кусками проволоки в виде петли вокруг шеи каждой статуи. Дело не двигалось с места, и тогда наконец Иэйтс вышел вперед и, не признаваясь в том, что это он водрузил стихари на святых, вызвался взобраться наверх при свете дня и снять их. Что он и сделал, вызвав восхищение собравшейся толпы.

Мораль этой истории в том, что, возможно, статистика способна не только обслуживать интересы других наук, но самостоятельно ставить проблемы, которые вызывают живейший интерес широкой публики... (Наблюдательный прохожий заметит, что в настоящее время у двух статуй на башне часовни колледжа Святого Иоанна посохи окрашены в бледно-зеленый цвет, что, очевидно, не соответствует первоначально задуманному убранству. Возможно, так практиковались представители следующего поколения последователей школы Фишера, прежде чем внести свой вклад в развитие статистики XXI века.)

Таблицы случайных величин и вероятностных распределений

Некоторые полезные дискретные распределения

Семейство Обозначение Область определения	Вероятности $\mathsf{P}(X=r)$	Среднее	Дисперсия	Производящая функция $\mathsf{E}s^X$
Пуассона $\mathrm{Po}(\lambda)$ $0, 1, \dots$	$\dfrac{\lambda^r e^{-\lambda}}{r!}$	λ	λ	$e^{\lambda(s-1)}$
Геометрическое $\mathrm{Geom}(p)$ $0, 1, 2, \dots$	$p(1-p)^r$	$\dfrac{1-p}{p}$	$\dfrac{1-p}{p^2}$	$\dfrac{p}{1-s(1-p)}$
Биномиальное $\mathrm{Bin}(n, p)$ $0, \dots n$	$C_n^r p^r (1-p)^{n-r}$	np	$np(1-p)$	$[ps+(1-p)]^n$
Отрицательное биномиальное $\mathrm{NegBin}(p, k)$ $0, 1, \dots$	$C_{r+k-1}^r p^k (1-p)^r$	$\dfrac{k(1-p)}{p}$	$\dfrac{k(1-p)}{p^2}$	$\left[\dfrac{p}{1-s(1-p)}\right]^k$
Гипергеометрическое $\mathrm{Hyp}(N, D, n)$ $(n+D-N)_+, \dots, D\wedge n$	$\dfrac{C_D^r C_{N-D}^{n-r}}{C_N^n}$	$\dfrac{nD}{N}$	$\dfrac{nD(N-D)(N-n)}{N^2(N-1)}$	${}_2F_1\left(\begin{matrix}-D & -n \\ -N & 1-s\end{matrix}\right)$
Равномерное $\mathrm{U}[1, n]$ $1, \dots n$	$\dfrac{1}{n}$	$\dfrac{n+1}{2}$	$\dfrac{n^2-1}{12}$	$\dfrac{s(1-s^n)}{n(1-s)}$

Некоторые полезные непрерывные распределения, 1

Семейство Обозначение Область определения	Плотность $f_X(x)$	Среднее	Дисперсия	П. ф. м. $\mathsf{E}e^{tX}$ Х. ф. $\mathsf{E}e^{i}tX$
Равномерное $\mathrm{U}(a, b)$ (a, b)	$\dfrac{1}{b-a}$	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$	$\dfrac{e^{bt}-e^{at}}{(b-a)t}$ $\dfrac{e^{ibt}-e^{iat}}{(b-a)t}$

(окончание на следующей странице)

(*окончание*)

Показательное (экспоненциальное) $\mathrm{Exp}(\lambda)$ $\mathbb{R}_+$	$\lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{\lambda}{\lambda-t}$ $\frac{\lambda}{\lambda-it}$
Гамма $\mathrm{Gam}(\alpha, \lambda)$ $\mathbb{R}_+$	$\frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$	$(1-t/\lambda)^{-\alpha}$ $(1-it/\lambda)^{-\alpha}$
Нормальное $\mathrm{N}(\mu, \sigma^2)$ $\mathbb{R}$	$\frac{\exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right]}{\sqrt{2\pi\sigma^2}}$	μ	σ^2	$\exp\left(t\mu+\frac{1}{2}t^2\sigma^2\right)$ $\exp\left(it\mu-\frac{1}{2}t^2\sigma^2\right)$
Многомерное нормальное $\mathrm{N}(\mu, \Sigma)$ $\mathbb{R}^n$	$\frac{\exp\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]}{\sqrt{(2\pi)^n \det\Sigma}}$	μ	Σ	$\exp\left(t^T\mu+\frac{1}{2}t^T\Sigma t\right)$ $\exp\left(it^T\mu-\frac{1}{2}t^T\Sigma t\right)$
Коши $\mathrm{Ca}(\alpha, \tau)$ $\mathbb{R}$	$\frac{\tau}{\pi(\tau^2+(x-\alpha)^2)}$	не определено	не определено	$\mathrm{E}(e^{tX})=\infty$ при $t\neq 0$ $\mathrm{E}(e^{itX})=\exp(it\alpha-\lvert t\rvert\tau)$

Некоторые полезные непрерывные распределения, 2

Семейство Обозначение	Область	Распределение	Примечания
Хи-квадрат χ^2_n	$\mathbb{R}^+$	$\chi^2_n \sim \sum_1^n \mathrm{N}(0, 1)^2$	$\chi^2_n \sim \mathrm{Gam}\left(\frac{n}{2}, \frac{1}{2}\right)$
Стьюдента t_n	$\mathbb{R}$	$\mathrm{t}_n \sim \frac{\mathrm{N}(0, 1)}{\sqrt{\chi^2_n/n}}$	$\mathrm{t}_1 \sim \mathrm{Ca}(0, 1)$; $\mathrm{E}X=0$, $n>1$; $\mathrm{Var}\,X=\frac{n}{n-2}$, $n>2$
Фишера $\mathrm{F}_{m,n}$	$\mathbb{R}^+$	$\mathrm{F}_{m,n} \sim \frac{\chi^2_m/m}{\chi^2_n/n}$	$\mathrm{E}X=\frac{n}{n-2}$, $n>2$; $\mathrm{Var}\,X=\frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$, $n>4$
Вейбулла $\mathrm{Weib}(\alpha)$	$\mathbb{R}^+$	$f(x)=\alpha x^{\alpha-1}e^{-x^\alpha}$	$\mathrm{Weib}(\alpha) \sim \mathrm{Exp}(1)^{1/\alpha}$
Бета $\mathrm{Bet}(r, s)$	$[0, 1]$	$f(x)=\frac{x^{r-1}(1-x)^{s-1}}{\mathrm{B}(r, s)}$	$\mathrm{Bet}(r, s) \sim \left(1+\frac{\mathrm{Gam}(s, \lambda)}{\mathrm{Gam}(r, \lambda)}\right)^{-1}$; $\mathrm{E}X=\frac{r}{r+s}$
Логистическое Logist	$\mathbb{R}$	$f(x)=\frac{e^x}{(1+e^x)^2}$	$\mathrm{Logist} \sim \log(\mathrm{Par}(1)-1)$; $\mathrm{E}(e^{tX})=\mathrm{B}(1+t, 1-t)$
Парето $\mathrm{Par}(\alpha)$	$[1, \infty)$	$f(x)=\alpha x^{-(\alpha+1)}$	$\mathrm{Par}(\alpha) \sim \mathrm{Bet}(\alpha, 1)^{-1}$; $\mathrm{E}X=\infty$, $\alpha\leqslant 1$; $\mathrm{Var}\,X=\infty$, $\alpha\leqslant 2$

Список литературы

[A] *D. Applebaum.* Probability and Information: an Integrated Approach. Cambridge: Cambridge University Press, 1996.

[ArBN] *B. C. Arnold, N. Balakrishnan, H. N. Nagaraja.* A First Course in Order Statistics. New York: Wiley, 1992.

[As] *R. B. Ash.* Probability and Measure Theory / With contributions from C. Doléans-Dade. 2nd ed. San Diego, CA: Harcourt/Academic; London: Academic, 2000.

[AuC] *J. Auñón, V. Chandrasekar.* Introduction to Probability and Random Processes. New York—London: McGraw-Hill, 1997.

[Az] *A. Azzalini.* Statistical Inference: Based on the Likelihood. London: Chapman and Hall, 1996.

[BE] *L. J. Bain, M. Engelhardt.* Introduction to Probability and Mathematical Statistics. 2nd ed. Boston, MA: PWS-KENT, 1992.

[BaN] *N. Balakrishnan, V. B. Nevzorov.* A Primer on Statistical Distributions. Hoboken, New York—Chichester: Wiley, 2003.

[BarC] *O. E. Barndorff-Nielsen, D. R. Cox.* Inference and Asymptotics. London: Chapman and Hall, 1994.

[BartN] *R. Bartoszynski, M. Niewiadomska-Bugaj.* Probability and Statistical Inference. New York—Chichester: Wiley, 1996.

[Be] *J. O. Berger.* Statistical Decision Theory and Bayesian Analysis. 2nd ed. New York: Springer, 1985.

[BerL] *D. A. Berry, B. W. Lindgren.* Statistics, Theory and Methods. Pacific Grove, CA: Brooks/Cole, 1996.

[Bi] *P. Billingsley.* Probability and Measure. 3rd ed. New York—Chichester: Wiley, 1995. Перев. на рус. яз.: *П. Биллингсли.* Сходимость вероятностных мер. М.: Наука, 1977.

[Bio] Biometrika Tables for Statisticians / Editors E. Pearson, H. O. Hartley. Vol. 1. Cambridge: Cambridge University Press, 1966.

[Bo] *V. S. Borkar.* Probability Theory: an Advanced Course. New York—London: Springer, 1995.

[Bor] *А. А. Боровков.* Математическая статистика. М.: Наука, 1984.

[BD] *G. E. P. Box, N. R. Draper.* Evolutionary Operation: a Statistical Method for Process Improvement. New York—Chichester: Wiley, 1998.

[BL] *G.E.P.Box, A.Luceño.* Statistical Control by Monitoring and Feedback Adjustment. New York—Chichester: Wiley, 1997.

[BTi] *G.E.P.Box, G.C.Tiao.* Bayesian Inference in Statistical Analysis. New York: Wiley, 1992.

[Box] Box on Quality and Discovery: with Design, Control, and Robustness / Ed.-in-chief G. C. Tiao; Editors S. Bisgaard et al. New York—Chichester: Wiley, 2000.

[CK] *M.Capinski, E.Kopp.* Measure, Integral and Probability. London: Springer, 1999.

[CZ] *M.Capinski, T.Zastawniak.* Probability through Problems. New York: Springer, 2000.

[CaB] *G.Casella, J.O.Berger.* Statistical Inference. Pacific Grove, CA: Brooks/Cole, 1990.

[CaL] *G.Casella, E.L.Lehmann.* Theory of Point Estimation. New York—London: Springer, 1998.

[ChKB] *C.A.Charalambides, M.V.Koutras N.Balakrishnan.* Probability and Statistical Models with Applications. Boca Raton, FL—London: Chapman and Hall/CRC, 2001.

[ChaHS] *S.Chatterjee, M.S.Handcock, J.S.Simonoff.* A Casebook for a First Course in Statistics and Data Analysis. New York—Chichester: Wiley, 1995.

[ChaY] *L.Chaumont, M.Yor.* Exercises in Probability. A Guided Tour from Measure Theory to Random Processes, via Conditioning. Cambridge: Cambridge University Press, 2003.

[ChoR] *Y.S.Chow, H.Robbins, D.Siegmund.* The Theory of Optimal Stopping. New York—Dover—London: Constable, 1991. Перев. на рус. яз.: *Г.Роббинс, Д.Сигмунд, И.Чао.* Теория оптимальных правил остановки. М.: Наука, 1977.

[ChoT] *Y.S.Chow, H.Teicher.* Probability Theory: Independence, Interchangeability, Martingales. 3rd ed. New York—London: Springer, 1997.

[Chu] *K.L.Chung.* A Course in Probability Theory. 3rd ed. San Diego, CA—London: Academic Press, 2001.

[ClC] *G.M.Clarke, D.Cooke.* A Basic Course in Statistics. 4th ed. London: Arnold, 1998.

[CoH1] *D.R.Cox, D.V.Hinkley.* Theoretical Statistics. London: Chapman and Hall, 1979. Перев. на рус. яз.: *Д.Кокс, Л.Хинкли.* Теоретическая статистика. М.: Мир, 1978.

[CoH2] *D.R.Cox, D.V.Hinkley.* Problems and Solutions in Theoretical Statistics. London: Chapman and Hall, 1978. Перев. на рус. яз.: *Д.Кокс, Л.Хинкли.* Задачи по теоретической статистике с решениями. М.: Мир, 1981.

[CouR] *R.Courant, H.Robbins.* What Is Mathematics? An Elementary Approach to Ideas and Methods. Oxford: Oxford University Press, 1996. Перев. на рус. яз.: *Р.Курант, Г.Роббинс.* Что такое математика? Элементарный очерк идей и методов. Изд. 2-е. М.: Просвещение, 1967. Изд. 3-е, испр. и доп. М.: МЦНМО, 2001.

[CrC] *J.Crawshaw, J.Chambers.* A Concise Course in Advanced Level Statistics: with Worked Examples. 4th ed. Cheltenham: Nelson Thornes, 2001.

[DS] *M. N. DeGroot, M. J. Schervish.* Probability and Statistics. 3rd ed. Boston, MA—London: Addison-Wesley, 2002.

[De] *J. L. Devore.* Probability and Statistics for Engineering and the Sciences. 4th ed. Belmont, CA—London: Duxbury, 1995.

[DorSSY] *А. Я. Дороговцев, Д. С. Сильвестров, А. В. Скороход, М. Й. Ядренко.* Теория вероятностей: Сборник задач. К.: «Выща школа», 1980.

[Du] *R. M. Dudley.* Real Analysis and Probability. Cambridge: Cambridge University Press, 2002.

[Dur1] *R. Durrett.* The Essentials of Probability. Belmont, CA: Duxbury Press, 1994.

[Dur2] *R. Durrett.* Probability: Theory and Examples. 2nd ed. Belmont, CA—London: Duxbury Press, 1996.

[E1] *B. Efron.* R. A. Fisher in the 21st century // Statistical Science. 1998. V. 13. P. 95—122.

[E2] *B. Efron.* Robbins, empirical Bayes and microanalysis // Annals of Statistics. 2003. V. 31. P. 366—378.

[ETh1] *B. Efron, R. Thisted.* Estimating the number of unseen species: How many words did Shakespeare know? // Biometrika. 1976. V. 63. P. 435—447.

[ETh2] *B. Efron, R. Thisted.* Did Shakespeare write a newly-discovered poem? // Biometrika. 1987. V. 74. P. 445—455.

[ETi] *B. Efron, R. Tibshirani.* An Introduction to the Bootstrap. New York; London: Chapman and Hall, 1993.

[EHP] *M. Evans, N. Hastings, B. Peacock.* Statistical Distributions. 3rd ed. New York—Chichester: Wiley, 2000.

[FV] *R. M. Feldman, C. Valdez-Flores.* Applied Probability and Stochastic Processes. Boston, MA—London: PWS Publishing Company, 1996.

[Fe] *W. Feller.*An Introduction to Probability Theory and Its Applications. V. 1, 2. 2nd ed. New York: Wiley; London: Chapman and Hall, 1957—1971. Перев. на рус. яз.: *В. Феллер.* Введение в теорию вероятностей и ее приложения. Т. 1, 2. М.: Мир, 1984.

[FiB] *J. Fisher Box.* R. A. Fisher. The Life of a Scientist. New York—Chichester—Brisbane—Toronto: Wiley, 1978.

[Fer] *T. S. Ferguson.* Mathematical Statistics: a Decision Theoretic Approach. New York—London: Academic Press, 1967.

[Fr] *J. E. Freund.* Introduction to Probability. New York: Dover, 1993.

[FreP] *J. E. Freund, B. M. Perles.* Statistics: a First Course. 7th ed. Upper Saddle River, NJ: Prentice Hall; London: Prentice-Hall International, 1999.

[FriG] *B. Fristedt, L. Gray.* A Modern Approach to Probability Theory. Boston, MA: Birkhäuser, 1997.

[G] *J. Galambos.* Advanced Probability Theory. 2nd ed., rev. and expanded. New York: M. Dekker, 1995.

[Gh] *S. Ghahramani.* Fundamentals of Probability. 2nd ed. International ed. Upper Saddle River, NJ: Pearson/Prentice-Hall, 2000.

[Gi] *J. D. Gibbons.* Nonparametric Statistics: an Introduction. Newbury Park, CA—London: Sage, 1993.

[Gn] *Б. В. Гнеденко*. Курс теории вероятностей. 6-е изд. М.: Наука, 1988.

[Go] *H. Gordon*. Discrete Probability. New York—London: Springer, 1997.

[Gr] *A. Graham*. Statistics: an Introduction. London: Hodder and Stoughton, 1995.

[GriS1] *G. R. Grimmett, D. R. Stirzaker*. Probability and Random Processes: Problems and Solutions. Oxford: Clarendon Press, 1992.

[GriS2] *G. R. Grimmett, D. R. Stirzaker*. Probability and Random Processes. 3rd ed. Oxford: Oxford University Press, 2001.

[GriS3] *G. R. Grimmett, D. R. Stirzaker*. One Thousand Exercises in Probability. Oxford: Oxford University Press, 2003.

[GrinS] *C. M. Grinstead, J. L. Snell*. Introduction to Probability. 2nd rev. ed. Providence, RI: American Mathematical Society, 1997.

[H] *J. Haigh*. Probability Models. London: Springer, 2002.

[Ha] *A. Hald*. A History of Mathematical Statistics from 1750 to 1930. New York—Chichester: Wiley, 1998.

[Has] *K. J. Hastings*. Probability and Statistics. Reading, MA—Harlow: Addison-Wesley, 1997.

[Haw] *A. G. Hawkes*. On the development of statistical ideas and their applications: Inaugural lecture at the University of Wales, Swansea, 1975.

[Hay] *A. J. Hayter*. Probability and Statistics for Engineers and Scientists. Boston, MA—London: PWS, 1996.

[He] *L. L. Helms*. Introduction to Probability Theory: with Contemporary Applications. New York: W. H. Freeman, 1997.

[Ho] *J. Hoffmann-Jorgensen*. Probability with a View toward Statistics. New York—London: Chapman and Hall, 1994.

[HogT] *R. V. Hogg, E. A. Tanis*. Probability and Statistical Inference. 6th ed. Upper Saddle River, NJ: Prentice Hall International, 2001.

[JP] *J. Jacod, P. Protter*. Probability Essentials. 2nd ed. Berlin—London: Springer, 2003.

[JaC] *R. Jackson, J. T. Callender*. Exploring Probability and Statistics with Spreadsheets. London: Prentice Hall, 1995.

[Je] *H. Jeffreys*. Theory of Probability. 3rd ed. Oxford: Oxford University Press, 1998.

[K] *O. Kallenberg*. Foundations of Modern Probability. 2nd ed. New York—London: Springer, 2002.

[KSR] *M. Kelbert, I. Sazonov, A. G. Wright*. Exact expression for the variance of the photon emission process in scintillation counters // Nucl. Instruments and Methods in Phys. Research, ser. A. 2006. V. 564, № 1. P. 185—189.

[Ki] *S. Kim*. Statistics and Decisions: an Introduction to Foundations. New York: Van Nostrand Reinhold; London: Chapman and Hall, 1992.

[Kinn] *J. J. Kinney*. Probability: an Introduction with Statistical Applications. New York—Chichester: Wiley, 1997.

[Kit] *L. J. Kitchens*. Exploring Statistics: a Modern Introduction to Data Analysis and Inference. 2nd ed. Pacific Grove, CA: Duxbury Press, 1998.

[KlR] *D. A. Klain, G.-C. Rota*. Introduction to Geometric Probability. Cambridge: Cambridge University Press, 1997.

[Ko] *А. Н. Колмогоров*. Основные понятия теории вероятностей. М.—Л.: ОНТИ, 1936; 2-е изд. М.: Наука, 1974; 3-е изд. М.: Фазис, 1998.

[KS] *L. B. Koralov, Y. G. Sinai*. Theory of Probability and Random Processes. 2nd ed. New York: Springer-Verlag, 2006.

[Kr] *N. Krishnankutty*. Putting Chance to Work. A Life in Statistics. A biography of C. R. Rao. State College, PA: Dialogue, 1996.

[LS] *T. L. Lai, D. Siegmund*. The contributions of Herbert Robbins to Mathematical Statistics // Statistical Science. 1986. V. 1. P. 276—284.

[La] *J. W. Lamperti*. Probability: a Survey of the Mathematical Theory. 2nd ed. New York—Chichester: Wiley, 1996. Перев. на рус. яз.: *Дж. Ламперти*. Вероятность. М.: Наука, 1973.

[Lan] *W. H. Lange*. Study Guide for Mason, Lind and Marchal's Statistics: an Introduction. 5th ed. Pacific Grove, CA—London: Duxbury Press, 1998.

[LarM] *R. J. Larsen, M. L. Marx*. An Introduction to Mathematical Statistics and its Applications. 3rd ed. Upper Saddle River, NJ: Prentice Hall; London: Prentice Hall International, 2001.

[LawC] *G. Lawler, L. N. Coyle*. Lectures on Contemporary Probability. Providence, RI: American Mathematical Society, 1999.

[Le] *P. M. Lee*. Bayesian Statistics: an Introduction. 3rd ed. London: Arnold, 2004.

[Leh1] *E. L. Lehmann*. Testing Statistical Hypotheses. 2nd ed. New York—London: Springer, 1997. Перев. на рус. яз.: *Э. Леман*. Проверка статистических гипотез. 2-е изд., испр. М.: Наука, 1979.

[Leh2] *E. L. Lehmann*. Theory of point estimation. New York: Wiley, 1983. Перев. на рус. яз.: *Э. Леман*. Теория точечного оценивания. М.: Наука, 1991.

[LiS] *D. V. Lindley, W. F. Scott*. New Cambridge Statistical Tables. 2nd ed. Cambridge: Cambridge University Press, 1995.

[M] *P. Malliavin*. Integration and Probability / In cooperation with H. Airault, L. Kay and G. Letac. New York—London: Springer-Verlag, 1995.

[MaLM1] *R. D. Mason, D. A. Lind, W. G. Marchal*. Statistics: an Introduction. 5th ed. Pacific Grove, CA—London: Duxbury Press, 1998.

[MaLM2] *R. D. Mason, D. A. Lind, W. G. Marchal*. Instructor's Manual for Statistics: an Introduction. 5th ed. Pacific Grove, CA—London: Duxbury Press, 1998.

[McClS] *J. T. McClave, T. Sincich*. A First Course in Statistics. 7th ed. Upper Saddle River, NJ: Prentice Hall; London: Prentice-Hall International, 2000.

[McCo] *J. H. McColl*. Probability. London: Edward Arnold, 1995.

[MeB] *W. Mendenhall, R. J. Beaver, B. M. Beaver*. Introduction to Probability and Statistics. 10th ed. Belmont, CA—London: Duxbury, 1999.

[MiM] *I. Miller, M. Miller*. John Freund's Mathematical Statistics with Applications. 7th ed. Upper Saddle River, NJ: Pearson Prentice-Hall Education, 2004.

[MT] *F. Mosteller, J. W. Tukey*. Data Analysis and Regression: a Second Course in Statistics. Reading, MA—London: Addison-Wesley, 1977. Перев. на рус. яз.: *Ф. Мостеллер, Дж. Тьюки*. Анализ данных и регрессия: В 2-х вып. М.: Финансы и статистика, 1982.

[N] E. Nelson. *Radically Elementary Probability Theory*. Princeton, NJ: Princeton University Press, 1987.

[NiFY] *Е.П. Никитина, В.Д. Фрейдлина, А.В. Ярхо*. Коллекция определений термина «статистика». М.: Изд-во Московского государственного университета им. М.В. Ломоносова, 1972. (Межфакультетская лаборатория статистических методов; Вып. 37).

[OF] *A. O'Hagan, J. Forster*. Kendall's Advanced Theory of Statistics. Vol. 2B. London: Hodder Arnold, 2004.

[Oc] *M.K. Ochi*. Applied Probability and Stochastic Processes: in Engineering and Physical Sciences. New York—Chichester: Wiley, 1990.

[OtL] *R.L. Ott, M. Longnecker*. An Introduction to Statistical Methods and Data Analysis. 5th ed. Australia—United Kingdom: Duxbury, 2001.

[Ox] The Oxford Dictionary of Statistical Terms / Editors Y. Dodge et al. 6th ed. Oxford: Oxford University Press, 2003.

[Pi] *J. Pitman*. Probability. New York: Springer-Verlag, 1993.

[Pol] *D. Pollard*. A User's Guide to Measure Theoretic Probability. Cambridge: Cambridge University Press, 2002.

[Por] *S.C. Port*. Theoretical Probability for Applications. New York—Chichester: Wiley, 1994.

[Ra] *C.R. Rao*. Linear Statistical Inference and Its Applications. 2nd ed. New York: Wiley, 2002. Перев. на рус. яз.: *С.Р. Рао*. Линейные статистические методы и их применения. М.: Наука, 1968.

[R] *S. Rasmussen*. An Introduction to Statistics with data Analysis. International student ed. Pacific Grove, CA: Brooks/Cole, 1992.

[Re] *A.G. Rencher*. Linear Models in Statistics. New York—Chichester: Wiley, 2000.

[Rei] *C. Reid*. Neyman — from Life. New York—Hedelberg—Berlin: Springer-Verlag, 1982.

[Res] *S.I. Resnick*. A Probability Path. Boston, MA: Birkhäuser, 1999.

[Ri] *J.A. Rice*. Mathematical Statistics and Data Analysis. 2nd ed. Pacific Grove, CA—London: Duxbury Press, 1995.

[Rob] *G. Robbins*. A remark on Stirling's formula // Amer. Math. Monthly. 1955. V. 62. P. 26—29.

[Ro] *J. Rosenblat*. Basic Statistical Methods and Models for the Sciences. Boca Raton, FL—London: Chapman and Hall/CRC, 2002.

[Ros1] *S.M. Ross*. Solutions Manual for Introduction to Probability Models. 4th ed. Boston, MA: Academic, 1989.

[Ros2] *S.M. Ross*. Introductory Statistics. New York—London: McGraw-Hill, 1996.

[Ros3] *S.M. Ross*. Introduction to Probability Models. 7th ed. San Diego, CA—London: Harcourt/Academic, 2000.

[Ros4] *S.M. Ross*. Introduction to Probability and Statistics for Engineers and Scientists. 2nd ed. San Diego, CA—London: Harcourt/Academic, 2000.

[Ros5] *S. Ross*. A First Course in Probability. 6th ed. Upper Saddle River, NJ—London: Prentice Hall, 2002.

[Ros6] *S. M. Ross*. Probability Models for Computer Science. San Diego, CA—London: Harcourt Academic Press, 2002.

[RotE] *V. K. Rothagi, S. M. Ehsanes*. An Introduction to Probability and Statistics. 2nd ed. New York—Chichester: Wiley, 2001.

[Rou] *G. G. Roussas*. A Course in Mathematical Statistics. 2nd ed. San Diego, CA—London: Academic, 1997.

[Roy] *R. M. Royall*. Statistical Evidence: a Likelihood Paradigm. London: Chapman and Hall, 1997.

[S] *R. L. Scheaffer*. Introduction to Probability and Its Applications. 2nd ed. Belmont, CA—London: Duxbury, 1995.

[Sc] *R. B. Schinazi*. Probability with Statistical Applications. Boston, MA: Birkhäuser, 2001.

[SeS] *P. K. Sen, J. M. Singer*. Large Sample Methods in Statistics: an Introduction with Applications. New York—London: Chapman and Hall, 1993.

[Sh] *А. Н. Ширяев*. Вероятность: В 2-х кн. 3-е изд. М.: МЦНМО, 2004.

[SiM] *A. F. Siegel, C. J. Morgan*. Statistics and Data Analysis: an Introduction. 2nd ed. New York—Chichester: Wiley, 1996.

[Sin] *Я. Г. Синай*. Курс теории вероятностей. М.: Изд-во МГУ, 1985. 2-е изд. 1986.

[SpSS] *M. R. Spiegel, J. J. Schiller, R. A. Srinivasan*. Schaum's Outline of Theory and Problems of Probability and Statistics. 2nd ed. New York—London: McGraw-Hill, 2000.

[St1] *D. Stirzaker*. Elementary Probability. Cambridge: Cambridge University Press, 1994.

[St2] *D. Stirzaker*. Solutions Manual for Stirzaker's Elementary Probability. Cambridge: Cambridge University Press, 1994.

[St3] *D. Stirzaker*. Probability and Random Variables: a Beginner's Guide. New York: Cambridge University Press, 1999.

[St4] *D. Stirzaker*. Probability Vicit Expectation // Probability, Statistics and Optimization. Chichester: Wiley, 1994. P. 31—41.

[Sto] *J. M. Stoyanov*. Counterexamples in Probability. 2nd ed. Chichester: Wiley, 1997. Перев. на рус. яз.: *Й. Стоянов*. Контрпримеры в теории вероятностей. М.: Факториал, 1999.

[Str] *D. W. Strook*. Probability Theory: an Analytic View. Cambridge: Cambridge University Press, 1993.

[T] *K. Trivedi*. Probability and Statistics with Reliability, Queueing, and Computer Science Applications. 2nd ed. New York—Chichester: Wiley, 2002.

[Tu] *H. C. Tuckwell*. Elementary Applications of Probability Theory: with an Introduction to Stochastic Differential Equations. 2nd ed. London: Chapman and Hall, 1995.

[Tuk] *J. W. Tukey*. Exploratory Data Analysis. Reading, MA: Addison-Wesley, 1977. Перев. на рус. яз.: *Дж. Тьюки*. Анализ результатов наблюдений. Разведочный анализ. М.: Мир, 1981.

[WMM] *R. E. Walpole, R. H. Myers, S. H. Myers.* Probability and Statistics for Engineers and Scientists. 6th ed. Upper Saddle River, NJ—London: Prentice Hall International, 1998.

[We] *R. Weber.* IB Statistics: Lecture Notes, 2003. Электронная версия доступна по адресу `http://www.statslab.cam.ac.uk/Dept/People/weber.html`.

[Wh] *P. Whittle.* Probability via Expectation. 4th ed. New York: Springer, 2000. Перев. на рус. яз.: *П. Уиттл.* Вероятность. М.: Наука, 1982.

[Wi] *M. V. Wilkes.* Memoirs of a Computer Pioneer. Cambridge, MA: MIT Press, 1985.

[Wil] *W. F. Willcox.* Definitions of Statistics // ISI Rev. 1935. V. 3, № 4. P. 388—399.

[Will] *D. Williams.* Weighing the Odds: a Course in Probability and Statistics. Cambridge: Cambridge University Press, 2001.

[WisH] *G. L. Wise, E. B. Hall.* Counterexamples in Probability and Real Analysis. New York—Oxford: Oxford University Press, 1993.

Предметный указатель

Кельберт Марк Яковлевич
Сухов Юрий Михайлович

ВЕРОЯТНОСТЬ И СТАТИСТИКА В ПРИМЕРАХ И ЗАДАЧАХ

Том I. Основные понятия теории вероятностей
и математической статистики

Подписано в печать 04.06.2007 г. Формат 60 × 90 1/16. Бумага офсетная № 1.
Печать офсетная. Печ. л. 28,5. Тираж 1000 экз. Заказ № 1104

Издательство Московского центра непрерывного математического образования
119002, Москва, Большой Власьевский пер., д. 11.

Отпечатано с готовых диапозитивов в ППП «Типография „Наука“».
121099, Москва, Шубинский пер., д. 6.

Книги издательства МЦНМО можно приобрести в магазине «Математическая книга»,
Большой Власьевский пер., д. 11. Тел. (495) 241-72-85. E-mail: biblio@mccme.ru