БАЙЕСОВСКАЯ СТАТИСТИКА

STAR WARS®, LEGO®,
РЕЗИНОВЫЕ УТОЧКИ
И МНОГОЕ ДРУГОЕ

УИЛЛ КУРТ

no starch press

BAYESIAN STATISTICS THE FUN WAY

Understanding Statistics and Probability with Star Wars®, LEGO®, and Rubber Ducks

by Will Kurt

no starch press

San Francisco

УИЛЛ КУРТ

БАЙЕСОВСКАЯ СТАТИСТИКА

STAR WARS®, LEGO®,
РЕЗИНОВЫЕ УТОЧКИ
И МНОГОЕ ДРУГОЕ

ПИТЕР®

Санкт-Петербург • Москва • Минск

2021

ББК 32.973.2-018.1+22.172
УДК 004.43:519.226.3
К93

Курт Уилл

К93 Байесовская статистика: Star Wars®, LEGO®, резиновые уточки и многое другое. — СПб.: Питер, 2021. — 304 с.: ил. — (Серия «Библиотека программиста»).

ISBN 978-5-4461-1655-3

Нужно решить конкретную задачу, а перед вами куча непонятных данных, в которой черт ногу сломит? «Байесовская статистика» расскажет, как принимать правильные решения, задействуя свою интуицию и простую математику.

Пора забыть про заумные и занудные университетские лекции! Эта книга даст вам полное понимание байесовской статистики буквально «на пальцах» — с помощью простых объяснений и ярких примеров.

Чтобы узнать, как применить байесовские подходы к реальной жизни, вы отправитесь на охоту за НЛО, поиграете в «Лего», рассчитаете вероятность выживания Хана Соло при полете через поле астероидов, а также узнаете, как оценить вероятность того, что вы не заболели (ковидом?!), несмотря на то, что нагуглили все симптомы родильной горячки.

Прикладные задачи и упражнения помогут закрепить материал и заложить фундамент для работы с широким спектром задач: от невероятных текущих событий до ежедневных сюрпризов делового мира.

16+ (В соответствии с Федеральным законом от 29 декабря 2010 г. № 436-ФЗ.)

ББК 32.973.2-018.1+22.172
УДК 004.43:519.226.3

ISBN 978-1593279561 англ.

ISBN 978-5-4461-1655-3

Краткое содержание

Часть III. Оценка параметров

Часть IV. Проверка гипотез: сердце статистики

Приложения

Оглавление

Часть I. Введение в теорию вероятностей

Часть III. Оценка параметров

Часть IV. Проверка гипотез: сердце статистики

Приложения

Мелани, которая пробудила во мне страсть к писательству

Об авторе

Уилл Курт (Will Kurt) — специалист по данным в Wayfair, уже больше пяти лет использует байесовскую статистику для решения реальных задач бизнеса. Он часто пишет о вероятности на своем веб-сайте CountBayesie.com. Курт является автором «Программируй на Haskell» (изд. «ДМК Пресс»), в настоящее время живет в Бостоне, штат Массачусетс.

О научном редакторе

Челси Парлетт-Пеллерити (Chelsea Parlett-Pelleriti) — аспирантка в области науки о вычислениях и data science и давняя любительница всего забавного и статистического. Автор текстов о статистике, участница различных проектов, включая серию «Краткий курс статистики» (*Crash Course Statistics*) на YouTube и «Подготовка к экзамену по статистике (углубленная программа)» (*Cracking the AP Statistics Exam*) в The Princeton Review. В настоящее время живет в Южной Калифорнии.

Благодарности

Написание книги — это действительно невероятное усилие, которое складывается из труда многих людей. Ниже я упомянул лишь некоторых людей, благодаря которым появилась эта книга. Хотел бы начать с благодарности сыну Арчеру за то, что он всегда сохранял во мне любопытство и вдохновлял меня.

Я очень люблю книги No Starch, и для меня большая честь поработать вместе с удивительной командой, которая теперь издает и мою книгу. Я очень благодарен своим редакторам, рецензентам и невероятной команде No Starch. Лиз Чедвик (Liz Chadwick) в самом начале обратилась ко мне с просьбой о создании этой книги и давала рекомендации на протяжении всего процесса ее написания. Лорел Чан (Laurel Chun) позаботилась о том, чтобы весь процесс перехода от черновых записей с заметками по R к полноценной книге прошел гладко. Челси Парлетт-Пеллерити вышла далеко за рамки требований научного редактора и постаралась сделать эту книгу как можно лучше. Фрэнсис Соу (Frances Saux) добавила много важных комментариев к последующим главам.

Отдельная благодарность Биллу Поллоку (Bill Pollock) за создание такого восхитительного издательства.

Будучи специалистом по английской литературе, я и представить себе не мог, что напишу книгу, связанную с математикой. Хотел бы упомянуть нескольких людей, которые помогли мне увидеть чудо математики. Я всегда буду благодарен моему соседу по комнате в колледже Грегу Мюллеру (Greg

Muller), который показал сумасшедшему знатоку английской филологии, насколько захватывающим и интересным может быть мир математики. Профессор Анатолий Темкин (Anatoly Temkin) из Бостонского университета открыл мне двери в математическое мышление, научив меня всегда отвечать на вопрос «что это значит». И, конечно же, огромное спасибо Ричарду Келли (Richard Kelley), который, когда я потерялся в пустыне, стал оазисом математических разговоров и подсказок. Также хотел бы поблагодарить команду по науке о данных в Bombora, особенно Патрика Келли (Patrick Kelley) — с ним мы провели множество увлекательных бесед, и некоторые из них нашли свое отражение в книге. Я также всегда буду благодарен читателям своего блога *Count Bayesie*, которые подбрасывали замечательные вопросы и идеи. Среди читателей я особенно хотел бы поблагодарить комментатора под ником Nevin, который помог исправить некоторые недоразумения, возникшие вначале.

Наконец хочу поблагодарить выдающихся авторов текстов о байесовской статистике, книги которых во многом способствовали моему собственному развитию в этой области. «Doing Bayesian Data Analysis» Джона Крушке (John Kruschke) и «Bayesian Data Analysis» Эндрю Гельмана (Andrew Gelman) — отличные книги, которые стоит прочитать каждому. Безусловно, книга, которая больше всего повлияла на мое мышление, — это феноменальная «Probability Theory: The Logic of Science» Э. Т. Джейнса (E. T. Jaynes). И еще я хотел бы поблагодарить Обри Клейтона (Aubrey Clayton) за серию лекций об этой сложной книге, которая действительно помогла мне разобраться в непонятных темах.

Введение

Практически все в жизни в некоторой степени неопределенно. Это может показаться преувеличением, но чтобы в этом убедиться, проведите быстрый эксперимент. В начале дня запишите то, что, по вашему мнению, произойдет в следующие полчаса, час, три часа и шесть часов. Затем проверьте, какие из этих пунктов осуществились именно так, как вы себе представляли. Вы быстро поймете, что ваш день полон неопределенностей. Даже что-то такое предсказуемое, как «я почищу зубы» или «я выпью чашку кофе», может по тем или иным причинам не произойти, вне зависимости от ваших ожиданий.

В большинстве случаев, даже несмотря на неопределенность, мы можем достаточно хорошо спланировать свой день. Например, даже если из-за пробок вы будете добираться на работу дольше, чем обычно, то можете довольно точно оценить, во сколько нужно выходить из дома, чтобы успеть. Если у вас очень важная утренняя встреча, можете выйти раньше, чтобы учесть возможные задержки. У всех нас есть врожденное чувство того, как справляться с неопределенными ситуациями и рассуждать о неопределенности. Когда вы думаете так, вы думаете *вероятностно*.

Зачем изучать статистику?

Байесовская статистика помогает лучше рассуждать о неопределенности, так же как изучение логики в школе помогает увидеть ошибки

в повседневном логическом мышлении. Учитывая, что практически каждый имеет дело с неопределенностью в жизни, о чем мы только что говорили, аудитория читателей этой книги становится довольно широкой. Специалисты по работе с данными и исследователи, уже использующие статистику, извлекут выгоду из более глубокого понимания и интуиции насчет работы этих инструментов. Инженеры и программисты узнают много нового о том, как лучше количественно оценивать решения, которые им приходится принимать (я даже использовал байесовский анализ для определения причин ошибок программного обеспечения!). Маркетологи и продавцы могут применять идеи, изложенные в этой книге, при проведении A/B-тестов, пытаясь понять свою аудиторию и лучше оценить возможные сделки. Любой, кто принимает решения на высоком уровне, должен иметь хотя бы базовое чувство вероятности, чтобы можно было быстро сделать предварительные оценки затрат и выгод от неопределенных решений. Я бы хотел, чтобы генеральный директор мог изучить эту книгу во время полета. К моменту приземления у него будет прочный фундамент в статистике, позволяющий лучше оценивать варианты, связанные с вероятностями и неопределенностью.

Я искренне верю, что всем будет полезно думать о проблемах байесовским способом. С помощью байесовской статистики вы можете использовать математику для моделирования неопределенности, чтобы сделать лучший выбор, учитывая ограниченную информацию. Допустим, вам нужно вовремя прийти на работу к особенно важной встрече и вы можете выбрать два разных маршрута. Первый маршрут обычно более короткий, но из-за оживленного движения могут возникнуть пробки. Второй маршрут в целом занимает больше времени, но пробок там не предвидится. Какой маршрут выбрать? Какого типа информация понадобится, чтобы принять решение? И насколько вы можете быть уверены в своем выборе? Даже небольшая добавленная сложность требует дополнительных размышлений и техники. Обычно, когда люди думают о статистике, они думают об ученых, работающих над новым лекарством, экономистах, следящих за тенденциями на рынке, аналитиках, предсказывающих следующие выборы, менеджерах по бейсболу, пытающихся создать лучшую команду, и т. д. Хотя все это, безусловно, увлекательное использование статистики, понимание основ байесовских рассуждений поможет в гораздо большем количестве областей жизни. Если вы когда-нибудь сомневались в новостях, не спали ночами и шерстили интернет в поисках ответа на вопрос «есть ли у вас редкое заболевание» или спорили с родственником по поводу их иррациональных

убеждений о мире, изучение байесовской статистики поможет рассуждать лучше.

Что такое байесовская статистика?

Что это за байесовский метод? Если вы когда-либо посещали занятия по статистике, скорее всего, они основывались на *частотной статистике*. Частотная статистика базируется на идее, что вероятность представляет собой частоту, с которой что-то происходит. Если вероятность выпадения орла при броске одной монетки равна 1 к 2, это означает, что после броска одной монетки мы можем получить половину орла (после двух бросков мы можем получить целого орла, что имеет больше смысла).

Байесовская статистика, с другой стороны, связана с тем, как вероятности отражают неопределенность полученной нами информации. С точки зрения Байеса, если вероятность выпадения орла при подбрасывании монетки равна 0,5, это означает, что мы в равной степени не уверены в том, получим мы орла или решку. Для таких проблем, как подбрасывание монеток, и частотный, и байесовский подходы кажутся разумными, но при выражении уверенности в том, что ваш кандидат победит на следующих выборах, байесовская интерпретация имеет гораздо больший смысл. В конце концов, выборы всего одни, поэтому говорить о том, как часто будет побеждать этот кандидат, не имеет смысла. При проведении байесовской статистики мы просто пытаемся точно описать, что мы думаем об окружающем мире, учитывая имеющуюся у нас информацию.

Поскольку мы можем рассматривать байесовскую статистику просто как рассуждение о неопределенных вещах, то все инструменты и методы имеют интуитивный смысл. Байесовская статистика — это поиск проблемы, с которой вы столкнулись, выяснение того, как можно описать ее математически, а затем использование причины возникновения проблемы для ее решения. Нет никаких загадочных тестов, дающих результаты, в которых вы не совсем уверены, нет распределений, которые нужно запомнить, и нет традиционных экспериментов, которые вы должны идеально воспроизвести. Хотите ли вы выяснить вероятность того, что новый дизайн сайта привлечет больше клиентов, что ваша любимая команда победит в следующей игре или что мы действительно одни во Вселенной, байесовская статистика позволит начать рассуждать об этом математически, используя всего несколько простых правил и новый взгляд на проблемы.

Структура книги

Вот краткое описание книги.

Часть I. Введение в теорию вероятностей

Глава 1. Байесовские рассуждения в обычной жизни

Первая глава знакомит вас с байесовскими рассуждениями и показывает, насколько они схожи с критическим мышлением. Основываясь на своих знаниях о мире, мы рассмотрим вероятность того, что яркий свет в окне ночью — это НЛО.

Глава 2. Измеряем неопределенность

В этой главе будем подбрасывать монетку, чтобы выразить фактические значения неопределенности в виде вероятностей: это будут числа в интервале 0 и 1, которые показывают степень уверенности в своем мнении относительно чего-либо.

Глава 3. Логика неопределенности

В логике для объединения истинных и ложных выражений используются операторы И, НЕ и ИЛИ. Оказывается, для этих операторов вероятность имеет схожие понятия. Мы рассмотрим, как обосновать выбор транспорта, чтобы добраться до места встречи, и шансы на получение штрафа.

Глава 4. Как получить биноминальное распределение вероятностей

Используя правила для вероятностей как логику, в этой главе вы построите свое собственное распределение вероятностей — биномиальное распределение, которое можно будет применить ко многим вероятностным задачам, имеющим схожую структуру. Мы попытаемся предсказать вероятность получения определенной известной коллекционной карточки из игры «гача»[1].

[1] Изначально автомат по продаже игрушек для получения различных коллекционных предметов. Зародилась в Японии. — *Примеч. ред.*

Глава 5. Бета-распределение

Здесь вы впервые столкнетесь с непрерывным распределением вероятностей и узнаете, чем статистика отличается от теории вероятности. Практическая часть статистики включает в себя попытки выяснить, какие неизвестные вероятности могут быть основаны на данных. Мы рассмотрим загадочную коробочку для раздачи монет и шансы заработать денег больше, чем потерять.

Часть II. Байесовские и априорные вероятности

Глава 6. Условная вероятность

В этой главе определим вероятности на основе имеющейся информации. Например, если мы знаем, мужчина перед нами или женщина, это позволяет предположить, страдает ли этот человек дальтонизмом. Вы также познакомитесь с теоремой Байеса, которая позволяет «обратить» условные вероятности.

Глава 7. Теорема Байеса и Lego

Здесь визуализируем теорему Байеса на примерах деталек Lego! Эта глава даст вам пространственное представление о том, что теорема Байеса делает математически.

Глава 8. Априорная, апостериорная вероятности и правдоподобие в теореме Байеса

Теорема Байеса обычно разбивается на три части, каждая из которых в байесовских рассуждениях имеет свою цель. В этой главе вы узнаете, как они называются и как их использовать, на примере изучения ограбления со взломом: было ли это преступлением или просто серией совпадений.

Глава 9. Байесовские априорные вероятности и распределение вероятностей

В этой главе посмотрим, как можно использовать теорему Байеса, чтобы лучше понять классическую сцену с астероидом из «Звездных войн: Империя наносит ответный удар». Здесь мы углубимся в априорные вероятности в байесовской статистике. Вы также увидите, как можно использовать целые распределения как априорные вероятности.

Часть III. Оценка параметров

Глава 10. Введение в усреднение и оценку параметров

Оценка параметров — это метод, который применяется для формулирования наилучшего предположения для неопределенного значения. Основной инструмент в оценке параметров — простое усреднение наблюдений. В этой главе мы проанализируем уровни снегопада и увидим, почему это работает.

Глава 11. Измерение разброса данных

Поиск среднего значения — полезный первый шаг в оценке параметров, но нам также нужен способ для учета разброса наблюдений. Здесь вы познакомитесь со средним абсолютным отклонением (Mean Absolute Deviation, MAD), дисперсией и стандартным отклонением как способами измерения разброса наблюдений.

Глава 12. Нормальное распределение

Комбинируя среднее значение и стандартное отклонение, мы получаем очень полезный инструмент для оценки: нормальное распределение. В этой главе вы узнаете, как использовать нормальное распределение, чтобы не только оценить неизвестные значения, но и узнать степень уверенности в оценках. Применим эти новые навыки, чтобы рассчитать время побега при ограблении банка.

Глава 13. Инструменты оценки параметров: PDF, CDF и квантильная функция

Здесь вы узнаете о функции плотности вероятности (PDF), кумулятивной функции распределения (CDF) и квантильной функции, чтобы лучше понять выполняемые вами оценки параметров. С помощью этих инструментов вы оцените коэффициенты конверсии рассылки и увидите, на что они влияют.

Глава 14. Оценка параметров с априорными вероятностями

Хороший способ улучшить оценки параметров — добавить априорную вероятность. В этой главе вы узнаете, как добавление априорной информации об успешном использовании коэффициента переходов в письме поможет лучше оценить реальный коэффициент конверсии для новых рассылок.

Часть IV. Проверка гипотез: сердце статистики

Глава 15. От оценки параметров к проверке гипотез: построение байесовских A/B-тестов

Теперь, когда мы можем оценивать неопределенные значения, нужно найти способ их сравнения для проверки гипотез. Вы создадите A/B-тест, чтобы определить степень уверенности в новом методе электронного маркетинга.

Глава 16. Введение в коэффициент Байеса и апостериорные шансы: конкуренция идей

Было ли у вас такое, что вы не спали полночи, гугля симптомы редкой болезни, которая, как вам кажется, у вас есть? В этой главе мы представим другой подход к проверке идей, который поможет определить, стоит ли волноваться.

Глава 17. Байесовские рассуждения в «Сумеречной зоне»

Вы верите в экстрасенсорные способности? В этой главе будем развивать собственные навыки чтения мыслей, проанализировав ситуацию из эпизода «Сумеречной зоны».

Глава 18. Когда данные не убеждают

Иногда кажется, что данных недостаточно, чтобы изменить чье-то мнение или выиграть спор. Узнайте, как переубедить друга в том, с чем вы не согласны, и почему не стоит тратить время на споры с воинственным дядей!

Глава 19. От проверки гипотез к оценке параметров

Здесь мы вернемся к оценке параметров и узнаем, как сравнить ряд гипотез. Вы рассмотрите первый пример статистики — бета-распределение, используя инструменты, которые мы изучили для простых проверок гипотез, чтобы проанализировать честность конкретной игры.

Приложение A. Краткое введение в язык R

В этом небольшом приложении даны основы языка программирования *R*.

Приложение Б. Математический минимум

Здесь мы рассмотрим математику на уровне, достаточном для того, чтобы понимать расчеты, приведенные в книге.

Приложение В. Ответы к упражнениям

Здесь вы найдете все упражнения и ответы к ним.

Для некоторых упражнений есть несколько способов решения, поэтому я дам как минимум один вариант.

Что стоит знать, прежде чем приступить к чтению

Единственным требованием к читателю является знание основ алгебры средней школы. Далее в книге вы увидите несколько примеров вычислений, но не особенно сложных. Мы будем использовать немного кода на языке программирования R. Необходимости изучать R заранее нет, я расскажу обо всем по ходу дела. Мы также коснемся высшей математики, но опять же никакого опыта не требуется: в приложениях дано достаточно информации для понимания темы.

Другими словами, эта книга призвана помочь вам начать думать математически, не требуя значительных математических знаний. Когда вы закончите чтение, то сможете даже написать уравнения для описания проблем, с которыми сталкиваетесь в жизни!

Если у вас действительно есть серьезный опыт в статистике (даже в байесовской статистике), думаю, вы все равно весело проведете время с этой книгой. Я считаю, что лучший способ хорошо понять тему — пересматривать основы раз за разом, каждый раз в ином свете. Даже я, автор этой книги, обнаружил в процессе работы много всего нового, что меня удивило!

Отправляемся в приключение!

Вы увидите, что байесовская статистика не только очень полезна, но и может доставлять массу удовольствия! Чтобы изучить байесовские рассуждения, поговорим о Lego, «Сумеречной зоне», «Звездных войнах»

и о многом другом. Вскоре вы обнаружите, что начинаете думать о задачах вероятностно и повсюду использовать байесовскую статистику. Эта книга — для быстрого и приятного чтения, поэтому смело переворачивайте страницу и отправляйтесь в путешествие по миру байесовской статистики!

От издательства

Ваши замечания, предложения, вопросы отправляйте по адресу comp@piter.com (издательство «Питер», компьютерная редакция).

Мы будем рады узнать ваше мнение!

На веб-сайте издательства www.piter.com вы найдете подробную информацию о наших книгах.

ЧАСТЬ I

ВВЕДЕНИЕ В ТЕОРИЮ ВЕРОЯТНОСТЕЙ

1

Байесовские рассуждения в обычной жизни

В этой главе мы обсудим *байесовские рассуждения* — формальный способ уточнить наши представления об окружающем мире, основываясь на наблюдаемых данных. Мы проработаем сценарии и поймем, как связать повседневный опыт с байесовскими рассуждениями.

Хочу вас обрадовать! Все вы уже встречались с байесовскими рассуждениями еще до того, как взяли в руки эту книгу. Байесовская статистика очень хорошо согласуется с тем, как люди рассуждают в обычной жизни, делая выводы из имеющихся сведений. Сложность лишь в том, чтобы разбить этот процесс на шаги и привести к строгой математической форме.

В статистике, чтобы правильно оценивать вероятности, надо строить модели и производить вычисления. Но сейчас мы не будем обращаться к математическим моделям, а просто познакомимся с основными понятиями и узнаем, что такое вероятность, обращаясь к интуиции. Уже потом, в следующих главах, мы присвоим вероятностям точные численные значения. В дальнейшем вы узнаете, как использовать математические методы для построения формальных моделей и для строгих рассуждений о понятиях, вводимых в этой главе.

Рассуждения о странных происшествиях

Как-то ночью вы вдруг просыпаетесь от бьющего в окно яркого света. Вы вскакиваете с кровати, выглядываете на улицу и видите в небе большой объект в форме — да-да — тарелки. Вообще-то, вы скептик и никогда не верили в истории о встречах с инопланетянами, но в растерянности от увиденного и невольно думаете: «А может, это НЛО?!»

Байесовские рассуждения повторяют ход ваших мыслей, когда вы столкнулись с новой ситуацией — заметить сделанные вероятностные предположения и, основываясь на этих предположениях, обновить представления о мире.

В ситуации с НЛО вы уже прошли весь цикл байесовского анализа, а именно:

1) получили данные;
2) сформулировали гипотезу;
3) пересмотрели свои представления, основываясь на новых данных.

Эти рассуждения обычно происходят так быстро, что вы не успеваете проанализировать собственные мысли. Вы обновили представления о мире, не задаваясь лишними вопросами: хотя до этого вы и не верили в НЛО, после этого происшествия вы пересмотрели свои взгляды и теперь уверены, что увидели летающую тарелку.

В этой главе мы сосредоточимся на том, как упорядочить свои представления о мире, и на том, как возникают новые суждения, чтобы взглянуть на них более строго. К числам мы перейдем в следующих главах.

Рассмотрим все этапы наших рассуждений, начиная с получения данных.

Получение данных

Ключевая идея байесовских рассуждений — делать выводы, исходя из имеющихся данных. Перед тем как сделать какие-либо выводы о ситуации (например, заявить, что вы видели НЛО), нужно понять полученные данные. В нашем случае:

- ослепительный свет за окном;
- висящий в воздухе объект в форме тарелки.

На основании прошлого опыта вы можете описать картину за окном как неожиданную. На вероятностном языке это можно записать так:

$$P(\text{яркий свет за окном, тарелкообразный объект в небе}) = \text{очень низкая},$$

где P — обозначение для вероятности, а данные перечислены в скобках. Это равенство можно прочитать как: «Вероятность наблюдать яркий свет за окном и тарелкообразный объект в небе очень низкая». Рассматривая совместную вероятность нескольких событий, перечисляем эти события через запятую. Заметим (это важно, как мы увидим дальше!), что в этих данных нет ни слова об НЛО: они состоят только из наблюдений. Можно также рассматривать вероятности отдельных событий, они будут записываться так:

$$P(\text{дождь}) = \text{весьма высокая},$$

что расшифровывается как: «Вероятность дождя весьма высокая».

В сценарии про НЛО мы должны определить вероятность того, что произойдут одновременно оба события. Вероятность только одного из этих событий будет совсем другой. Например, источником яркого света легко может оказаться проезжающая машина, так что вероятность одного этого события гораздо больше, чем совместно с наблюдением «тарелки» («тарелка» весьма неожиданна сама по себе).

Так как же определить вероятности? Пока обратимся к интуиции — общим представлениям о том, насколько ожидаемы события. В следующей главе мы увидим, как придать вероятностям точные числовые значения.

Априорные предположения и условная вероятность

Встать утром, заварить кофе и поехать на работу — задачи, не требующие от вас аналитических усилий. У вас есть *априорные предположения* (prior beliefs) о том, как устроен мир. Наши априорные предположения — набор представлений, сформированных за годы жизни (то есть на основе наблюдения за данными!). Вы уверены, что взойдет солнце — оно восходило каждый день, начиная с вашего рождения. Вы можете также предполагать, что если на перекрестке для вас зеленый свет, а для перпендикулярного потока — красный, то можно безопасно проезжать перекресток. Без априорных предположений мы каждый вечер ложились бы спать с ужасом, что

завтра солнце может не взойти, а на каждом перекрестке останавливались бы, пристально вглядываясь в приближающиеся машины.

Наши априорные предположения подсказывают, что одновременно увидеть за окном яркий свет и нечто вроде тарелки в небе — весьма редкий случай. По крайней мере, на Земле. Однако живи вы на далекой планете, кишащей летающими тарелками и постоянно посещаемой космическими пришельцами, вероятность увидеть огни и тарелки была бы гораздо выше.

Поэтому мы вводим в формулу наши априорные предположения, отделяя их вертикальной чертой |:

$$P\left(\begin{array}{c}\text{яркий свет за окном, тарелкообразный} \\ \text{объект в небе} \mid \text{дело происходит на Земле}\end{array}\right) = \text{очень низкая.}$$

Это равенство читается так: «Вероятность наблюдать яркий свет за окном и тарелкообразный объект в небе при условии, что дело происходит на Земле, очень низкая».

Такая вероятность называется условной — мы оцениваем вероятность события при некотором условии. В данном случае условия — наш прошлый опыт.

Для вероятности использовалось обозначение P. Часто мы также используем короткие обозначения для событий и условий. Если вы не привыкли к уравнениям, они могут казаться слишком сжатыми. Но через некоторое время вы увидите, как короткие названия переменных упрощают чтение и помогают обобщать равенства в целые классы задач.

Так, все наши данные мы будем обозначать одной буквой D:

$$D = \text{яркий свет за окном, тарелкообразный объект в небе.}$$

С этого момента, когда мы говорим о вероятности нашего набора данных, то пишем просто $P(D)$. Аналогично для априорных предположений мы будем использовать переменную X, например:

$$X = \text{дело происходит на Земле.}$$

Теперь мы можем обозначать вероятность как $P(D|X)$. Смысл не поменялся, а запись стала намного проще.

Множественные условия

Если на вероятность могут влиять несколько факторов, можно использовать более одного априорного предположения. Допустим, дело происходит под Новый год и ваш опыт говорит вам, что в Новый год часто запускают фейерверки. Если дело происходит на Земле, а на календаре при этом 1 января, увидеть в небе огни уже не так неожиданно, да и сама тарелка может оказаться причудливым фейерверком. Теперь уравнение выглядит так:

$$P\left(\begin{array}{c}\text{яркий свет за окном, тарелкообразный} \\ \text{объект в небе} \mid \text{1 января, дело происходит на Земле}\end{array}\right) = \text{низкая}.$$

При учете обоих условий наша условная вероятность превращается из «очень низкой» в просто «низкую».

Априорные предположения на практике

В статистике весь наш прошлый опыт обычно не вводится как явное условие, его существование предполагается неявно. Поэтому здесь мы не будем вводить для этого условия отдельную переменную. Однако в байесовском анализе очень важно помнить, что наше понимание мира всегда обусловлено прошлым опытом. Всю эту главу переменная «дело происходит на Земле» будет сохраняться.

Построение гипотезы

Итак, у нас имеются данные D (мы видели яркий свет и тарелкообразный объект) и наш прошлый опыт X. Чтобы объяснить, что же мы увидели, следует выдвинуть некоторую гипотезу — модель мира, которая даст какое-то предсказание. Гипотезы бывают разными. По сути, все наши основные представления о мире — гипотезы:

- если вы верите, что Земля вертится, вы можете предсказать, что Солнце будет всходить и заходить в определенное время;
- если вы верите, что бейсбольная команда, за которую вы болеете, — самая сильная, то можете предсказать, что они будут выигрывать чаще других команд;
- если вы верите в астрологию, то можете предсказать, что расположение звезд говорит о людях и событиях.

Гипотезы могут быть и более формальными или сложными:

- ученый может строить гипотезу, что некоторое лекарство замедлит развитие рака;
- финансовый аналитик может строить модель ситуации на рынке;
- глубокая нейронная сеть может определять, на каких картинках изображены животные, а на каких — растения.

Все это — примеры гипотез, в них заложен некоторый способ понимания мира, и он используется для предположения о том, что будет происходить. Говоря о гипотезах в байесовской статистике, мы обычно интересуемся, насколько хорошо они предсказывают наблюдаемые нами данные. Когда после увиденного вы думаете: «НЛО!» — то выдвигаете гипотезу. Гипотеза об НЛО, скорее всего, основана на бесчисленных фильмах и телепередачах, просмотренных ранее. Обозначим нашу первую гипотезу так:

$$H_1 = \text{НЛО у меня во дворе!}$$

Но что же предсказывает эта гипотеза? «Задним числом» можно спросить: «Что вы ожидали бы увидеть, если бы у вас во дворе приземлилось НЛО?» И ответ был бы таким: «Яркий свет и объект в форме тарелки». Так как гипотеза H_1 предсказывает данные D, то, когда мы наблюдаем эти данные при условии верности гипотезы, их вероятность повышается. Формально это записывается как:

$$P(D|H_1, X) >> P(D|X).$$

Это равенство читается так: «Вероятность увидеть яркий свет за окном и тарелкообразный объект в небе при условии, что это НЛО, и при моем прошлом опыте намного больше (что показано двумя знаками «больше»: >>), чем просто увидеть яркий свет за окном и тарелкообразный объект в небе без объяснений». Здесь используется язык теории вероятностей, чтобы показать, что гипотеза объясняет имеющиеся данные.

Гипотезы в обычной речи

Легко заметить связь вероятности с тем, как мы говорим в обычной жизни. Сказать, что нечто «неожиданно» — это как сказать, что эти данные имеют низкую вероятность на основании нашего прошлого опыта. Слова, что нечто «правдоподобно», могут означать, что данные имеют большую

вероятность на основании наших априорных предположений. Сейчас такие переформулировки кажутся очевидными, но суть вероятностных рассуждений — следить, как вы интерпретируете данные, строите гипотезы и меняете представления даже в обычной жизни. Без гипотезы H_1 вы были бы в растерянности и не смогли бы объяснить наблюдаемые данные.

Сбор дополнительных доказательств и обновление представлений

Итак, у вас есть данные и гипотеза. Однако с учетом вашего предыдущего опыта (а вы всегда были скептиком) гипотеза все еще смотрится диковато. Чтобы прийти к более надежным выводам, нужно собрать больше данных. Это следующий шаг в статистических рассуждениях (впрочем, в жизни мы интуитивно делаем то же самое). Чтобы собрать больше данных, надо провести новые наблюдения. В нашем сценарии вы выглядываете в окно, чтобы осмотреться.

Вы видите, что источников света вокруг уже несколько, что «тарелка» удерживается канатами, замечаете оператора с камерой, слышите хлопок и крик: «Стоп! Снято!»

Наверняка вы тут же поменяли мнение о том, что случилось. До этого вы думали, что видите НЛО. Но новые данные говорят, что, кажется, рядом снимают кино. Ваш мозг только что за секунды провел сложный байесовский анализ! Разберем подробнее, что же произошло.

Исходная гипотеза:

$$H_1 = \text{Приземлилось НЛО!}$$

Сама по себе, при условии вашего прошлого опыта, такая гипотеза крайне маловероятна:

$$P(H_1|X) = \text{очень-очень низкая.}$$

Но это была единственная толковая гипотеза, которую можно было построить при имеющихся данных. После получения дополнительных данных вы немедленно приходите к другой возможной гипотезе — рядом снимают кино:

$$H_2 = \text{За окном снимают кино.}$$

Вероятность этой гипотезы самой по себе также представляется очень низкой (если вы не живете рядом с киностудией):

$$P(H_2|X) = \text{очень низкая}.$$

Заметим, что мы присвоили H_1 «очень-очень низкую» вероятность, а H_2 просто «очень низкую». Это согласуется с житейской интуицией. Если бы у вас спросили (без всяких дополнительных данных), что более правдоподобно: ночное появление рядом НЛО или съемки фильма по соседству, вы бы ответили, что съемки правдоподобнее визита пришельцев.

Теперь нам нужно понять, как учитывать новые данные при пересмотре представлений.

Сравнение гипотез

Сначала вы приняли гипотезу об НЛО, несмотря на ее неправдоподобие, поскольку иных объяснений не было. Но теперь есть другое возможное объяснение — киносъемки, так что появилась *альтернативная гипотеза*. Рассмотреть альтернативную гипотезу — значит сравнить теории, используя имеющиеся данные.

Когда вы видите канаты, съемочную группу и свет, меняются данные. Обновленные данные выглядят так:

$$D_{\text{обнов.}} = \text{яркий свет, объект в форме тарелки,}$$
$$\text{канаты, съемочная группа, другие источники света и т. д.}$$

Получив дополнительные данные, вы меняете мнение о том, что происходит.

Разобьем этот процесс на байесовские шаги. Ваша исходная гипотеза, H_1, сначала объясняла все данные, но после дополнительных наблюдений H_1 уже не может это сделать. Это можно записать так:

$$P(D_{\text{обнов.}} \mid H_1, X) = \text{очень-очень низкая}.$$

Теперь у вас есть новая гипотеза, H_2, объясняющая данные гораздо лучше, то есть:

$$P(D_{\text{обнов.}} \mid H_2, X) \gg P(D_{\text{обнов.}} \mid H_1, X).$$

Ключевой момент — сравнить, насколько хорошо две гипотезы объясняют наблюдаемые данные. Говоря, что вероятность наших данных при условии второй гипотезы намного больше, чем при условии первой, мы сообщаем, что вторая гипотеза объясняет наблюдения лучше. Это подводит нас к сути байесовского анализа: *проверкой убеждений является то, насколько хорошо они объясняют мир*. Мы считаем, что некоторые представления правильнее других, поскольку они лучше объясняют наблюдаемые вокруг явления. Математически мы выражаем нашу идею как отношение двух вероятностей:

$$\frac{P(D_{\text{обнов.}} \mid H_2,\ X)}{P(D_{\text{обнов.}} \mid H_1,)}.$$

Большое отношение, например 1000, означает, что «H_2 объясняет данные в 1000 раз лучше, чем H_1». Так как H_2 объясняет данные во много раз лучше, чем H_1, мы меняем наши представления с H_1 на H_2. Именно это произошло, когда вы поменяли мнение о наблюдаемом явлении. Теперь вы считаете, что увидели за окном киносъемки, и это более правдоподобное объяснение для имеющихся данных.

Данные влияют на представления, но не наоборот

Напоследок стоит подчеркнуть: абсолютны и неоспоримы во всех наших примерах только данные. Гипотезы меняются, опыт X различен для разных людей, но данные D одинаковы для всех. Рассмотрим две формулы. Первую мы использовали на протяжении всей главы:

$$P(D \mid H, X).$$

Она означает вероятность данных с учетом гипотезы и опыта, проще говоря, «насколько хорошо мои представления объясняют наблюдаемое».

Но можно обратить ее (что мы часто делаем в обычной жизни):

$$P(H \mid D, X).$$

Получим «вероятность моих представлений при условии данных и опыта», то есть «насколько хорошо то, что я вижу, согласуется с моими убеждениями».

В первом случае мы меняем представления о мире в соответствии с собранными данными. Во втором — собираем данные для поддержки имеющихся представлений.

Байесовский стиль мышления основан на пересмотре и изменении представлений о мире. Реальны только данные, а наши представления о мире должны с ними согласовываться. В жизни нужно быть готовым поменять свое мнение. Когда съемочная группа собирается уезжать, вы замечаете, что на всех машинах армейская символика. Группа снимает куртки, под ними — военная форма, и кто-то говорит: «Если кто-то это видел, то мы точно его обдурили. Отличная работа!» С такими новыми данными вы наверняка еще раз поменяете мнение!

Заключение

Повторим, что мы узнали. Наши представления о мире исходно основаны на имеющемся опыте X. Полученные данные X либо согласуются с опытом, $P(D \mid X)$ = очень высокая, либо оказываются неожиданными, $P(D \mid X)$ = очень низкая.

Пытаясь объяснить окружающий мир, вы выдвигаете мнение об увиденном, или гипотезу, H. Нередко новая гипотеза позволяет объяснить неожиданные данные, $P(D \mid H, X) >> P(D \mid X)$. Получив новые данные или придумав новые идеи, вы можете выдвинуть больше гипотез, H_1, H_2, H_3, ... Вы меняете представления о мире, когда новая гипотеза объясняет данные лучше старой:

$$\frac{P(D_{\text{обнов.}} \mid H_2 X)}{P(D_{\text{обнов.}} \mid H_1 X)} = \text{большое число.}$$

Наконец, важно обращать больше внимания на данные, меняющие представления, а не на поддержку имеющихся представлений, $P(H \mid D)$.

Итак, мы изучили основы и теперь можем добавить цифры. Далее в части I вы построите математическую модель своих представлений о мире, чтобы точно определить, когда и как их менять.

Упражнения

Попробуйте ответить на эти вопросы, чтобы понять, насколько хорошо вы научились байесовским рассуждениям. Решения можно найти здесь: https://nostarch.com/learnbayes/.

1. Перепишите утверждения ниже, используя математическую нотацию из этой главы:
   - вероятность дождя низкая;
   - вероятность дождя при условии облачности высокая;
   - вероятность, что вы с зонтом при условии дождя, выше, чем просто вероятность, что вы с зонтом.
2. Запишите, используя математические обозначения из этой главы, данные из такой истории. Придумайте гипотезу, объясняющую эти данные.

   Вы приходите домой с работы и замечаете, что дверь открыта, а окно разбито. Войдя, вы видите, что вашего ноутбука нет на месте.
3. Дополним историю выше новыми данными. Покажите, как новая информация меняет ваши представления, и придумайте новую гипотезу для объяснения данных. Используйте обозначения из этой главы!

   К вам подбегает соседский ребенок и долго извиняется, что случайно попал камнем в ваше окно. Он говорит, что заметил ноутбук и испугался, что его украдут. Открыв дверь, он унес его к себе до вашего прихода.

2

Измеряем неопределенность

В первой главе мы рассмотрели основные приемы рассуждений, которыми пользуемся интуитивно, и поняли, как данные влияют на наши представления о мире.

Но важный вопрос остался нерешенным: как измерять? В теории вероятностей недостаточно просто слов о «высокой» и «низкой» вероятности — нужны числа. Тогда можно создавать численные модели мира и видеть, насколько данные меняют наши представления, решать, когда поменять мнение, и четко понимать, в чем и насколько мы уверены. В этой главе событиям будут присвоены численные вероятности.

Что такое вероятность?

С идеей вероятности мы встречаемся ежедневно. Мы говорим: «Это маловероятно!», или «Уж наверняка!», или «Не уверен». Вероятность — мера нашей убежденности в чем-либо. В предыдущей главе мы описывали наши убеждения размытыми формулировками. Но чтобы по-настоящему понять, как возникают и меняются наши представления о мире, надо формально определить $P(X)$ как число. Это число покажет, насколько мы убеждены в X. В каком-то смысле вероятность — расширение логики. В логике у нас есть истина и ложь — обе выражают абсолютную убежденность. Мы говорим, что

нечто истинно, когда совершенно уверены в этом. Логика полезна во многих задачах, но мы редко считаем нечто стопроцентно истинным или ложным, почти в каждом нашем решении есть момент неуверенности. Вероятности расширяют логику до промежуточных значений между истиной и ложью.

Компьютеры обычно представляют истину единицей, а ложь — нулем. Воспользуемся этой системой для вероятностей. $P(X) = 0$ означает, что X = ложь, а $P(X) = 1$, что X = истина. Между нулем и единицей лежит бесконечно много возможных значений. Значение ближе к 0 показывает, что мы скорее считаем нечто ложным, значение ближе к 1 — что мы скорее считаем это истинным. Заметим, что значение 0,5 говорит, что мы совершенно не в состоянии понять, истинно нечто или ложно.

Важная логическая операция — *отрицание*. «Не истина» — это ложь, «не ложь» — истина. Мы хотим действовать с вероятностями подобным образом, так что вероятности X и «не X» в сумме должны дать единицу:

$$P(X) + \neg P(X) = 1.$$

ПРИМЕЧАНИЕ

Символ ¬ означает «отрицание» или «не».

Таким образом, мы всегда можем найти вероятность отрицания X, вычитая $P(X)$ из единицы. Например, при $P(X) = 1$ вероятность отрицания равна 0, что согласуется с правилами логики. Аналогично при $P(X) = 0$ вероятность отрицания $1 - P(X) = 1$.

Теперь зададимся вопросом, как же измерить неопределенность. Можно взять произвольные значения: например, 0,95 для очень большой уверенности и 0,05 для очень маленькой. Однако это ненамного лучше размытых слов, с которыми мы имели дело раньше. Нужно вычислять вероятности формальными методами.

Вычисление вероятностей через подсчет исходов

Самый простой способ вычислить вероятность — посчитать возможные исходы. Понадобятся два множества. Первое — это множество всех возможных исходов некоторого события. Когда мы бросаем монетку, возможные

исходы — орел и решка. Второе — исходы, которые нам интересны. Например, выпадение «орла» (если мы бросаем монетку один раз, такой исход всего один). Нас может интересовать вероятность выпадения орла при бросании монеты, заражения гриппом, того, что за окном приземлится НЛО. У нас есть два множества исходов, интересные и неинтересные нам, и важно нам только отношение числа интересных исходов к числу всех возможных исходов.

Рассмотрим простой пример с бросанием монетки, где все возможные исходы — это выпадение орла и выпадение решки. Сначала посчитаем все возможные события — их только два. В теории вероятностей большая греческая буква омега (Ω) используется для множества всех событий:

$$\Omega = \{\text{орел, решка}\}.$$

Нужно узнать вероятность получить орла при одном броске монеты, запишем ее как $P(\text{орел})$. Смотрим на число интересных нам исходов — такой всего один — и делим его на общее число возможных исходов, 2:

$$\frac{\{\text{орел}\}}{\{\text{орел, решка}\}}.$$

При одном броске монеты нас интересует один исход из двух возможных, так что вероятность выпадения орла — это

$$P(\text{орел}) = \frac{1}{2}.$$

Теперь зададимся более сложным вопросом: какова вероятность выпадения хотя бы одного орла, когда мы бросаем две монеты? Список возможных событий становится сложнее — это уже не просто {орел, решка}, а все возможные пары из орла и решки:

$$\Omega = \{(\text{орел, орел}), (\text{орел, решка}), (\text{решка, орел}), (\text{решка, решка})\}.$$

Чтобы вычислить вероятность выпадения хотя бы одного орла, посмотрим, какие пары соответствуют этому условию:

$$\{(\text{орел, орел}), (\text{орел, решка}), (\text{решка, орел})\}.$$

Как можно заметить, множество интересных нам событий содержит 3 элемента, а всего у нас 4 возможные пары. Таким образом, $P(\text{хотя бы один орел}) = \frac{3}{4}$.

Это очень простые примеры, но, умея подсчитывать интересующие вас исходы и все исходы, можно быстро и легко вычислять вероятности. Когда примеры усложняются, подсчет исходов вручную становится невозможным. При решении подобных, но более трудных задач задействуют комбинаторику. В главе 4 мы увидим, как использовать комбинаторику для несколько более сложной задачи.

Вычисление вероятности как соотношения предположений

Подсчет событий полезен, когда речь идет о физических объектах, но не для большей части обыденных вопросов о вероятности:

- Какова вероятность, что завтра будет дождь?
- Думаешь, она правда президент компании?
- Это НЛО?!

Почти каждый день вы принимаете решения, основываясь на вероятности, но если вас спросят: «Насколько вероятно, что вы не опоздаете на поезд?», — вы не сможете посчитать ее описанным только что способом.

Таким образом, нужен другой подход к вероятности, который позволит рассуждать о более абстрактных задачах. Представьте, что вы болтаете с другом, и он спрашивает, слышали ли вы об эффекте Манделы. Вы не слышали, и друг рассказывает: «Это такой странный эффект ложных воспоминаний. Например, множество людей вспоминало, что Нельсон Мандела умер в тюрьме в 1980-х. Но на самом деле он был освобожден, стал президентом ЮАР и умер только в 2013-м!» Вы смотрите на друга скептически и отвечаете: «Ну, это какая-то диванная психология из интернета. Вряд ли кто-то всерьез вспоминал о смерти Манделы. Готов спорить, об этом даже нет статьи в Википедии».

Итак, вы хотите измерить P (в англоязычной Википедии нет статьи об эффекте Манделы[1]). Предположим, что сотовая связь не ловит, и быстро

[1] В русскоязычном сегменте Википедии такая статья есть: https://ru.wikipedia.org/wiki/Эффект_Манделы. В англоязычном сегменте эффект Манделы упоминается в статье *False memory*, https://en.wikipedia.org/wiki/False_memory. — *Примеч. ред.*

это не проверить. Вы уверены, что статьи нет, и хотите присвоить этому предположению высокую вероятность. Но надо присвоить вероятности численное значение от 0 до 1, с чего же начать?

Вы решаете заключить пари и говорите другу: «Это наверняка выдумка. Давай так: если статьи об эффекте Манделы нет, ты отдаешь мне пять долларов, если статья есть — я тебе 100 долларов!» Пари — способ на практике выразить нашу убежденность в чем-либо. Вы уверены, что существование статьи настолько маловероятно, что готовы отдать другу 100 долларов, если ошиблись, и получить всего 5 долларов за свою правоту. И теперь мы можем начать оценивать вероятность вашего предположения, что статьи про эффект Манделы в Википедии нет.

Использование ставок для определения вероятности

Гипотеза вашего друга состоит в том, что об эффекте Манделы есть статья. А у вас есть альтернативная гипотеза $H_{\text{статьи нет}}$.

Мы еще не знаем конкретных вероятностей, но ваша ставка показывает сильную уверенность в своей гипотезе. Ставки часто используют как показатель уверенности, рассматривая как отношение суммы, которую вы готовы заплатить за ошибку, к той, которую вы получите за верный прогноз. Например, пусть ставки на лошадь на скачках — 12 к 1. Это означает, что, поставив 1 доллар, вы получите от букмекера 12, если лошадь выиграет. Ставки часто произносят как «m к n», но можно смотреть на них просто как на дробь: m/n.

Между ставками и вероятностями существует прямая связь.

Мы можем записать ставки в вашем пари: «100 к 5». Как извлечь отсюда вероятность?

Ваша ставка показывает, насколько больше ваша уверенность в том, что статьи нет, чем в том, что она есть. Запишем это как отношение вашей уверенности в отсутствии статьи, $P(H_{\text{нет статьи}})$ к уверенности друга, что статья есть, $P(H_{\text{статья есть}})$:

$$\frac{P(H_{\text{статьи нет}})}{P(H_{\text{статья есть}})} = \frac{100}{5} = 20.$$

Из отношения этих двух гипотез мы видим, что ваша убежденность в отсутствии статьи в 20 раз больше, чем в гипотезе друга. Можно использовать это для вычисления точной вероятности — понадобится лишь немного алгебры.

Вычисление вероятности

Запишем уравнение, где выразим то, что хотим узнать — вероятность вашей гипотезы:

$$P(H_{\text{статьи нет}}) = 20 \times P(H_{\text{статья есть}})$$

(читается как «Вероятность того, что статьи нет, в 20 раз больше, чем того, что статья есть»).

Но возможностей всего две: в Википедии либо есть статья про эффект Манделы, либо нет. Наши две гипотезы покрывают все возможности, так что вероятность *наличия* статьи — это 1 минус вероятность ее *отсутствия*, и можно заменить $P(H_{\text{статья есть}})$ на ее выражение через $P(H_{\text{статьи нет}})$:

$$P(H_{\text{статьи нет}}) = 20 \times (1 - P(H_{\text{статья есть}})).$$

Раскроем скобки в выражении $20 \times (1 - P(H_{\text{статьи нет}})$ и получим:

$$P(H_{\text{статьи нет}}) = 20 - 20 \times P(H_{\text{статьи нет}}).$$

Мы можем избавиться от $P(H_{\text{статьи нет}})$ в правой части уравнения, прибавив $20 \times P(H_{\text{статьи нет}})$ к обеим частям. $P(H_{\text{статьи нет}})$ остается только в левой части:

$$21 \times P(H_{\text{статьи нет}}) = 20.$$

Поделив обе части на 21, приходим к:

$$P(H_{\text{статьи нет}}) = \frac{20}{21}.$$

Получается прекрасное точное численное значение между 0 и 1, выражающее вашу уверенность в гипотезе, что статьи об эффекте Манделы нет. Можно обобщить этот способ преобразования ставок в вероятности так:

$$P(H) = \frac{O(H)}{1 + O(H)},$$

где O — ставка (от «*odd*» — ставка).

Столкнувшись на практике с каким-то абстрактным представлением, спрашивайте себя, сколько вы бы поставили на его верность. Вы наверняка согласились бы на ставку миллиард к одному на то, что завтра взойдет солнце, но не на выигрыш любимой бейсбольной команды. В любом случае можно присвоить этим событиям вероятности, пользуясь только что описанным способом.

Измеряем уверенность при бросании монеты

Итак, у нас есть способ определить вероятности абстрактных идей с использованием ставок. Но настоящей проверкой метода станет то, сработает ли он с броском монеты, про который мы все знаем, посчитав исходы. Зададим себе вопрос: «Насколько я уверен, что при следующем броске выпадет орел?» Теперь мы говорим не о P(орел), но о гипотезе $P(H_{\text{орел}})$. Как и прежде, нужна альтернативная гипотеза, с которой мы сравним нашу. Можно сказать, что альтернативная гипотеза — «выпадет не орел», но чаще мы скажем проще: «Выпадет решка». Важно, что по сути это одно и то же:

$$H_{\text{решка}} = H_{\neg\text{орел}}, \text{ и } P(H_{\text{решка}}) = 1 - P(H_{\text{орел}}).$$

Теперь мы смотрим на отношение

$$\frac{P(H_{\text{орел}})}{P(H_{\text{решка}})} = ?$$

Выражение читается как «Насколько сильнее я уверен в выпадении орла, чем в выпадении решки?» Но ни один из исходов не выглядит предпочтительнее, так что единственная разумная ставка — 1 к 1. Конечно, можно использовать и другие равные друг другу значения: 2 к 2, 5 к 5 или 10 к 10. Отношение всегда одно:

$$\frac{P(H_{\text{орел}})}{P(H_{\text{решка}})} = \frac{10}{10} = \frac{5}{5} = \frac{2}{2} = \frac{1}{1} = 1.$$

Учитывая, что отношение всегда одно и то же, мы просто повторяем способ, которым считали вероятность отсутствия статьи об эффекте

Манделы. Мы знаем, что в сумме вероятности орла и решки дают 1 и отношение этих вероятностей — тоже 1. Итак, вероятности описываются двумя уравнениями:

$$P(H_{\text{орел}}) + P(H_{\text{решка}}) = 1 \text{ и } \frac{P(H_{\text{орел}})}{P(H_{\text{решка}})} = 1.$$

Повторив весь процесс рассуждений про эффект Манделы, вы найдете, что единственное возможное значение для $P(H_{\text{орел}})$ будет равно 1/2. Мы пришли к тому же результату, что и при подсчете исходов, поэтому такой метод вычисления вероятностей как меры уверенности достаточно надежен!

Если есть два способа вычислять вероятности, то разумно спросить, когда какой их них стоит использовать. К счастью, раз они эквивалентны, можно выбирать тот, который проще применить к имеющейся задаче.

Заключение

В этой главе мы рассмотрели два взгляда на вероятность: вероятность исходов и вероятность как мера уверенности. Мы определили вероятность как отношение числа интересных нам исходов к общему числу исходов. Это самое популярное определение вероятности, но его трудно применить к представлениям о мире: для большинства повседневных задач нет четко определенного набора исходов, которым легко присвоить числовые значения. Поэтому, чтобы оценить вероятность наших убеждений, следует оценить, насколько сильнее мы уверены в одной гипотезе, чем в другой.

Хорошей проверкой будет готовность сделать ставки на свои гипотезы. Например, вы поспорили с другом и платите ему 1000 долларов за доказательство существования НЛО, а он вам — всего один доллар за доказательство, что НЛО не существует. Вы фактически сообщаете, что в 1000 раз более уверены в том, что НЛО не существует, чем в обратном. Вооружившись этим методом, можно вычислять вероятности в самых разных ситуациях. В следующей главе я расскажу, как применять основные логические операции И и ИЛИ к вероятностям. Но прежде чем двигаться дальше, попробуйте попрактиковаться.

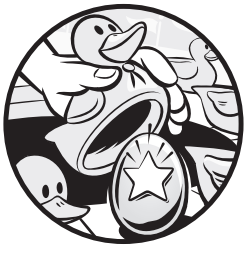
Упражнения

Чтобы убедиться, что вы понимаете, как присвоить вероятностям значения от 0 до 1, попробуйте ответить на эти вопросы.

1. Какова вероятность бросить два шестигранных кубика и получить в сумме больше 7?
2. Какова вероятность бросить три шестигранных кубика и получить в сумме больше 7?
3. Играют команды «Янки» и «Ред Сокс». Вы — преданный фанат «соксов» и заключаете с другом пари на их выигрыш. Если «Сокс» проиграет, вы платите другу 30 долларов, если выиграет — друг платит вам 5 долларов. Какую вероятность вы присвоите гипотезе, что выиграет «Ред Сокс»?

3

Логика неопределенности



В главе 2 мы обсудили, что вероятности — это расширение логических понятий истины и лжи и выражаются как числа между 1 и 0. Сила вероятностей — в их способности принимать бесконечное число значений между этими полюсами. В этой главе мы обсудим, как правила логики и ее операции применяются к вероятностям. В классической логике важнейшую роль играют три операции:

- И;
- ИЛИ;
- НЕ.

С помощью этих трех простых операций можно построить любое высказывание в классической логике. Рассмотрим, например, высказывание: *«Если идет дождь И я собираюсь на улицу, мне нужен зонтик»*. В этом высказывании всего одна логическая операция: И. Благодаря ей мы знаем, что если истинно утверждение, что идет дождь, И так же истинно утверждение, что я собираюсь на улицу, то мне нужен зонтик.

Мы можем переформулировать это утверждение, используя другие операции: *«Если НЕТ дождя или я НЕ собираюсь на улицу, мне НЕ нужен зонтик»*. Здесь используются простейшие логические операции и факты для решения о том, нужно ли брать зонт.

Но такой способ логических рассуждений работает лишь тогда, когда факты либо абсолютно истинны, либо абсолютно ложны. Он годится, чтобы понять, нужен ли зонт прямо сейчас — когда я точно знаю, идет ли в этот момент дождь и собираюсь ли я на улицу. Зададимся вопросом: «Понадобится ли мне зонт завтра?» Тогда факты теряют определенность, ведь прогноз погоды дает только вероятность завтрашнего дождя, да и я не могу точно знать, нужно ли мне завтра выходить на улицу.

В этой главе мы объясним, как применять логические операции при работе с вероятностями и в ситуациях неопределенности рассуждать по аналогии с классической логикой. Вы уже знаете, как определить операцию НЕ для вероятностных рассуждений:

$$\neg P(X) = 1 - P(X).$$

Дальше я расскажу, как использовать две другие операции, И и ИЛИ, для сочетаний вероятностей и делать точные и полезные выводы.

Вероятность и операция И

В статистике И используется, когда речь идет о вероятности одновременных событий. Например, вероятности:

- бросив кубик и монету, выкинуть шестерку И орла;
- попасть под дождь И забыть зонт;
- выиграть в лотерею И получить удар молнии.

Чтобы понять, как определить операцию И для вероятностей, начнем с простого примера про монету и кубик.

Вычисление совместной вероятности

Допустим, нужно узнать вероятность выпадения орла и шестерки при бросании монеты и кубика. Нам известны вероятности обоих событий по отдельности:

$$P(\text{орел}) = \frac{1}{2},\ P(\text{шестерка}) = \frac{1}{6}.$$

Требуется найти вероятность, что оба события произойдут, то есть:

$$P(\text{орел, шестерка}) = ?$$

Можно вычислить ее тем же способом, что в главе 2: посчитать интересные нам исходы и поделить на общее число исходов. Пусть сначала мы бросаем монету. Как показано на рис. 3.1, возможных исходов два.

Далее после каждого результата броска монеты возможны шесть исходов броска кубика, как показано на рис. 3.2.

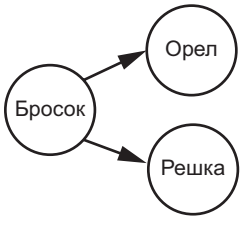


Рис. 3.1. Два возможных исхода броска монеты как два пути

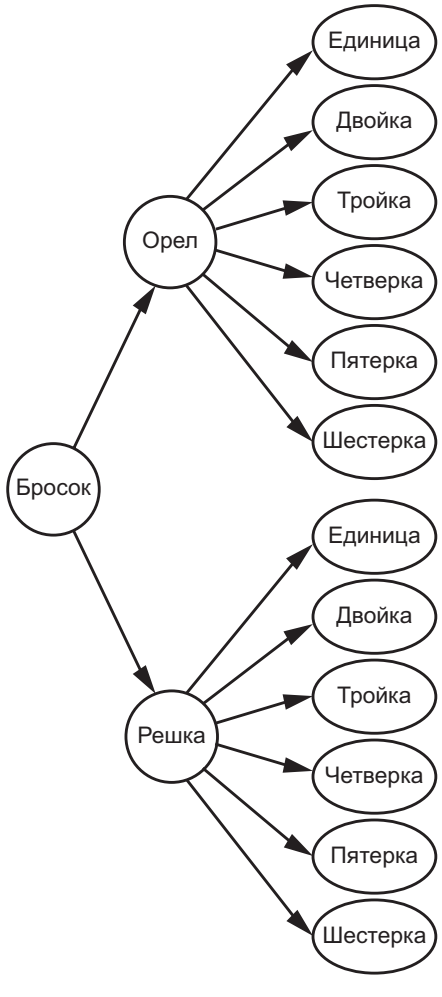


Рис. 3.2. Возможные исходы броска монеты и броска кубика

Посмотрев на рисунок, легко посчитать возможные исходы бросков монеты и кубика. Их 12, а нас интересует только один, поэтому

$$P(\text{орел, решка}) = \frac{1}{12}.$$

Мы решили одну конкретную задачу. Но нужно иметь общее правило, позволяющее считать вероятность любой комбинации событий. Посмотрим, как обобщить это решение.

Применяем правило произведения вероятностей

Вернемся к той же задаче: какова вероятность выкинуть орла И шестерку?

Сначала найдем вероятность выпадения орла. Посмотрев на ветвящиеся пути на рисунке, вы поймете, сколько будет путей при данных вероятностях. Нам нужны только пути, включающие выпадение орла. Вероятность выпадения орла 1/2, поэтому половина возможностей отпадает. Посмотрим только на оставшуюся ветку возможностей. Шансы получить желаемый результат (выкинуть шестерку) — всего 1/6. Эти рассуждения наглядно показаны на рис. 3.3, очевидно, что нам интересен всего один исход.

Перемножив эти вероятности, мы видим, что

$$\frac{1}{2} \times \frac{1}{6} = \frac{1}{12}.$$

Получается в точности такой же ответ, что и раньше, но вместо подсчета всех возможных событий мы

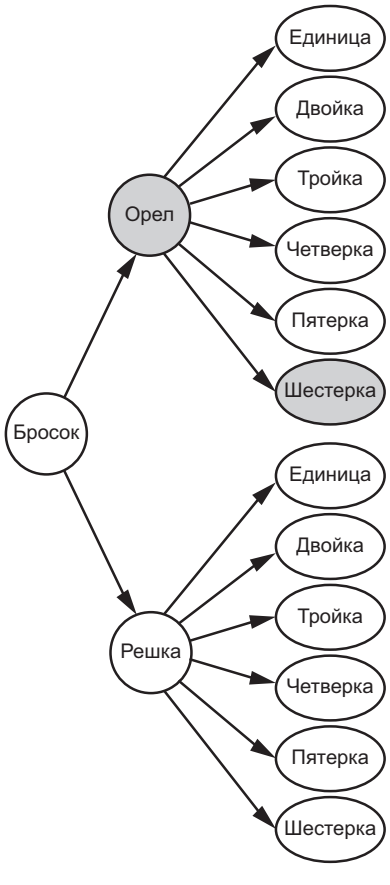


Рис. 3.3. Вероятность одновременно выкинуть орла и шестерку

считали только вероятности интересующих нас событий, двигаясь по веткам «дерева». С помощью рисунка легко решить эту простую задачу, но важнее другое — здесь заложено общее правило, как соединять вероятности операцией И:

$$P(A, B) = P(A) \times P(B).$$

Умножим результаты, посчитаем произведение и назовем эту формулу правилом произведения вероятностей. Можно обобщить ее и на большее число событий.

Рассмотрим комбинацию событий A и B как одно событие и посчитаем вероятность его комбинации с C:

$$P(P(A, B), C) = P(A, B) \times P(C) = P(A) \times P(B) \times P(C).$$

Можно пользоваться правилом произведения для любого количества событий.

Пример: вероятность опоздать

Рассмотрим пример применения правил произведения для чуть более сложной задачи, чем броски монет и кубиков. Допустим, вы условились встретиться с другом за чашечкой кофе в 16:30 на другом конце города, куда вы можете добраться на трамвае или автобусе. Сейчас 15:30. Ближайший автобус приедет в 15:45 и за 45 минут довезет вас до кофейни. Ближайший трамвай придет в 15:50, доедет за 30 минут, но затем придется идти еще 10 минут пешком. В обоих случаях вы окажетесь на месте ровно в 16:30, а при любой задержке уже опоздаете. К счастью, так как автобус приходит раньше, то в случае его задержки вы можете сесть на трамвай и успеть (если не задержится и трамвай!). Все хорошо и в случае, когда автобус приходит вовремя, а трамвай задерживается. Опоздаете вы, только если задержатся и автобус, и трамвай. Как же найти вероятность опоздания?

Найдем вероятность, что и трамвай, и автобус задержатся. Пусть транспортная компания пишет, что

$$P(\text{задержка}_{\text{трамвая}}) = 0{,}15;$$

$$P(\text{задержка}_{\text{автобуса}}) = 0{,}2$$

(потом мы научимся оценивать такие вероятности на основе данных).

Эти данные говорят нам, что трамваи задерживаются в 15 % случаев, а автобусы — в 20 % случаев. Вы опоздаете, только если задержатся *оба* вида транспорта, так что применим к задаче правило произведения:

$$P(\text{опоздание}) = P(\text{задержка}_{\text{трамвая}}) \times P(\text{задержка}_{\text{автобуса}}) = 0{,}15 \times 0{,}2 = 0{,}03.$$

Даже если по отдельности шансы задержки трамвая и автобуса довольно велики, вероятность, что опоздают оба, значительно меньше — всего 0,03. Можно сказать, что шанс задержки обоих составит 3 %. Теперь вы можете сильно не переживать, что опоздаете.

Вероятность и операция ИЛИ

Другое важнейшее правило логики описывает, как вероятности сочетаются при помощи операции ИЛИ, например:

- заболеть гриппом ИЛИ простудой;
- выбросить орла на монете ИЛИ шесть на кубике;
- проколоть колесо ИЛИ столкнуться с нехваткой бензина.

Вероятность одного ИЛИ другого события чуть сложнее, так как события могут быть взаимоисключающими, а могут и не быть. Взаимоисключающими события называются, когда из наступления одного следует невозможность другого. Например, исходы броска кубика взаимоисключающие, так как за один бросок вы не в состоянии получить и единицу, и шестерку. С другой стороны, бейсбольный матч могут отменить, если будет дождь или заболеет тренер. И это не взаимоисключающие события — вполне возможно, что и тренер болен, и дождь идет.

ИЛИ для взаимоисключающих событий

Комбинировать два события, используя ИЛИ, кажется интуитивным действием. Спроси вас, какова вероятность, бросив монету, получить орла или решку, и вы ответите: 1. Мы знаем, что:

$$P(\text{орел}) = \frac{1}{2},\ P(\text{решка}) = \frac{1}{2}.$$

Мы интуитивно складываем вероятности этих событий и знаем, что это сработает, поскольку других исходов, кроме орла и решки, не бывает,

а вероятность всех возможных исходов равна 1. Если вероятность всех возможных исходов не равна 1, значит, мы не учли какой-то исход. Как это понять? Допустим, вероятность выпадения орла $P(\text{орел}) = 1/2$, но некто утверждает, что вероятность выпадения решки $P(\text{решка}) = 1/3$. Мы уже знаем, что вероятность невыпадения орла равна:

$$P(\text{не орел}) = 1 - \frac{1}{2} = \frac{1}{2}.$$

Так как вероятность не выкинуть орла равна $1/2$, а заявленная вероятность решки всего $1/3$, то либо мы не учли какое-то событие, либо вероятность решки ошибочна. Пока события взаимоисключающие, можно просто сложить все вероятности, чтобы получить вероятность того, что произошло одно из них — ИЛИ то, ИЛИ другое. Еще одним примером будет бросок кубика. Мы знаем, что вероятность выпадения единицы равна $1/6$, как и вероятность выпадения двойки:

$$P(\text{единица}) = \frac{1}{6},\ P(\text{двойка}) = \frac{1}{6}.$$

Можно проделать то же самое — сложить две вероятности и увидеть, что вероятность выкинуть 1 ИЛИ 2 равна $2/6$, или $1/3$:

$$P(\text{единица}) + P(\text{двойка}) = \frac{2}{6} = \frac{1}{3}.$$

Все кажется интуитивно понятным.

Правило сложения верно только для взаимоисключающих событий. На языке вероятностей взаимоисключающие события такие, когда

$$P(A) \text{ И } P(B) = 0.$$

То есть вероятность наступления одновременно A И B равна 0. Для наших примеров это так:

- невозможно бросить одну монету и одновременно выкинуть орла и решку;
- невозможно бросить кубик один раз и одновременно выкинуть 1 и 2.

Чтобы действительно понять, как объединять события операцией ИЛИ, рассмотрим не взаимоисключающие события.

Правило суммы для не взаимоисключающих событий

Вернемся к примеру с кубиком и монетой и найдем вероятность выбросить орла ИЛИ шестерку. Часто новички наивно полагают, что сложение вероятностей сработает и здесь. Мы знаем, что $P(\text{орел}) = 1/2$ и $P(\text{шестерка}) = 1/6$, и кажется правдоподобным, что вероятность одного из этих событий равна $4/6$. Но если рассмотреть вероятность выпадения орла или числа меньше шести, станет ясно, что сложение не работает. Так как $P(\text{меньше шестерки}) = 5/6$, сложение вероятностей даст $8/6$, что больше 1! Нарушено правило, что вероятность находится в диапазоне между 0 и 1, а значит, где-то ошибка.

Беда в том, что выкинуть орла и выкинуть шестерку — события не взаимоисключающие. Мы знаем, что $P(\text{орел, шестерка}) = 1/2$. Так как вероятность наступления обоих событий одновременно не равна 0, они по определению не взаимоисключающие. Сложение вероятностей для не взаимоисключающих событий не работает, так как мы дважды считаем исходы, когда происходят оба события. В качестве примера рассмотрим все исходы бросков монеты и кубика, в которых выпадает орел:

Орел — 1;
Орел — 2;
Орел — 3;
Орел — 4;
Орел — 5;
Орел — 6.

Это 6 из 12 возможных исходов, что ожидаемо, так как $P(\text{орел}) = 1/2$. Посмотрим теперь на все исходы, в которых выпадает шестерка:

Орел — 6;
Решка — 6.

Это два из 12 возможных исходов, и снова ничего удивительного, потому что $P(\text{шестерка}) = 1/6$. Так как шесть исходов удовлетворяют условию выпадения орла и два — выпадения шестерки, очень хочется сказать, что орел или шестерка появляются в восьми исходах. Однако *орел — шестерка* есть в обоих списках, мы посчитали его дважды. Так что разных исходов только семь из 12. Наивно сложив $P(\text{орел})$ и $P(\text{шестерка})$, мы получим больше чем надо.

Чтобы исправить расчеты, сложим обе вероятности и вычтем вероятность возникновения обоих событий. Получаем правило, как объединять с помощью ИЛИ не взаимоисключающие события, — правило суммы вероятностей:

$$P(A \text{ ИЛИ } B) = P(A) + P(B) - P(A, B).$$

Мы складываем вероятности обоих событий по отдельности и вычитаем вероятность обоих событий одновременно, чтобы не учитывать дважды вероятности некоторых исходов, включенные и в $P(A)$, и в $P(B)$. В примере с монетой и кубиком вероятность выкинуть орла или число, меньшее шести, равна

$$P(\text{орел ИЛИ шестерка}) = P(\text{орел}) + P(\text{шестерка}) - P(\text{орел, шестерка}) =$$
$$= \frac{1}{2} + \frac{1}{6} - \frac{1}{12} = \frac{7}{12}.$$

Для закрепления материала рассмотрим последний пример с ИЛИ.

Пример: вероятность большого штрафа

Представьте, что вас тормознули за превышение скорости. Вас уже давно не останавливали, и вы понимаете, что, возможно, забыли взять новые права или новую страховку. Если у вас нет хотя бы одного из этих документов, штраф будет гораздо выше. Как, еще не проверив это, оценить вероятность большего штрафа? Вы почти уверены, что брали права, поэтому присваиваете их наличию вероятность 0,7. Но вы почти уверены, что забыли страховку дома на столе, и присваиваете ее наличию в машине вероятность 0,2. Итак,

$$P(\text{права}) = 0{,}7;$$
$$P(\text{страховка}) = 0{,}2.$$

Но это вероятности наличия документов, а вас беспокоят вероятности их отсутствия. Посчитать их просто, используя отрицание:

$$P(\text{нет}_{\text{прав}}) = 1 - P(\text{права}) = 0{,}3;$$
$$P(\text{нет}_{\text{страховки}}) = 1 - P(\text{страховка}) = 0{,}8.$$

Если не использовать настоящее правило суммы, а просто сложить вероятности, мы получим вероятность больше 1:

$$P(\text{нет}_{\text{прав}}) + P(\text{нет}_{\text{страховки}}) = 1{,}1.$$

События не взаимоисключающие: возможно, вы забыли оба документа, так что мы посчитали этот исход дважды. Надо найти вероятность отсутствия обоих документов, чтобы затем вычесть ее. Применим правило произведения:

$$P(\text{нет}_{\text{прав}}, \text{нет}_{\text{страховки}}) = 0{,}24.$$

Теперь применим правило суммы, чтобы определить вероятность отсутствия одного из документов, подобно тому как считали вероятность выпадения орла или шестерки:

$$P(\text{нет}_{\text{документа}}) = P(\text{нет}_{\text{прав}}) + P(\text{нет}_{\text{страховки}}) - P(\text{нет}_{\text{прав}}, \text{нет}_{\text{страховки}}) = 0{,}86.$$

Что ж, с вероятностью 0,86 одного из документов нет. Будьте повежливее с инспектором!

Заключение

В этой главе мы изучили логику неопределенности, добавив правила для комбинации вероятностей через операции И и ИЛИ. Вспомним изученные логические правила. В главе 2 вы узнали, что вероятности измеряются по шкале от 0 до 1, где 0 — ложь (невозможное событие), а 1 — истина (то, что точно случится). Следующее важное правило касается комбинации вероятностей с использованием И. Для этого используется правило произведения, гласящее, что для нахождения вероятности одновременного наступления двух событий, A и B, нужно умножить их вероятности:

$$P(A, B) = P(A) \times P(B).$$

Последнее правило касается комбинации вероятностей с использованием ИЛИ по правилу суммы. Тонким моментом правила суммы является то, что при сложении вероятностей не взаимоисключающих событий мы дважды учтем вероятность исходов, когда они происходят одновременно, и будем обязаны ее вычесть. Правило суммы при этом использует правило произведения (вспомним, что для взаимоисключающих событий $P(A, B) = 0$):

$$P(A \text{ ИЛИ } B) = P(A) + P(B) - P(A, B).$$

Эти правила вместе с изученными в главе 2 позволяют решать широкий спектр задач. На них будут основаны вероятностные рассуждения на протяжении всей книги.

Упражнения

Чтобы убедиться, что вы понимаете, как применять правила логики к вероятностям, попробуйте ответить на эти вопросы.

1. Какова вероятность выбросить двадцать на двадцатигранной игральной кости три раза подряд?
2. Прогноз погоды сообщает, что завтра с 10 %-ной вероятностью пойдет дождь. Вы забываете зонтик дома в половине случаев. Какова вероятность, что завтра вы окажетесь под дождем без зонта?
3. Сырые яйца с вероятностью 1/20 000 заражены сальмонеллой. Вы съели два сырых яйца, какова вероятность, что вы съели яйцо с сальмонеллой?
4. Какова вероятность выкинуть два орла за два броска монеты или три шестерки за три броска шестигранного кубика?

4

Как получить биноминальное распределение

Из главы 3 вы узнали некоторые правила для вероятностей, относящиеся к логическим операциям: И, ИЛИ, НЕ. Используем их в этой главе, чтобы получить свое первое вероятностное распределение — способ описать все возможные события и их вероятности. Вероятностные распределения часто изображают графически для большей наглядности. Мы придем к распределению, определив функцию для обобщения целого класса задач о вероятности. Это распределение позволит считать вероятности для многих ситуаций, а не только для одной конкретной. Для этого мы найдем общие элементы во всех задачах и выделим их. Статистики часто так делают, это позволяет легче находить решения для целых классов задач. Особенно полезен такой подход в случае сложных задач или когда подробности неизвестны. Тогда можно использовать хорошо изученные и понятные распределения как приблизительное описание не вполне понятного реального мира.

Распределения вероятностей очень полезны, когда нас интересует диапазон возможных значений. Например, мы можем использовать их для определения вероятности, что покупатель зарабатывает от 30 000 до 45 000 долларов в год. Или вероятности, что взрослый человек будет выше двух метров. Или вероятности, что от 25 до 35 % посетивших веб-страницу зарегистрируются. Многие вероятностные распределения описываются сложными

формулами, к которым надо привыкнуть. Однако все эти формулы выводятся из базовых правил, описанных в предыдущих главах.

Структура биномиального распределения

Сейчас вы познакомитесь с биномиальным распределением, которое используется для подсчета числа успешных исходов, когда мы знаем число попыток и вероятность успеха.

Приставка «би» означает, что возможных исходов два: событие происходит или нет. Если возможных исходов более двух, распределение называется мультиномиальным. Биномиальное распределение описывает вероятности:

- дважды выбросить орла за три броска монеты;
- купить миллион лотерейных билетов и выиграть хотя бы один раз;
- выбросить двадцать меньше трех раз за 10 бросков двадцатигранной кости.

Все эти задачи похожи по структуре. Действительно, каждое биномиальное распределение имеет три параметра:

- k — число интересующих нас исходов;
- n — число попыток;
- p — вероятность интересующего нас исхода.

Эти параметры подаются на вход распределения. Например, когда мы ищем вероятность выкинуть двух орлов за три броска монеты:

- $k = 2$ — число интересующих нас исходов (здесь — орлов);
- $n = 3$ — общее число бросков;
- $p = 1/2$ — вероятность выкинуть орла при броске.

Можно построить биномиальное распределение, чтобы обобщить задачи такого рода, и после этого легко решить любую задачу с такими тремя параметрами. Такое распределение обозначается

$$B(k; n, p).$$

В примере с тремя бросками монеты мы пишем $B(2; 3, 1/2)$. B — сокращение от «биномиальный» (binomial). Заметим, что k отделено от других

параметров точкой с запятой, так как нас обычно интересует распределение для всех k при фиксированных n и p. Поэтому $B(k; n, p)$ обозначает одно из значений распределения, но распределение в целом обычно обозначается просто $B(n, p)$.

Рассмотрим подробнее, как найти функцию, обобщающую все перечисленные задачи.

Выделение главного в задаче

Один из лучших способов увидеть, как переход к распределениям облегчает подсчет вероятностей, — начать с конкретного примера, попробовать его решить, а потом понять, что можно изменить. Продолжим на примере трех бросков монеты, в которых мы хотим получить два орла. Так как количество возможных исходов невелико, мы можем легко записать все интересующие нас исходы с двумя орлами:

ООР, ОРО, РОО.

Хочется перечислить и все другие возможные исходы, а потом поделить число интересующих нас исходов на общее количество. Такой способ сработает для этой задачи, но наша цель — решить любую задачу, в которой есть некоторое количество попыток и вероятность желательного исхода. Без обобщения, если мы поменяем параметры, придется решать задачу снова. Уже сам вопрос «какова вероятность получить двух орлов за четыре броска монеты» потребует нового отдельного решения.

Воспользуемся правилами для вероятностей и постараемся обобщить задачу. Разобьем ее на кусочки, которые можно легко решить и записать в виде формул, а потом соберем эти формулы вместе в функцию биномиального распределения.

В первую очередь заметим, что все желательные исходы имеют одинаковую вероятность. Каждый из них — просто перестановка другого:

$$P(\{\text{орел, орел, решка}\}) = P(\{\text{орел, решка, орел}\}) = \\ = P(\{\text{решка, орел, орел}\}).$$

Назовем эту вероятность так:

$$P(\text{желательный исход}).$$

Исхода три, но реализуется только один (нам неважно какой). Поэтому исходы являются взаимоисключающими:

$$P(\{\text{орел, орел, решка}\}, \{\text{орел, решка, орел}\}, \{\text{решка, орел, орел}) = 0.$$

Так что нам легко применить правило суммы вероятностей и заключить, что

$$P(\{\text{орел, орел, решка}\} \text{ или } \{\text{орел, решка, орел}\} \text{ или } \{\text{решка, орел, орел}\}) =$$
$$= P(\text{желательный исход}) + P(\text{желательный исход}) +$$
$$+ P(\text{желательный исход}).$$

Конечно же, сложение даст нам

$$3 \times P(\text{желательный исход}).$$

Итак, мы умеем коротко записывать интересующие нас исходы, но проблема в том, что число 3 относится именно к этой задаче. Исправим это, заменив 3 на коэффициент $N_{\text{исходов}}$. Получаем неплохое обобщение:

$$B(k; n, p) = N_{\text{исходов}} \times P(\text{желательный исход}).$$

Теперь надо решить две подзадачи: как найти число интересующих нас исходов и как определить вероятность одного исхода. Выяснив это, мы решим задачу!

Подсчет исходов через биномиальные коэффициенты

Сначала надо найти, сколько исходов соответствуют данным k (числу желательных исходов) и n (числу попыток). Для маленьких чисел это несложно посчитать. Желая получить четырех орлов за пять бросков, мы имеем пять желательных исходов:

ООООР, ОООРО, ООРОО, ОРООО, РОООО.

Но очень быстро перечисление становится слишком трудоемким — например, «какова вероятность выкинуть две шестерки за три броска шестигранного кубика». Это все еще биномиальная задача — мы либо выкидываем шесть, либо нет, но событий, описываемых как «не шестерка», гораздо

больше. Попытки перечислить их все быстро утомляют, даже если бросков кубика всего три:

6 — 6 — 1
6 — 6 — 2
6 — 6 — 3
...
4 — 6 — 6
...
5 — 6 — 6
...

Очевидно, что перечисление всех исходов быстро перестает работать. Спасет нас комбинаторика.

Комбинаторика: умный подсчет через биномиальные коэффициенты

Идею решения задачи подскажет раздел математики под названием «комбинаторика». По сути — это способы умного подсчета. В комбинаторике существует понятие биномиального коэффициента — количества способов выбрать k вещей из n — например, интересующих нас исходов из общего числа попыток. Обозначение для биномиального коэффициента выглядит так:

$$\binom{n}{k}$$

Это выражение читается как «эн по ка». В нашем примере запишем «два орла из трех бросков» как

$$\binom{3}{2}.$$

Определяется эта величина как

$$\binom{n}{k} = \frac{n!}{k! \times (n-k)!}.$$

Восклицательный знак обозначает факториал, то есть произведение всех чисел от единицы до числа перед восклицательным знаком включительно, например: 5! = (5 × 4 × 3 × 2 × 1).

Большинство языков программирования, применяющихся в области математики, имеют функцию `choose()` для вычисления биномиальных коэффициентов. Например, в языке R мы вычислим биномиальный коэффициент для двух орлов за три броска монеты так:

```
choose(3,2)
>>3
```

Пользуясь биномиальными коэффициентами для подсчета числа интересующих нас исходов, мы перепишем общую формулу так:

$$B(k; n, p) = \frac{n}{k} \times P(\text{желательный исход}).$$

Вспомним, что P(желательный исход) — вероятность любой из комбинаций трех бросков, в которой есть два орла. В предыдущем равенстве мы использовали эту величину, не зная, чему она равна. Теперь нам осталось найти P(один исход), и мы будем готовы решать целый класс задач!

Вычисляем вероятность желательного исхода

Нам осталось найти P(желательный исход), вероятность любого из интересных нам событий. До этого мы использовали P(желательный исход) как переменную для удобства записи, теперь надо вычислить ее значение. Рассмотрим вероятность получить два орла за пять бросков. Сосредоточимся на одном подходящем исходе: OOPPP. Мы знаем, что вероятность выкинуть орла за один бросок равна 1/2, но для большей общности обозначим ее как P(орел), не привязываясь к конкретному значению. Используя правило произведения и свойство отрицания, переформулируем задачу как

$$P(\text{орел, орел, не орел, не орел, не орел}).$$

Или, словами, «вероятность выкинуть орла, орла, не орла, не орла, не орла».

Свойство отрицания говорит, что P(не орел) = 1 – P(орел). Теперь достаточно применить правило произведения:

$$P(\text{орел, орел, не орел, не орел, не орел}) =$$
$$= P(\text{орел}) \times P(\text{орел}) \times (1 - P(\text{орел})) \times (1 - P(\text{орел})) \times (1 - P(P(\text{орел})).$$

Запишем проще, используя степени:

$$P(\text{орел})^2 \times (1 - P(\text{орел}))^3.$$

В итоге

$$P(\text{два орла за пять бросков}) = P(\text{орел})^2 \times (1 - P(\text{орел}))^3.$$

Заметим, что степени при $P(\text{орел})$ и $1 - P(\text{орел})$ — это просто количество орлов и не орлов, то есть k (число интересующих нас исходов) и $n - k$ (число попыток минус число интересующих нас исходов). Все вместе приводит к общей формуле, в которой нет конкретных чисел:

$$\binom{n}{k} \times P(\text{орел})^k \left(1 - P(\text{орел})\right)^{n-k}.$$

Мы не всегда говорим именно о вероятности выпадения орла, поэтому заменим $P(\text{орел})$ на p. Мы получили общее решение для числа интересующих нас исходов k, числа попыток n и вероятности одного интересующего нас исхода p:

$$B(k; n, p) = \binom{n}{k} \times p^k \times (1 - p)^{n-k}.$$

Имея такую формулу, мы в состоянии решить любую задачу про броски монетки. Например, посчитаем вероятность выбросить 12 орлов за 24 броска:

$$B\left(12; 24, \frac{1}{2}\right) = \binom{24}{12} \times \frac{1}{2}^{12} \times \left(1 - \frac{1}{2}\right)^{24-12} = 0{,}1612.$$

До того как вы узнали о биномиальном распределении, решить эту задачу было бы гораздо сложнее!

Функция, описывающая распределение, называется функцией вероятности (*PMF, probability mass function*). С ней мы можем вычислить вероятность для любого k при фиксированных n и p. Например, можно подставить все возможные значения k для 10 бросков монеты и изобразить биномиальное распределение, как на рис. 4.1.

Можно рассмотреть аналогичное распределение вероятностей выбросить шестерку при десяти бросках шестигранного кубика, показанное на рис. 4.2.

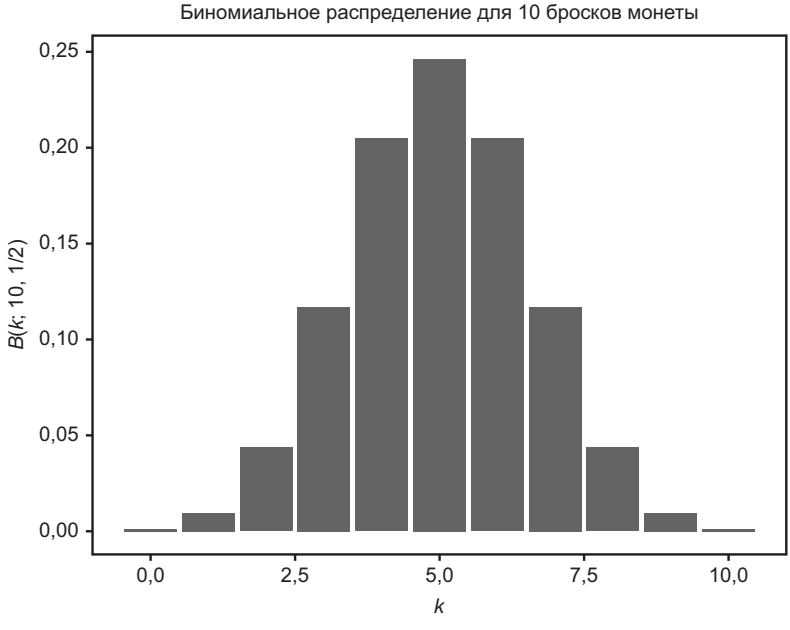


Рис. 4.1. Гистограмма вероятности получения k орлов из 10 бросков монеты

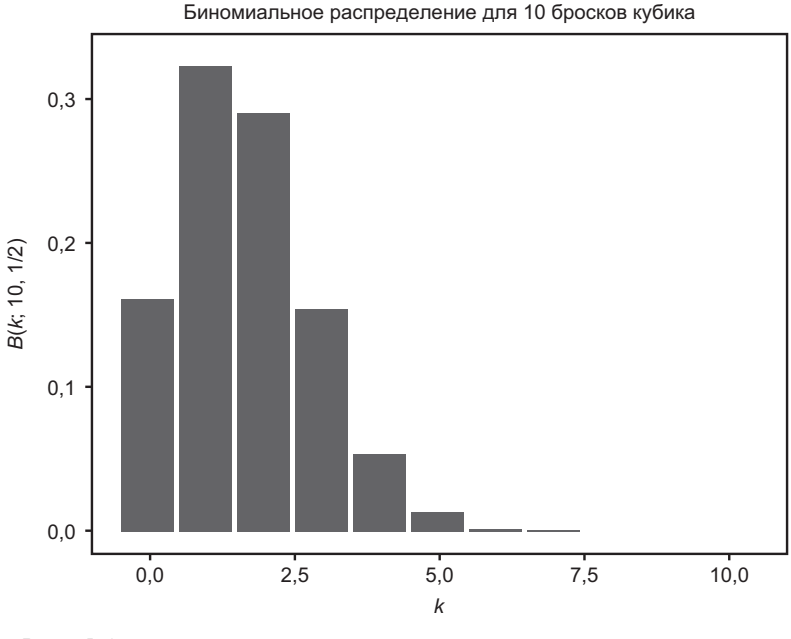


Рис. 4.2. Вероятность получения шестерки за 10 бросков кубика

Как мы видим, вероятностное распределение позволяет обобщить целый класс задач.

Теперь мы можем применять наше распределение в самых разных ситуациях. Но помните, что мы вывели его, пользуясь простыми правилами для вероятностей. Проверим, как оно работает.

Пример: игра «гача»

В Японии очень популярны разновидности мобильной игры «гача». Игрокам за игровую валюту нужно покупать виртуальные карты, которые выдаются случайным образом, и игрок не может повлиять на то, что получит. Карты неравноценны, и, как в автоматах, игрок стремится брать новые карты из колоды, пока не получит желаемую. Посмотрим, как биномиальное распределение поможет решить, рисковать или нет в такой игре.

Итак, вы установили новую игру «Байесовские бойцы». Текущий набор карт, из которых вы можете вытянуть новую, называют баннером. В нем есть и обычные карты, и так называемые суперкарты. Как можно догадаться, все персонажи карт в «Байесовских бойцах» — знаменитые специалисты по теории вероятностей и статистике. На баннере лежат следующие карты (рядом указана вероятность их вытащить):

- Томас Байес: 0,721 %;
- Эдвин Томпсон Джейнс: 0,720 %;
- Гарольд Джеффрис: 0,718 %;
- Эндрю Гельман: 0,718 %;
- Джон Крушке: 0,714 %.

Эти суперкарты соответствуют вероятности всего в 0,03591. Но общая вероятность равна 1, так что шанс вытянуть менее желанную карту составляет 0,96409. Мы также считаем, что колода бесконечна — то есть вытаскивание одной карты не меняет вероятностей, вытянутая карта «остается в колоде». С реальной колодой, если вы не возвращаете карту, дело обстояло бы иначе!

Вы очень хотите заполучить в свою элитную байесовскую коллекцию Эдвина Джейнса. Увы, чтобы тянуть карты, нужна валюта — байес-баксы. За одну карту вы обычно платите один байес-бакс, и сейчас проходит акция — 100 байес-баксов за 10 долларов. Вы не хотите тратить больше этой суммы,

да и ее — только если есть хороший шанс заполучить желанную карту. Вы купите байес-баксы, только если вероятность получить великолепного Джейнса будет не менее 0,5.

Конечно, мы можем подставить вероятность вытянуть Джейнса в нашу формулу для биномиального распределения:

$$\binom{100}{1} \times 0{,}00720^1 \times (1-0{,}00720)^{99} = 0{,}352.$$

Результат меньше 0,5, вроде бы, надо сдаться. Но стоп! Мы забыли, что в этой формуле мы считаем вероятность вытянуть только одну карточку с Джейнсом. А их может быть две или даже три! Так что нас интересует вероятность вытянуть одну или более карт, то есть

$$\binom{100}{1} \times 0{,}00720^1 \times (1-0{,}00720)^{99} + \binom{100}{2} \times 0{,}00720^2 \times (1-0{,}00720)^{98} +$$

$$+ \binom{100}{3} \times 0{,}00720^3 \times (1-0{,}00720)^{97} \ldots$$

и так далее, до 100 карт, которые можно купить на байес-баксы. Записывать всю сумму утомительно, воспользуемся специальным обозначением:

$$\sum_{k=1}^{100} \binom{100}{k} \times 0{,}00720^k \times (1-0{,}00720)^{n-k}.$$

Σ — обозначение суммы, число снизу показывает, откуда мы начинаем складывать, сверху — когда заканчиваем. Так что выше записана сумма всех значений биномиального распределения для k от 1 до n, при p, равном 0,00720.

Мы смогли упростить это выражение, но надо вычислить его значение. Не доставайте калькулятор — давайте воспользуемся языком R. В R можно вызвать функцию `pbinom()`, чтобы просуммировать значения плотности распределения для всех k. На рис. 4.3 показано, как мы используем `pbinom()` для этой задачи.

Функция `pbinom()` принимает три обязательных аргумента и один опциональный `lower.tail` (по умолчанию он равен `TRUE`). Когда четвертый аргумент равен `TRUE`, мы суммируем вероятности для всех k, *меньших или равных* первому аргументу. Когда `lower.tail` принимает значение `FALSE`,

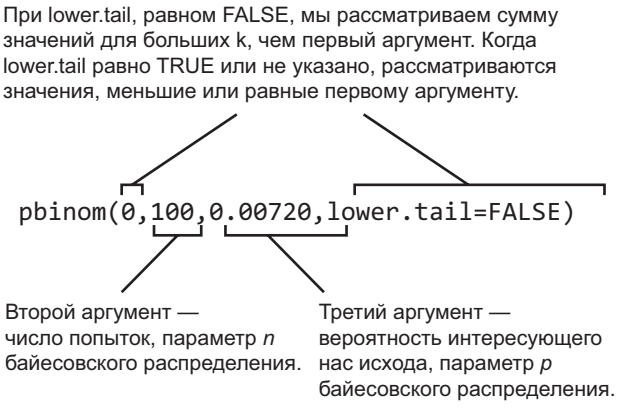


Рис. 4.3. Использование pbinom() для задачи о байесовских бойцах

мы суммируем вероятности для всех k, *строго больше* первого аргумента. Передав в качестве первого аргумента `0`, мы ищем вероятность получить не менее одной карты с Джейнсом. Мы присвоили `lower.tail` значение `FALSE`, так как это значит, что мы хотим перебирать значения, большие первого аргумента (а по умолчанию перебирали бы меньшие). Следующий аргумент соответствует n, числу попыток, а третий — p, вероятности успеха.

Подставив наши числа и `lower.tail`, равное `FALSE`, как показано на рис. 4.3, мы получим от R вероятность вытянуть за наши 100 байес-баксов *хотя бы одну* карту с Джейнсом:

$$\sum_{k=1}^{100}\binom{100}{k}\times 0{,}00720^{k}\times(1-p)^{n-k}=0{,}515.$$

Хотя вероятность вытянуть ровно одну карту с Эдвином Джейнсом составляет всего 0,352, вероятность получить не менее одной карты достаточно высока, чтобы рискнуть. Доставайте десять баксов и пополняйте команду элитных байесовцев!

Заключение

В этой главе мы научились применять правила вычисления вероятностей (а также комбинаторику), чтобы вывести общую формулу для целого класса задач. Любая задача, состоящая в определении вероятности k исходов за

n испытаний, где вероятность исхода равна p, легко решается с использованием биномиального распределения:

$$B(k;\, n,\, p) = \binom{n}{k} \times p^k \times (1-p)^{n-k}.$$

Как ни странно, но в выводе этой формулы нет ничего, кроме подсчета и применения базовых правил вероятности.

Упражнения

Чтобы убедиться, что вы понимаете биномиальное распределение, попробуйте ответить на эти вопросы.

1. Каковы параметры биномиального распределения для вероятности выкинуть один или двадцать на двадцатигранной кости, если бросить кость 12 раз?
2. В колоде из 52 карт четыре туза. Вы вытягиваете карту, возвращаете ее обратно, тасуете колоду и снова вытягиваете карту. Сколькими способами можно вытянуть только одного туза за пять попыток?
3. Продолжая предыдущую задачу: какова вероятность вытянуть пять тузов за десять попыток (помните, что карта возвращается в колоду!)?
4. При поиске новой работы полезно иметь больше одного предложения — это открывает возможность поторговаться. Пусть вероятность получить после собеседования предложение о работе равна1/5, и за месяц вы проходите семь собеседований. Какова вероятность, что к концу месяца вы получите хотя бы два предложения?
5. Вы получили немало писем от рекрутеров и обнаружили, что в следующем месяце у вас 25 собеседований. Ох, это утомительно, а вероятность получить предложение о работе, когда проходишь собеседование усталым, падает до 1/10. Вы готовы пройти 25 собеседований, только если это в два раза повысит вероятность получить хотя бы два предложения. Надо ли проходить 25 собеседований или остановиться на семи?

5

Бета-распределение

В этой главе на основе тех же идей, что стоят за биномиальным распределением из прошлой главы, мы вводим новое распределение — *бета-распределение*. Оно используется для оценки вероятности события, когда вы уже пронаблюдали некоторое количество испытаний и успешных исходов. Например, вы наблюдали 100 бросков монетки, из них 40 раз выпал орел — для оценки вероятности выпадения орла вы будете использовать бета-распределение.

Изучая бета-распределение, мы также обсудим разницу между теорией вероятностей и статистикой. В специализированных книгах вероятности событий часто явно заданы. В реальности так бывает редко. Но мы используем имеющиеся данные для оценки вероятностей, и поможет нам в этом статистика, которая позволяет давать оценки вероятностей по данным.

Странная история: получение данных

Представим такую ситуацию. Однажды вы заходите в магазин диковинок. Владелец любезно приветствует вас и, посмотрев, как вы слоняетесь по магазину, спрашивает, ищете ли вы что-то конкретное. Вы отвечаете, что хотите посмотреть на самую странную вещь в магазине. Он улыбается и вытаскивает черную коробочку размером с кубик Рубика, но невозможно

тяжелую. Вы заинтересованно спрашиваете, что это. Владелец указывает на прорези сверху и снизу и говорит: «Бросьте 25-центовую монету в верхнее отверстие — и снизу могут появиться две!» Решив попробовать, вы достаете монету из кармана, бросаете в коробочку, но ничего не происходит. Владелец замечает: «А иногда она просто съедает монетку! Коробочка у меня давно, но она никогда не оставалась без монеток и никогда не была заполнена так, чтобы монета не влезала». Вы в замешательстве, но, решив блеснуть свежими знаниями по теории вероятностей, спрашиваете: «А какова вероятность получить две монетки?» Владелец загадочно отвечает: «Не знаю. Это черный ящик, и инструкции к нему нет. Я знаю только, как он себя ведет. Иногда дает две монетки, а иногда съедает вашу».

Теория вероятностей, статистика и статистический вывод

Хотя задача с загадочной коробкой весьма необычна, на самом деле это чрезвычайно распространенный тип вероятностной задачи. До этого, кроме первой главы, мы знали вероятности всех возможных событий или хотя бы свою готовность поставить на них. В реальности мы почти никогда не знаем точной вероятности событий, у нас есть только наблюдения и данные. В этом и есть основное различие между теорией вероятностей и статистикой. В теории вероятностей мы точно знаем, какова вероятность всех событий, и интересуемся, насколько вероятно получить тот или иной результат наблюдений. Например, мы можем знать, что вероятность выкинуть орла при броске монеты равна 1/2, и мы можем интересоваться вероятностью получить ровно семь орлов за 20 бросков.

В статистике мы решаем обратную задачу: если вы наблюдаете, что за 20 бросков выпало семь орлов, какова вероятность выпадения орла при одном броске? Как можно видеть, в этом примере вероятности неизвестны. Статистика в каком-то смысле — теория вероятностей наоборот. Задача нахождения вероятностей по данным называется статистическим выводом и лежит в основе статистики.

Сбор данных

Главное в статистическом выводе — данные. Пока мы только один раз испробовали странную коробочку: бросили монетку и не получили ничего. В этот момент мы знаем только, что можем потерять деньги. Но владелец говорит, что мы можем и выиграть, однако уверенности в этом пока нет.

Мы хотим оценить вероятность того, что загадочная коробочка выдаст две монеты. Для этого надо провести еще несколько испытаний и посмотреть, насколько часто мы будем выигрывать. Владелец магазина также заинтересован и готов внести десять долларов 25-центовыми монетами — 40 монет, при условии, что вы отдадите ему выигрыш. Вы бросаете монетку, и, ура, выскакивают две! Теперь у нас есть результат двух испытаний: действительно, иногда коробочка выдает дополнительную монету, а иногда съедает брошенную. Можно наивно предположить, что если вы один раз проиграли и один раз выиграли, то $P(\text{две монеты}) = 1/2$. Но данных слишком мало, чтобы понять, насколько часто коробочка выдает лишнюю монету.

Желая собрать побольше данных, вы израсходовали все 40 монет. В итоге, с учетом первого опыта, получилось следующее:

14 выигрышей;
27 проигрышей.

Быть может, вам хочется теперь изменить мнение с $P(\text{две монеты}) = 1/2$ на $P(\text{две монеты}) = 14/41$. Но означают ли новые данные, что первоначальная догадка не может быть верной?

Вычисляем вероятность вероятностей

Чтобы решить эту задачу, рассмотрим две возможные вероятности — наши гипотезы о том, как часто коробочка возвращает две монеты:

$$P(\text{две монеты}) = 1/2,\ P(\text{две монеты}) = 14/41.$$

Присвоим обозначение каждой гипотезе:

$$H_1 \text{ — это } P(\text{две монеты}) = 1/2;$$
$$H_2 \text{ — это } P(\text{две монеты}) = 14/41.$$

Большинство людей интуитивно скажут, что гипотеза H_2 вероятнее, так как более точно соответствует наблюдениям. Но это надо доказать математически! Рассмотрим задачу в контексте того, насколько хорошо каждая гипотеза объясняет наблюдения, а проще говоря, «насколько вероятно то, что мы наблюдаем, при H_1? А при H_2? Оказывается, мы можем легко вычислить это, применив биномиальное распределение из главы 4. Мы знаем, что $n = 41$, $k = 14$, и примем пока, что p соответствует H_1 или

H_2. Обозначим наши данные через D. Подставив числа в формулу биномиального распределения (напомним, что ее можно найти в главе 4), мы получим такие результаты:

$$P(D \mid H_1) = B\left(14; 41, \frac{1}{2}\right) \approx 0{,}016,$$

$$P(D \mid H_2) = B\left(14; 41, \frac{14}{41}\right) \approx 0{,}130.$$

Иными словами, если верна гипотеза H_1 и вероятность получить две монеты равна 1/2, то вероятность 14 случаев получения двух монет за 41 попытку составляет 0,016.

Но если верна гипотеза H_2 и вероятность получить две монеты равна 14/41, то вероятность такого же результата наблюдений составляет 0,130.

Таким образом, при наших данных (14 случаев получения двух монет за 41 попытку) H_2 почти в 10 раз вероятнее, чем H_1. Но также мы показали, что обе гипотезы возможны, и, конечно же, можно выдвинуть много других. Например, можно выдвинуть гипотезу $H_3 P$(две монеты) = 15/42. В поисках закономерности мы можем проверять каждую вероятность от 0,1 до 0,9 с шагом 0,1, вычисляя вероятность наблюдаемых данных для каждого распределения, и, исходя из этого, строить гипотезы. Рисунок 5.1 показывает все такие значения вероятности.

Все возможные гипотезы рассмотреть нельзя — их бесконечно много. Но можно проверить больше распределений и получить больше информации. Повторим эксперимент, проверяя все вероятности от 0,01 до 0,99 с шагом всего 0,01. И получим результаты с рис. 5.2.

Хотя мы и не можем проверить все возможные гипотезы, явно прослеживается закономерность: что-то похожее на распределения вероятностей для поведения черной коробочки. Это ценная информация, и легко увидеть, где вероятность выше. Но наша цель — оценить уверенность во всех возможных гипотезах (распределение вероятностей на множестве гипотез). У нашего подхода две проблемы. Во-первых, гипотез все же бесконечно много, и каким бы маленьким мы ни делали шаг, всех возможностей не перебрать (неохваченных останется бесконечно много). Это не столь важно на практике, нас часто не заботят значения вроде 0,000001 или 0,0000011, но все же хотелось бы точнее представлять весь спектр гипотез. Посмотрев на график, вы заметите вторую, более важную проблему: по крайней мере

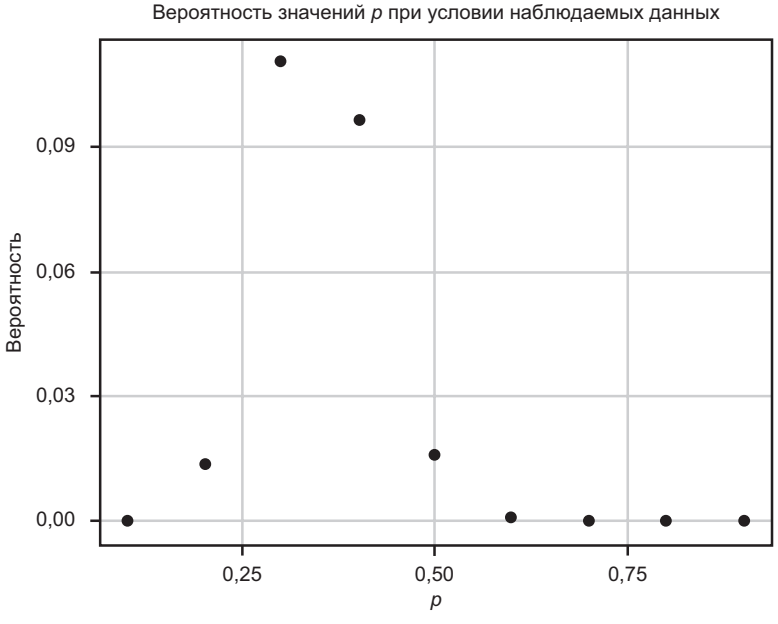


Рис. 5.1. Диаграмма гипотез о шансах получить две монеты

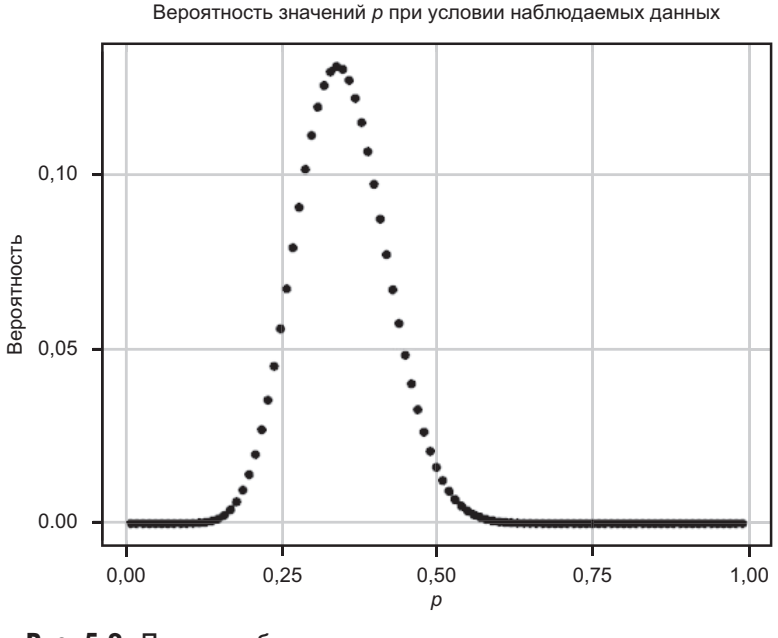


Рис. 5.2. Проверив больше гипотез, мы видим закономерность

10 точек лежат выше 0,1, а ведь нам не хватает еще бесконечного множества точек! Таким образом, наши вероятности в сумме не дают 1! Но правила гласят, что вероятности всех гипотез должны в сумме давать 1. Если это не так, часть гипотез не учтена, либо, если сумма больше 1, нарушается правило о том, что вероятность лежит между 0 и 1. Сумма должна быть равна 1 даже при бесконечном числе гипотез! И тут на сцену вступает бета-распределение.

Бета-распределение

Справиться с этими задачами нам поможет бета-распределение. В отличие от биномиального распределения, распадающегося на дискретный набор значений, бета-распределение определено на сплошном интервале, что позволяет представить все бесконечное множество гипотез.

Определим бета-распределение через плотность вероятности (*probability density function*, *PDF*), очень похожую на функцию вероятности для биномиального распределения, но определенную на сплошном интервале. Плотность вероятности бета-распределения выглядит так:

$$\text{Beta}(p;\ \alpha,\ \beta) = \frac{p^{\alpha-1} \cdot (1-p)^{\beta-1}}{\text{beta}(\alpha,\ \beta)}.$$

Формула выглядит пугающе в отличие от формулы биномиального распределения! Но на самом деле различаются они не столь сильно. Не будем выводить ее с нуля, как функцию вероятности биномиального распределения, но что происходит, разберемся.

Разбираемся с плотностью распределения

Посмотрим на параметры: p, α (строчная греческая буква «альфа») и β (строчная греческая буква «бета»).

- P обозначает вероятность события, что соответствует разным гипотезам о вероятности выигрыша у черного ящичка.
- α показывает, сколько раз произошло интересующее нас событие, например получение двух монет.
- β — сколько раз оно не произошло (в нашем примере — сколько раз коробочка съела монету).

Общее число испытаний равно $\alpha + \beta$. Здесь видна разница с биномиальным распределением, где имеется k интересных нам исходов и конечное число n испытаний.

Числитель плотности распределения выглядит знакомым — он почти совпадает с функцией вероятности биномиального распределения, которая выглядит как

$$B(k;\, n,\, p) = \binom{n}{k} \times p^k \times (1-p)^{n-k}.$$

Но в плотности распределения на месте $p^k(1-p)^{n-k}$ стоит $p^{\alpha-1}(1-p)^{\beta-1}$, мы вычитаем 1 из показателей степени. В знаменателе стоит другая функция, бета-функция (заметьте, что ее обозначение начинается со строчной буквы), в честь которой и названо бета-распределение. Мы вычитаем 1 из показателя степени и делим на бета-функцию для нормализации (то есть для того, чтобы распределение суммировалось в 1). Бета-функция — это интеграл от 0 до 1 от $p^{\alpha-1}(1-p)^{\beta-1}$. Мы поговорим об интегралах дальше, а сейчас можно думать о нем как о сумме всех возможных значений $p^{\alpha-1}(1-p)^{\beta-1}$ при p, принимающем значения от 0 до 1. Обсуждение, почему вычитание единицы из показателя и деление на бета-функцию приводит к нормализации, сильно выходит за пределы этой главы, но сейчас достаточно знать, что они позволяют всем значениям суммироваться в 1 и, таким образом, дают нам разумное определение вероятности. В итоге получается функция, описывающая вероятности всех возможных гипотез о шансах получить две монеты — при условии, что мы наблюдали α примеров одного исхода и β примеров другого. Помните, что к бета-распределению мы пришли, сравнив, насколько хорошо разные биномиальные распределения, каждое со своей собственной вероятностью p, описывают наши данные. Другими словами, бета-распределение показывает, насколько хорошо все возможные биномиальные распределения описывают наблюдаемые данные.

Применение плотности вероятности к задаче

Подставив значения наших данных о черном ящичке и изобразив бета-распределение (как на рис. 5.3), мы видим, что это просто гладкая версия рис. 5.2. Так выглядит плотность вероятности Beta(14, 27).

Как видите, большие значения плотности соответствуют значениям p, меньшим 0,5, — что ожидаемо, ведь мы получали две монеты меньше чем в половине случаев. Видно, что вероятность получить две монеты хотя бы

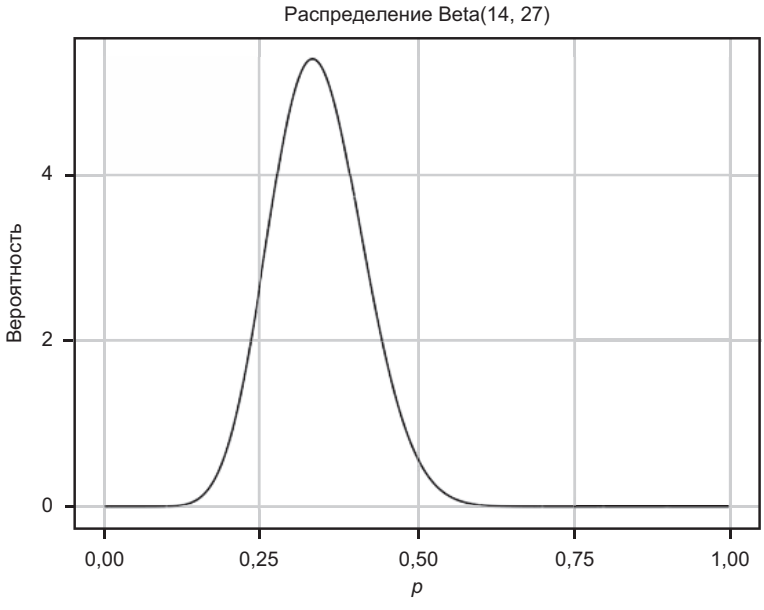


Рис. 5.3. Бета-распределение для данных о черной коробочке

в половине случаев очень мала, поэтому стоить заканчивать пихать монеты в коробочку. Не успев потратить слишком много, мы все же выяснили, что потерять деньги вероятнее, чем заработать. Мы смотрим на график, видим распределение для наших гипотез и можем точно ответить, насколько уверены, что шансы получить две монеты меньше 0,5, воспользовавшись азами матанализа и языком R.

Интегрируем непрерывные распределения

Бета-распределение принципиально отличается от биномиального: в последнем мы ищем распределение k, число интересующих нас исходов, которое легко посчитать. Однако в случае бета-распределения мы имеем дело с распределением параметра p, который может принимать бесконечно много значений. Это приводит к следующей задаче, знакомой тем, кто уже изучал матанализ (но не пугайтесь, если у вас не было такого опыта!). В примере с $\alpha = 14$ и $\beta = 27$ мы хотим узнать, какова вероятность, что шансы получить две монеты равны $1/2$. В случае биномиального распределения число возможных исходов конечно, и легко найти вероятность одного конкретного исхода. Для непрерывного распределения все сложнее. Мы знаем основное

правило — сумма всех значений вероятности должна быть равной 1, но каждое отдельное значение бесконечно мало — вероятность фактически равна 0. Для тех, кто не знаком с непрерывными функциями, это все прозвучало странно, так что понадобится небольшое пояснение. Пусть нечто составлено из бесконечного числа кусочков — представьте, например, большую шоколадку весом в один фунт (453 г). Вы делите ее на два куска — каждый весит по 1/2 фунта. Если кусков будет 10, каждый будет весить 1/10 фунта. Чем больше кусочков, тем меньше каждый — вы их уже и не увидите. Когда число кусочков стремится к бесконечности, каждый из них фактически исчезает!

Кусочки шоколада исчезли, но общая масса осталась. Даже поделив плитку на бесконечное число кусочков, мы можем сложить веса всех кусочков в одной половине шоколадки. Аналогично, рассуждая о вероятности для непрерывного распределения, по-прежнему можно суммировать значения из некоторого интервала. Но разве мы не получим 0, когда каждое конкретное значение равно 0? Здесь и возникает интегральное исчисление: способ суммировать бесконечно маленькие значения называется *интегрированием*. Желая узнать, меньше ли вероятность получить две монеты, чем 0,5 (принимает ли она значение от 0 до 0,5), мы вычисляем

$$\int_0^{0,5} \frac{p^{14-1} \times (1-p)^{27-1}}{\text{beta}(14,\ 27)}.$$

Пояснение для тех, кто не знаком с матанализом: вытянутая S — аналог значка $\sum$, применяющегося не к дискретным, а к непрерывным функциям. Таким образом, мы просто хотим просуммировать все «кусочки» функции (в приложении Б можно найти краткое изложение основных постулатов матанализа). Не пугайтесь формул — считать все равно будет R. В нем есть функция `dbeta()` — плотность вероятности для бета-распределения. Она принимает три аргумента, соответствующие p, α и β. Мы также воспользуемся функцией `integrate()` для интегрирования. Так мы посчитаем вероятность того, что шансы получить две монеты меньше 0,5:

```
> integrate(function(p) dbeta(p, 14, 27),0, 0,5)
```

Результат:

```
0,9807613 with absolute error < 5,9e-06
```

Сообщение указывает максимальное значение допущенной ошибки — ведь компьютеры не умеют вычислять интегралы с идеальной точностью, но

ошибки обычно так малы, что беспокоиться не о чем. Таким образом, при наших данных с вероятностью 0,98 истинная вероятность получить две монеты меньше 0,5. Так что продолжать бросать монеты почти наверняка невыгодно.

Реверс-инжиниринг игры «гача»

В реальной жизни мы практически никогда не знаем настоящих вероятностей событий. Поэтому бета-распределения — один из главных инструментов для понимания данных. В игре «гача» из главы 4 вероятность вытянуть интересную карту была известна. Но на самом деле разработчики игр почти никогда не дают игрокам такой информации — по многим причинам (например, чтобы игроки не поняли, насколько маловероятно вытянуть нужную карту). Обратимся к новой игре — «Бойцы-фреквентисты»[1]! Снова с картами знаменитых статистиков. Теперь мы охотимся за картой Брэдли Эфрона.

Мы не знаем, каковы наши шансы, но хотим вытянуть эту карту, а лучше и не одну. Потратив немало денег и вытянув 1200 карт, мы получаем лишь 5 карт с Эфроном. Наш друг тоже хотел бы сыграть, но готов тратить деньги только в том случае, если с вероятностью более 0,7 шансы вытянуть Эфрона больше 0,005. Поэтому он просит нас посчитать, стоит ли ему играть. Данные говорят, что из 1200 карт только 5 были с Эфроном — поэтому мы обращаемся к распределению Beta(5, 1195), изображенному на рис. 5.4 (как мы помним, $\alpha + \beta$ — это общее число вытянутых карт).

На графике видим, что практически вся плотность вероятности сосредоточена при p, меньших 0,01. Надо найти, сколько приходится на интервал от 0,005 — как и раньше, для этого достаточно проинтегрировать в R:

```
integrate(function(x) dbeta(x,5,1195), 0,005, 1)
0,29
```

Таким образом, вероятность того, что шансы вытянуть карту с Брэдли Эфроном не меньше 0,005 — при наших данных, — всего 0,29. Друг же согласен играть лишь при вероятности не менее 0,7, так что, по нашим данным, ему не стоит и пытаться.

---

[1] Фреквентистский подход к статистике: вероятность — это предел частоты при увеличении числа экспериментов. — *Примеч. ред.*

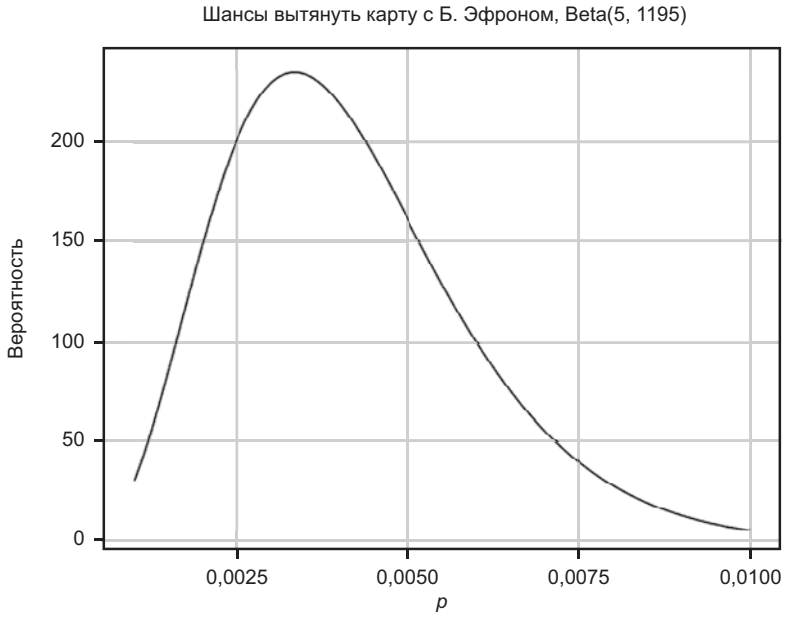


Рис. 5.4. Бета-распределение шансы получить карту с Брэдли Эфроном при наших данных

Заключение

В этой главе мы познакомились с бета-распределением, тесно связанным с биномиальным и при этом во многом не похожем на него. Мы пришли к бета-распределению, наблюдая, насколько хорошо все большее и большее число биномиальных распределений объясняют имеющиеся данные. Чтобы описать бесконечное число возможных гипотез, требуется непрерывное распределение. Бета-распределение описывает, насколько мы уверены в каждой из возможных гипотез об имеющихся данных. Таким образом, мы можем производить статистические выводы на основе данных — определять, какие вероятности мы присвоим событиям и насколько можно быть в них уверенными (каковы вероятности вероятностей!).

Главное отличие бета-распределения от биномиального — в его непрерывности. Так как оно определено в бесконечном числе точек, мы не можем просто суммировать все значения, как при дискретном распределении. Вместо этого приходится применять математический анализ — но, к счастью, для вычисления интегралов можно использовать R.

Упражнения

Чтобы убедиться, что вы понимаете бета-распределение, попробуйте ответить на эти вопросы.

1. Вы хотите использовать бета-распределение, чтобы определить, честная ли монетка — то есть равны ли для нее вероятности выкинуть орел и решку. Вы подбрасываете монетку 10 раз и получаете 4 орла и 6 решек. Используя бета-распределение, найдите вероятность того, что орел выпадает в более чем 60 % бросков.
2. Вы еще 10 раз подбрасываете монетку и в итоге получаете 9 орлов и 11 решек. Какова вероятность того, что монетка честная, используя наше определение честности плюс-минус 5 %?
3. Данные — лучший способ убедиться в верности своих утверждений. Вы еще 200 раз подбрасываете монетку и в итоге получаете 109 орлов и 111 решек. Какова теперь вероятность того, что монетка честная (плюс-минус 5 %)?

ЧАСТЬ II

БАЙЕСОВСКИЕ И АПРИОРНЫЕ ВЕРОЯТНОСТИ

6

Условная вероятность

Пока мы обсуждали только вероятности *независимых* событий. События независимы, если исход одного не влияет на исход второго. Например, выпадение орла при броске монеты никак не влияет на то, выпадет ли шестерка на кубике. Подсчет вероятностей независимых событий гораздо проще, чем зависимых, но предположение о независимости часто не отражает реального положения дел. Например, вероятности того, что не прозвенит будильник, и того, что вы не опоздаете на работу, независимыми *не* являются. При несработавшем будильнике ваши шансы опоздать гораздо больше.

В этой главе вы научитесь обращаться с условными вероятностями — зависящими от исхода некоторых событий. Вы также познакомитесь с одним из важнейших приложений условной вероятности — теоремой Байеса.

Определение условной вероятности

Наш первый пример условной вероятности будет посвящен прививкам от гриппа и их побочным эффектам. Обычно пациент получает информацию о возможных побочных эффектах — одним из них является повышенный риск синдрома Гийена — Барре (СГБ), очень редкого и опасного для жизни состояния, при котором иммунная система атакует собственные

нервные клетки. По данным Центра по контролю и профилактике заболеваний США (*Centers for Disease Control and Prevention, CDC*), СГБ возникает в год у двух из 100 000 жителей. Мы можем представить эту вероятность так:

$$P(\text{СГБ}) = \frac{2}{100\ 000}.$$

Обычно повышение риска СГБ из-за прививки ничтожно мало. Однако в 2010 году, во время вспышки свиного гриппа, вероятность возникновения СГБ после прививки возросла до 3/100 000. Таким образом, вероятность развития СГБ стала ощутимо зависеть от наличия прививки — пример условной вероятности. Условные вероятности записываются так: $P(A|B)$, вероятность A при условии B. Можно математически записать вероятность развития СГБ как

$$P(\text{СГБ} \mid \text{прививка от гриппа}) = \frac{3}{100\ 000}.$$

В переводе на человеческий язык: «Вероятность развития СГБ, если вам сделали прививку от гриппа, 3 из 100 000».

Почему условные вероятности важны

Условные вероятности — важнейший инструмент статистики, они позволяют показать, как наши представления изменяются на основании поступившей информации. Если вы не знаете, прививался ли некто от гриппа, можно сказать, что вероятность развития у него синдрома Гийена — Барре 2/100 000 (это вероятность для произвольно выбранного человека). Но если дело происходит в 2010 году и вы знаете, что человек делал прививку, то, как вам известно, вероятность составляет уже 3/100 000. Рассмотрим отношение этих вероятностей:

$$\frac{P(\text{СГБ} \mid \text{прививка от гриппа})}{P(\text{СГБ})} = 1{,}5.$$

Так что, если вы прививались от гриппа в 2010-м, ваш риск синдрома Гийена — Барре, скорее всего, на 50 % выше, чем у случайного человека. К счастью, он все равно очень низок — если нас волнует индивидуальный риск. Но если мы посмотрим на все население, то среди привитых будет на 50 % больше людей с СГБ, чем в целом по популяции.

Впрочем, есть и другие факторы, повышающие риск СГБ. Например, вероятность выше у мужчин и у пожилых людей. Используя условные вероятности, мы можем учесть всю эту информацию и лучше оценить риск СГБ для индивида.

Зависимость: пересматриваем правила

В качестве второго примера условной вероятности рассмотрим дальтонизм — нарушение цветовосприятия. В популяции 4,25 процента дальтоников, и в большинстве случаев причина в наследственности. Дальтонизм возникает из-за дефекта в одном из генов *X*-хромосомы. Так как у мужчин одна *X*-хромосома, а у женщин — две, мужчины в 16 раз чаще страдают дальтонизмом из-за дефектного гена. Так что, хотя в целом по популяции доля дальтоников составляет 4,25 %, среди женщин их всего 0,5 %, а среди мужчин — 8 %. Во всех наших вычислениях мы будем для простоты предполагать, что доля мужчин и женщин в популяции одинакова. Запишем известные нам факты через условные вероятности:

$$P(\text{дальтоник}) = 0{,}0425;$$
$$P(\text{дальтоник}\,|\,\text{женщина}) = 0{,}005;$$
$$P(\text{дальтоник}\,|\,\text{мужчина}) = 0{,}08.$$

Выберем из популяции случайного человека — какова вероятность, что это мужчина-дальтоник?

В главе 3 мы научились комбинировать вероятности с использованием И по правилу произведения. Согласно правилу произведения, результат должен был бы составлять:

$$P(\text{мужчина, дальтоник}) = P(\text{мужчина}) \times P(\text{дальтоник}) = \\ = 0{,}5 \times 0{,}0425 = 0{,}02125.$$

Но при использовании правила произведения для условных вероятностей возникают проблемы. Это заметно при попытке найти вероятность того, что выбрана женщина-дальтоник:

$$P(\text{женщина, дальтоник}) = P(\text{женщина}) \times P(\text{дальтоник}) = 0{,}5 \times 0{,}0425 = \\ = 0{,}02125.$$

Получились равные вероятности — такого не может быть! Мы знаем, что, хотя вероятности выбрать мужчину и женщину равны, но при выборе

женщины вероятность дальтонизма у нее должна быть сильно ниже, чем у мужчины. Формула должна учитывать, что при выборе случайного человека вероятность дальтонизма зависит от того, мужчина это или женщина. Правило произведения из главы 3 работает только для независимых вероятностей. Но принадлежность к тому или иному полу и дальтонизм — события зависимые. Так что на самом деле вероятность выбрать мужчину-дальтоника — это вероятность выбрать мужчину, умноженная на вероятность того, что мужчина — дальтоник. Это можно записать формулой

$$P(\text{мужчина, дальтоник}) = P(\text{мужчина}) \times P(\text{дальтоник} \mid \text{мужчина}) = \\ = 0{,}5 \times 0{,}08 = 0{,}04.$$

В общем случае правило произведения меняется так:

$$P(A, B) = P(A) \times P(B \mid A).$$

Такое определение работает и для независимых вероятностей — для них $P(B) = P(B \mid A)$. Это интуитивно понятно: представьте подбрасывание монетки и кубика, где $P(\text{шестерка})$ равна $1/6$ независимо от того, какой стороной выпала монета, так что $P(\text{шестерка} \mid \text{орел})$ также равна $1/6$.

Обновится и формулировка правила суммы:

$$P(A \text{ ИЛИ } B) = P(A) + P(B) - P(A) \times P(B \mid A).$$

Теперь можно работать с условными вероятностями, используя правила вероятностной логики из части I.

Рассуждая об условных вероятностях и зависимости, важно помнить, что на практике связь между двумя событиями часто неясна. Например, рассмотрим вероятности, что человек владеет грузовиком и что он добирается на работу дольше часа. Можно придумать множество причин, по которым одно могло бы зависеть от другого — быть может, владельцы грузовиков чаще живут в сельской местности и далеко не ездят, — но данных, чтобы это подтвердить, у нас нет. Предположение о независимости событий (даже если на самом деле это не так) — обычная практика в статистике. Но иногда, как в нашем примере с мужчинами-дальтониками, такое предположение приводит к грубым ошибкам. Так что, хотя часто и приходится предполагать независимость, помните о том, какой эффект может оказать наличие зависимости!

Переворачиваем условную вероятность: теорема Байеса

Один из самых замечательных трюков при работе с условными вероятностями — перемена мест условия и зависящего от него события, то есть использование вероятности $P(A|B)$ для вычисления $P(B|A)$. Допустим, вы пишете продавцу из компании, продающей очки для дальтоников. Очки весьма дороги, и вы опасаетесь, что они бесполезны. Тот отвечает: «Я сам дальтоник и ношу такие очки — они отлично работают!»

Найдем вероятность, что продавец — мужчина. К сожалению, у нас нет о нем никакой информации, кроме идентификационного номера. Что делать? Мы знаем, что $P(\text{дальтоник}|\text{мужчина}) = 0{,}08$, а $P(\text{дальтоник}|\text{женщина}) = 0{,}005$, как определить $P(\text{мужчина}|\text{дальтоник})$? Интуитивно мы понимаем: скорее всего, это мужчина. Но как вычислить вероятность? К счастью, имеющейся информации достаточно. Мы ищем вероятность, что человек, страдающий дальтонизмом, является мужчиной:

$$P(\text{мужчина}|\text{дальтоник}) = ?$$

Главное в байесовской статистике — данные, но данные, кроме известных нам вероятностей, состоят только из одного факта: продавец страдает дальтонизмом.

Рассмотрим теперь из всего населения только дальтоников и определим, какова среди них доля мужчин. Для простоты введем переменную N — численность населения. Как мы уже говорили, надо найти, сколько дальтоников среди населения. Мы знаем $P(\text{дальтоник})$, так что можем записать:

$$P(\text{мужчина} \mid \text{дальтоник}) = \frac{?}{P(\text{дальтоник}) \times N}.$$

Теперь вычислим количество мужчин-дальтоников. Это легко: мы знаем $P(\text{мужчина})$ и $P(\text{мужчина}|\text{дальтоник})$ и можем пользоваться уточненным правилом произведения. Так что мы просто умножаем вероятность на численность населения:

$$P(\text{мужчина}) \times P(\text{мужчина} \mid \text{дальтоник}) \times N.$$

Итак, вероятность того, что продавец — мужчина, если он дальтоник, равна

$$P(\text{мужчина} \mid \text{дальтоник}) = \frac{P(\text{мужчина}) \times P(\text{мужчина} \mid \text{дальтоник}) \times N}{P(\text{дальтоник}) \times N}.$$

Численность населения N присутствует и в числителе, и в знаменателе, так что ее можно сократить:

$$P(\text{мужчина} \mid \text{дальтоник}) = \frac{P(\text{мужчина}) \times P(\text{мужчина} \mid \text{дальтоник})}{P(\text{дальтоник})}.$$

Теперь мы знаем все:

$$P(\text{мужчина} \mid \text{дальтоник}) = \frac{P(\text{мужчина}) \times P(\text{мужчина} \mid \text{дальтоник})}{P(\text{дальтоник})} =$$

$$= \frac{0{,}5 \times 0{,}08}{0{,}0425} = 0{,}941.$$

С вероятностью 94,1 % представитель — мужчина.

Теорема Байеса

В формуле выше нет ничего специфического для описания дальтонизма — она обобщается для вероятностей любых событий A и B. Таким образом, мы приходим к главной формуле этой книги, теореме Байеса:

$$P(A \mid B) = \frac{P(A)P(B \mid A)}{P(B)}.$$

Чтобы понять, чем так важна теорема Байеса, сформулируем задачу в общем виде. У нас есть какие-то представления о мире. При наблюдениях условная вероятность показывает, насколько *увиденное вероятно при условии наших представлений*:

$$P(\text{наблюдения} \mid \text{представления}).$$

Предположим, что вы верите в глобальное потепление и ожидаете, что в вашем регионе в течение 10 лет засух будет больше, чем ранее. Ваши представления основаны на том, что происходит глобальное потепление, а ваши наблюдения — это количество засух; пусть их было 5 за 10 лет. Определить вероятность того, что за 10 лет будет 5 засух при условии глобального потепления, весьма сложно. Можно спросить у эксперта-климатолога

о вероятности засух в случае глобального потепления. Но пока мы задались только вопросом, какова вероятность наших наблюдений, если мы верим в наличие глобального потепления. На самом деле мы хотим понять, насколько уверенными можно быть в потеплении — при условии имеющихся данных. Теорема Байеса позволяет нам «обратить» $P(\text{наблюдения} \mid \text{представления})$, которую мы узнали у климатолога, и найти вероятность правильности своих представлений при условии имеющихся наблюдений $P(\text{представления} \mid \text{наблюдения})$.

В этом примере теорема Байеса позволит получить из наблюдений за пятью засухами за десятилетие меру нашей уверенности в глобальном потеплении на основании этих данных. Вся необходимая для этого дополнительная информация — обычная вероятность пяти засух за десять лет (которую можно оценить по историческим данным) и мера первоначальной уверенности в наличии потепления. Последняя будет разной для разных людей, но теорема Байеса позволяет оценить, насколько данные изменят наши представления. Например, после сообщения эксперта, что 5 засух за 10 лет очень вероятны при условии глобального потепления, большинство людей начнет чуть более склоняться к его наличию — скептики ли они или такие борцы против изменения климата, как Ал Гор.

Но пусть эксперт отвечает, что 5 засух за 10 лет при условии глобального потепления маловероятны. Ваша первоначальная уверенность в его наличии несколько ослабеет. Именно в соответствии с теоремой Байеса данные изменяют наши исходные представления. Теорема Байеса позволяет нам получить из данных и исходных представлений о мире оценку нашей уверенности в своих представлениях при этих данных. Часто наши представления $P(A)$ в теореме Байеса взяты «с потолка». Мы ожесточенно спорим, уменьшит ли больший контроль продажи оружия число насильственных преступлений, помогает ли тестирование улучшению качества образования и хороша ли очередная реформа здравоохранения. Но мы редко думаем, как нас — или оппонентов — должны переубеждать данные. Теорема Байеса помогает понять, как данные меняют нашу уверенность в той или иной идее.

Далее мы увидим, как сравнивать вероятность предположений, и увидим, как иногда данные не заставляют нас изменить свое мнение (впрочем, кто же из нас не знает это из споров с родственниками!).

В следующей главе мы еще немного поговорим о теореме Байеса. Мы выведем ее снова с помощью кубиков *Lego*; разберемся, как она работает и как смоделировать наши априорные предположения и их изменение.

Заключение

В этой главе вы познакомились с условными вероятностями, то есть вероятностями событий, зависящих от других событий. Работать с ними сложнее, чем с вероятностями независимых событий, — понадобилось учесть зависимость в правиле произведения. Зато мы получили теорему Байеса, главный инструмент для понимания, как менять наши представления о мире на основании данных.

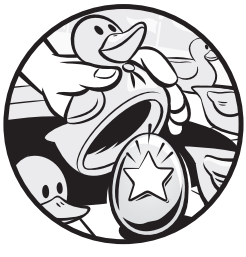
Упражнения

Чтобы убедиться, что вы понимаете условные вероятности и теорему Байеса, попробуйте ответить на эти вопросы.

1. Мы хотим использовать теорему Байеса для определения вероятности того, что в 2010 году пациент с синдромом Гийена — Барре был привит от гриппа. Какая информация нам нужна?
2. Какова вероятность того, что случайно выбранный из всей популяции человек — женщина и не дальтоник?
3. Какова вероятность того, что мужчина, привитый от гриппа в 2010 году, будет страдать либо от дальтонизма, либо от синдрома Гийена — Барре?

7

Теорема Байеса и Lego



В предыдущей главе мы познакомились с условными вероятностями и пришли к важнейшей из идей теории вероятностей — теореме Байеса, гласящей:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}.$$

Заметим, что по сравнению с главой 6 мы внесли одно маленькое изменение — вместо $P(A)P(B|A)$ написали $P(B|A)P(A)$ — результат не поменялся, но иногда перемена мест множителей упрощает понимание. С помощью теоремы Байеса мы можем «обращать» условные вероятности — зная вероятность $P(B|A)$, вычислять $P(A|B)$. Теорема Байеса лежит в основе статистики потому, что позволяет перейти от вероятности наблюдения при условии неких априорных предположений к мере нашей уверенности в этом предположении при условии наблюдений. Например, зная вероятность чихания при простуде, можно определить вероятность, что вы простужены, если чихнули. Таким образом, мы используем наблюдения для обновления представлений о мире.

В этой главе мы будем использовать Lego для наглядного и конкретного объяснения теоремы Байеса. Возьмем кирпичики и будем задавать вопросы. На рис. 7.1 изображен прямоугольник 6×10 — с 60 «выступами», которыми соединяются кирпичики.

Рис. 7.1. Прямоугольник 6 × 10 из Lego, выступающий в роли пространства возможных событий

Его можно представить как пространство из 60 возможных взаимоисключающих событий. Например, синие выступы могут представлять 40 студентов (в группе из 60 человек), сдавших экзамен, а красные — 20 студентов, не сдавших его. В прямоугольнике с 60 выступами 40 синих, так что если мы ткнем пальцем в случайное место, то попадем на синий кирпичик с вероятностью

$$P(\text{синий}) = \frac{40}{60} = \frac{2}{3}.$$

Вероятность ткнуть в красный кирпичик:

$$P(\text{красный}) = \frac{20}{60} = \frac{1}{3}.$$

Вероятность ткнуть либо в синий, либо в красный кирпичик ожидаемо равна 1:

$$P(\text{синий}) + P(\text{красный}) = 1,$$

то есть синий и красный кирпичики вместе представляют все множество возможных событий.

Теперь положим сверху желтый кирпичик, представляющий некоторое новое множество — например, студентов, которые готовились всю ночь и не спали. Получится конструкция с рис. 7.2.

Рис. 7.2. Положим кирпичик 2 × 3 на прямоугольник 6 × 10

Теперь, если мы ткнем в случайный выступ, вероятность попасть на желтый кирпичик равна

$$P(\text{желтый}) = \frac{6}{60} = \frac{1}{10}.$$

Но если сложить $P(\text{желтый}) + P(\text{синий}) + P(\text{красный})$, мы получим вроде бы невозможный результат, больший 1! Дело, конечно же, в том, что желтый кирпичик лежит поверх красного и синего, так что вероятность ткнуть в желтый кирпичик — условная, зависящая от того, над красной или синей областью мы оказались. Как мы знаем из предыдущей главы, эту условную вероятность можно записать как $P(\text{желтый} \mid \text{красный})$ — вероятность желтого при условии, что мы оказались над красной областью. В примере выше это будет вероятностью, что студент не спал всю ночь при условии, что он не сдал экзамен.

Наглядное представление условных вероятностей

Вернемся к кирпичикам Lego и найдем $P(\text{желтый} \mid \text{красный})$. Рисунок 7.3 поможет в визуализации.

Рис. 7.3. Наглядно представляем P(желтый | красный)

Рассмотрим весь процесс нахождения $P(\text{желтый} \mid \text{красный})$ по нашей наглядной модели:

1. Разделим красную и синюю области.
2. Вычислим площадь красной области: $2 \times 10 = 20$ выступов.
3. Вычислим площадь желтого кусочка над красной областью: 4 выступа.
4. Поделим площадь желтого кусочка на площадь красной области. Получим $P(\text{желтый} \mid \text{красный}) = \frac{4}{20} = \frac{1}{5}$.

Ура! Мы нашли условную вероятность желтого при условии красного. Прекрасно. Почему бы не обратить эту вероятность, чтобы найти $P(\text{красный} \mid \text{желтый})$? Проще говоря, если мы знаем, что ткнули в желтый выступ, какова вероятность, что внизу — красная область? Или какова вероятность, что не спавший всю ночь студент провалил экзамен?

Взглянув на рис. 7.3, вы, быть может, уже определили $P(\text{красный} \mid \text{желтый})$, рассуждая так: «Желтых выступов 6, 4 из них над красной областью, так что вероятность ткнуть в желтый над красным 4/6». Если вы это поняли, поздравляем! Вы только что сами пришли к теореме Байеса. Но подкрепим рассуждения вычислениями.

Формулы

Переход от интуитивных представлений к теореме Байеса потребует некоторых усилий. Начнем с того, как *вычислить* количество желтых выступов (6). Умножим вероятность попасть на желтый выступ на общее количество выступов:

$$\text{желтыеВыступы} = P(\text{желтый}) \times \text{всеВыступы} = \frac{1}{10} \times 60 = 6.$$

Как показать, что 4 из желтых выступов лежат над красной областью? Сначала найдем количество красных выступов (так же, как и желтых):

$$\text{красныеВыступы} = P(\text{красный}) \times \text{всеВыступы} = \frac{1}{3} \times 60 = 20\,.$$

Мы уже вычислили долю красных выступов, покрытых желтым кирпичиком $P(\text{желтый}\,|\,\text{красный})$. Чтобы найти их количество, умножим эту вероятность на общее количество красных выступов:

$$\text{красныеПодЖелтым} = P(\text{желтый} \mid \text{красный}) \times \text{красныеВыступы} = \frac{1}{5} \times 20 = 4.$$

Наконец, вычислим долю красных выступов, накрытых желтым кирпичиком, от общей площади желтого кирпичика:

$$P(\text{красный} \mid \text{желтый}) = \frac{\text{красныеПодЖелтым}}{\text{желтыеВыступы}} = \frac{4}{6} = \frac{2}{3},$$

что согласуется с интуицией. Однако эта формула не похожа на теорему Байеса, имеющую вид

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}.$$

Чтобы прийти к ней, произведем подстановки:

$$P(\text{красный} \mid \text{желтый}) = \frac{P(\text{желтый} \mid \text{красный}) \times \text{красныеВыступы}}{P(\text{желтый}) \times \text{всеВыступы}},$$

то есть

$$P(\text{красный} \mid \text{желтый}) = \frac{P(\text{желтый} \mid \text{красный}) \times P(\text{красный}) \times \text{всеВыступы}}{P(\text{желтый}) \times \text{всеВыступы}}.$$

Сократив общее количество выступов, получим:

$$P(\text{красный} \mid \text{желтый}) = \frac{P(\text{желтый} \mid \text{красный}) \times P(\text{красный})}{P(\text{желтый})}.$$

От наглядных представлений мы пришли к теореме Байеса!

Заключение

Идеи теоремы Байеса интуитивны, но ее формальный вывод не столь очевиден. Преимущество работы с формулами — в выделении логического костяка из интуитивных рассуждений. Мы показали, что наши интуитивные представления разумны, и получили новый мощный инструмент для задач о вероятностях — в том числе более сложных, чем задачи о детальках Lego. В следующей главе мы увидим, как использовать теорему Байеса для обновления представлений на основе данных.

Упражнения

Чтобы убедиться, что вы хорошо понимаете использование теоремы Байеса в задачах об условных вероятностях, попробуйте ответить на эти вопросы.

1. Канзас-Сити, вопреки названию, стоит на границе двух штатов, Миссури и Канзаса. Агломерация Канзас-Сити состоит из 15 округов: 9 в штате Миссури и 6 в Канзасе. В штате Канзас всего 105 округов, в Миссури — 114. Используя теорему Байеса, вычислите вероятность, что человек, переехавший в агломерацию Канзас-Сити, окажется в штате Канзас. Используйте $P(\text{Канзас})$, $P(\text{Канзас-Сити})$ и $P(\text{Канзас-Сити} \mid \text{Канзас})$.
2. В колоде 52 красные и черные карты, в том числе четыре туза: два красных и два черных. Вы вынули из колоды черный туз и перемешали ее. Ваш друг вытянул карту черной масти. Какова вероятность, что это туз?

8

Априорная и апостериорная вероятности и правдоподобие в теореме Байеса

Теперь, узнав, как вывести теорему Байеса из пространственных соображений, давайте поймем, как ее использовать для рассуждений о вероятностях. В этой главе мы применим ее для вычисления, насколько правдоподобны наши предположения при имеющихся данных. При этом мы рассмотрим три компонента этой теоремы — априорную вероятность, апостериорную вероятность и правдоподобие. С ними со всеми мы будем часто встречаться в нашем путешествии по просторам байесовской статистики.

Три компонента

Теорема Байеса позволяет нам выразить численно, насколько наблюдаемые данные влияют на наши представления. При этом мы хотим знать $P(\text{предположения} \mid \text{данные})$ — то есть насколько сильно мы держимся за наши предположения при условии имеющихся данных. Эта часть формулы называется *апостериорной вероятностью*, именно для ее поиска мы будем применять теорему Байеса. Для этого нам понадобится следующий компонент: вероятность имеющихся данных при условии наших предположений,

P(данные|предположения). Он известен как *правдоподобие*, поскольку показывает, насколько правдоподобны данные (при условии имеющихся предположений).

Наконец, мы хотим оценить вероятность априорных предположений, P(предположения). Этот компонент называется априорной вероятностью и характеризует нашу убежденность до того, как мы увидели данные. Правдоподобие и априорная вероятность позволяют вычислить апостериорную вероятность. Чтобы апостериорная вероятность лежала между 0 и 1, нам надо поделить на вероятность P(данные). На практике эту величину часто можно игнорировать, поэтому специального имени у нее нет.

Как вы уже знаете, предположения обычно называют гипотезами, а данные мы будем обозначать D. На рис. 8.1 показаны все компоненты теоремы Байеса.

Рис. 8.1. Компоненты теоремы Байеса

В этой главе мы будем расследовать преступления, собирая кусочки теоремы и делая выводы о произошедшем.

Осмотр места происшествия

Предположим, что как-то, вернувшись с работы, вы обнаруживаете разбитое окно, открытую входную дверь и пропажу ноутбука. Первая ваша мысль: «Ограбили!» Но как вы пришли к такому выводу и, главное, как численно оценить это предположение?

Итак, ваша первая гипотеза H = вас ограбили. Мы хотим оценить, насколько она похожа на правду, то есть найти апостериорную вероятность

$$P(\text{ограбление} \mid \text{разбитое окно, открытая входная дверь, пропавший ноутбук}).$$

Для этого надо подставить недостающие кусочки теоремы Байеса.

Находим правдоподобие

Сначала нам нужно найти правдоподобие — в нашей ситуации вероятность, что мы увидим ту же картину при ограблении (показывающую, насколько данные согласуются с гипотезой):

$$P(\text{разбитое окно, открытая входная дверь, пропавший ноутбук} \mid \text{ограбление}).$$

По сути, мы задаем вопрос: если бы вас ограбили, насколько вероятно было бы увидеть такую картину? Можно представить множество ситуаций, в которых не было бы чего-нибудь из описанного. Например, опытный вор может вскрыть замок, забрать ноутбук и запереть за собой дверь без всяких разбитых окон. С другой стороны, грабитель мог бы разбить окно, взять ноутбук и улизнуть через окно. Однако то, что мы видим, кажется весьма типичной сценой при ограблении, так что давайте примем, что вероятность такой картины при ограблении составляет 3/10.

Важно заметить, что, хотя в нашем примере мы просто прикинули вероятность, можно изучить вопрос и сделать оценку получше. Можно обратиться в местное отделение полиции и запросить статистику по ограблениям или изучить новостные заметки. Таким образом, мы получим более точную оценку правдоподобия: при ограблении вы увидите именно такую картину.

Потрясающее свойство теоремы Байеса — возможность использовать ее как в привычных ситуациях, так и для огромных массивов данных и очень точных вероятностей. Даже если вам кажется, что 3/10 — не очень хорошая оценка, всегда можно вернуться к вычислениям — что мы и сделаем после — и посмотреть, как меняется результат при других предположениях. Например, если вы считаете, что вероятность увиденного вами при условии ограбления — всего 3/100, легко можно все пересчитать, используя это предположение. Благодаря байесовской статистике мы фактически можем измерить наше несогласие с другими! Так как предположения связаны с численными значениями вероятности, можно повторить все вычисления этой главы с другими вероятностями и посмотреть, отразится ли это на выводах.

Вычисляем априорную вероятность

Теперь нужно определить, какова вообще вероятность, что вас ограбят, — то есть априорная вероятность. Априорные вероятности очень важны, они позволяют использовать общую информацию об обстоятельствах. Например, пусть описанная ситуация происходит на пустынном острове,

где вы — единственный житель. В таком случае вторжение грабителя (по крайней мере, двуногого) практически невозможно.

А быть может, наоборот, ваш дом расположен в довольно криминальном районе, где часты ограбления. Для простоты предположим, что вероятность быть ограбленным

$$P(\text{ограбление}) = \frac{1}{1000}.$$

Важно, что мы всегда можем скорректировать эти цифры на основании новой информации.

Теперь почти все готово для вычисления апостериорной вероятности, осталась только нормализация. Прежде чем двигаться дальше, посмотрим на ненормализованную апостериорную вероятность:

$$P(\text{ограбление}) \times P\left(\begin{array}{c}\text{разбитое окно, открытая входная дверь,} \\ \text{пропавший ноутбук} \mid \text{ограбление}\end{array}\right) = \frac{3}{10\,000}.$$

Это очень маленькое значение, что неожиданно. Казалось бы, вероятность ограбления с учетом увиденной картины весьма высока. Но мы еще не учли вероятность увиденного.

Нормализация данных

В нашем выражении не хватает вероятности наблюдаемых данных (безотносительно того, ограбили вас или нет). В данном примере это вероятность того, что окно разбито, дверь открыта, а ноутбук пропал — неважно, по какой причине. Теперь выражение выглядит так:

$$P\left(\begin{array}{c}\text{ограбление} \mid \text{разбитое окно,} \\ \text{открытая входная дверь, пропавший ноутбук}\end{array}\right) = \frac{\frac{1}{1000} \times \frac{3}{10}}{P(D)}.$$

Числитель очень мал, но мы просто не нормализовали его вероятностью имеющихся данных. В табл. 8.1 показано, как при изменении $P(D)$ меняется апостериорная вероятность.

Когда вероятность имеющихся данных уменьшается, апостериорная вероятность увеличивается. Это связано с тем, что по мере того как наблюдаемые нами данные становятся все более маловероятными, то обычно маловероятное объяснение лучше объясняет событие (рис. 8.2).

Таблица 8.1. Как апостериорная вероятность зависит от P(D)

P(D)	Апостериорная вероятность
0,050	0,006
0,010	0,030
0,005	0,060
0,001	0,300

Рассмотрим такой исключительный пример: ваш друг может стать миллионером, только выиграв в лотерею или получив наследство от доселе неизвестного родственника. Это крайне маловероятно, но вы узнаете, что он таки стал миллионером. Теперь вероятность, что друг выиграл лотерею, становится гораздо выше — ведь это один из всего двух способов, которыми друг мог получить миллион.

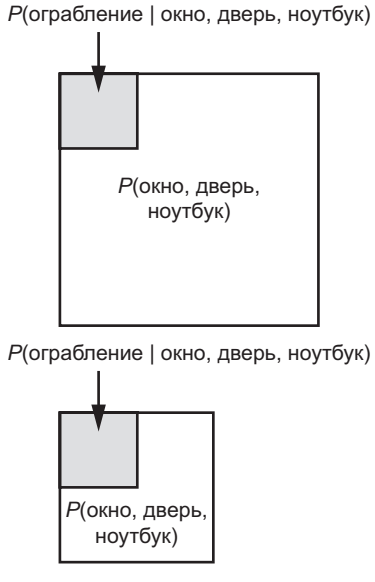


Рис. 8.2. Когда вероятность имеющихся данных уменьшается, апостериорная вероятность увеличивается

Конечно же, ограбление — лишь одно из возможных объяснений увиденного вами, есть и множество других. Но если мы не знаем вероятность того, что наблюдаем, мы не можем нормализовать остальные вероятности. Итак, чему же равно $P(D)$? Это сложный вопрос. В реальных задачах $P(D)$ часто очень трудно вычислить точно. Для всех других компонентов формулы можно собрать настоящие сведения (хотя сейчас мы и взяли значения наугад). Чтобы узнать априорную вероятность P(ограбление), можно посмотреть на данные о преступности и зафиксировать вероятность, с которой в заданный день дом на вашей улице будет ограблен. Мы также теоретически можем исследовать прошлые ограбления и точнее оценить правдоподобие увиденной картины при условии, что произошло ограбление. Но как можно даже примерно оценить P(разбитое окно, открытая входная дверь, пропавший ноутбук)?

Вместо выяснения вероятности наблюдаемых данных можно посчитать вероятность всех прочих событий, объясняющих наблюдаемую картину.

Вероятность всех объяснений должна в сумме давать 1, так что потом мы сможем найти $P(D)$. Но в нашей ситуации возможных объяснений неограниченное количество. Без $P(D)$ мы в затруднении. В главах 6 и 7, когда мы считали вероятность, что продавец — мужчина, и вероятность ткнуть в разноцветные кирпичики, мы знали $P(D)$. Это позволяло точно вычислить вероятность, что гипотеза верна при условии наблюдаемых данных. Без $P(D)$ нам не посчитать $P(\text{ограбление} \mid \text{разбитое окно, открытая входная дверь, пропавший ноутбук})$. Но не все потеряно.

К счастью, иногда не нужно знать $P(D)$, поскольку достаточно просто сравнить гипотезы. В нашем примере мы сравним вероятность ограбления и других возможных объяснений. Это можно сделать, рассмотрев отношение ненормализованных апостериорных вероятностей. Так как $P(D)$ не меняется, ее можно сократить.

Итак, вместо вычисления $P(D)$ мы посвятим остаток главы формулировке альтернативной гипотезы, вычислению ее апостериорной вероятности и сравнению апостериорных вероятностей двух гипотез. Мы не можем вычислить точную вероятность того, что были ограблены, но по-прежнему, благодаря теореме Байеса, можем поиграть в детективов.

Рассматриваем альтернативную гипотезу

Сформулируем альтернативную гипотезу и сравним с исходной. Новая гипотеза будет состоять из трех событий:

1. Соседский ребенок разбил мячом окно.
2. Вы сами забыли закрыть дверь.
3. Вы забыли ноутбук на работе.

Будем обращаться к этим объяснениям по их номеру, а все их вместе обозначим через H_2, то есть $P(H_2) = P(1, 2, 3)$. Найдем их правдоподобие и априорную вероятность.

Правдоподобие альтернативной гипотезы

Напомним, что правдоподобие — это вероятность имеющихся данных при условии гипотезы, то есть $P(D|H_2)$. Интересно (и логично), что оно окажется равным единице: $P(D|H_2) = 1$.

Если произойдут все события из нашей гипотезы, вы непременно получите разбитое окно, открытую дверь и отсутствующий ноутбук.

Априорная вероятность альтернативной гипотезы

Априорная вероятность характеризует возможность того, что произошли все три события. Значит, надо сначала выяснить вероятность каждого из них, а потом воспользоваться правилом произведения. Мы примем, что эти три события независимы. Первая часть нашей гипотезы — соседский ребенок разбил окно мячом. Так часто бывает в фильмах, но в жизни лично я не слышал о таких происшествиях, зато часто слышал про ограбления. Предположим, что попадание мячом в окно в два раза менее вероятно, чем кража.

$$P(1) = \frac{1}{2000}.$$

Вторая часть гипотезы состоит в том, что вы оставили дверь незапертой. Предположим, что это случается раз в месяц, то есть

$$P(2) = \frac{1}{30}.$$

Наконец, забытый ноутбук. Принести ноутбук на работу и оставить его там может быть обычным делом, но совсем не помнить об этом — случай более редкий. Предположим, что такое происходит раз в год:

$$P(3) = \frac{1}{365}.$$

Теперь, присвоив вероятности всем событиям гипотезы H_2, мы можем вычислить априорную вероятность по правилу произведения:

$$P(H_2) = \frac{1}{2000} \times \frac{1}{30} \times \frac{1}{365} = \frac{1}{21\ 900\ 000}.$$

Как можно видеть, априорная вероятность всех трех событий чрезвычайно мала. Теперь нам надо сравнить апостериорные вероятности гипотез.

Апостериорная вероятность альтернативной гипотезы

Мы знаем, что правдоподобие $P(D \mid H_2)$ равно 1, и если вторая гипотеза верна, мы точно получим имеющуюся картину. Без учета априорной вероятности кажется, что апостериорная вероятность второй гипотезы должна быть гораздо больше первоначальной гипотезы об ограблении (ведь при ней вовсе не обязательно мы будем наблюдать все данные). Теперь посмотрим, как априорная вероятность кардинально меняет ненормализованную апостериорную вероятность:

$$P(D|H_2)\times P(H_2)=1\times\frac{1}{21\ 900\ 000}=\frac{1}{21\ 900\ 000}.$$

Теперь надо сравнить наши апостериорные вероятности (а значит, убедительность гипотез), вычислив отношение. И для этого не требуется $P(D)$.

Сравнение ненормализованных апостериорных вероятностей

Нам нужно отношение двух апостериорных вероятностей, которое покажет, во сколько раз одна гипотеза правдоподобнее другой. Исходную гипотезу мы обозначим за H_1, и отношение будет выглядеть так:

$$\frac{P(H_1|D)}{P(H_2|D)}.$$

Теперь распишем числитель и знаменатель по теореме Байеса как $P(H)\times$ $\times P(D|H)\times 1/P(D)$:

$$\frac{P(H_1)\times P(D|H_1)}{P(H_2)\times P(D|H_2)}=\frac{\frac{1}{P(D)}}{\frac{1}{P(D)}}.$$

Заметим, что и числитель, и знаменатель содержат $1/P(D)$, а значит, мы можем сократить этот множитель, и отношение не изменится. Именно поэтому $P(D)$ не имеет значения при сравнении гипотез. Мы получили отношение ненормализованных апостериорных вероятностей. Так как апостериорная вероятность — мера нашей уверенности в гипотезе, вычисленное отношение говорит нам, насколько лучше объясняет данные H_1, чем H_2 (и не требует знания $P(D)$). Сократим $P(D)$ и подставим числа:

$$\frac{P(H_1)\times P(D|H_1)}{P(H_2)\times P(D|H_2)}=\frac{\frac{3}{10\ 000}}{\frac{1}{21\ 900\ 000}}=6570.$$

Это значит, что H_1 объясняет увиденное в 6570 раз лучше, чем H_2. Иными словами, наша исходная гипотеза (H_1) объясняет данные гораздо лучше, чем альтернативная (H_2). Это хорошо согласуется с интуицией — учитывая наблюдаемую картину, ограбление кажется более правдоподобным вариантом.

Хочется строго сформулировать свойства ненормализованной вероятности и в дальнейшем использовать их. Для этого понадобится следующая версия теоремы Байеса:

$$P(H \mid D) \propto P(H) \times P(D \mid H).$$

Ее можно прочитать так: «Апостериорная вероятность — вероятность гипотезы при условии данных — пропорциональна априорной вероятности H, умноженной на вероятность данных при условии H».

Эта форма теоремы Байеса очень полезна, когда нужно сравнить вероятность двух идей, но нет возможности легко узнать $P(D)$. Невозможно найти само по себе значение вероятности для гипотезы, но все еще можно сравнивать гипотезы по теореме Байеса. Сравнение гипотез означает, что мы всегда можем увидеть, во сколько раз одно объяснение лучше, чем другое.

Заключение

В этой главе мы узнали, как теорема Байеса становится инструментом для моделирования наших представлений о мире с учетом имеющихся данных. Теорема Байеса содержит три важнейших компонента: апостериорную вероятность $P(H|D)$, априорную вероятность $P(H)$ и правдоподобие $P(D|H)$. Вероятность самих данных $P(D)$ в этом списке отсутствует, поскольку не нужна нам в случаях, когда необходимо только сравнить гипотезы.

Упражнения

Попробуйте ответить на эти вопросы, чтобы оценить свое понимание компонентов теоремы Байеса.

1. Как уже говорилось, вы можете не согласиться с нашей оценкой правдоподобия для первой гипотезы. Как это повлияет на меру нашей убежденности в превосходстве H_1 над H_2?

$$P(\text{разбитое окно, открытая входная дверь,}$$
$$\text{пропавший ноутбук} \mid \text{ограбление}) = \frac{3}{10}.$$

2. Насколько малой должна быть априорная вероятность ограбления, чтобы гипотезы H_1 и H_2 при имеющихся данных были равновероятны?

9

Байесовские априорные вероятности и распределение вероятностей

Априорные вероятности являются наиболее спорным аспектом теоремы Байеса, потому что их часто считают субъективными. Но на практике они нередко иллюстрируют, как применять жизненно важную справочную информацию, чтобы полностью обосновать неопределенную ситуацию.

В этой главе мы рассмотрим, как использовать априорные вероятности для решения проблемы, а также способы использования распределений вероятностей для численного описания наших убеждений как диапазона возможных значений, а не отдельных значений. Использование вероятностных распределений вместо отдельных значений полезно по двум основным причинам.

Во-первых, в действительности часто существует широкий спектр возможных убеждений, которые можно было бы иметь и рассматривать. Во-вторых, представление диапазонов вероятностей позволяет заявить об уверенности в ряде гипотез. Мы рассмотрели оба примера при изучении таинственной коробочки из главы 5.

Сомнения C-3PO насчет области астероидов

В качестве примера возьмем одну из самых запоминающихся ошибок статистического анализа из эпизода «Звездные войны: Империя наносит ответный удар». Когда Хан Соло, пытаясь уклониться от вражеских истребителей, направляет «Тысячелетний Сокол» в астероидное поле, всезнающий C-3PO сообщает Хану, что вероятность не на его стороне. C-3PO говорит: «Сэр, возможность успешного преодоления области астероидов — 3720 к 1!»

«Никогда больше не сообщай мне соотношение!» — отвечает Хан.

На первый взгляд, это просто забавный эпизод, в котором игнорируется «скучный» анализ данных, но на самом деле здесь присутствует интересная дилемма. Мы, зрители, знаем, что Хан справится с препятствием, но при этом с анализом C-3PO не согласиться не можем. Даже Хан считает, что это опасно, говоря: «Они, должно быть, сумасшедшие, раз последовали за нами». А еще ни один из преследующих бойцов *TIE* не проходит астероидную область, что является довольно убедительным доказательством того, что цифры C-3PO не лишены здравого смысла.

Однако C-3PO упускает из своих расчетов тот факт, что Хан — крутой сорвиголова! C-3PO не ошибается, он просто забывает добавить важную информацию. Теперь возникает вопрос: можем ли мы найти способ избежать ошибки C-3PO без полного исключения вероятности, как предлагает Хан? Чтобы ответить на этот вопрос, нужно смоделировать как мышление C-3PO, так и то, что мы думаем о Хане, а затем смешать эти модели, используя теорему Байеса.

Мы начнем с рассуждений C-3PO в следующем разделе, а затем разберемся с крутостью Хана.

Определение убеждений C-3PO

C-3PO не просто составляет цифры. Он свободно владеет более чем шестью миллионами форм общения, и для его поддержки требуется много данных. Поэтому можно предположить, что он владеет фактами, подтверждающими его заявление о шансах преодолеть астероиды «приблизительно 3720 к 1». Однако C-3PO предоставляет лишь *приблизительные* шансы, а значит, его данные дают достаточно информации только для того, чтобы предложить диапазон возможных коэффициентов успеха. Чтобы представить этот диапазон, нужно взглянуть на *распределение* утверждений относительно вероятности успеха, а не на одно значение, представляющее вероятность.

Для С-3РО единственным возможным результатом является успех или провал при перемещении по области астероидов. Мы определим различные возможные вероятности успеха, учитывая данные С-3РО и используя бета-распределение, о котором шла речь в главе 5. Бета-распределение правильно моделирует диапазон возможных вероятностей для события с учетом доступной информации о соотношении успехов и неудач.

Напомним, что бета-распределение имеет параметры α (количество наблюдаемых успехов) и β (количество наблюдаемых отказов):

$$P(\text{КоэффициентУспеха} \mid \text{Успехи и Неудачи}) = \text{Beta}(\alpha, \beta).$$

Это распределение говорит, какие показатели успеха наиболее возможны с учетом предоставленных данных. Чтобы определить убеждения С-3РО, мы сделаем некоторые предположения о том, откуда берутся его данные. Допустим, у С-3РО есть записи о том, что два человека выжили в астероидной области, а 7440 человек завершили свое путешествие из-за впечатляющего взрыва! На рис. 9.1 показан график функции плотности вероятности, которая представляет убеждение С-3РО об истинном коэффициенте успеха.

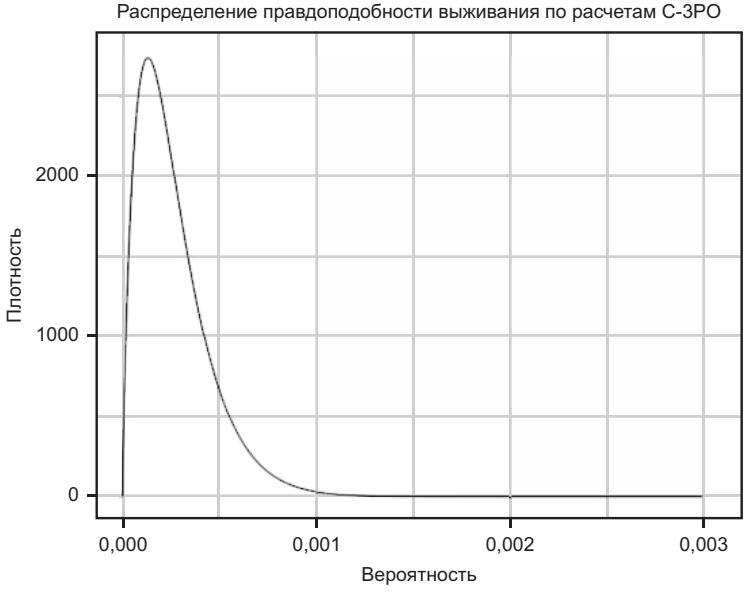


Рис. 9.1. Бета-распределение, представляющее убеждение С-3РО в том, что Хан выживет

Для любого обычного пилота, входящего в область астероидов, все выглядит плохо. В байесовских терминах оценка С-3РО истинной вероятности успеха с учетом наблюдаемых данных, 3720:1, является *правдоподобностью*, о которой говорилось в главе 8. Затем нужно определить априорную вероятность.

Расчеты для преследователей Хана

Проблема анализа С-3РО состоит в том, что его данные относятся ко *всем* пилотам, но Хан сильно отличается от среднего пилота. Если мы не сможем рассчитать крутость Хана, анализ будет сорван — не только потому, что Хан прошел через область астероидов, но и потому, что мы твердо *верим*, что он это сделает. Статистика — это инструмент, который развивает и организует наши рассуждения и представления о мире. Если статистический анализ не только противоречит рассуждениям и убеждениям, но и не может их изменить, тогда с ним что-то не так.

У нас есть *априорное убеждение*, что Хан пройдет через астероидную область, потому что до сих пор он смог пережить все невероятные ситуации. Легендарность Хана Соло заключается в том, что несмотря на малую вероятность выживания, он всегда добивается успеха.

Априорная вероятность часто очень противоречива для аналитиков данных вне байесовского анализа. Многие считают, что просто «придумать» априорную вероятность — это необъективно. Но эта сцена — предметная глава о том, почему отстранение от наших априорных убеждений еще более абсурдно. Представьте, что вы впервые смотрите «Империю». Вы доходите до этой сцены, и ваш друг говорит: «Что ж, Хану конец». Нет никаких шансов, что вы ему поверите. Помните, что С-3РО не совсем ошибается в том, насколько маловероятно выживание. Поэтому если ваш друг скажет: «Что ж, истребителям TIE конец», — вы, вероятно, рассмеетесь и согласитесь.

Сейчас есть множество причин полагать, что Хан выживет, но нет цифр, подтверждающих данное убеждение. Давайте попробуем собрать все вместе.

Начнем с какого-либо верхнего предела для крутости Хана. Если бы мы поверили, что Хан абсолютно не может умереть, фильм стал бы предсказуемым и скучным. С другой стороны, вера в то, что Хан пройдет астероидную область, сильнее, чем вера С-3РО в то, что он этого не сделает, поэтому предположим, что наша вера в то, что Хан выживет, составляет 20 000 к 1.

На рис. 9.2 показано распределение для априорной вероятности того, что Хан сделает это.

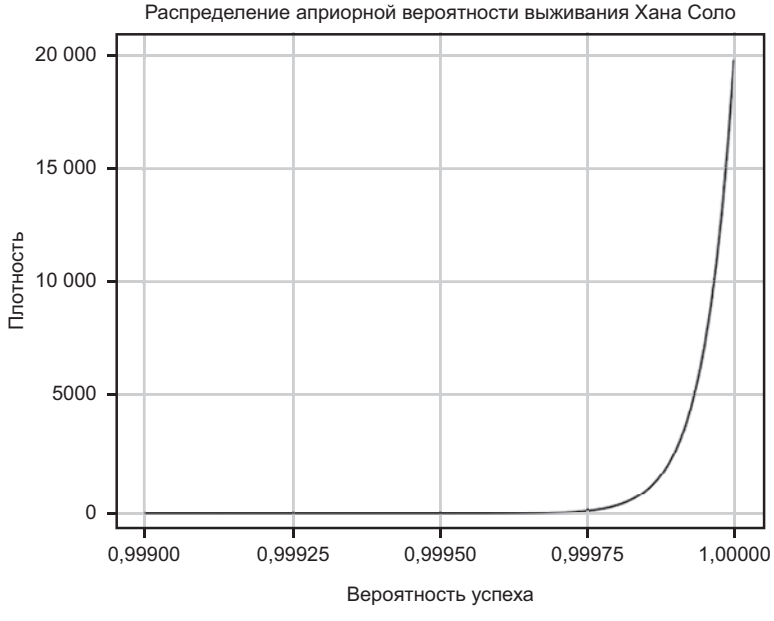


Рис. 9.2. Бета-распределение, представляющее диапазон априорной вероятности выживания Хана Соло

Это еще одно бета-распределение, которое используется по двум причинам. Во-первых, наши убеждения очень приблизительны, поэтому стоит признать переменный коэффициент успеха. Во-вторых, бета-распределение значительно облегчит будущие вычисления.

Теперь, учитывая правдоподобность и априорную вероятность, можно вычислить апостериорную вероятность в следующем разделе.

Создание неопределенности с апостериорной вероятностью

Мы установили, во что верит С-3РО (правдоподобность), и смоделировали наши собственные убеждения о Хане (априорная вероятность) — нужен способ объединить их. Объединяя убеждения, мы создаем *апостериорное*

распределение. В этом случае апостериорная вероятность моделирует чувство неопределенности при изучении правдоподобности С-3РО: цель анализа С-3РО состоит в том, чтобы подшутить над его аналитическим мышлением, но также и создать ощущение реальной опасности. Одно только наше предварительное решение оставило бы нас совершенно безразличными к Хану, но после коррекции его на основе данных С-3РО мы развиваем новое убеждение, которое учитывает реальную опасность.

Формула для апостериорной вероятности на самом деле очень проста и интуитивно понятна. Учитывая, что у нас есть только правдоподобность и априорная вероятность, можно использовать пропорциональную форму теоремы Байеса, о которой говорилось в предыдущей главе:

$$\text{Апостериорная вероятность} \propto \text{Правдоподобность} \times \text{Априорная вероятность}.$$

Помните, что использование этой пропорциональной формы теоремы Байеса означает, что сумма апостериорных распределений необязательно равна 1. Но нам повезло, потому что есть простой способ объединить бета-распределения, которые дадут *нормализованную* апостериорную вероятность при наличии только правдоподобности и априорной вероятности. Таким образом, объединение двух бета-распределений — данные С-3РО (правдоподобность) и наше прежнее убеждение в способности Хана выжить в любой ситуации (априорная вероятность) — становится удивительно легким:

$$\text{Beta}(\alpha_{\text{апостериорная вероятность}}, \beta_{\text{апостериорная вероятность}}) =$$
$$= \text{Beta}(\alpha_{\text{правдоподобность}} + \alpha_{\text{априорная вероятность}}, \beta_{\text{правдоподобность}} + \beta_{\text{априорная вероятность}}).$$

Мы просто добавляем альфы для априорной и апостериорной вероятностей и беты для априорной и апостериорной вероятностей — и получаем нормализованную апостериорную вероятность. Это просто, поэтому работа с бета-распределением очень удобна в байесовской статистике. Чтобы определить апостериорную вероятность для Хана, проходящего через область астероидов, мы можем выполнить этот простой расчет:

$$\text{Beta}(20\,002, 7401) = \text{Beta}(2 + 20\,000, 7400 + 1).$$

Теперь мы можем визуализировать новое распределение для наших данных. На рис. 9.3 показано последнее апостериорное убеждение.

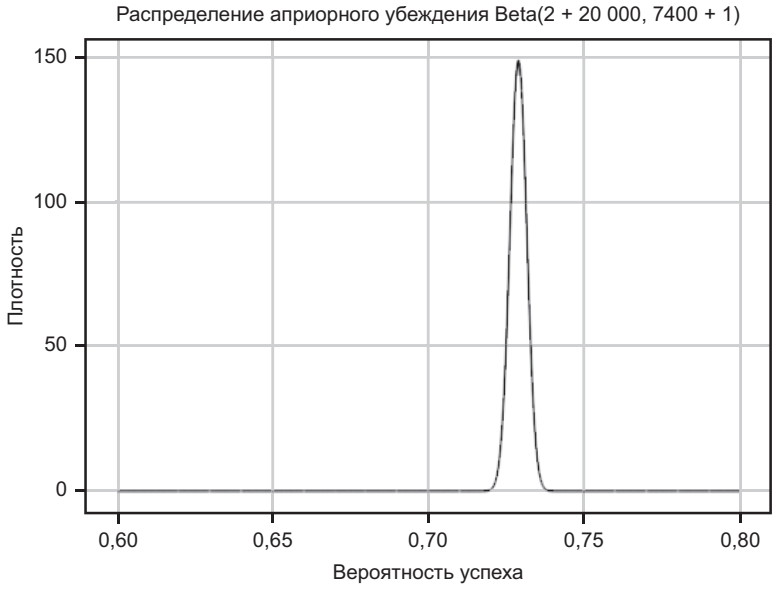


Рис. 9.3. Объединение правдоподобности и априорной вероятности дает более интригующую апостериорную вероятность

Объединив убеждения С-3РО с нашими убеждениями о преследователях Хана, мы получаем гораздо более разумную позицию. Наше апостериорное убеждение — шанс выживания примерно 73 %. Это значит, что Хан, скорее всего, выживет, однако мы все еще напряженно ждем развязки эпизода.

Действительно полезно, что у нас есть не просто грубая вероятность того, как Хан сможет это сделать, а скорее полное распределение возможных убеждений. Для многих примеров в книге мы придерживались простого использования единственного значения для вероятностей, но на практике использование полного распределения помогает проявлять гибкость относительно силы наших убеждений.

Заключение

В этой главе вы узнали, насколько важна исходная информация для анализа имеющихся данных. Данные С-3РО предоставили функцию правдоподобия, которая не соответствовала нашему априорному пониманию способностей Хана. Вместо того чтобы просто отклонить утверждение

С-3PO, что делает Хан, мы объединили правдоподобность С-3PO с нашей априорной вероятностью, чтобы прийти к скорректированному убеждению о возможности успеха Хана. В эпизоде «Империя наносит ответный удар» эта неопределенность жизненно важна для напряженности, которую создает сцена. Если мы полностью поверим данным С-3PO или нашим собственным априорным убеждениям, то будем почти уверены, что Хан либо умрет, либо выживет.

Вы также видели, что можно использовать распределение вероятностей, а не одну вероятность, чтобы выразить диапазон возможных убеждений. В последующих главах эти распределения будут рассмотрены более подробно, чтобы мы могли изучить нюансы неопределенности убеждений.

Упражнения

Чтобы убедиться, что вы понимаете, как объединить априорные распределения вероятности и правдоподобность для получения точного апостериорного распределения, попробуйте ответить на эти вопросы.

1. Ваш друг находит монетку, подбрасывает ее и получает шесть орлов подряд, а затем одну решку. Найдите бета-распределение, которое описывает этот случай. Используйте интегрирование, чтобы определить вероятность того, что истинная вероятность выпадения орла находится в диапазоне от 0,4 до 0,6. Это значит, что монетка является относительно честной.
2. Придумайте априорную вероятность того, что монетка *честная*. Используйте бета-распределение таким образом, чтобы с вероятностью не менее 95 процентов истинная вероятность выпадения орла составляла от 0,4 до 0,6.
3. Теперь посмотрите, сколько еще орлов (без решек) потребуется, чтобы убедить вас в существовании реальной вероятности того, что монетка нечестная. В этом случае наша вера в то, что вероятность нечестности монетки составляет от 0,4 до 0,6, падает ниже 0,5.

Часть III

ОЦЕНКА ПАРАМЕТРОВ

10

Введение в усреднение и оценку параметров

В этой главе вы познакомитесь с *оценкой параметров* — важной частью статистического вывода, где используются данные, чтобы угадать значение неизвестной переменной. Например, может понадобиться оценить вероятность того, что посетитель на веб-странице совершит покупку, узнать предположительное количество драже в банке или местоположение и импульс частицы. Во всех этих случаях у нас есть неизвестное значение, которое нужно оценить, и мы можем использовать наблюдаемую информацию, чтобы сделать предположение. Эти неизвестные значения называются *параметрами*, а процесс выбора наилучшего значения этих параметров — оценкой параметров.

Мы сосредоточимся на *усреднении* (averaging), которое является основной формой оценки параметров. Почти все понимают, что усреднение набора наблюдений — лучший способ оценить истинное значение, но лишь немногие действительно пытаются разобраться, почему это работает и верно ли это вообще. Нужно доказать, что мы можем доверять усреднению, потому что в последующих главах оно будет встраиваться в более сложные формы оценки параметров.

Оценка глубины снежного покрова

Представьте, что прошлой ночью шел сильный снег, и необходимо точно определить, сколько снега выпало в дюймах на вашем дворе. К сожалению, у вас нет снежного датчика, который предоставил бы точное измерение. Посмотрев на улицу, вы увидите, что ветер разметал снег за ночь, что означает, что он не равномерно ровный. Решено использовать линейку для измерения глубины в семи случайных местах во дворе. Вы получаете следующие измерения (в дюймах):

$$6{,}2;\ 4{,}5;\ 5{,}7;\ 7{,}6;\ 5{,}3;\ 8{,}0;\ 6{,}9.$$

Снег заметно сместился, и двор тоже не совсем ровный, поэтому все измерения отличаются друг от друга. Учитывая это, как можно использовать измерения, чтобы сделать правильное предположение о фактическом снегопаде?

Такая задача является отличным примером для оценки параметров. Оцениваемый параметр — это фактическая глубина снегопада предыдущей ночью. Обратите внимание: поскольку ветер разметал снег, а снежного датчика нет, мы никогда не сможем узнать *точное* количество выпавшего снега. Но у нас есть набор данных, которые можно объединить, используя вероятность, чтобы определить вклад каждого наблюдения в оценку и сделать наилучшее возможное предположение.

Усреднение измерений для минимизации ошибки

Вероятно, в первую очередь эти измерения хочется усреднить. В начальной школе мы учимся усреднять элементы, складывая их и деля сумму на общее количество элементов. Поэтому, если есть n измерений, каждое из которых помечено как m_i, где i — это i-е измерение, получаем:

$$\text{среднее} = \frac{m_1 + m_2 + m_3 \ldots m_n}{n}.$$

Подставив данные, получаем следующее решение:

$$\frac{(6{,}2 + 4{,}5 + 5{,}7 + 7{,}6 + 5{,}3 + 8{,}0 + 6{,}9)}{7} = 6{,}31.$$

Итак, учитывая семь наблюдений, лучшее предположение состоит в том, что выпало около 6,31 дюймов (16 см) снега. Усреднение — метод, знакомый

нам с детства, поэтому его применение к этой проблеме кажется очевидным, но на самом деле трудно понять, почему он работает и как связан с вероятностью. В конце концов, каждое из измерений отличается, и все они, вероятно, отличаются от истинного значения выпавшего снега. Даже великие математики боялись, что усреднение данных объединяет все эти ошибочные измерения, что приводит к очень неточной оценке.

При оценке параметров очень важно понять, почему мы принимаем то или иное решение; в противном случае мы рискуем использовать оценку, которая может быть непреднамеренной или систематической ошибкой. В статистике обычно допускают одну ошибку — слепое применение процедур без их понимания, что часто приводит к неправильному решению проблемы. Вероятность — это наш инструмент для рассуждений о неопределенности, а оценка параметров, возможно, является наиболее распространенным процессом для решения проблем неопределенности. Давайте подробно изучим усреднение, чтобы понять, действительно ли это правильный путь.

Решение упрощенной версии задачи

Давайте немного упростим задачу о снегопаде: вместо того чтобы представлять все возможные глубины снега, представьте, что снег падает в красивые однородные блоки, так что двор образует простую двумерную сетку. На рис. 10.1 показан этот идеальный снежный покров глубиной 6 дюймов, визуализированный сбоку (а не с высоты птичьего полета).

Это идеальный сценарий. У нас нет неограниченного количества возможных измерений; вместо этого мы выбираем шесть возможных местоположений, и у каждого местоположения есть только одно возможное измерение — 6 дюймов. Очевидно, что усреднение работает в этом случае, потому что, какие бы данные ни были выбраны, ответ всегда будет равен 6 дюймам.

Сравните это с рис. 10.2, где показаны данные при включении в них сметенного ветром снега с левой стороны дома.

Теперь вместо красивой гладкой поверхности появилась некоторая неопределенность. Конечно, это не совсем верно, потому что можно легко сосчитать каждый блок снега и точно узнать, сколько снега выпало. Этот пример используется в учебных целях, чтобы понять ход рассуждений

Рис. 10.1. Визуализация равномерного дискретного снежного покрова

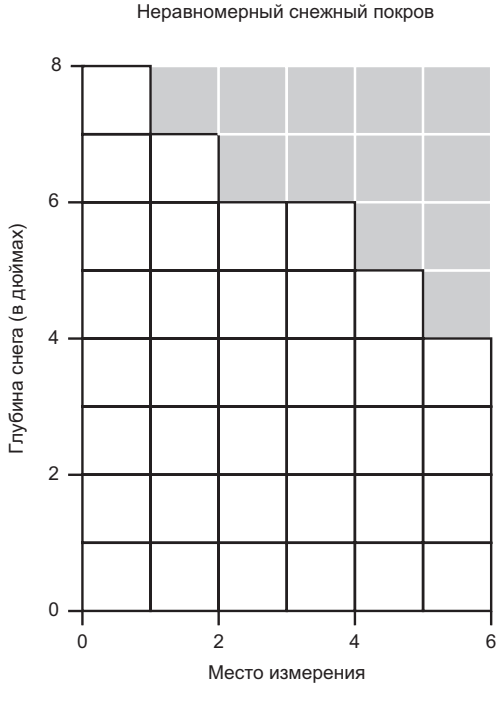


Рис. 10.2. Визуализация снега, который сдул ветер

относительно неопределенной ситуации. Начнем с измерения каждого из блоков во дворе:

$$8, 7, 6, 6, 5, 4.$$

Далее нужно связать вероятности с каждым значением. Поскольку мы жульничаем и знаем, что истинное значение глубины снежного покрова составляет 6 дюймов, запишем также разницу между наблюдением и истинным значением, известную как значение *ошибки* (табл. 10.1).

Таблица 10.1. Наблюдения,а также их частоты и отклонения от истины

Наблюдение	Отклонение от истины	Вероятность
8	2	1/6
7	1	1/6
6	0	2/6
5	–1	1/6
4	–2	1/6

Взглянув на расстояние от истинного измерения для каждого возможного наблюдения, можно увидеть, что вероятность завышения определенного значения уравновешивается вероятностью заниженного измерения. Например, существует вероятность 1/6 выбора измерения, которое на 2 дюйма выше истинного значения, но та же вероятность и у выбора измерения, которое на 2 дюйма ниже истинного значения. Это приводит к первому ключевому пониманию того, почему усреднение работает: ошибки в измерении имеют тенденцию взаимно компенсировать друг друга.

Решение более экстремального случая

При таком плавном распределении ошибок предыдущий сценарий мог не убедить вас в том, что в более сложных ситуациях ошибки устраняются. Чтобы показать, как этот эффект сохраняется в других случаях, рассмотрим более экстремальный пример. Предположим, что ветер надул 21 дюйм снега на один из шести квадратов и оставил только 3 дюйма на каждом из оставшихся квадратов, как показано на рис. 10.3.

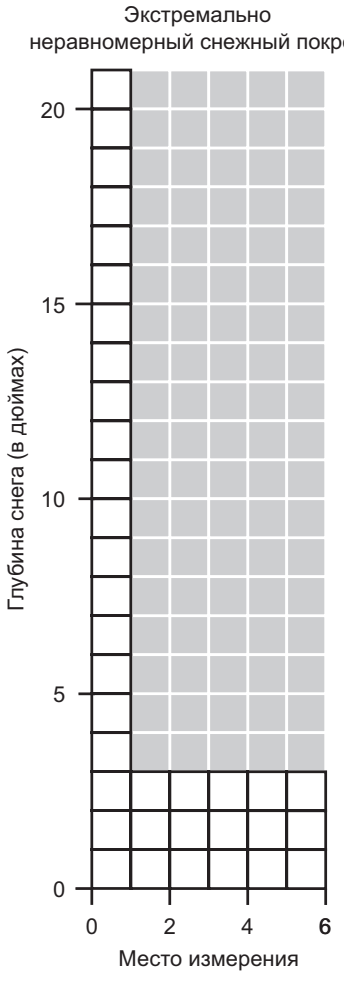


Рис. 10.3. Экстремальный случай смещения снега ветром

Теперь мы видим совершенно иное распределение снежного покрова. В отличие от предыдущего примера ни одно из значений, из которых мы можем сделать выборку, не равно истинному уровню выпавшего снега. Кроме того, наши ошибки больше не распределяются должным образом: существует куча измерений ниже ожидаемых и одно чрезвычайно высокое. В табл. 10.2 показаны возможные измерения, отличие от истинного значения и вероятность каждого измерения.

Таблица 10.2. Наблюдения, различия и вероятности для экстремального примера

Наблюдение	Отклонение от истины	Вероятность
21	15	1/6
3	–3	5/6

Очевидно, что мы не можем просто сопоставить значение ошибки одного наблюдения с другим и заставить их уравновесить друг друга. Тем не менее можно использовать вероятность, чтобы показать, что даже в этом экстремальном распределении ошибки по-прежнему компенсируют друг друга. Это возможно, если представлять каждое измерение ошибки как значение, за которое проголосовали наши данные. Вероятность каждой наблюдаемой ошибки заключается в том, насколько сильно мы верим в эту ошибку. При необходимости объединить наблюдения можно рассматривать вероятность наблюдения как значение, представляющее силу голоса в отношении окончательной оценки. В этом случае погрешность –3 дюйма в пять раз более вероятна, чем погрешность 15 дюймов, поэтому значение –3 становится более весомым. Таким образом, если бы мы принимали участие в голосовании, –3 получило бы пять голосов, тогда как 15 получило бы только один голос. Мы объединяем все голоса, умножая каждое значение на его вероятность и складывая их вместе, в результате чего получается *взвешенная сумма*. В крайнем случае, когда все значения одинаковы, мы просто умножим 1 на наблюдаемое значение, и результатом будет само это значение. В нашем примере мы получаем следующее:

$$\frac{5}{6}\times -3+\frac{1}{6}\times 15=0\,.$$

Ошибки в каждом наблюдении сводятся к нулю! Еще раз: мы обнаруживаем, что неважно, является ли ни одно из возможных значений истинным измерением или равномерно ли распределение ошибок. При взвешивании наблюдений по нашим убеждениям ошибки, как правило, взаимоисключаются.

Оценка истинного значения с помощью взвешенных вероятностей

Теперь мы достаточно уверены, что ошибки истинных измерений взаимоисключаются. Но все еще есть проблема: мы работали с ошибками из

истинного наблюдения, но для их применения нужно знать истинное значение. Когда мы не знаем истинного значения, все, с чем можно работать, — это наши собственные наблюдения, поэтому нужно посмотреть, все ли ошибки устраняются при наличии взвешенной суммы исходных наблюдений.

Чтобы продемонстрировать, что метод работает, нужны «неизвестные» истинные значения. Начнем со следующих ошибок:

$$2, 1, -1, -2.$$

Поскольку истинное измерение неизвестно, мы представим его переменной t, а затем добавим ошибку. Теперь можно взвесить каждое из этих наблюдений по вероятности:

$$\frac{1}{4}(2+t)+\frac{1}{4}(1+t)+\frac{1}{4}(-1+t)+\frac{1}{4}(-2+t).$$

Все, что мы здесь сделали, — добавили ошибку к постоянному значению t, которое представляет истинную меру, а затем взвесили каждый из результатов по его вероятности. Мы делаем это, чтобы посмотреть, можно ли будет уравновесить ошибки и оставить только значение t. Если это так, то можно ожидать, что ошибки будут устраняться даже при простом усреднении обычных наблюдений.

Следующий шаг — применить вес вероятности к значениям величин:

$$\frac{2}{4}+\frac{1}{4}t+\frac{1}{4}+\frac{1}{4}t+\frac{-1}{4}+\frac{1}{4}t+\frac{-2}{4}+\frac{1}{4}t=0+t.$$

Теперь, если мы переупорядочим эти величины так, чтобы все ошибки были вместе, то увидим, что ошибки все равно будут аннулированы, и взвешенное значение t суммируется до просто t, неизвестного истинного значения:

$$\left(\frac{2}{4}+\frac{1}{4}+\frac{-1}{4}+\frac{-2}{4}\right)+\left(\frac{1}{4}t+\frac{1}{4}t+\frac{1}{4}t+\frac{1}{4}t\right)=0+t.$$

Это показывает, что даже при определении измерений как неизвестного истинного значения t и добавлении некоторого значения ошибки все ошибки все равно взаимоисключаются! В конце остается только t. Даже когда мы не знаем, каково истинное измерение или истинная ошибка, при усреднении значений ошибки, как правило, сводятся на нет.

На практике не всегда можно отобрать все пространство возможных измерений, но чем больше выборка, тем большее количество ошибок будет устранено и тем ближе оценка будет к истинному значению.

Определение ожидания, среднего значения и усреднения

То, к чему мы пришли, формально называется *ожиданием*, или *средним значением* данных. Это просто сумма каждого значения, взвешенного по его вероятности. Если обозначить каждое из измерений как x_i, а вероятность каждого измерения как p_i, среднее значение, которое обычно обозначается как μ (строчная греческая буква «мю»), математически будет определено следующим образом:

$$\mu = \sum_{1}^{n} p_i x_i \, .$$

Для ясности, это *в точности* такое же вычисление, как усреднение, которое мы выучили в начальной школе, просто с нотацией, чтобы сделать использование вероятности более явным. В качестве примера, в школе усреднение четырех чисел мы записали бы как:

$$\frac{x_1 + x_2 + x_3 + x_4}{4},$$

что идентично записи:

$$\frac{1}{4}x_1 + \frac{1}{4}x_2 + \frac{1}{4}x_3 + \frac{1}{4}x_4.$$

Можно просто сказать, что $p_i = 1/4$ и записать это следующим образом:

$$\mu = \sum_{1}^{4} p_i x_i.$$

Так что, хотя среднее значение действительно просто то же среднее, с которым мы уже знакомы, основываясь на принципах вероятности, мы видим, почему усреднение данных работает. Независимо от того, как распределены ошибки, вероятность ошибок в одном экстремуме компенсируется вероятностями в другом экстремуме. По мере получений большего количества данных в выборке средние значения с большей вероятностью сводятся на нет, и мы начинаем приближаться к необходимому истинному измерению.

Средние значения измерений и суммы

Мы использовали среднее значение для оценки истинного измерения по распределению наблюдений с некоторой добавленной ошибкой. Но среднее часто используется как способ *суммирования* набора данных. Например, можно сослаться на такие вещи, как:

- средний рост человека;
- средняя цена дома;
- средний возраст учащегося.

Во всех этих случаях среднее значение используется не в качестве оценки параметра для одного истинного измерения; вместо этого суммируются свойства населения. Ради точности мы оцениваем параметр некоторого абстрактного свойства этих групп, которое может даже не быть реальным. Несмотря на то что среднее значение является очень простой и общеизвестной оценкой параметров, им можно легко злоупотребить, что приведет к весьма странным результатам.

Фундаментальный вопрос, который всегда нужно задавать себе при усреднении данных: «Что именно я пытаюсь измерить и что на самом деле означает это значение?» В примере со снегопадом ответ прост: мы пытаемся оценить, сколько снега выпало прошлой ночью, прежде чем ветер разметал его. Однако при измерении «среднего роста» ответ не так ясен. Нет «нормального» человека, и различия в росте, которые мы наблюдаем, не являются ошибками — это действительно разные величины. Человек имеет рост 1,67 метра не потому, что какая-то его часть сместилась на человека ростом 1,92 метра!

Если вы строите парк развлечений и хотите знать, какие ограничения по высоте накладывать на американские горки, чтобы покататься на них могла по крайней мере половина посетителей, то в этом случае есть определенное значение, которое нужно измерить. И в этом случае среднее значение вдруг становится менее полезным. Лучшим измерением для оценки является вероятность того, что кто-то, входящий в парк, будет выше x, где x — минимальный рост для катания на американских горках.

Все утверждения в этой главе предполагают, что мы говорим о попытке измерить определенное значение и использовать среднее значение для устранения ошибок. То есть усреднение используется как форма оценки параметров, где параметр является фактическим значением, которое мы

просто никогда не узнаем. Хотя усреднение также может быть полезно для суммирования больших наборов данных, нельзя использовать интуицию «устранения ошибок», поскольку изменение в данных является подлинным, значимым изменением, а не ошибкой в измерении.

Заключение

В этой главе вы узнали, что можно доверять интуиции в усреднении измерений, чтобы получить наилучшую оценку неизвестного значения. Это работает, потому что ошибки имеют тенденцию к взаимоисключению. Можно формализовать это понятие усреднения в идею ожидания или среднего значения. При вычислении среднего значения все наблюдения взвешиваются по вероятности их появления. Наконец, даже если усреднение является простым инструментом для рассуждений, всегда стоит определять и понимать то, что именно мы пытаемся определить путем усреднения; в противном случае результаты могут оказаться недействительными.

Упражнения

Чтобы убедиться, что вы понимаете усреднение для оценки неизвестного измерения, попробуйте ответить на эти вопросы.

1. Можно получить ошибки, не в полной мере взаимоисключающие. По шкале Фаренгейта 98,6 градуса — это нормальная температура тела, а 100,4 градуса — типичный порог лихорадки. Скажем, вы ухаживаете за ребенком, которому жарко и который кажется больным, но все повторные показания термометра находятся между 99,5 и 100,0 градуса: высоковато, но не совсем лихорадка. Вы ставите термометр самому себе и получаете несколько показаний между 97,5 и 98. Что может быть не так с термометром?
2. Учитывая, что вы чувствуете себя здоровым и у вас всегда стабильная нормальная температура, как можно изменить измерения 100, 99,5, 99,6 и 100,2, чтобы оценить, есть ли у ребенка температура?

11

Измерение разброса данных

В этой главе вы изучите три различных метода — среднее абсолютное отклонение, дисперсию и стандартное отклонение — для количественной оценки *разброса* или различных экстремумов наблюдений.

В предыдущей главе вы узнали, что среднее значение — лучший способ угадать значение неизвестного измерения и что чем больше разброс наблюдений, тем более неопределенными будут оценки среднего значения. Например, если мы пытаемся выяснить место столкновения двух машин, основываясь только на распространении оставшегося мусора после буксировки, то чем больше будет мусора, тем меньше мы будем уверены, что это именно то самое место, где они столкнулись.

Поскольку разброс наблюдений связан с неопределенностью в измерении, нужно иметь возможность количественно оценить его, чтобы делать вероятностные заявления об оценках (как это сделать, будет описано в следующей главе).

Бросаем монетку в колодец

Представьте, что вы с другом гуляете по лесу и натыкаетесь на странно выглядящий старый колодец. Вы заглядываете внутрь и вам кажется, что там нет дна. Чтобы проверить это, вы берете монетку и бросаете ее в колодец.

Через несколько секунд доносится всплеск. Вы делаете вывод, что колодец глубокий, но не бездонный.

Несказанно удивившись, вы с другом загораетесь любопытством — насколько глубок колодец на самом деле? Чтобы собрать больше данных, вы берете еще пять монет и бросаете их, получая следующие измерения в секундах:

3,02; 2,95; 2,98; 3,08; 2,97.

Как и ожидалось, в результатах обнаруживаются некоторые различия; это в первую очередь связано с необходимостью убедиться, что вы бросили монету с той же высоты, а затем правильно записали время, когда послышался всплеск.

Затем ваш друг хочет попробовать свои силы в получении некоторых измерений. Вместо того чтобы набрать пять монет одинакового размера, он берет разные предметы — от мелкой гальки до прутьев. Бросив их в колодец, друг получает следующие измерения:

3,31; 2,16; 3,02; 3,71; 2,80.

Оба этих примера имеют среднее значение (μ) около 3 секунд, но ваши измерения и измерения друга разбросаны в разной степени. Наша цель в этой главе — найти способ количественно оценить разницу между разбросом ваших измерений и разбросом измерений вашего друга. Мы будем использовать этот результат в следующей главе, чтобы вычислить вероятность определенных диапазонов значений для нашей оценки.

В оставшейся части этой главы мы рассмотрим, когда речь идет о первой группе значений (ваши наблюдения) с переменной a и о второй группе (наблюдения вашего друга) с переменной b. Для каждой группы наблюдения обозначаются индексами; например, a_2 — второе наблюдение из группы a.

Находим среднее абсолютное отклонение

Начнем с измерения разброса каждого наблюдения по среднему значению (μ). Среднее значение и для a, и для b равно 3. Поскольку μ является наилучшей оценкой истинного значения, имеет смысл начать количественную

оценку разницы в двух разбросах путем измерения отклонения каждого из значений от среднего. Таблица 11.1 отображает каждое наблюдение и его отклонение от среднего значения.

Таблица 11.1. Наблюдения ваши и вашего друга и их отклонения от среднего значения

Наблюдение	Отклонение от среднего значения
Группа a	
3,02	0,02
2,95	–0,05
2,98	-0,02
3,08	0,08
2,97	–0,03
Группа b	
3,31	0,31
2,16	–0,84
3,02	0,02
3,71	0,71
2,80	-0,16

ПРИМЕЧАНИЕ

Отклонение от среднего значения отличается от значения ошибки, которое является отклонением от истинного значения и в этом случае неизвестно.

Как количественно определить разницу между двумя бросками? Во-первых, попробовать суммировать их отличия от среднего значения. Но при попытке это сделать мы обнаружим, что сумма отличий для обоих наборов наблюдений абсолютно одинакова, что странно, учитывая заметную разницу в разбросе двух наборов данных:

$$\sum_{i=1}^{5} a_1 - \mu_a = 0 \quad \sum_{i=1}^{5} b_1 - \mu_b = 0.$$

Причина, по которой мы не можем просто суммировать отличия от среднего значения, связана в первую очередь с тем, как работает среднее значение: как мы знаем из главы 10, ошибки имеют тенденцию взаимно исключать друг друга. Требуется математический метод, который гарантирует, что различия не устраняются, не влияя на достоверность измерений.

Различия сводятся на нет потому, что некоторые из них являются отрицательными, а некоторые — положительными. Таким образом, если мы конвертируем все различия в положительные значения, то сможем устранить эту проблему, не аннулируя значения.

Самый очевидный способ сделать это — взять *абсолютную величину* различий; это отклонение числа от 0, поэтому абсолютное значение 4 равно 4, а абсолютное значение –4 также равно 4. Это дает положительную версию отрицательных чисел без их фактического изменения. Чтобы представить абсолютное значение, мы заключаем значение в вертикальные линии, как в $|-6| = |6| = 6$.

Если взять абсолютные значения различий из табл. 11.1 и вместо этого использовать их в расчетах, мы получим результат, с которым можно работать:

$$\sum_{1}^{5}|a_i - \mu_a| = 0{,}2 \quad \sum_{1}^{5}|b_i - \mu_b| = 2{,}08\,.$$

Попробуйте сделать это вручную, и получите те же результаты. Это более разумный подход для нашей конкретной ситуации, но он применяется только тогда, когда две группы выборок имеют одинаковый размер.

Представьте, что у нас было еще 40 наблюдений для группы *a* — скажем, 20 наблюдений из 2,9 и 20 из 3,1. Даже с этими дополнительными наблюдениями данные в группе *a* кажутся менее разбросанными, чем данные в группе *b*, но абсолютная сумма группы *a* теперь составляет 85,19 просто потому, что у нее больше наблюдений!

Чтобы это исправить, можно нормализовать значения путем деления на общее количество наблюдений. Вместо того чтобы делить, мы просто умножим на 1 общую сумму. Это называется *умножением обратной величины* и выглядит следующим образом:

$$\frac{1}{5}\times\sum_{1}^{5}|a_i - \mu_a| = 0{,}04 \quad \frac{1}{5}\times\sum_{1}^{5}|b_i - \mu_b| = 2{,}08\,.$$

Теперь у нас есть измерение разброса, которое не зависит от размера выборки! Обобщение этого подхода обозначается следующим образом:

$$\mathrm{MAD}(x) = \frac{1}{n} \times \sum_{1}^{n} |x_i - \mu| .$$

Мы вычислили среднее абсолютных отличий между нашими наблюдениями и средним значением. Это означает, что для группы *a* среднее наблюдение отклоняется на 0,04 секунды от среднего значения, а для группы *b* — на около 0,416 секунды. Мы называем результат этой формулы *средним абсолютным отклонением (mean absolute deviation, MAD)*. MAD — очень полезная и интуитивно понятная мера разброса наблюдений. Учитывая, что группа *a* имеет MAD 0,04, а группа *b* — около 0,4, теперь можно сказать, что группа *b* примерно в 10 раз больше, чем группа *a*.

Поиск величины расхождения

Другой способ математически сделать все наши различия положительными без аннулирования данных состоит в возведении их в квадрат: $(x_i - \mu)^2$. Этот метод имеет как минимум два преимущества по сравнению с использованием MAD.

Первое преимущество несколько академическое: со значениями, возведенными в квадрат, гораздо проще работать математически, чем брать их абсолютное значение. В книге мы не будем это использовать, но математиков функция абсолютного значения может раздражать.

Вторая и более практическая причина — возведение в квадрат приводит к *экспоненциальному штрафу* (exponential penalty), а это означает, что измерения, очень далекие от среднего, штрафуются гораздо больше. Другими словами, маленькие различия не так важны, как большие, как кажется интуитивно. Например, если кто-то запланировал встречу с вами не в том кабинете, вы не расстроитесь, если окажетесь по соседству с нужным кабинетом, но почти наверняка расстроитесь, если вас отправят в офис в другой части страны.

Если подставить абсолютную величину для возведенных в квадрат отличий, мы получим следующее:

$$\mathrm{Var}(x) = \frac{1}{n} \times \sum_{1}^{n} (x_i - \mu)^2 .$$

Эта формула, которая занимает особое место в изучении вероятности, называется *расхождением*. Обратите внимание, что уравнение для расхождения точно такое же, как MAD, за исключением того, что функция абсолютного значения в MAD была заменена на возведение в квадрат. Поскольку квадрат обладает более хорошими математическими свойствами, расхождение используется гораздо чаще при изучении вероятности, чем MAD. Мы можем видеть, как различаются результаты при вычислении их расхождения:

$$\text{Var}(\text{группа } a) = 0{,}002,\ \text{Var}(\text{группа } b) = 0{,}269.$$

Поскольку мы возводим в квадрат, интуитивного понимания результатов расхождения больше нет. MAD дает интуитивное определение: это среднее отклонение от среднего значения. Расхождение, с другой стороны, говорит: это средняя квадратическая разница. Напомним, что при использовании MAD группа *b* была примерно в 10 раз больше группы *a*, но в случае с расхождением группа *b* теперь в 100 раз больше!

Нахождение стандартного отклонения

Хотя в теории расхождение имеет множество свойств, которые делают его полезным, на практике бывает сложно интерпретировать результаты. Людям не всегда понятно, что означает разница в 0,002 секунды в квадрате. Как уже упоминалось, отличительной чертой MAD является то, что результат интуитивен. Если MAD группы *b* составляет 0,4, это означает, что среднее расстояние между данным наблюдением и средним значением составляет буквально 0,4 секунды. Но усреднение по квадратным различиям не позволяет так же хорошо анализировать результат.

Чтобы исправить это, можно взять квадратный корень из расхождения, чтобы уменьшить его до числа, которое немного лучше подходит для интуиции. Корень квадратный из расхождения называется *стандартным отклонением* и представлен строчной греческой буквой сигма (σ). Он определяется следующим образом:

$$\sigma = \sqrt{\frac{1}{n} \times \sum_{1}^{n} (x_i - \mu)^2}.$$

Формула стандартного отклонения не такая страшная, как может показаться на первый взгляд. Цель состоит в том, чтобы численно представить, как распределены данные. Поэтому:

1. Нужно получить разницу между нашими данными и средним значением, $x_i - \mu$.
2. Нужно преобразовать отрицательные числа в положительные, поэтому мы берем квадрат, $(x_i - \mu)^2$.
3. Нужно сложить все различия:

$$\sum_i^n (x_i - \mu)^2.$$

4. Нам не нужно, чтобы сумма зависела от количества наблюдений, поэтому мы нормализуем ее с помощью $1/n$.
5. Наконец, берем квадратный корень из всего, чтобы числа были ближе к тому же результату, что и в случае с более интуитивным абсолютным отклонением.

Если посмотреть на стандартное отклонение для двух групп, то можно заметить, что оно очень похоже на MAD:

$$\sigma(\text{группа } a) = 0{,}046,\ \sigma(\text{группа } b) = 0{,}519.$$

Стандартное отклонение является золотой серединой между интуитивностью MAD и математической легкостью расхождения. Обратите внимание, что, как и в случае с MAD, разница в разбросе между b и a составляет 10. Стандартное отклонение настолько полезно и повсеместно используемо, что в литературе по вероятности и статистике расхождение определяется просто как σ^2, или сигма в квадрате!

Итак, теперь у нас есть три разных способа измерить разброс данных. Результаты отражены в табл. 11.2.

Таблица 11.2. Измерение разброса по методу

Метод измерения разброса	Группа a	Группа b
Средние абсолютные отклонения	0,040	0,416
Расхождение	0,002	0,269
Стандартное отклонение	0,046	0,519

Ни один из этих методов измерения разброса не является более правильным, чем другой. Безусловно, наиболее часто используемый метод — стандартное отклонение, потому что можно использовать его вместе со средним для определения нормального распределения, которое, в свою очередь, позволяет определять явные вероятности для возможных истинных значений измерений. В следующей главе мы рассмотрим нормальное распределение и посмотрим, как оно может помочь понять уровень нашей уверенности в измерениях.

Заключение

В этой главе вы изучили три метода количественной оценки распространения группы наблюдений. Наиболее интуитивным измерением разброса значений является среднее абсолютное отклонение (MAD), которое представляет собой среднее отклонение каждого наблюдения от среднего значения. Интуитивно понятный MAD не так полезен математически, как другие варианты.

Математически предпочтительным методом является расхождение — квадрат отклонений наших наблюдений. Но при вычислении расхождения мы теряем интуитивное понимание того, что означает расчет.

Третий вариант — стандартное отклонение, которое является квадратным корнем из расхождения. Стандартное отклонение математически полезно, а также дает результаты, которые являются интуитивными.

Упражнения

Чтобы убедиться, что вы понимаете различные методы измерения разброса данных, попробуйте ответить на эти вопросы.

1. Одним из преимуществ расхождения является то, что возведение в квадрат различий делает штрафы экспоненциальными. Приведите несколько примеров, когда это будет полезно.
2. Рассчитайте среднее значение, расхождение и стандартное отклонение для следующих значений: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.

12

Нормальное распределение

В предыдущих двух главах вы узнали об очень важных понятиях: среднее значение (μ), которое позволяет оценивать измерения по различным наблюдениям, и стандартное отклонение (σ), которое позволяет измерять разброс наблюдений. Каждое понятие полезно уже само по себе, но вместе они сила: их можно использовать в качестве параметров для наиболее известного распределения вероятностей из всех — *нормального распределения*.

В этой главе вы узнаете, как использовать нормальное распределение для определения точной вероятности степени уверенности в том, что одна оценка окажется верной по сравнению с другими. Истинная цель оценки параметров заключается не просто в оценке значения, а в том, чтобы назначить вероятность для *диапазона* возможных значений. Это позволяет проводить более сложные рассуждения с неопределенными значениями.

В предыдущей главе мы установили, что вычисление среднего значения является надежным методом оценки неизвестного значения на основе существующих данных и что стандартное отклонение можно использовать для измерения разброса этих данных. Измеряя разброс наблюдений, можно определить, насколько мы уверены в средних значениях. Чем больше разбросаны наблюдения, тем меньше мы уверены в своих силах. Нормальное

распределение позволяет точно определить, *насколько* мы уверены в различных убеждениях, принимая во внимание наблюдения.

Зажигательные шнуры для гадких делишек

Представьте, что усатый мультяшный злодей хочет бросить бомбу, чтобы взорвать стену в банковском хранилище. К сожалению, у него всего одна бомба, но довольно большая. Он знает, что если отойдет от бомбы на 200 футов (60 метров), то окажется в безопасности. Бег до укрытия занимает 18 секунд. Если злодей не успеет добежать, то рискует жизнью.

У него только одна бомба и шесть зажигательных шнуров одинакового размера, поэтому он решает проверить пять из шести шнуров, оставив последний для бомбы. Все шнуры одинаковой длины, и для их поджигания требуется одинаковое количество времени. Злодей поджигает каждый шнур и измеряет, сколько нужно времени, пока шнуры полностью не прогорят, чтобы убедиться, что у него есть 18 секунд для побега. Конечно, спешка приводит к противоречивым измерениям. Вот время, которое он записал (в секундах) для каждого перегоревшего шнура: 19, 22, 20, 19, 23.

Пока все хорошо: ни один из шнуров не сгорает раньше чем за 18 секунд. Вычисление среднего дает нам $\mu = 20{,}6$, а стандартного отклонения — $\sigma = 1{,}62$.

Теперь нужно определить конкретную вероятность того, что предохранитель сработает менее чем за 18 секунд. Поскольку злодей дорожит своей жизнью даже больше, чем деньгами, то хочет быть на 99,9 % уверен, что переживет взрыв. Иначе он даже не станет пытаться грабить банк.

Из главы 10 мы знаем, что среднее значение является хорошей оценкой истинного значения с учетом набора измерений, но способа выразить, насколько *сильно* мы считаем это значение истинным, мы пока не нашли.

Из главы 11 мы также знаем, что можно количественно оценить, насколько разбросаны наблюдения, рассчитав стандартное отклонение. Кажется закономерным, что это поможет выяснить, насколько вероятны альтернативы среднему значению. Предположим, что вы уронили стакан и он разбился вдребезги. В зависимости от того, как разлетелись осколки, возможно, вам придется убрать и в соседней комнате. Если осколки

находятся близко друг к другу (рис. 12.1), то, скорее всего, другую комнату убирать не придется.

Но если осколки раскиданы так, как на рис. 12.2, то стоит проверить другую комнату. Так и у злодея со шнуром: если значения времени сгорания фитиля сильно разбросаны, несмотря на то что сгорающих быстрее чем за 18 секунд шнуров не обнаружено, вполне вероятно, что какой-то шнур все еще может сгореть менее чем за 18 секунд.

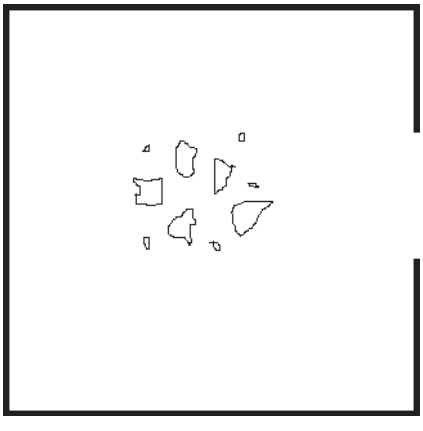


Рис. 12.1. Когда осколки расположены близко друг к другу,вы знаете, где нужно убраться

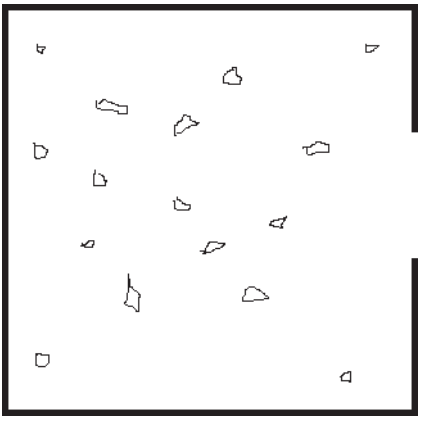


Рис. 12.2. Когда осколки разбросаны, вы точно не знаете,где они могут находиться

Когда наблюдения визуально разбросаны, мы интуитивно чувствуем, что в крайних пределах видимости могут быть и другие значения. Также мы менее уверены в том, где именно находится центр разброса. В примере со стаканом точно не понятно, куда упали осколки, если вы сами не видели падение и осколки разлетелись.

Можно количественно определить эту интуицию с помощью наиболее изученного и известного распределения вероятностей: нормального распределения.

Нормальное распределение

Нормальное распределение — это непрерывное распределение вероятностей (например, бета-распределение в главе 5), которое наилучшим образом

описывает силу возможных убеждений в значении неопределенного измерения, учитывая известное среднее значение и стандартное отклонение. Оно принимает значения μ и σ (среднее значение и стандартное отклонение соответственно) в качестве двух параметров. Нормальное распределение с $\mu = 0$ и $\sigma = 1$ имеет форму колокола, как показано на рис. 12.3.

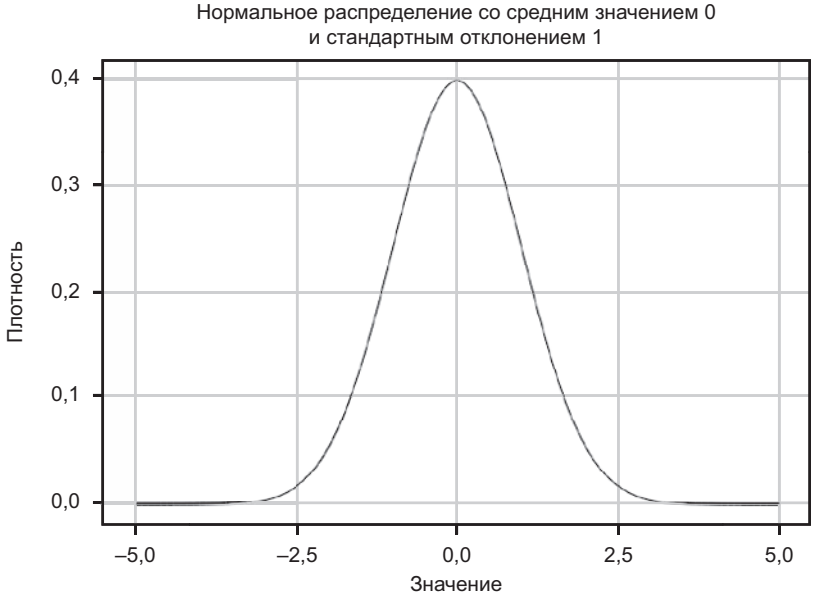


Рис. 12.3. Нормальное распределение с = 0 и = 1

Как видите, центр нормального распределения — это среднее значение. Ширина нормального распределения определяется его стандартным отклонением. На рис. 12.4 и 12.5 показаны нормальные распределения с $\mu = 0$ и $\sigma = 0{,}5$ и 2 соответственно.

По мере того как стандартное отклонение уменьшается, уменьшается и ширина нормального распределения.

Как уже говорилось, нормальное распределение отражает то, насколько сильно мы верим в среднее значение. Таким образом, если наши наблюдения разбросаны сильнее, мы верим в более широкий диапазон возможных значений и меньше доверяем среднему значению. И наоборот, если все наши наблюдения более или менее одинаковы (имеется в виду небольшое σ), мы считаем оценку довольно точной.

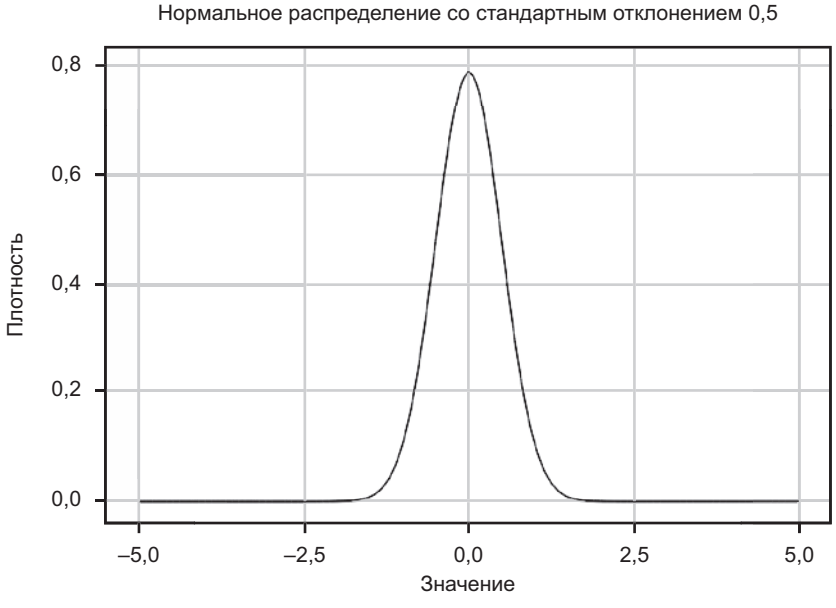


Рис. 12.4. Нормальное распределение с = 0 и = 0,5

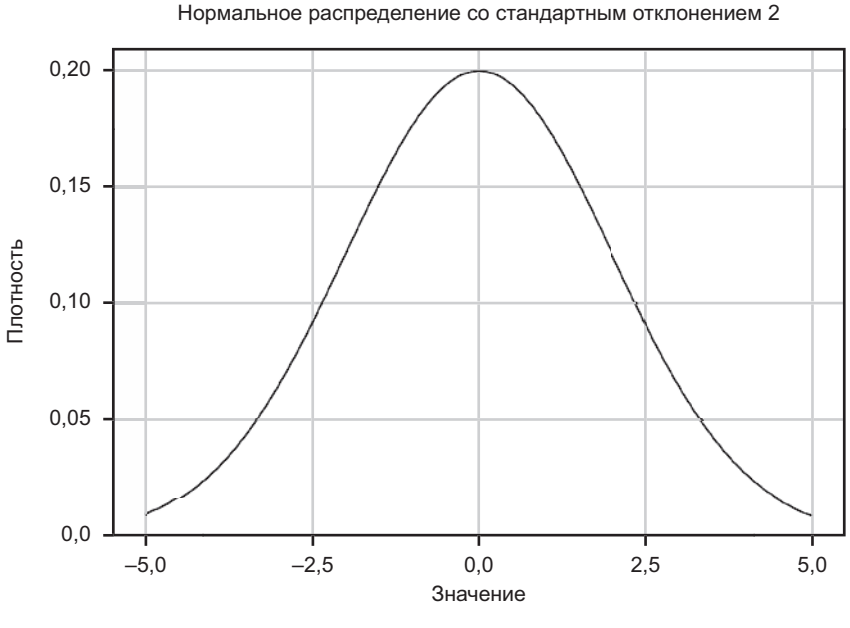


Рис. 12.5. Нормальное распределение с = 0 и = 2

Когда *единственное*, что мы знаем о проблеме, — это среднее значение и стандартное отклонение наблюдаемых данных, то нормальное распределение является наиболее достоверным представлением состояния убеждений.

Решение задачи с зажигательным шнуром

Вернемся к исходной задаче. Имеется нормальное распределение с $\mu = 20{,}6$ и $\sigma = 1{,}62$. Мы ничего не знаем о свойствах зажигательных шнуров, кроме зарегистрированного времени сгорания, поэтому можем моделировать данные с нормальным распределением, используя наблюдаемое среднее значение и стандартное отклонение (рис. 12.6).

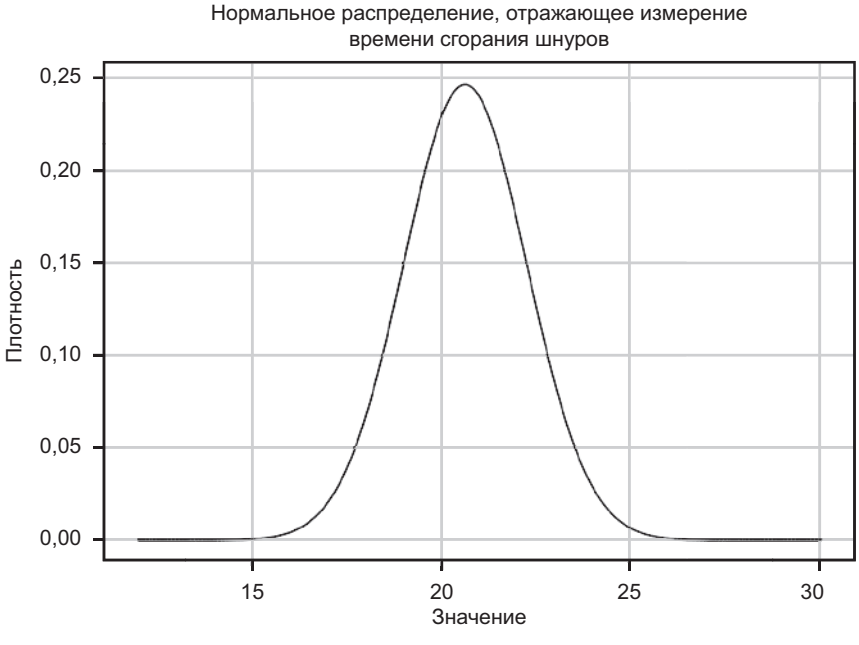


Рис. 12.6. Нормальное распределение с = 20,6 и = 1,62

Нужно ответить на вопрос: учитывая наблюдаемые данные, какова вероятность того, что шнур сгорит в течение 18 секунд или за меньшее время? Воспользуемся функцией плотности вероятности (*probability density function*, *PDF*), это концепция, о которой вы впервые узнали в главе 5. PDF для нормального распределения такая:

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

Чтобы получить вероятность, нужно *интегрировать* эту функцию по значениям, меньшим чем 18:

$$\int_{-\infty}^{18} N(\mu=20{,}6,\ \sigma=1{,}62).$$

Интегрирование можно представить как простое взятие области под кривой для отрезка, который вас интересует (рис. 12.7).

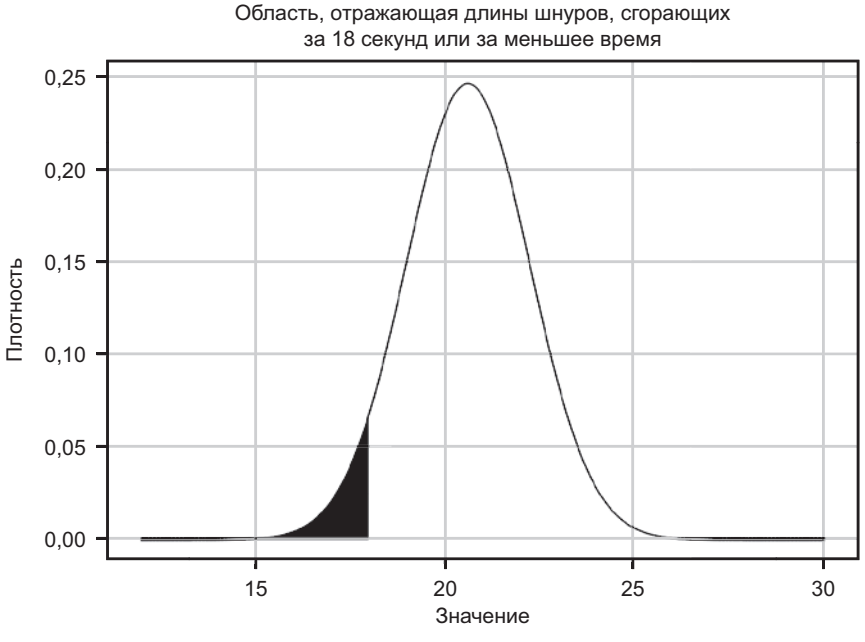


Рис. 12.7. Интересующая нас область под кривой

Закрашенная область представляет собой вероятность того, что шнур прогорит за 18 секунд или меньше, учитывая проведенные наблюдения. Обратите внимание, что хотя ни одно из наблюдаемых значений не было меньше 18, из-за разброса наблюдений нормальное распределение на рис. 12.6 показывает, что появление значения 18 или меньше все еще возможно. Интегрируя по всем значениям меньше 18, мы можем рассчитать вероятность того, что зажигательный шнур *не* продержится так долго, как нужно злодею.

Интегрирование этой функции вручную — непростая задача. К счастью, есть язык R, который все сделает за нас.

Но для начала нужно определить, с какого числа начать интегрирование. Нормальное распределение определяется в диапазоне всех возможных значений от минус бесконечности ($-\infty$) до бесконечности (∞). Итак, теоретически нужно получить следующее:

$$P(\text{время сгорания} < 18) = \int_{-\infty}^{18} N(\mu, \sigma).$$

Очевидно, что мы не можем интегрировать функцию из минус бесконечности на компьютере! К счастью, как можно увидеть на рис. 12.6 и 12.7, функция плотности вероятности очень быстро становится чрезвычайно малым значением. Можно заметить, что линия в PDF почти плоская на значении 10, а это означает, что в данной области практически нет вероятности, поэтому можно просто интегрировать от 10 до 18. Мы могли бы также выбрать более низкое значение, например 0, потому что в этой области действительно нет вероятности, но это не изменит наш результат каким-либо значимым образом. В следующем разделе мы обсудим эвристику, которая облегчает выбор нижней или верхней границы.

Интегрируем эту функцию с помощью методов `integrate()` в *R* и `dnorm()` (который является функцией R для PDF с нормальным распределением), вычисляя PDF нормального распределения следующим образом:

```
integrate(function(x) dnorm(x, mean=20,6, sd=1,62), 10, 18)
0,05425369 с абсолютной погрешностью < 3e-11.
```

Округлив значение, видно, что $P(\text{время сгорания} < 18) = 0{,}05$. Это говорит о пятипроцентной вероятности того, что шнур сгорит меньше чем за 18 секунд. Наш злодей ценит свою жизнь и грабить банк станет, только если на 99,9 % уверен, что сможет избежать взрыва. Так что сегодня банк в безопасности!

Сила нормального распределения заключается в том, что мы можем рассуждать вероятностно о широком диапазоне возможных альтернатив среднему значению, что дает представление о том, насколько реалистичным является среднее значение. Можно использовать нормальное распределение в любое время, когда нужно рассуждать о данных, для которых известно только среднее значение и стандартное отклонение.

Но здесь заключается и опасность нормального распределения. Если у вас есть информация о проблеме, кроме среднего значения и стандартного

отклонения, обычно стоит ее использовать. Пример показан в следующем разделе.

Немного хитрости и интуиции

Хотя R значительно упрощает интегрирование нормального распределения по сравнению с попытками взять интеграл вручную, есть полезная фишка, которая может *еще* больше упростить положение вещей при работе с нормальным распределением. Для любого нормального распределения с известным средним значением и стандартным отклонением можно оценить площадь под кривой вокруг μ в терминах σ.

Например, площадь под кривой для диапазона от μ – σ (одно стандартное отклонение меньше среднего) до μ + σ (одно стандартное отклонение больше среднего) содержит 68 % массы распределения.

Это означает, что 68 % возможных значений находятся в пределах ± одного стандартного отклонения от среднего значения, как показано на рис. 12.8.

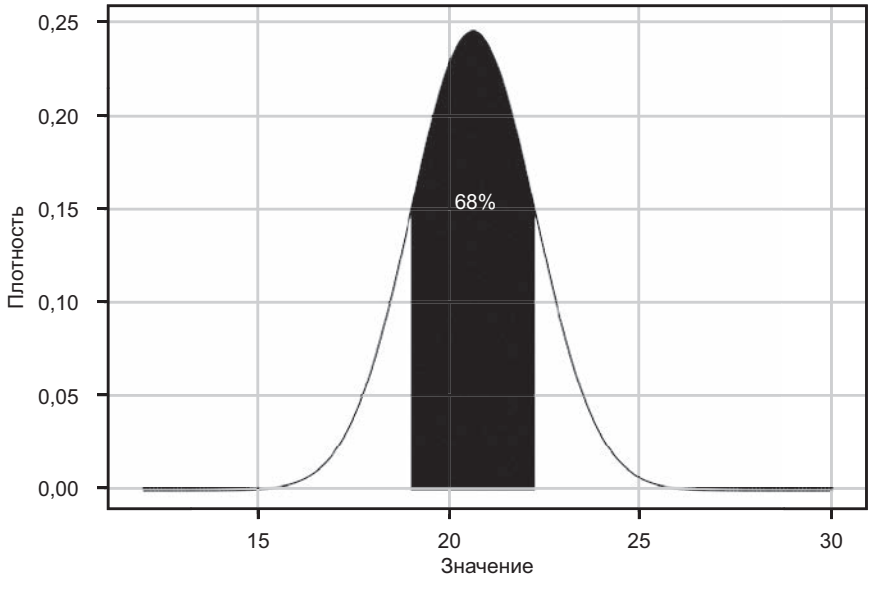


Рис. 12.8. 68 % плотности вероятности (площадь под кривой) лежит между одним стандартным отклонением от среднего значения в любом направлении

Можно продолжить, увеличив расстояние от среднего на отрезки, кратные σ. В табл. 12.1 даны вероятности для этих областей.

Таблица 12.1. Области под кривой для различных средних значений

Расстояние от среднего значения	Вероятность
σ	68 %
2σ	95 %
3σ	99,7 %

Эта хитрость полезна для быстрой оценки вероятности значения даже для небольшой выборки. Все, что вам нужно, — это калькулятор, чтобы легко вычислить μ и σ. Это значит, что вы можете делать довольно точные оценки, выполнив только половину измерений!

Например, при измерении глубины снежного покрова в задачах главы 10 у нас были следующие измерения: 6,2; 4,5; 5,7; 7,6; 5,3; 8,0; 6,9. Для этих измерений среднее значение составляет 6,31, а стандартное отклонение — 1,17. Это означает, что мы можем быть на 95 % уверены, что истинное значение глубины снежного покрова было где-то между 3,97 дюйма (6,31 – 2 × 1,17) и 8,65 дюйма (6,31 + 2 × 1,17). Не нужно вручную вычислять интеграл или нагружать компьютер, чтобы использовать R!

Даже *при использовании R* этот прием может быть полезен для определения минимального или максимального значения пределов интегрирования. Например, если нужно узнать вероятность того, что зажигательный шнур бомбы злодея продержится дольше 21 секунды, не нужно интегрировать от 21 до бесконечности. Что использовать в качестве верхней границы? Можно интегрировать от 21 до 25,46 (что составляет 20,6 + + 3 × 1,62) — это три стандартных отклонения от среднего значения. Три стандартных отклонения от среднего значения будут составлять 99,7 % от общей вероятности. Остальные 0,3 % лежат по обе стороны от распределения, поэтому только половина этого, 0,15 % от плотности вероятности, находится в области, превышающей 25,46. Так что если мы проведем интегрирование в пределах от 21 до 25,46, то упустим лишь небольшую вероятность в результате. Ясно, что можно было бы легко использовать R для интегрирования от 21 до чего-то действительно безопасного, например 30, но этот трюк позволяет выяснить, что такое «действительно безопасный».

События «n сигм»

Наверняка вы слышали о событии, описываемом в терминах *событий сигм*, например «падение цены акций было событием в восемь сигм». Это выражение означает, что наблюдаемые данные представляют собой восемь стандартных отклонений от среднего значения. Мы наблюдали прогрессирование одного, двух и трех стандартных отклонений от среднего значения в табл. 12.1, которые составляли значения 68, 95 и 99,7 % соответственно.

Исходя из этого легко догадаться, что событие с восьмью сигмами должно быть крайне маловероятным. Фактически, если вы наблюдаете данные, которые на пять стандартных отклонений отдалены от среднего значения, это, вероятно, является хорошим признаком того, что нормальное распределение не моделирует базовые данные точно.

В качестве примера растущей редкости возникновения события по мере его возрастания на *n* сигм предположим, что вы рассматриваете события, которые можете наблюдать в этот день. Некоторые очень распространены, например проснуться до восхода солнца. Другие встречаются реже, например проснуться в день рождения.

Таблица 12.2 показывает, сколько дней потребуется, чтобы ожидать увеличения события на одну сигму.

Таблица 12.2. Редкость события по мере его увеличения на n сигм

(- / +) Отклонение от среднего значения	Ожидается каждый (-е)
σ	3 дня
2σ	3 недели
3σ	Год
4σ	4 десятилетия
5σ	5 тысячелетий
6σ	1,4 миллиона лет

Таким образом, событие трех сигм — вы просыпаетесь и понимаете, что сегодня ваш день рождения, а событие шести сигм — вы просыпаетесь и понимаете, что на Землю летит гигантский астероид!

Бета-распределение и нормальное распределение

Из главы 5 вы помните, что бета-распределение позволяет оценить истинную вероятность с учетом наблюдения α желаемых результатов и β нежелательных, где общее количество результатов составляет $\alpha + \beta$. Можно не согласиться с тем, что нормальное распределение является действительно лучшим методом моделирования оценки параметров, учитывая, что мы знаем только среднее значение и стандартное отклонение любого заданного набора данных. В конце концов, можно было бы описать ситуацию, когда $\alpha = 3$ и $\beta = 4$, просто наблюдая три значения 1 и четыре значения 0. Это даст нам $\mu = 0{,}43$ и $\sigma = 0{,}53$. Затем можно сравнить бета-распределение при $\alpha = 3$ и $\beta = 4$ с нормальным распределением при $\mu = 0{,}43$ и $\sigma = 0{,}53$, как показано на рис. 12.9.

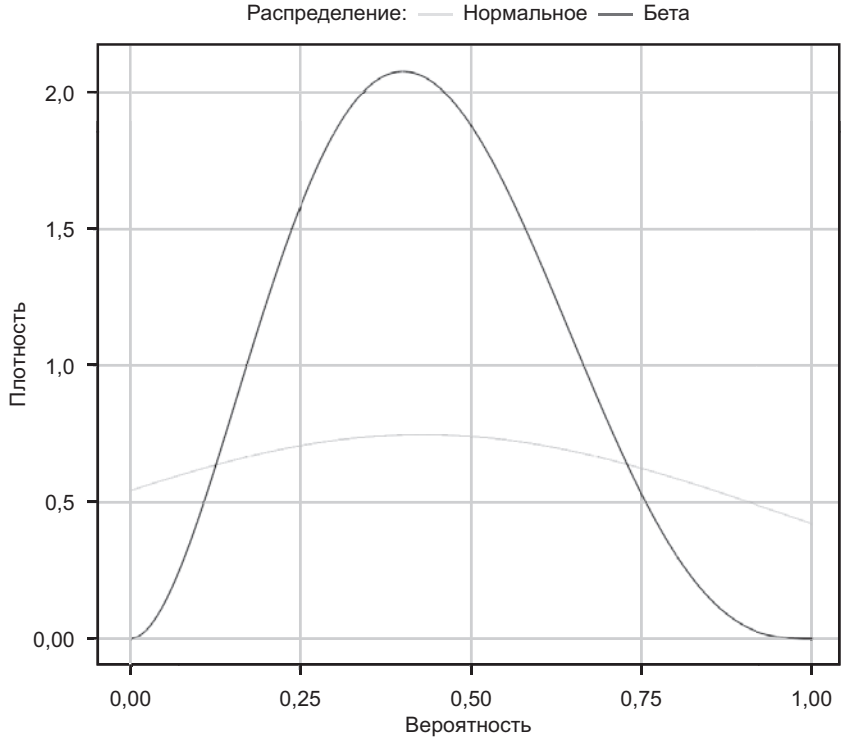


Рис. 12.9. Сравнение бета-распределения с нормальным распределением

Понятно, что эти распределения совершенно разные. Для обоих распределений центр масс появляется примерно в одном и том же месте, но границы нормального распределения выходят далеко за пределы нашего графика. Здесь скрыт ключевой момент: только когда вы ничего не знаете о данных, кроме их среднего значения и дисперсии, безопасно предполагать нормальное распределение.

Для бета-распределения мы знаем, что искомое значение должно лежать в диапазоне от 0 до 1. Нормальное распределение определяется от $-\infty$ до ∞ и часто включает значения, которые не могут существовать. Тем не менее в большинстве случаев это не является практически важным, поскольку такие измерения почти невозможны в вероятностных терминах. Но для нашего примера измерения вероятности наступления события эта недостающая информация важна для моделирования проблемы.

Хотя нормальное распределение и является очень мощным инструментом, оно не заменяет необходимости сбора дополнительной информации о проблеме.

Заключение

Нормальное распределение является продолжением использования среднего значения для оценки числа, полученного из наблюдений. Нормальное распределение объединяет среднее значение и стандартное отклонение, чтобы смоделировать, насколько наши наблюдения отличаются от среднего значения. Это важно, потому что это позволяет рассуждать об ошибке в измерениях вероятностным способом. Мы не только можем использовать среднее значение, чтобы сделать лучшее предположение, но и можем сделать вероятностные заявления о диапазонах возможных значений для оценки.

Упражнения

Для закрепления темы нормального распределения попробуйте ответить на эти вопросы.

1. Какова вероятность наблюдения значения на пять сигм большего или меньшего, чем среднее значение?

2. Лихорадка — это любая температура выше 100,4 градуса по шкале Фаренгейта. Учитывая следующие измерения, какова вероятность того, что у пациента жар?

   100,0; 99,8; 101,0; 100,5; 99,7.

3. Предположим, что в главе 11 мы попытались измерить глубину колодца по времени падения монет и получили следующие значения:

   2,5, 3, 3,5, 4, 2.

   Расстояние, на которое падает объект, может быть рассчитано (в метрах) по следующей формуле:

   $$\text{расстояние} = 1/2 \times G \times \text{время}^2,$$

   где G составляет 9,8 м/с(м/с). Какова вероятность того, что глубина колодца превышает 500 метров?

4. Какова вероятность того, что колодца нет (то есть колодец имеет фактическую глубину 0 метров)? Вы заметите, что вероятность выше, чем можно было бы ожидать, учитывая наблюдения, что колодец есть. Есть два хороших объяснения того, что эта вероятность выше, чем должна быть. Во-первых, нормальное распределение является плохой моделью для измерений; во-вторых, при составлении чисел для примера я выбрал значения, которые вы вряд ли увидите в реальной жизни. Что для вас более вероятно?

13

Инструменты оценки параметров: PDF, CDF и квантильная функция

До сих пор мы были сосредоточены на стандартных блоках нормального распределения и их использовании при оценке параметров. В этой главе мы еще немного углубимся в изучение математических инструментов, которые можно использовать, чтобы лучше давать оценки параметров. Возьмем задачу из реального мира и посмотрим, как по-разному подойти к ней, используя различные метрики, функции и визуализации.

В этой главе поговорим о функции плотности вероятности (*probability density function*, *PDF*) и накопительной функции распределения (*cumulative distribution function*, *CDF*), которая помогает легче определять вероятность диапазонов значений. Также затронем квантили, которые делят распределения вероятностей на части с равными вероятностями. Например, *процентиль* — это 100-я квантиль, то есть он делит распределение вероятностей на 100 равных частей.

Оценка коэффициента конверсии рассылки

Предположим, вы ведете блог и хотите знать вероятность того, что посетитель блога подпишется на вашу рассылку. В маркетинге побуждение пользователя выполнить желаемое действие называется *событием конверсии*, или просто *конверсией*, а вероятность того, что пользователь подпишется, называется *коэффициентом конверсии*.

Мы будем использовать бета-распределение для оценки p, вероятности подписки, при наличии k, количества подписавшихся людей, и n, общего количества посетителей. Двумя параметрами, необходимыми для бета-распределения, являются α, которая в этом случае представляет общее количество подписавшихся (k), и β, представляющая общее количество неподписавшихся людей ($n - k$).

В главе про бета-распределение вы узнали вводную информацию: как оно выглядит и как себя ведет. Теперь вы увидите, как использовать его в качестве основы для оценки параметров. Мы хотим не только создать единую оценку для коэффициента конверсии, но и предложить диапазон возможных значений, в котором, как мы можем быть уверены, располагается реальный коэффициент конверсии.

Функция плотности вероятности

Первым инструментом станет функция плотности вероятности. Мы уже встречались с PDF в этой книге: в главе 5, когда говорили о бета-распределении; в главе 9, когда использовали PDF для объединения байесовских априорных значений; и еще раз в главе 12, когда обсуждали нормальное распределение. PDF — это функция, которая принимает значение и возвращает вероятность этого значения.

В случае оценки истинного коэффициента конверсии для вашего списка рассылки, скажем, для первых 40 000 посетителей, вы получите 300 подписчиков. PDF в этом примере — это бета-распределение, где $\alpha = 300$ и $\beta = 39\,700$:

$$\text{Beta}(x;\ 300,\ 39\ 700) = \frac{x^{300-1}(1-x)^{39\ 700-1}}{\text{beta}(300,\ 39\ 700)}.$$

Мы потратили много времени, обсуждая среднее значение в качестве хорошей оценки для измерения, учитывая некоторую неопределенность.

У большинства PDF есть среднее значение, которое специально рассчитывается для бета-распределения следующим образом:

$$\mu_{\text{Beta}} = \frac{\alpha}{\alpha + \beta}.$$

Эта формула интуитивно понятна: разделите число результатов, которые нас интересуют (300), на общее количество результатов (40 000). Это то же самое среднее значение, которое получилось бы, если бы мы просто считали каждое письмо наблюдением 1, а все остальные наблюдения — 0, а затем усредняли их.

Среднее значение — это первая попытка оценить параметр для истинного коэффициента конверсии. Но нужно узнать и другие возможные значения коэффициента конверсии.

Визуализация и интерпретация PDF

PDF — это обычно полезная функция для понимания распределения вероятностей. На рис. 13.1 показано PDF для бета-распределения коэффициента конверсии блога.

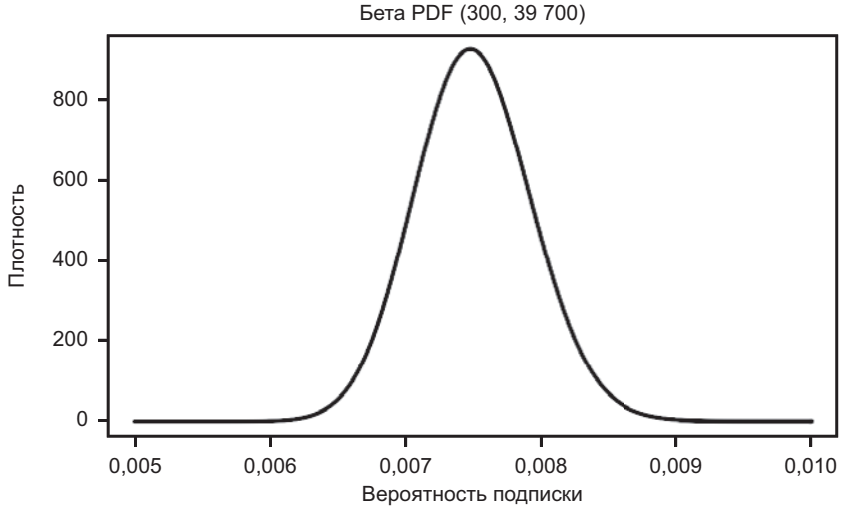


Рис. 13.1. Визуализация бета-PDF для наших убеждений об истинном коэффициенте конверсии

Что представляет собой PDF? Учитывая данные, мы знаем, что средний коэффициент конверсии блога:

$$\frac{\text{подписчики}}{\text{посетители}} = \frac{300}{40\ 000} = 0{,}0075\,,$$

или является *средним* показателем нашего распространения. Кажется маловероятным, что коэффициент конверсии составляет *ровно* 0,0075, а не 0,00751. Мы знаем, что общая площадь области под кривой PDF должна составлять до 1, поскольку эта PDF представляет вероятность всех возможных оценок. Мы можем оценить диапазоны значений для истинного коэффициента конверсии, посмотрев на область под кривой для диапазонов, которые нас интересуют. В математике эта область под кривой является *интегралом* и говорит о том, сколько общей вероятности находится в интересующей нас области PDF. Это то же самое, что и использование интегрирования с нормальным распределением в предыдущей главе.

У нас есть неопределенность в измерениях и есть среднее значение. Теперь было бы полезно выяснить, насколько более вероятно, что истинный коэффициент конверсии на 0,001 выше или ниже, чем среднее значение 0,0075, которое мы наблюдали. Это даст нам приемлемый предел погрешности (то есть мы будем рады любым значениям в этом диапазоне). Для этого можно рассчитать вероятность того, что фактический коэффициент будет ниже 0,0065, и вероятность того, что он будет выше 0,0085, а затем сравнить их. Вероятность того, что коэффициент конверсии на самом деле намного ниже, чем наблюдения, вычисляется следующим образом:

$$P(\text{намного ниже}) = \int_0^{0{,}0065} \text{Beta}(300,\ 39\ 700) = 0{,}008\,.$$

Помните, что при взятии интеграла функции мы просто складываем все маленькие части нашей функции. Итак, если взять интеграл от 0 до 0,0065 для бета-распределения с α 300 и β 39 700, то, сложив все вероятности для значений в этом диапазоне, мы определим вероятность того, что истинный коэффициент конверсии расположен где-то в диапазоне от 0 до 0,0065.

Мы можем задавать вопросы и о других экстремумах: какова вероятность того, что на самом деле была получена необычно плохая выборка и наш истинный коэффициент конверсии намного выше? Например, значение

больше, чем, скажем, 0,0085 (что означает лучший коэффициент конверсии, чем мы надеялись)?

$$P(\text{намного выше}) = \int_{0,0085}^{1} \text{Beta}(300,\ 397\ 000) = 0{,}012\,.$$

Здесь мы интегрируем от 0,0085 до максимально возможного значения, равного 1, чтобы определить вероятность того, что истинное значение находится где-то в этом диапазоне. Коэффициент конверсии выше на 0,001 или больше наблюдаемого — вероятность этого исхода выше, чем вероятность того, что он на 0,001 меньше или ниже наблюдаемого. Если бы пришлось принимать решение с ограниченными имеющимися у нас данными, то все равно можно было бы подсчитать, насколько один экстремум вероятнее, чем другой:

$$\frac{P(\text{намного выше})}{P(\text{намного ниже})} = \frac{\int_{0,0085}^{1} \text{Beta}(300,\ 397\ 000)}{\int_{0}^{0,0065} \text{Beta}(300,\ 39\ 700)} = \frac{0{,}012}{0{,}008} = 1{,}5.$$

Таким образом, вероятность того, что истинный коэффициент конверсии превысит 0,0085, на 50 % выше, чем вероятность того, что он ниже 0,0065.

Работа с PDF в R

В этой книге мы уже использовали две функции R для работы с PDF: `dnorm()` и `dbeta()`. Для большинства известных распределений вероятности R поддерживает эквивалентную функцию `dfunction()` для вычисления PDF.

Такие функции, как `dbeta()`, полезны и для аппроксимации непрерывного PDF, например, когда нужно быстро получить такие значения:

```
xs <- seq(0.005,0.01,by=0.00001)
xs.all <- seq(0,1,by=0.0001)
plot(xs,dbeta(xs,300,40000-300),type='l',lwd=3,
     ylab="плотность",
     xlab="вероятность подписки",
     main="Бета PDF (300,39700)").
```

ПРИМЕЧАНИЕ

Чтобы понять код построения графика, см. приложение А.

В этом примере кода мы создаем последовательность значений, каждое из которых равно 0,00001 — маленькое, но не бесконечно малое, как это действительно было бы в непрерывном распределении. При нанесении этих значений на график мы видим нечто, достаточно близкое к действительно непрерывному распределению (см. рис. 13.1).

Введение в кумулятивную функцию распределения

Наиболее распространенное математическое использование PDF — это интегрирование для определения вероятностей, связанных с различными диапазонами. Тем не менее можно сэкономить много усилий с помощью *кумулятивной функции распределения (CDF)*, которая суммирует все части распределения, заменяя большую часть вычислений.

CDF принимает значение и возвращает вероятность получения этого или меньшего значения. Например, CDF для Beta (300, 397 000) при $x = 0{,}0065$ составляет приблизительно 0,008. Это означает, что вероятность действительного коэффициента конверсии, равного 0,0065 или менее, равна 0,008.

CDF получает эту вероятность, принимая совокупную площадь области под кривой для PDF (для тех, кто знаком с высшей математикой, CDF является *антипроизводной* PDF). Этот процесс можно объединить в два этапа: (1) определить совокупную площадь области под кривой для каждого значения PDF и (2) построить эти значения. Это и будет наша CDF. Значение кривой при любом данном значении x представляет собой вероятность получения значения x или меньшего. При 0,0065 значение кривой будет равно 0,008, как мы рассчитывали ранее.

Чтобы понять, как все работает, разберем PDF для нашей задачи на части по 0,0005 и сосредоточимся на области PDF, которая имеет наибольшую плотность вероятности: области от 0,006 до 0,009.

На рис. 13.2 показана совокупная область под кривой для бета-PDF (300,39700). Как видите, кумулятивная область под кривой учитывает все области в частях слева.

Говоря математически, на рис. 13.2 представлена следующая последовательность интегралов:

$$\int_0^{0{,}0065} \mathrm{Beta}(300,\ 397\ 000),$$

$$\int_0^{0{,}0065} \mathrm{Beta}(300,\ 397\ 000) + \int_{0{,}0065}^{0{,}007} \mathrm{Beta}(300,\ 397\ 000),$$

$$\int_0^{0{,}0065} \mathrm{Beta}(300,\ 397\ 000) + \int_{0{,}0065}^{0{,}007} \mathrm{Beta}(300,\ 397\ 000) +$$

$$+ \int_{0{,}007}^{0{,}0075} \mathrm{Beta}(300,\ 397\ 000)$$

(и так далее).

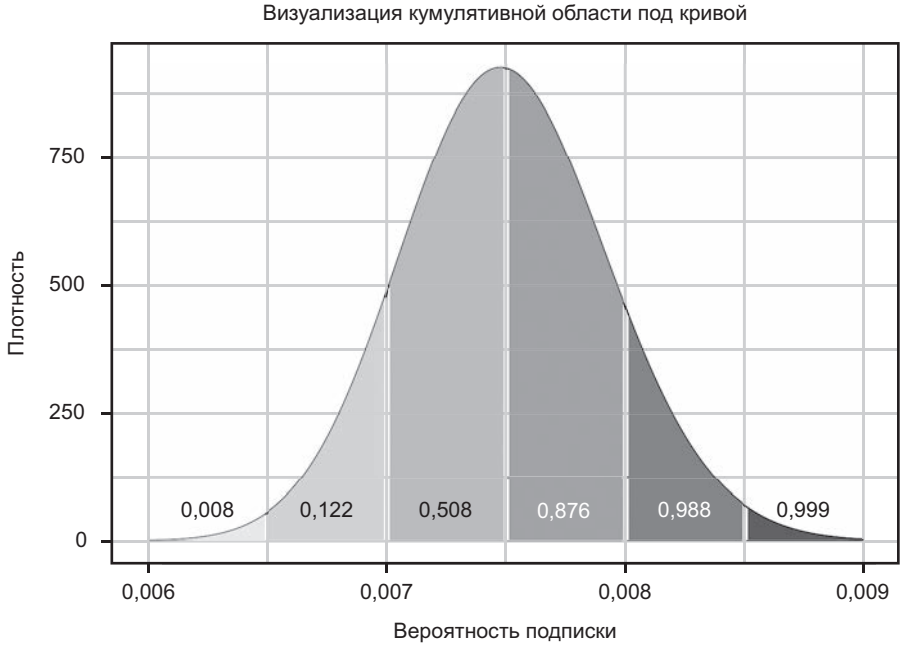


Рис. 13.2. Визуализация кумулятивной области под кривой

Используя этот подход, по мере продвижения по PDF мы учитываем все более высокую вероятность, пока кумулятивная область не станет 1, или полной уверенностью. Чтобы превратить это в CDF, можно представить функцию, которая просматривает только эти области под кривой. На рис. 13.3 показано, что произойдет, если мы нанесем область под кривой для каждой из наших точек, которые находятся на расстоянии 0,0005.

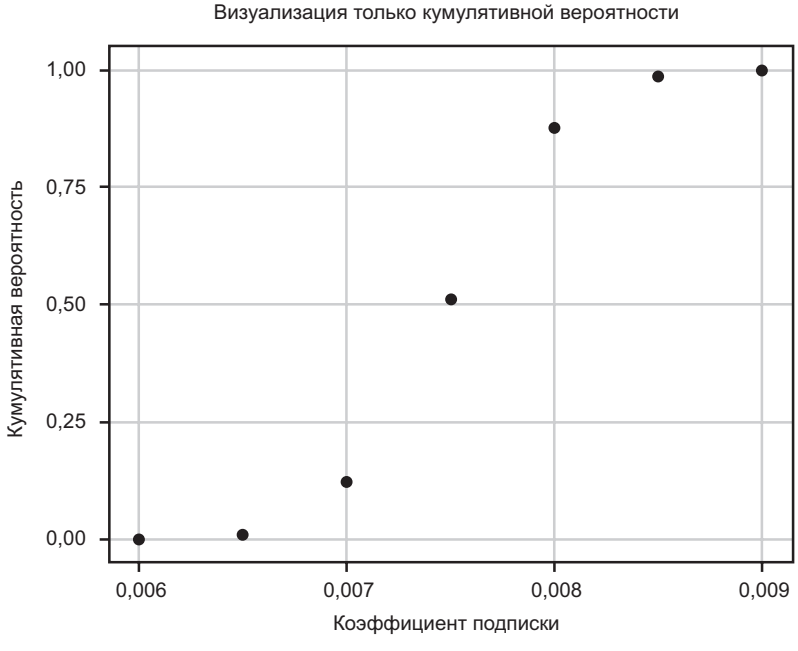


Рис. 13.3. Построение только кумулятивной вероятности из рис. 13.2

Теперь есть способ визуализировать то, как изменяется кумулятивная область под кривой при перемещении по значениям для нашей PDF. Конечно, проблема в том, что мы используем эти отдельные фрагменты. В действительности CDF просто использует бесконечно маленькие фрагменты PDF, поэтому мы получаем красивую плавную линию (рис. 13.4).

В нашем примере мы вывели CDF визуально и интуитивно. Математически получить CDF намного сложнее, и расчет часто приводит к очень сложным уравнениям. К счастью, обычно для работы с CDF используется код, что будет показано в следующих разделах.

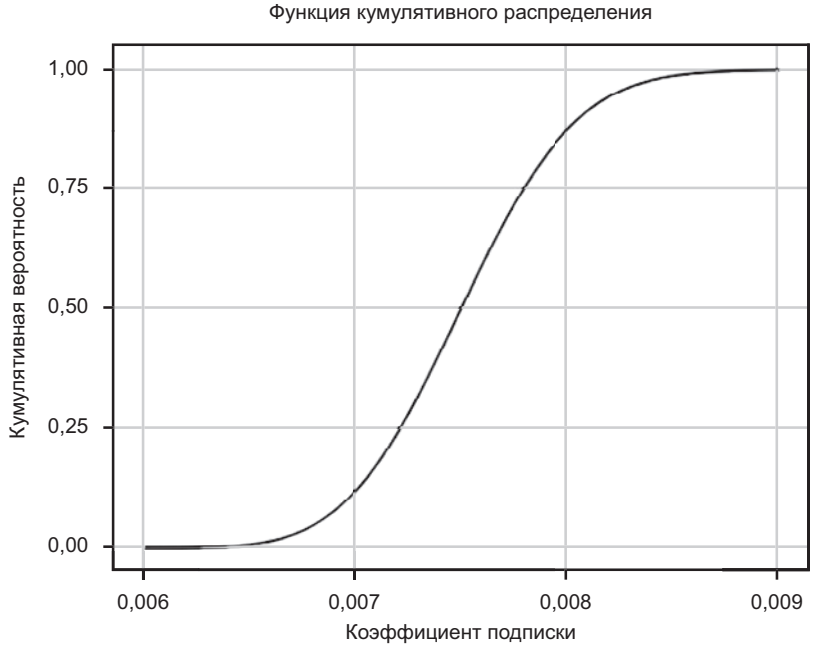


Рис. 13.4. CDF для нашей проблемы

Визуализация и интерпретация CDF

PDF наиболее полезна визуально для быстрой оценки того, где находится пик распределения, и для получения ширины (дисперсии) и формы распределения. Но с PDF очень сложно рассуждать о вероятности различных диапазонов, основываясь на визуальном представлении. CDF — намного более подходящий для этого инструмент. Например, можно использовать CDF (рис. 13.4), чтобы визуально обосновать гораздо более широкий диапазон вероятностных оценок для задачи, чем при использовании только PDF. Рассмотрим несколько примеров того, как можно использовать этот удивительный математический инструмент.

Нахождение медианы

Медиана — это точка в данных, в которой половина значений приходится на одну сторону, а половина на другую. Это точное *серединное значение*

наших данных. Другими словами, вероятность того, что значение больше медианы, и вероятность того, что оно меньше медианы, равна 0,5. Медиана особенно полезна для суммирования данных в тех случаях, когда они содержат экстремальные значения.

В отличие от среднего значения вычисление медианы может быть довольно сложным. Для небольших дискретных случаев это так же просто, как упорядочить свои наблюдения и выбрать значение в середине. Но для непрерывного распределения вроде бета-распределения это немного сложнее. К счастью, можно легко определить медиану по визуализации CDF. Просто проведите линию от точки, где совокупная вероятность равна 0,5; это означает, что 50 % значений располагается ниже этой точки, а 50 % — выше. Как показано на рис. 13.5, точка, где эта линия пересекает ось *X*, дает медиану!

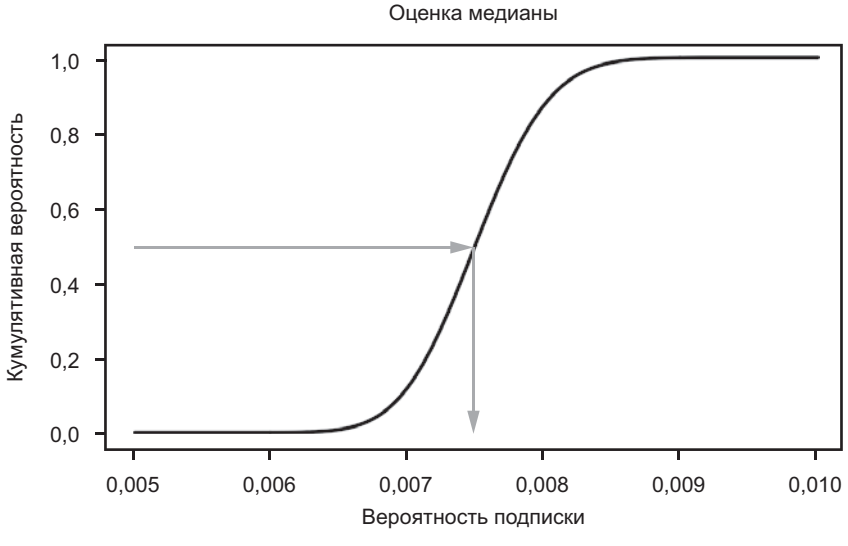


Рис. 13.5. Визуальная оценка медианы с помощью CDF

Можно заметить, что медиана для наших данных находится где-то между 0,007 и 0,008 (это очень близко к среднему значению 0,0075 и означает, что данные не особенно искажены).

Визуальное приближение интегралов

При работе с диапазонами вероятностей часто нужно знать вероятность того, что истинное значение находится где-то между некоторым значением y и некоторым значением x.

Можно решить такого рода проблемы с помощью интегрирования, но даже если R и упрощает решение интегралов, на понимание данных и постоянное использование R для вычисления интегралов уходит очень много времени. Нам нужно, чтобы приблизительная оценка вероятности подписки посетителя на блог попадала в определенный диапазон, и поэтому использовать интегрирование не требуется. CDF позволяет очень легко узнать, имеет ли определенный диапазон значений очень высокую или очень низкую вероятность появления.

Чтобы оценить вероятность того, что коэффициент конверсии находится между 0,0075 и 0,0085, можно отследить линии от оси X в этих точках, а затем посмотреть, где они встречаются с осью Y. Расстояние между двумя точками является приблизительным интегралом (рис. 13.6).

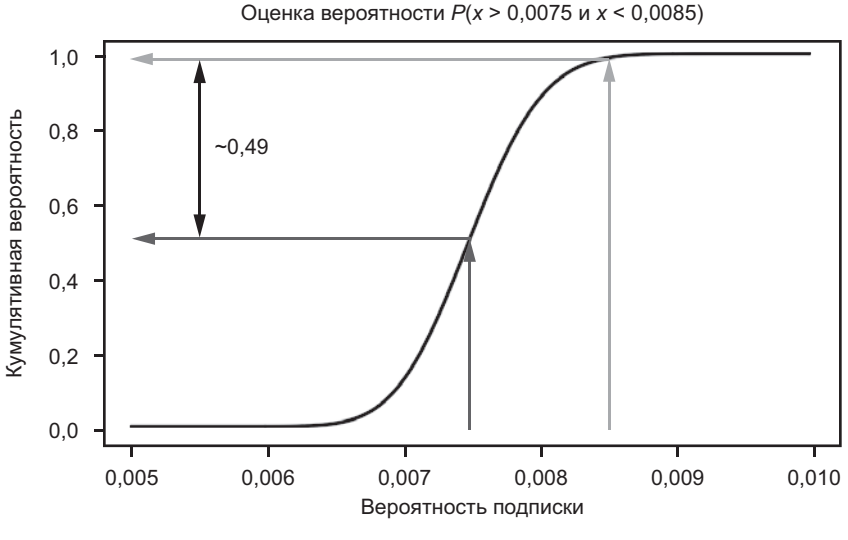


Рис. 13.6. Визуальное выполнение интегрирования с использованием CDF

Мы видим, что по оси *Y* эти значения находятся в диапазоне от 0,5 до 0,99, это означает, что имеется приблизительно 49 %-ная вероятность того, что истинный коэффициент конверсии находится где-то между этими двумя значениями. Самое приятное, что нам не нужно было заниматься интегрированием, потому что CDF представляет собой интеграл от минимума нашей функции до всех возможных значений.

Поскольку почти все вероятностные вопросы об оценке параметров включают знание вероятности, связанной с определенными диапазонами убеждений, CDF часто является гораздо более полезным визуальным инструментом, чем PDF.

Оценка доверительных интервалов

Анализ вероятности диапазонов значений приводит нас к очень важной концепции вероятности: *доверительному интервалу*. Доверительный интервал — это нижняя и верхняя границы значений, обычно центрированных по среднему значению, описывающих диапазон высокой вероятности, как правило, 95, 99 или 99,9 %. Когда мы говорим что-то вроде «95 %-ный доверительный интервал составляет от 12 до 20», мы имеем в виду, что существует 95 %-ная вероятность того, что наше истинное измерение находится где-то между 12 и 20. Доверительные интервалы — хороший способ описания диапазона возможностей, когда мы имеем дело с неопределенной информацией.

ПРИМЕЧАНИЕ

То, что мы называем доверительным интервалом, в байесовской статистике может называться по-другому, например «критическая область» или «критический интервал». В некоторых традиционных школах статистики «доверительный интервал» имеет несколько другое значение. Но эта тема выходит за рамки данной книги.

Оценить доверительные интервалы можно с помощью CDF. Допустим, нужно узнать диапазон, который охватывает 80 % возможных значений для истинного коэффициента конверсии. Решим эту задачу, комбинируя предыдущие подходы: рисуем линии на оси *Y* от 0,1 до 0,9, чтобы покрыть 80 %, а затем смотрим, где на оси *X* они пересекаются с CDF (рис. 13.7).

Ось *X* пересекается примерно с 0,007 и 0,008, это означает, что существует 80 %-ная вероятность того, что истинный коэффициент конверсии окажется где-то между этими двумя значениями.

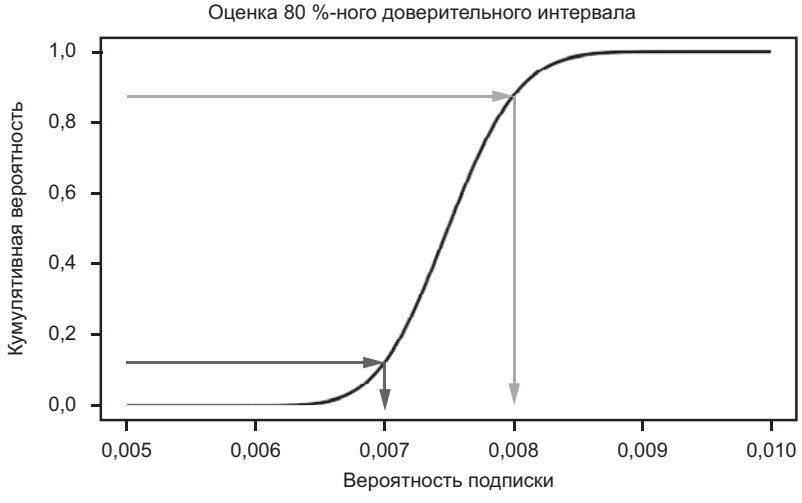


Рис. 13.7. Оценка доверительных интервалов визуально с использованием CDF

Использование CDF в R

Подобно тому как почти во всех основных PDF есть функция, начинающаяся с *d*, например `dnorm()`, функции CDF начинаются с *p*, например `pnorm()`. Чтобы вычислить вероятность того, что Beta(300,39700) меньше 0,0065, в R можно просто вызвать `pbeta()`:

```
pbeta(0.0065,300,39700)
> 0.007978686
```

Для вычисления истинной вероятности того, что коэффициент конверсии больше 0,0085, можно сделать следующее:

```
pbeta(1,300,39700) – pbeta(0.0085,300,39700)
> 0.01248151
```

Самое замечательное в CDF то, что не имеет значения, является ли ваше распределение дискретным или непрерывным. Например, если бы нужно было определить вероятность получения трех или менее орлов в бросках пяти монеток, мы бы использовали CDF для биномиального распределения так:

```
pbinom(3,5,0.5)
> 0.8125
```

Квантильная функция

Возможно, вы заметили, что средние и доверительные интервалы, взятые визуально с CDF, определить математически нелегко. С помощью визуализаций мы просто рисовали линии от оси *Y* и использовали их, чтобы найти точку на оси *X*.

Математически CDF похожа на любую другую функцию тем, что она принимает *x*, часто представляющее значение, которое нужно оценить, и дает значение *y*, которое представляет совокупную вероятность. Но нет очевидного способа сделать это в обратном порядке; то есть нельзя передать в одну и ту же функцию *y*, чтобы получить *x*. Представим, что есть функция, которая возводит значения в квадрат. Мы знаем, что `square (3) = 9`, но понадобится совершенно новая функция — функция квадратного корня, — чтобы узнать, что корень квадратный из 9 равен 3.

Однако обращение функции — это *именно то, что мы сделали* в предыдущем разделе для оценки медианы: выбрали 0,5 на оси *Y*, а затем проследили ее обратно до оси *X*. То, что мы сделали визуально, — вычислили *инверсию* CDF.

Хотя вычисление инверсии CDF визуально просто для получения оценок, нужна отдельная математическая функция для вычисления точных значений. Инверсия CDF — невероятно распространенный и полезный инструмент, называемый *квантильной функцией*. Чтобы вычислить точное значение для медианы и доверительного интервала, нужно использовать квантильную функцию для бета-распределения. Как и CDF, квантильную функцию часто очень сложно получить и использовать математически, поэтому призовем в помощь язык R, который сделает всю грязную работу.

Визуализация и понимание квантильной функции

Поскольку квантильная функция является инверсией CDF, она выглядит как CDF, повернутая на 90 градусов (рис. 13.8).

Всякий раз, когда вы слышите такие фразы, как:

«Лучшие 10 % студентов...»,

«Наименее обеспеченные 20 % работников зарабатывают меньше, чем...»,

«Верхний квартиль имеет заметно лучшую производительность, чем...» —

речь идет о значениях, которые находятся с помощью квантильной функции. Чтобы визуально найти квантиль, найдите интересующую вас величину по оси *X* и посмотрите, где она встречается с осью *Y*. Значение на оси *Y* является значением для этого квантиля. Имейте в виду, что если речь идет о «верхних 10 %», то нужен квантиль 0,9.

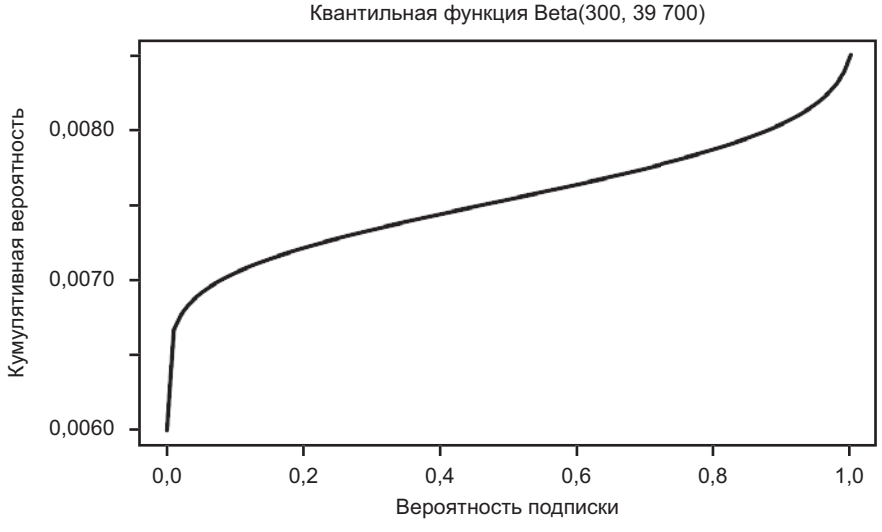


Рис. 13.8. Визуально квантильная функция это повернутая CDF

Вычисление квантилей в R

В R есть функция `qnorm()` для вычисления квантилей. Эта функция очень полезна, чтобы узнать, какие значения являются границами распределения вероятностей. Например, если требуется найти значение, которое меньше 99,9 % распределения, можно использовать `qbeta()` с квантилем, который требуется вычислить, в качестве первого аргумента, а также с параметрами альфа и бета нашего бета-распределения в качестве второго и третьего аргументов:

```
qbeta(0.999,300,39700)
> 0.008903462
```

Результат равен 0,0089, это означает, что мы можем быть на 99,9 % уверены, что истинный коэффициент конверсии для рассылки составляет менее 0,0089. Затем можно использовать квантильную функцию для быстрого вычисления точных значений доверительных интервалов наших оценок. Чтобы найти 95 %-ный доверительный интервал, мы можем найти значения, превышающие нижний квантиль на 2,5 %, и значения ниже, чем верхний квантиль, на 97,5 %, а интервал между ними — 95 %-ный доверительный интервал (неучтенная область составляет 5 % от плотности вероятности в обеих крайностях). Можно легко рассчитать их для наших данных с помощью `qbeta()`:

Наша нижняя граница — `qbeta(0,025,300,39700)=0,0066781`

Наша верхняя граница — `qbeta (0,975,300,39700)=0,0083686`

Теперь мы на 95 % уверены, что реальный коэффициент конверсии для посетителей блогов находится где-то между 0,67 и 0,84 %.

Можно, конечно, увеличить или уменьшить эти пороговые значения в зависимости от того, насколько велика должна быть уверенность. Теперь можем легко определить точный диапазон коэффициента конверсии с помощью этих инструментов. Хорошая новость в том, что их можно использовать и для прогнозирования диапазонов значений будущих событий.

Предположим, что статья в вашем блоге становится вирусной и привлекает 100 000 посетителей. Исходя из расчетов, мы знаем, что следует ожидать от 670 до 840 новых подписчиков на рассылку по электронной почте.

Заключение

Мы рассмотрели множество вопросов и затронули интересную взаимосвязь между функцией плотности вероятности (PDF), кумулятивной функцией распределения (CDF) и квантильной функцией. Это базовые инструменты для оценки параметров и расчета уверенности в этих оценках. Можно не только сделать правильное предположение о том, каким может быть неизвестное значение, но и определить доверительные интервалы, которые с высокой точностью представляют возможные значения для параметра.

Упражнения

Чтобы убедиться, что вы понимаете инструменты оценки параметров, попробуйте ответить на эти вопросы.

1. Используя пример кода для построения PDF на с. 155, постройте функции CDF и квантильную.
2. Возвращаясь к задаче измерения снежного покрова из главы 10, скажем, что у вас есть следующие измерения (в дюймах) снежного покрова:

   7,8, 9,4, 10,0, 7,9, 9,4, 7,0, 7,0, 7,1, 8,9, 7,4.

   Каков 99,9 %-ный доверительный интервал для истинного значения снежного покрова?
3. Девочка продает конфеты. Пока она посетила 30 домов и продала 10 конфет. Сегодня она посетит еще 40 домов. Каков 95 %-ный доверительный интервал для того, сколько конфет она продаст за остаток дня?

14

Оценка параметров с априорными вероятностями

В предыдущей главе мы рассмотрели использование некоторых важных математических инструментов оценки коэффициента конверсии для посетителей блога, подписавшихся на рассылку. Но мы еще не рассмотрели одну из самых важных частей оценки параметров: использование существующих представлений о задаче. В этой главе вы увидите, как можно использовать наши предыдущие вероятности в сочетании с данными наблюдений, чтобы получить более точную оценку, которая сочетает существующие знания с собранными данными.

Прогнозирование коэффициентов конверсии рассылки

Чтобы понять, как изменяется бета-распределение при получении информации, посмотрим на другой коэффициент конверсии. В этом примере мы попытаемся выяснить, с какой скоростью подписчики нажимают на ссылку после того, как они открыли ваше письмо. Большинство компаний, предоставляющих услуги по управлению рассылкой, в реальном времени сообщают вам, сколько людей открыли сообщение и нажали на ссылку.

Наши данные пока говорят, что из первых пяти человек, открывших письмо, двое нажимают на ссылку. На рис. 14.1 показано бета-распределение этих данных.

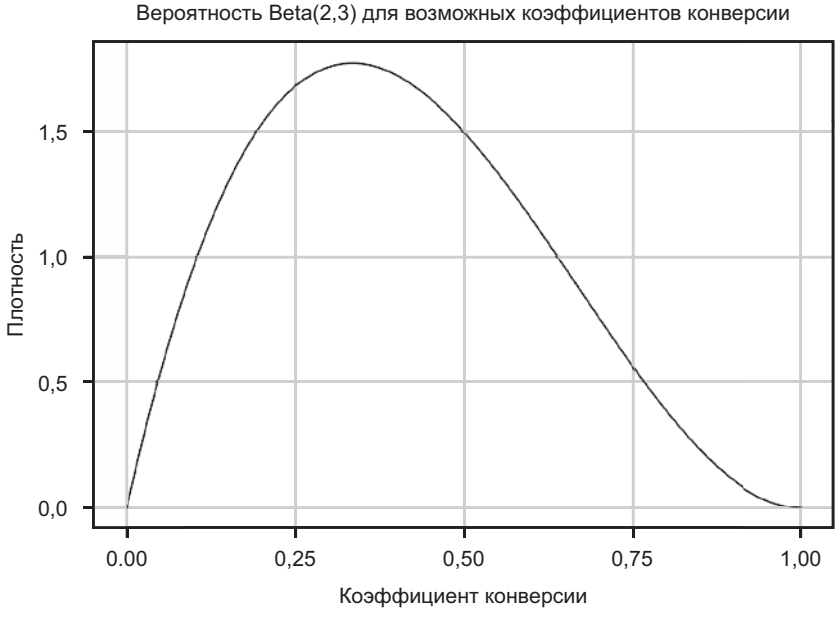


Рис. 14.1. Бета-распределение наших наблюдений

Рисунок 14.1 показывает распределение Beta(2,3). Мы использовали эти цифры, потому что два человека перешли по ссылке, а трое — нет. В отличие от предыдущей главы, где у нас был довольно узкий скачок возможных значений, здесь мы имеем огромный диапазон возможных значений для истинного коэффициента конверсии, потому что у нас очень мало информации для работы. Рисунок 14.2 показывает CDF для этих данных, чтобы помочь нам легче рассуждать об этих вероятностях.

95 %-ный доверительный интервал (то есть 95 %-ная вероятность того, что истинный коэффициент конверсии находится где-то в этом диапазоне) отмечен, чтобы его было легче увидеть. На данный момент наши данные говорят, что истинный коэффициент конверсии может располагаться где угодно между 0,05 и 0,8! Это отражение того, как мало информации мы на самом деле получили. Учитывая, что у нас было две конверсии, мы знаем, что истинная ставка не может быть равна 0, и поскольку у нас было три

не конверсии, мы также знаем, что она не может быть равна 1. Почти все остальное — справедливо.

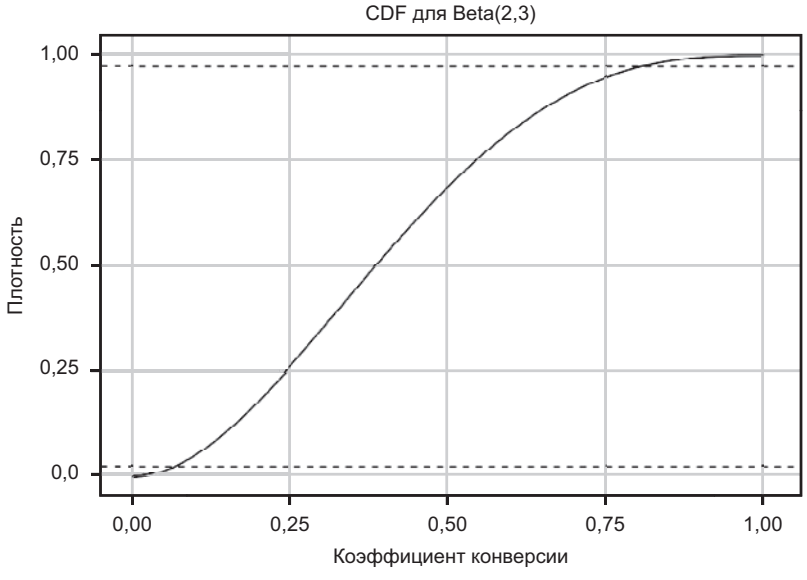


Рис. 14.2. CDF для нашего наблюдения

Использование широкого контекста с априорными вероятностями

Подождите секунду — вы можете ничего не знать о рассылках, но 80 %-ный рейтинг переходов по ссылке — это маловероятно. Я подписываюсь на множество рассылок, но определенно не перехожу к контенту в 80 % случаев, когда открываю письмо. Принимать эти 80 % за чистую монету кажется наивным, когда я рассматриваю собственное поведение.

Оказывается, ваш провайдер тоже считает это подозрительным. Давайте посмотрим на более широкий контекст. По данным вашего провайдера, для блогов, относящихся к той же категории, что и ваш, только 2,4 % людей, открывающих письма, переходят к контенту.

Из главы 9 вы узнали, как можно использовать полученную информацию, чтобы изменить убеждение в том, что Хан Соло может успешно перемещаться по астероидной области. Наши данные говорят одно, но исходная

информация утверждает другое. Как вы уже знаете, в байесовских терминах данные, которые мы наблюдали, являются нашей *правдоподобностью*, а информация внешнего контекста — в данном случае из личного опыта и от провайдера — *априорной вероятностью*. Наша задача сейчас состоит в том, чтобы выяснить, как моделировать априорные вероятности. К счастью, в отличие от случая с Ханом Соло у нас действительно имеются данные, чтобы упростить задачу.

Коэффициент конверсии от провайдера, равный 2,4 %, дает отправную точку: теперь мы знаем, что нужно бета-распределение со средним значением примерно 0,024. (Среднее значение бета-распределения составляет $\alpha/(\alpha + \beta)$.) Однако это все еще оставляет возможные варианты: Beta(1,41), Beta(2,80), Beta(5200), Beta(24 976) и т. д. Итак, что же из этого нужно использовать? Изобразим некоторые из них на графике (рис. 14.3).

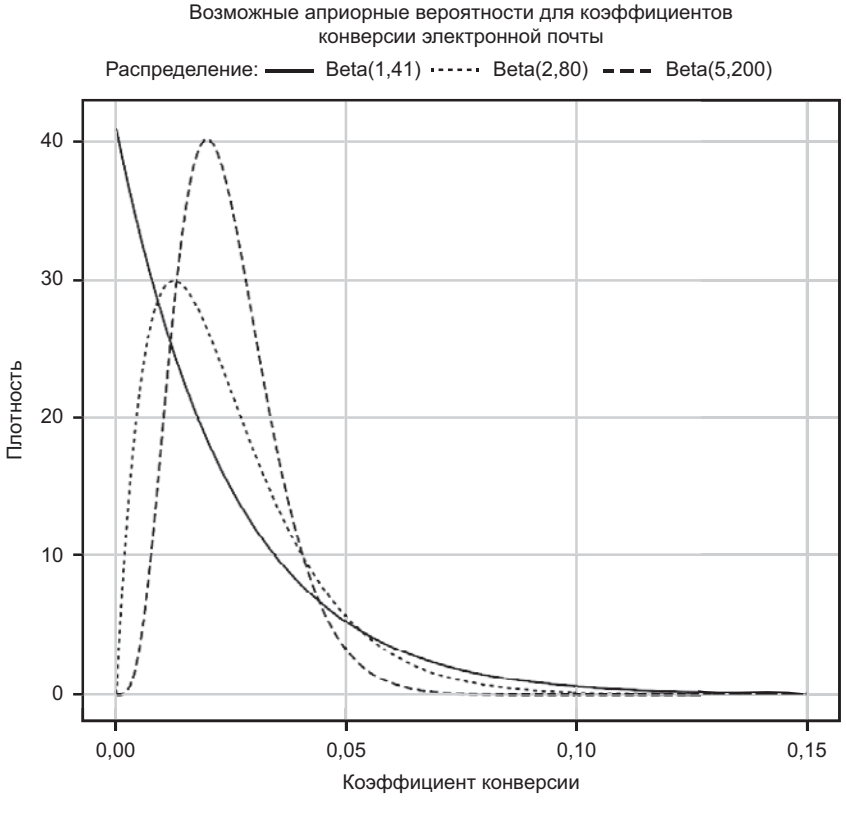


Рис. 14.3. Сравнение различных возможных априорных вероятностей

Как видите, чем меньше $\alpha + \beta$, тем шире распределение. Проблема заключается в том, что даже самый свободный вариант, который мы имеем, Beta(1,41), кажется слишком пессимистичным, так как большая часть плотности вероятности помещается в очень низкие значения. Но мы будем придерживаться этого распределения, поскольку оно основано на 2,4 %-ном коэффициенте конверсии в данных от провайдера и является самым слабым из приоритетов. «Слабая» априорная вероятность означает, что она будет легко переопределена фактическими данными, поскольку мы соберем еще больше информации. Более сильная априорная вероятность, такая как Beta(5200), потребовала бы больше доказательств для изменения (посмотрим, как это будет выглядеть дальше). Решение о том, следует ли использовать строгую априорную вероятность, является оценочным, исходя из того, насколько сильно вы ожидаете, что априорные данные описывают то, что вы делаете в данный момент. Как мы увидим, даже слабый априорный показатель может помочь сделать наши оценки более реалистичными при работе с небольшими объемами данных. Помните, что при работе с бета-распределением можно вычислить апостериорное распределение (сочетание нашей вероятности и априорной вероятности), просто сложив вместе параметры для двух бета-распределений:

$$\text{Beta}(\alpha_{\text{апостериорное}}, \beta_{\text{апостериорное}}) =$$

$$= \text{Beta}(\alpha_{\text{правдоподобности}} + \alpha_{\text{априорное}}, \beta_{\text{правдоподобности}} + \beta_{\text{априорное}}).$$

Используя эту формулу, мы можем сравнить свои убеждения с априорной вероятностью и без априорной вероятности, как показано на рис. 14.4.

Ого! Выглядит довольно отрезвляюще. Несмотря на то что мы работаем с относительно слабой априорной вероятностью, мы видим, что это оказало огромное влияние на то, что мы считаем реалистичными коэффициентами конверсии. Обратите внимание, что для правдоподобности без априорных данных мы считаем, что коэффициент конверсии может достигать 80 %. Как уже упоминалось, это очень подозрительно; любой опытный маркетолог, работающий с электронной почтой, скажет вам, что 80 %-ный коэффициент конверсии — это неслыханно. Добавление априорной вероятности к правдоподобности корректирует наши убеждения, так что они становятся намного более разумными. Но я все еще думаю, что наши обновленные убеждения немного пессимистичны. Может быть, истинный коэффициент конверсии не равен 40 %, но он все же может быть лучше, чем предполагает нынешнее апостериорное распределение.

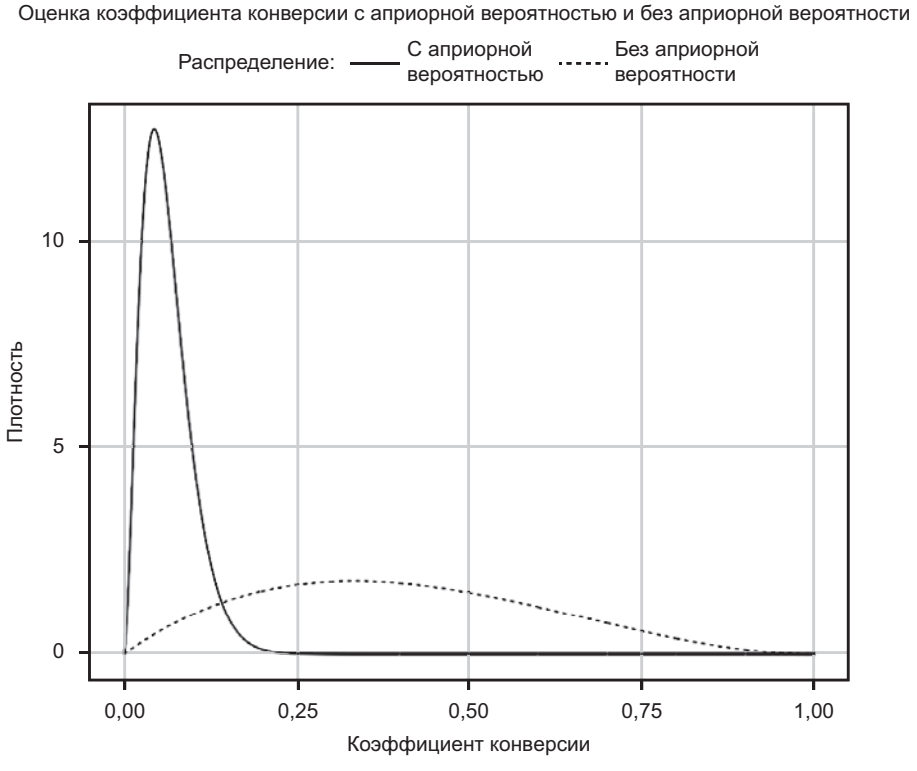


Рис. 14.4. Сравнение правдоподобия (без априорной вероятности) с апостериорной вероятностью

Как можно доказать, что блог имеет лучший коэффициент конверсии, чем сайты, указанные в данных провайдера, имеющие коэффициент 2,4 %? Как бы поступил любой рациональный человек? Предоставил больше данных! Мы ждем несколько часов, чтобы получить больше результатов, и выясняем, что из 100 человек, открывших письмо, 25 перешли по ссылке! Давайте посмотрим на разницу между нашей новой апостериорной вероятностью и правдоподобностью (рис. 14.5).

По мере того как мы продолжаем собирать данные, мы видим, что апостериорное распределение с использованием априорной вероятности начинает смещаться в сторону без априорной вероятности. Априорная вероятность по-прежнему контролирует наши данные, давая более консервативную оценку истинного коэффициента конверсии. Однако при добавлении доказательств к нашей правдоподобности она начинает оказывать большее

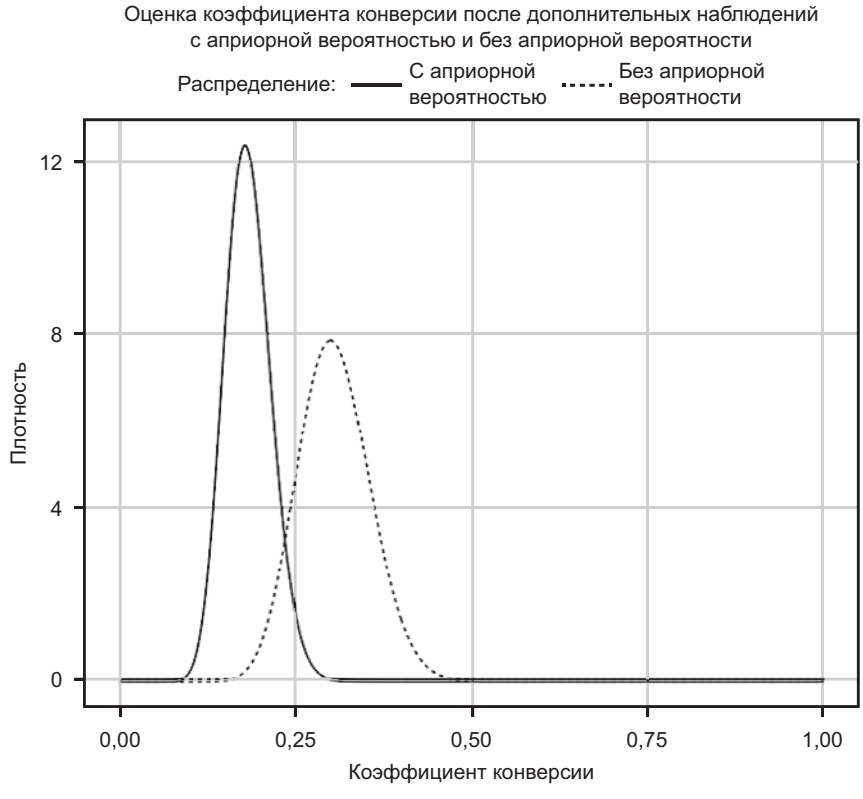


Рис. 14.5. Обновление убеждений с помощью большего количества данных

влияние на то, как выглядят апостериорные убеждения. Другими словами, дополнительные наблюдаемые данные делают то, что и должны: медленно раскачивают наши убеждения, чтобы соответствовать реальности. Так что давайте подождем еще ночь и вернемся, имея на руках еще больше данных!

Утром мы видим, что 300 подписчиков открыли письма и 86 из них нажали на ссылку. На рис. 14.6 показаны наши обновленные убеждения.

То, что мы наблюдаем здесь, является наиболее важным моментом в байесовской статистике: чем больше данных собирается, тем больше наши априорные убеждения уменьшаются в результате доказательств. Когда у нас почти не было доказательств, наша вероятность предложила некоторые варианты, которые, как мы знаем, абсурдны (например 80 % переходов) как интуитивно, так и из личного опыта. В свете небольшого количества доказательств наши априорные убеждения опровергли все имеющиеся данные.

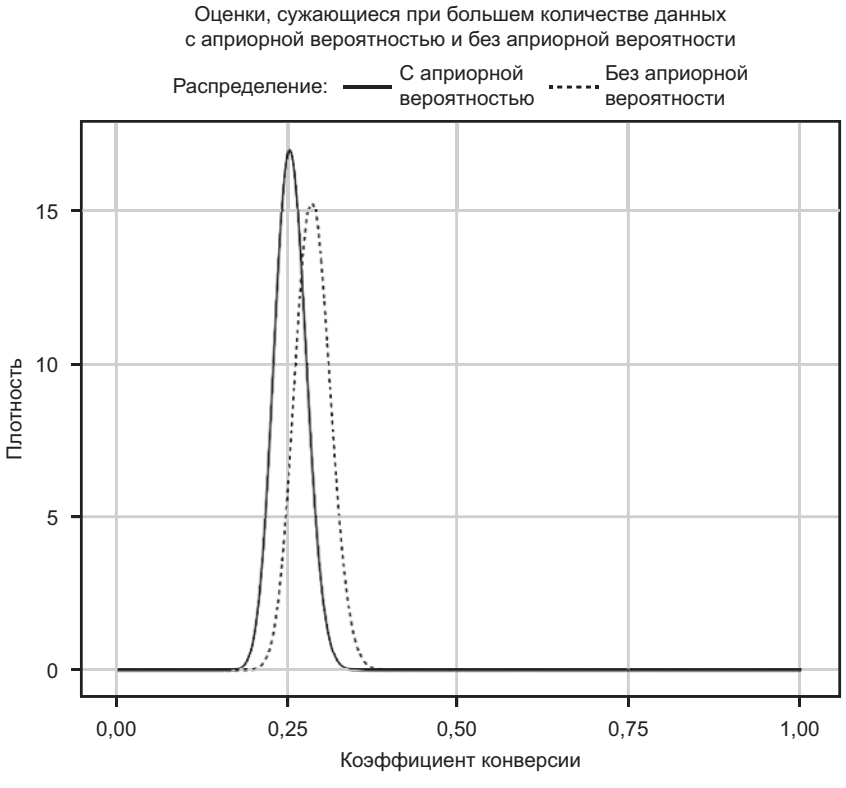


Рис. 14.6. Наши апостериорные убеждения с добавлением еще большего количества данных

Но по мере того как мы продолжаем собирать данные, которые не согласуются с априорными вероятностями, последующие убеждения смещаются в сторону того, что говорят нам собранные данные, и отходят от первоначальной априорной вероятности.

Другим важным выводом является то, что мы начали с довольно слабой априорной вероятности. Даже тогда, после всего лишь одного дня сбора сравнительно небольшого набора данных, мы смогли найти апостериорную вероятность, которая кажется гораздо более разумной.

Распределение априорных вероятностей в этом случае очень помогло сделать оценку намного более реалистичной при отсутствии данных. Это априорное распределение вероятностей было основано на реальных данных, поэтому мы могли быть вполне уверены, что оно поможет приблизить

оценку к реальности. Тем не менее во многих случаях никаких данных для сохранения априорных вероятностей обычно нет. Так что же делать?

Априорная вероятность как средство измерения опыта

Поскольку мы знали, что идея 80 %-ного коэффициента конверсии смехотворна, то использовали данные провайдера, чтобы составить более точную оценку априорной вероятности. Но даже если бы у нас не было данных, которые могли бы помочь установить априорную информацию, то мы все равно могли бы попросить кого-то, имеющего маркетинговый опыт, помочь сделать хорошую оценку. Например, опытный маркетолог знает, что стоит ожидать, к примеру, около 20 % коэффициента конверсии.

Учитывая эту информацию от опытного профессионала, можно выбрать относительно слабую априорную вероятность, такую как Beta(2,8), чтобы предположить, что ожидаемый коэффициент конверсии должен составлять около 20 %. Это распределение является лишь предположением, но важно то, что мы можем количественно оценить это предположение. Почти для каждого бизнеса эксперты часто могут предоставить мощную априорную информацию, основанную просто на предыдущем опыте и наблюдениях, даже если у них нет специальной подготовки по определению вероятности.

Количественно оценивая этот опыт, мы можем получить более точные оценки и посмотреть, как они могут меняться от эксперта к эксперту. Например, если маркетолог уверен, что истинный коэффициент конверсии должен составлять 20 %, мы можем смоделировать это убеждение как Beta(200 800). По мере сбора данных мы можем сравнивать модели и создавать несколько доверительных интервалов, которые количественно моделируют любые экспертные убеждения. Кроме того, по мере получения все большего и большего количества информации разница из-за этих априорных убеждений будет уменьшаться.

Существует ли справедливая априорная вероятность, если ничего не известно?

В некоторых школах статистики учат, что при оценке параметров без какой-либо другой априорной вероятности к α и β всегда нужно добавлять 1.

Это соответствует использованию очень слабой априорной вероятности, которая считает, что каждый результат одинаково вероятен: Beta(1,1). Аргумент заключается в том, что это «самая справедливая» (то есть самая слабая) априорная вероятность, которую можно придумать в отсутствие информации. Справедливая априорная вероятность называется *неинформативной априорной вероятностью* Beta(1,1) (рис. 14.7).

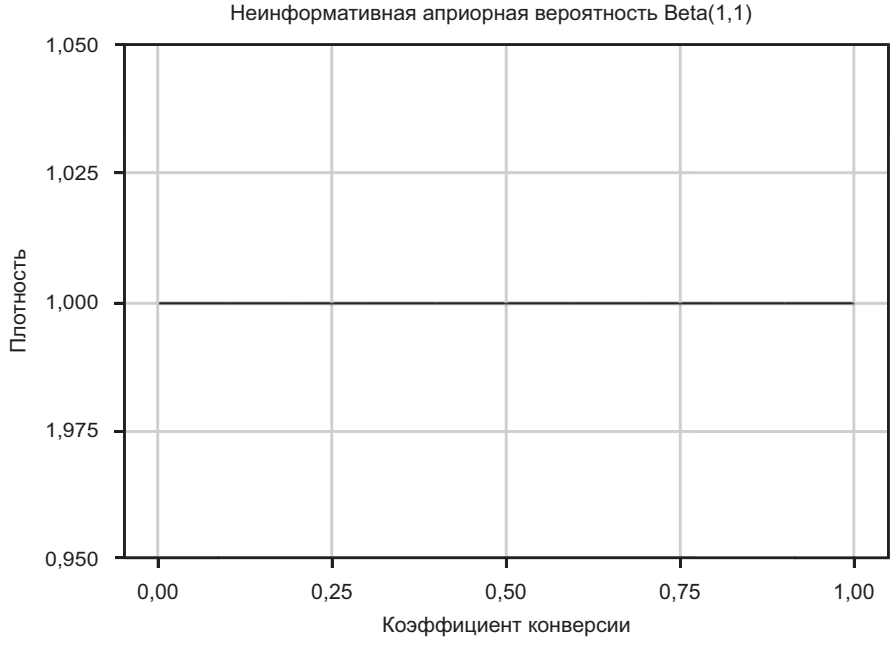


Рис. 14.7. Неинформативная априорная вероятность Beta(1,1)

Она представляет собой прямую линию, поэтому все результаты одинаково вероятны, а средняя вероятность равна 0,5. Идея использования неинформативной априорной вероятности заключается в том, что мы можем добавить априорную вероятность, чтобы сгладить оценку, но эта вероятность не смещена в сторону какого-либо конкретного результата. Хотя поначалу это может показаться наиболее справедливым способом решения, даже эта очень слабая априорная вероятность может привести к некоторым странным результатам при проверке.

Возьмем вероятность того, что солнце взойдет завтра. Скажем, вам 30 лет и вы пережили около 11 000 восходов солнца за свою жизнь. Теперь

предположим, что кто-то хочет узнать вероятность того, что солнце взойдет завтра. Вы хотите быть честным и использовать неинформативную априорную вероятность Beta(1,1). Распределение, которое представляет вашу уверенность в том, что солнце *не* взойдет завтра, будет Beta(1,11 001), что основывается на вашем опыте. Хотя это дает очень низкую вероятность того, что солнце не взойдет завтра, оно также предполагает, что мы ожидаем, что солнце *не* взойдет хотя бы один раз к тому времени, когда вам исполнится 60 лет. Так называемая «неинформативная» априорная вероятность дает довольно твердое мнение о том, как устроен мир!

Вы можете поспорить, что проблема только в том, как мы понимаем небесную механику, поскольку имеем сильную априорную информацию, которую не можем забыть. Но настоящая проблема в том, что *мы никогда не наблюдали случай, когда солнце не взошло*. Если мы вернемся к нашей функции правдоподобия без неинформативной априорной вероятности, то получим Beta(0,11 000).

Однако когда α или $\beta \leq 0$, бета-распределение *не определено*, и следовательно, ответа на вопрос, какова вероятность того, что солнце взойдет завтра, нет — вопрос не имеет смысла, потому что мы никогда не видели контрпример.

В качестве другого примера предположим, что вы нашли портал, который перенес вас и вашего друга в новый мир. Перед вами появляется пришелец и стреляет в вас из странно выглядящего пистолета, который просто не попадает. Друг спрашивает вас: «Какова вероятность того, что пистолет даст осечку?» Это совершенно чужой мир, пистолет выглядит причудливо, и вы ничего не знаете о его механике.

Теоретически это идеальный сценарий для использования неинформативной априорной вероятности, поскольку вы не имеете абсолютно никакой априорной информации об этом мире. Если вы добавите неинформативную априорную вероятность, то получите апостериорную вероятность Beta(1,2) того, что произойдет осечка (мы наблюдали $\alpha = 0$ осечек и $\beta = 1$ успешных выстрелов). Это распределение говорит, что средняя апостериорная вероятность осечки составляет 1/3, что кажется поразительно высоким уровнем, учитывая, что вы даже не знаете, может ли странное оружие дать осечку. Несмотря на то что Beta(0,1) не определена, ее применение выглядит как рациональный подход к этой проблеме. При отсутствии достаточных данных и какой-либо предварительной информации единственный честный вариант — поднять руки и сказать другу: «Не имею ни малейшего понятия, что вообще об этом сказать!»

Лучшие априорные вероятности подкреплены данными, и никогда не бывает настоящей «справедливости» при полном отсутствии данных. Каждый наблюдатель привносит в проблему свой собственный опыт и взгляд на мир. Ценность байесовских рассуждений, даже при субъективном назначении априорных вероятностей, заключается в том, что вы количественно определяете свои субъективные убеждения. Как мы увидим позже в книге, это означает, что вы можете сравнить свои априорные данные с данными других людей и увидеть, насколько хорошо эти данные объясняют мир вокруг вас. Априорная вероятность Beta(1,1) иногда используется на практике, но стоит применять ее только тогда, когда вы искренне уверены, что два возможных исхода, насколько вы знаете, одинаково вероятны. Точно так же никакое количество вычислений не может восполнить абсолютное невежество. Если у вас нет данных и предварительного понимания проблемы, единственный честный ответ — сказать, что вы ничего не можете сделать, пока не узнаете больше.

Стоит отметить, что вопрос, использовать Beta(1,1) или Beta(0,0), имеет давнюю историю, и многие великие умы обсуждают его. Томас Байес с горем пополам верил в Beta(1,1), великий математик Симон-Пьер Лаплас был совершенно уверен, что Beta(1,1) имеет право на жизнь, а известный экономист Джон Мейнард Кейнс считал, что использование Beta(1,1) настолько нелепо, что дискредитирует всю байесовскую статистику!

Заключение

Из этой главы вы узнали, как добавить априорную информацию, чтобы получить гораздо более точные оценки для неизвестных параметров. Когда информации мало, можно легко получить вероятностные оценки, которые кажутся невозможными. Но у нас может быть априорная информация, которая поможет сделать выводы из такого малого количества данных. Добавляя эту информацию к оценкам, мы получим гораздо более реалистичные результаты.

По возможности лучше использовать априорное распределение вероятностей на основе фактических данных. Но часто данных не хватает, поэтому можно либо привлечь личный опыт, либо обратиться к экспертам, у которых он есть. В этих случаях совершенно нормально оценить распределение вероятностей, соответствующее вашей интуиции. Даже если вы ошибаетесь, то будете не правы в том, что записано количественно. Самое

главное — даже если априорная вероятность неверна, она в конечном итоге будет отменена данными, когда вы соберете больше наблюдений.

Упражнения

Чтобы убедиться, что вы понимаете априорную вероятность, попробуйте ответить на эти вопросы.

1. Предположим, вы играете в аэрохоккей с друзьями и подбрасываете монетку, чтобы узнать, кто будет подавать шайбу. Проиграв 12 раз, вы понимаете, что друг, который приносит монету, почти всегда идет первым: 9 из 12 раз. Некоторые из ваших друзей начинают что-то подозревать. Определите априорное распределение вероятностей для следующих убеждений:
   - убеждения человека, который слабо верит, что друг обманывает и реальная скорость выпадения орла ближе к 70 %;
   - убеждения человека, который очень сильно верит, что монетка честная и дает 50 %-ную вероятность выпадения орла;
   - убеждения человека, который твердо верит, что монета склонна к выпадению орла в 70 % случаев.
2. Чтобы проверить монету, вы подбрасываете ее еще 20 раз и получаете 9 орлов и 11 решек. Используя априорные вероятности, которые вы рассчитали в предыдущем вопросе, определите обновленные апостериорные убеждения в истинной вероятности выпадения орла с точки зрения 95 %-ного доверительного интервала.

ЧАСТЬ IV

ПРОВЕРКА ГИПОТЕЗ: СЕРДЦЕ СТАТИСТИКИ

15

От оценки параметров к проверке гипотез: создание байесовских A/B-тестов

В этой главе мы создадим нашу первую проверку гипотезы — *A/B-тест*. Компании часто используют A/B-тесты, чтобы опробовать веб-страницы продукта, рассылки и другие маркетинговые материалы и понять, что лучше всего подойдет для клиентов. В этой главе мы проверим наше убеждение в том, что удаление картинки из мейла увеличит *коэффициент переходов* по сравнению с убеждением, что удаление картинки навредит кликабельности.

Поскольку мы уже знаем, как оценить один неизвестный параметр, все, что нужно сделать, — это оценить оба параметра, то есть коэффициенты конверсии каждого письма. Далее с помощью языка R мы запустим моделирование по методу Монте-Карло и определим, какая гипотеза, вероятно, будет работать лучше, то есть какой вариант — *A* или *B* — лучше. A/B-тесты проводятся с использованием классических статистических методов, таких как использование критерия *Стьюдента*, но построение теста байесовским способом поможет понять каждую его часть и даст более применимые результаты.

Мы уже хорошо знакомы с оценкой параметров, знаем, как использовать функции PDF, CDF и квантильную, чтобы узнать вероятность

определенных значений, и изучили, как добавить байесовскую априорную вероятность к своей оценке. Теперь используем наши оценки для сравнения двух известных параметров.

Настройка байесовского A/B-теста

Вспомните про электронную почту из прошлой главы и теперь представьте, что мы хотим узнать, увеличивает или уменьшает добавление картинки коэффициент конверсии. До этого в письме было изображение. Для теста мы отправим один вариант письма с картинкой, а другой без нее. Тест называется A/B-тестом, потому что мы сравниваем вариант *A* (с картинкой) и вариант *B* (без картинки), чтобы определить, какой из них работает лучше.

Предположим, что сейчас есть 600 подписчиков. Поскольку мы хотим использовать знания, полученные в ходе этого эксперимента, то проведем тест только на 300 из них; таким образом, мы можем отправить оставшимся 300 подписчиков письмо, которое считаем наиболее эффективным вариантом.

Триста человек, которых мы будем тестировать, будут разделены на две группы: *A* и *B*. Группа *A* получит обычное письмо с большой картинкой вверху, а группа *B* получит письмо без картинки. Гипотеза такая: более простое письмо меньше будет похоже на спам и побудит пользователей переходить по ссылке.

Нахождение априорной вероятности

Далее выясним, какую априорную вероятность нужно использовать. Кампания проводится каждую неделю, поэтому, исходя из этих данных, разумно ожидать, что вероятность перехода по ссылке на блог из любого конкретного письма должна составлять около 30 %. Для простоты мы будем использовать одну и ту же априорную вероятность для обоих вариантов. Мы также выберем довольно слабую версию нашего априорного распределения, а это означает, что в нем вероятен более широкий диапазон коэффициентов конверсии. Мы используем слабую априорную вероятность, потому что на самом деле не знаем, чего стоит ожидать от группы *B*, так как это новая электронная кампания и другие факторы могут привести к лучшей или худшей конверсии. Остановимся на Beta(3,7) для априорного распределения вероятностей. Это распределение позволяет представить бета-распределение, где 0,3 — среднее значение, но рассматривается широкий

диапазон возможных альтернативных показателей. Мы можем увидеть это распределение на рис. 15.1.

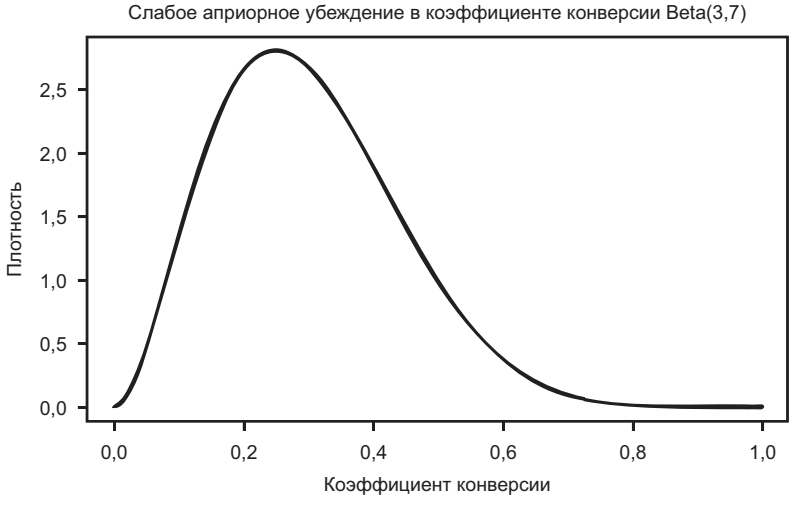


Рис. 15.1. Визуализация априорного распределения вероятностей

Все, что нам сейчас нужно, — это вероятность, а значит, нужно собрать больше данных.

Сбор данных

Мы отправляем электронные письма и получаем результаты, которые внесены в табл. 15.1.

Таблица 15.1. Показатели переходов по ссылке в письме

	Ссылка была открыта	Ссылка не была открыта	Наблюдаемый коэффициент конверсии
Вариант *A*	36	114	0,24
Вариант *B*	50	100	0,33

Каждый из этих вариантов можно рассматривать как отдельный параметр, который нужно оценить. Чтобы получить апостериорное распределение

для каждого, объединим их распределение по вероятности и априорное распределение. Мы уже решили, что априорной вероятностью для этих распределений должна быть Beta(3,7), представляющая относительно слабую веру в то, какими возможными значениями обладает коэффициент конверсии без дополнительной информации. Мы говорим, что это слабое убеждение, потому что мы не очень верим в конкретный диапазон значений и рассматриваем все возможные показатели с достаточно высокой вероятностью. Для вероятности каждого из них мы снова будем использовать бета-распределение, в котором α будет указывать количество нажатий на ссылку, а β — количество раз, когда нажатия не было. Напомним, что:

$$\text{Beta}(\alpha_{\text{апостериорное}}, \beta_{\text{апостериорное}}) =$$
$$= \text{Beta}(\alpha_{\text{априорное}} + \alpha_{\text{правдоподобности}}, \beta_{\text{априорное}} + \beta_{\text{правдоподобности}}).$$

Вариант *A* будет представлен как Beta(36 + 3,114 + 7), а вариант *B* — как Beta(50 + 3100 + 7). Рисунок 15.2 показывает оценки для каждого параметра соответственно.

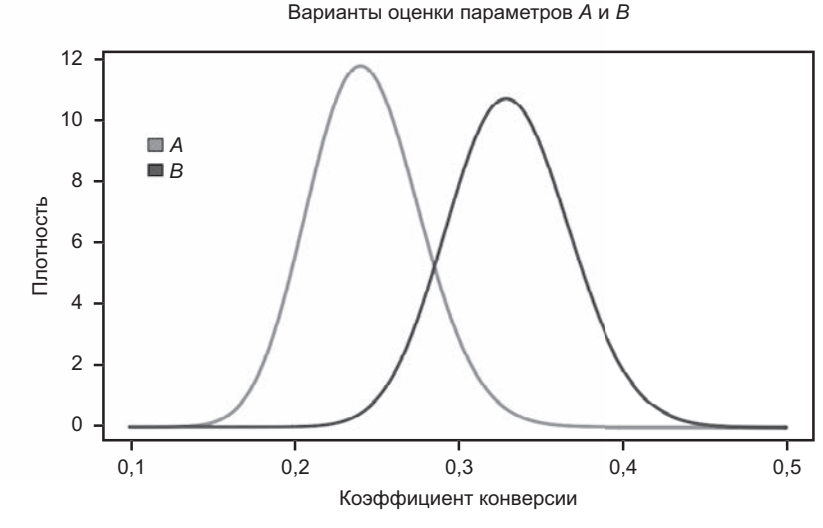


Рис. 15.2. Бета-распределения для наших оценок для обоих вариантов электронного письма

Данные показывают, что вариант *B* лучше, так как обеспечивает более высокий коэффициент конверсии. Однако из предыдущего обсуждения оценки

параметров мы знаем, что истинный коэффициент конверсии является одним значением из диапазона возможных. Здесь также видно, что есть перекрытие между возможными истинными коэффициентами конверсии для *A* и *B*. Что, если бы нам просто не повезло в наших ответах *A*, и истинный коэффициент конверсии *A* фактически намного выше? Что, если бы нам тоже просто повезло с *B*, а его коэффициент конверсии на самом деле намного ниже? Можно вообразить себе мир, где *A* на самом деле является лучшим вариантом, хотя в нашем тесте он показал себя хуже. Итак, вопрос: насколько мы можем быть уверены, что *B* — лучший вариант? Именно здесь начинается моделирование по методу Монте-Карло.

Моделирование по методу Монте-Карло

Ответ на вопрос, какой вариант письма приводит к более высокому коэффициенту переходов, находится где-то на пересечении распределений *A* и *B*. К счастью, есть способ выяснить это: моделирование по методу Монте-Карло. *Моделирование по методу Монте-Карло* — это техника, которая использует случайную выборку для решения проблемы. Мы будем случайным образом выбирать из двух распределений, где каждая выборка выбирается на основе ее вероятности в распределении, чтобы выборки в области с высокой вероятностью появлялись чаще. Например, на рис. 15.2 мы видим, что значение, *большее* чем 0,2, с большей вероятностью будет взято из *A*, чем значение, меньшее чем 0,2. Однако случайная выборка из распределения *B* почти наверняка будет выше 0,2. В нашей случайной выборке мы могли бы выбрать значение 0,2 для варианта *A* и 0,35 для варианта *B*. Каждая выборка является случайной и основана на относительной вероятности значений в распределениях *A* и *B*. Значения 0,2 для *A* и 0,35 для *B* могут быть истинным коэффициентом конверсии для каждого варианта на основе данных, которые мы наблюдали. Эта индивидуальная выборка из этих двух распределений подтверждает убеждение, что вариант *B* фактически превосходит *A*, поскольку 0,35 больше 0,2.

Мы бы могли также выбрать 0,3 для варианта *A* и 0,27 для варианта B, оба они с достаточной вероятностью будут отобраны из их соответствующих распределений. Это также реалистичные возможные значения для истинного коэффициента конверсии каждого варианта, но в данном случае они указывают, что вариант *B* на самом деле хуже, чем вариант *A*.

Основываясь на текущем состоянии убеждений в отношении каждого показателя конверсии, можно предположить, что апостериорное распределение

представляет все возможные миры. Всякий раз, когда мы получаем выборку из каждого распределения, мы видим, как может выглядеть один вероятный мир. Из рис. 15.1 можно визуально определить, что следует ожидать большего количества миров, где *B* действительно лучший вариант. Чем чаще проводится выборка, тем точнее можно сказать, в скольких мирах из всех выбранных миров *B* — лучший вариант. Получив образцы, можно посмотреть на соотношение миров, где *B* является лучшим, и общего количества наблюдаемых миров и получить точную вероятность того, что *B* на самом деле больше, чем *A*.

В скольких мирах B — лучший вариант?

Теперь напишем код, который будет выполнять эту выборку. Функция `rbeta()` в R позволяет автоматически делать выборки из бета-распределения. Можно считать каждое сравнение двух образцов одним испытанием. Чем больше испытаний мы запустим, тем более точным будет результат, поэтому начнем с 100 000 испытаний, присвоив это значение переменной `n.trials`:

```
n.trials <- 100000
```

Далее поместим наши априорные значения альфа и бета в переменные:

```
prior.alpha <- 3
prior.beta <- 7
```

Соберем образцы из каждого варианта и применим для этого `rbeta()`:

```
a.samples <- rbeta(n.trials,36+prior.alpha,114+prior.beta)
b.samples <- rbeta(n.trials,50+prior.alpha,100+prior.beta)
```

Сохраним результаты образцов `rbeta()` в переменные, чтобы было проще обращаться к ним. Для каждого варианта мы вводим количество людей, которые перешли в блог, и количество людей, которые этого не сделали.

Наконец, сравниваем, во сколько раз `b.samples` больше, чем `a.samples`, и делим это число на `n.trials`, что даст процент от общего числа испытаний, где вариант *B* был больше, чем вариант *A*:

```
p.b_superior <- sum(b.samples > a.samples)/n.trials
```

В результате мы получаем следующее:

```
p.b_superior
> 0.96
```

В 96 % из 100 000 испытаний вариант *B* был лучше. Можно представить это, анализируя 100 000 возможных миров. Исходя из распределения возможных коэффициентов конверсии для каждого варианта, в 96 % миров вариант *B* был лучшим из двух. Такой результат показывает, что даже при относительно небольшом количестве выборок мы имеем достаточно сильное убеждение, что *B* — лучший вариант. Если вы когда-либо делали проверки по критерию Стьюдента из классической статистики, это примерно эквивалентно — если мы использовали априорную вероятность Beta(1,1) — получению *p*-значения 0,04 из односторонней проверки по критерию Стьюдента (это значение часто считается «статистически значимым»). Но прелесть нашего подхода в том, что мы смогли создать этот тест с нуля, используя только знания о вероятности и простое моделирование.

Насколько каждый вариант B лучше, чем каждый вариант A?

Теперь можно точно сказать, насколько мы уверены, что *B* — лучший вариант. Но если бы это была реальная кампания, то просто сказать «*B* лучше» было бы недостаточно. Неужели вы не хотите знать, *насколько лучше*?

В этом заключается сила моделирования по методу Монте-Карло. Можно взять точные результаты последнего моделирования и проверить, насколько лучше будет вариант *B*, проанализировав, во сколько раз выборок *B* больше, чем выборок *A*. Другими словами, мы можем посмотреть на это соотношение:

$$\frac{B \text{ образцы}}{A \text{ образцы}}.$$

Если взять `a.samples` и `b.samples`, определенные ранее, то можно вычислить `b.samples/a.samples`. Это даст распределение относительных улучшений от варианта *A* к варианту *B*. Если представить это распределение в виде гистограммы (рис. 15.3), мы увидим, насколько велико ожидание, что вариант *B* окажется лучшим по числу переходов.

Из этой гистограммы мы можем видеть, что вариант *B*, скорее всего, будет лучше примерно на 40 % (в отношении 1,4) по сравнению с *A*, хотя существует

целый диапазон возможных значений. Как мы обсуждали в главе 13, кумулятивная функция распределения (CDF) гораздо более полезна, чем

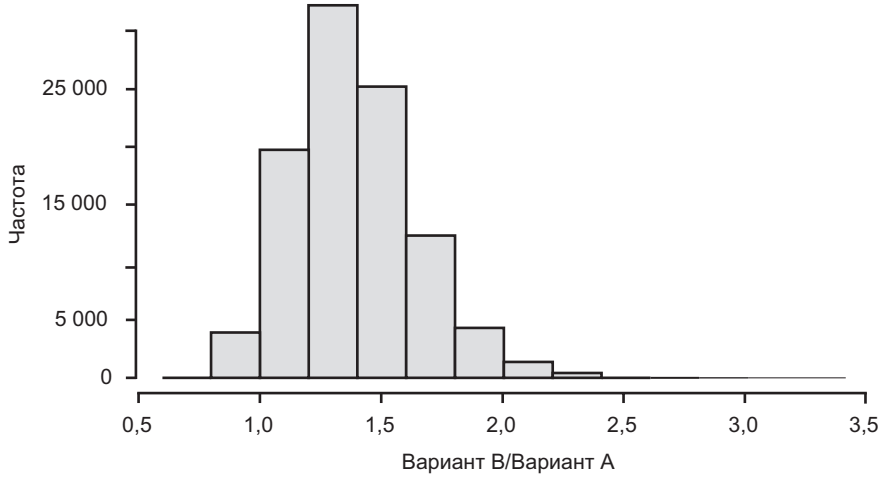


Рис. 15.3. Гистограмма возможных улучшений

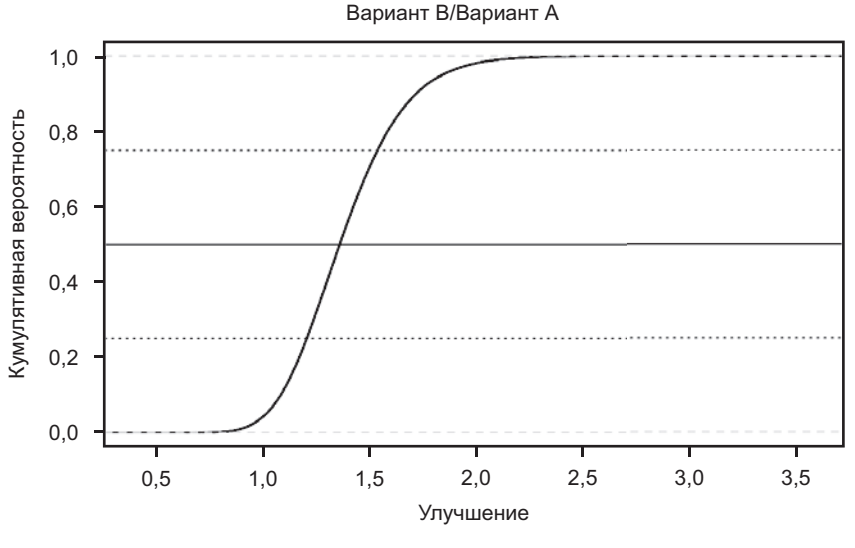


Рис. 15.4. Распределение возможных улучшений

гистограмма в контексте обсуждения результатов. Поскольку мы работаем с данными, а не с математической функцией, то вычислим *эмпирическую* кумулятивную функцию распределения с помощью функции `ecdf()`. Функция eCDF показана на рис. 15.4.

Результаты теперь видны более четко. Есть только крохотный шанс, что вариант *A* лучше, и даже если он лучше, то ненамного. Мы также видим, что вероятность того, что вариант *B* будет на 50 или более процентов лучше, чем *A*, составляет около 25 %, и существует даже разумный шанс, что коэффициент конверсии может более чем удвоиться! Теперь, выбирая *B* вместо *A*, можно говорить о своем риске: «Вероятность того, что *B* на 20 % хуже, примерно равна вероятности того, что он на 100 % лучше». Звучит как хорошая ставка и гораздо лучше, чем заявление: «Между *B* и *A* существует статистически значимая разница».

Заключение

В этой главе мы увидели, как оценка параметров естественным образом распространяется на проверку гипотез. Если гипотеза, которую нужно проверить, звучит так: «Вариант *B* имеет лучшую степень конверсии, чем вариант *A*», то можно начать с оценки параметров конверсии для каждого варианта. Как только мы узнаем эти оценки, можно будет использовать моделирование по методу Монте-Карло для выборки из них. Сравнивая эти образцы, мы можем прийти к вероятности, что гипотеза верна. Наконец, можно продвинуться еще на один шаг вперед, увидев, насколько хорошо наш новый вариант работает в этих возможных мирах, оценивая не только то, верна ли гипотеза, но и то, какое улучшение мы увидим.

Упражнения

Чтобы убедиться, что вы понимаете, что такое *A*/*B*-тесты, попробуйте ответить на эти вопросы.

1. Предположим, опытный директор по маркетингу говорит вам о своей уверенности в том, что вариант без картинок (*B*) не будет работать иначе, чем исходный вариант. Как это объяснить в нашей модели? Внедрите это изменение и посмотрите, как изменятся окончательные выводы.

2. Ведущий дизайнер видит ваши результаты и настаивает на том, что вариант *B* без картинок не будет работать лучше. Она считает, что вы должны принять коэффициент конверсии для варианта *B*, близкий к 20 %, а не к 30 %. Реализуйте решение для этого и снова просмотрите результаты анализа.
3. Предположим, что 95 %-ная уверенность означает, что вы более или менее «убеждены» в правильности гипотезы. Также предположим, что больше нет ограничений на количество писем, которые можно отправить в тесте. Если истинное преобразование для *A* составляет 0,25, а для *B* — 0,3, изучите, сколько выборок потребуется, чтобы убедить директора по маркетингу в том, что *B* на самом деле лучше. Изучите то же самое для ведущего дизайнера. Можно сгенерировать образцы конверсий с помощью следующего фрагмента R:

```
true.rate <- 0.25
number.of.samples <- 100
results <- runif(number.of.samples) <= true.rate
```

16

Введение в коэффициент Байеса и апостериорные шансы: конкуренция идей

В предыдущей главе мы увидели, что проверку гипотезы можно рассматривать как расширение оценки параметров. В этой главе подумаем о проверке гипотез как о способе сравнивать идеи, используя важный математический инструмент — *коэффициент Байеса*. Коэффициент Байеса — это формула, которая проверяет достоверность одной гипотезы, сравнивая ее с другой. В результате мы видим, во сколько раз одна гипотеза вероятнее, чем другая.

Далее мы научимся объединять коэффициент Байеса с априорными убеждениями, чтобы находить апостериорные шансы, которые указывают, насколько одно убеждение сильнее, чем другое, при объяснении данных.

Пересмотр теоремы Байеса

В главе 6 была представлена теорема Байеса, которая выглядит так:

$$P(H \mid D) = \frac{P(H) \times P(D \mid H)}{P(D)}.$$

Напомню, что существуют три части этой формулы, которые называются так:

- $P(H|D)$ — *апостериорная вероятность*, которая указывает, как сильно мы должны верить в гипотезу, учитывая данные;
- $P(H)$ — *априорное убеждение*, или вероятность гипотезы до просмотра данных;
- $P(D|H)$ — *правдоподобность* получения существующих данных в случае, если бы наша гипотеза была верной.

Последняя часть, $P(D)$, является вероятностью данных, наблюдаемых независимо от гипотезы. Эта часть нужна, чтобы убедиться, что апостериорная вероятность правильно размещена где-то между 0 и 1. Если у нас есть все эти фрагменты информации, мы можем точно рассчитать, насколько сильно следует верить в гипотезу в условиях наблюдаемых данных. Но как я говорил в главе 8, $P(D)$ очень трудно определить. Во многих случаях не очевидно, как можно выяснить вероятность наших данных. $P(D)$ также совершенно не нужна, если все, что нас волнует, — это сравнение относительной силы двух разных гипотез.

По этим причинам часто используется *пропорциональная форма* теоремы Байеса, которая позволяет анализировать силу гипотез без $P(D)$. Это выглядит так:

$$P(H \mid D) \propto P(H) \times P(D \mid H).$$

Пропорциональная форма теоремы Байеса говорит, что апостериорная вероятность нашей гипотезы пропорциональна априорной, умноженной на правдоподобность. Мы можем использовать это для сравнения двух гипотез, исследовав соотношение априорного убеждения, умноженное на вероятность для каждой гипотезы, и применив формулу *отношения апостериорных вероятностей*:

$$\frac{P(H_1) \times P(D \mid H_1)}{P(H_2) \times P(D \mid H_2)}.$$

Теперь есть отношение того, насколько хорошо каждая из гипотез объясняет полученные данные. Если отношение равно 2, то H_1 объясняет наблюдаемые данные дважды, так же как и H_2, а если отношение равно $1/2$, то H_2 объясняет данные дважды, так же как и H_1.

Создание проверки гипотезы с использованием отношения постериоров

Формула отношения постериоров дает *апостериорные шансы*, которые позволяют проверять гипотезы или представления об имеющихся данных. Даже когда мы знаем $P(D)$, апостериорные шансы — полезный инструмент, потому что позволяет сравнивать идеи. Чтобы лучше понять апостериорные шансы, мы разделим формулу отношения постериоров на две части: коэффициент правдоподобности, или коэффициент Байеса, и коэффициент априорных вероятностей. Это стандартная и очень полезная практика, которая значительно упрощает анализ правдоподобности и априорной вероятности в отдельности.

Коэффициент Байеса

Используя формулу отношения постериоров, давайте предположим, что $P(H_1) = P(H_2)$, то есть априорное убеждение в каждой гипотезе одинаково. В этом случае отношение априорных убеждений в гипотезах составляет всего 1, поэтому остается только:

$$\frac{P(D|H_1)}{P(D|H_2)}.$$

Это и есть коэффициент Байеса, отношение вероятностей двух гипотез.

Найдите минутку и подумайте о том, что говорит это уравнение. Когда мы собираемся спорить о нашем H_1, то есть о нашей вере в мир, то думаем о сборе доказательств, подтверждающих наши убеждения. Поэтому типичный аргумент включает в себя создание набора данных D_1, поддерживающего H_1, и затем уже спор с другом, который собрал набор данных D_2, поддерживающий его гипотезу, H_2.

Но в байесовских рассуждениях мы не собираем доказательства в поддержку наших идей, а смотрим, насколько хорошо наши идеи объясняют полученные доказательства. Это соотношение говорит о вероятности того, что мы видим, учитывая то, что принимаем за правду, по сравнению с убеждениями, которые *кто-то еще* считает правдой. Наша гипотеза побеждает, если объясняет мир лучше, чем гипотеза оппонента.

Но если гипотеза оппонента объясняет данные гораздо лучше, чем наша, возможно, пришло время сменить убеждения. Ключевым моментом здесь

является то, что в байесовских рассуждениях мы не беспокоимся о поддержке наших убеждений — мы сосредоточены на том, насколько хорошо убеждения поддерживают наблюдаемые данные. В конце концов, данные могут либо подтвердить наши идеи, либо заставить передумать.

Априорные шансы

До сих пор мы предполагали, что априорная вероятность каждой гипотезы одинакова. Это не всегда так: гипотеза может хорошо объяснять данные, даже если она маловероятна. Например, если вы потеряли телефон, то можете предположить, что либо оставили его в ванной, либо инопланетяне забрали его для изучения человеческих технологий, что достаточно хорошо объясняет данные. Тем не менее гипотеза с ванной явно более вероятна. Вот почему следует рассмотреть отношение априорных вероятностей:

$$\frac{P(H_1)}{P(H_2)}.$$

Это соотношение сравнивает вероятность двух гипотез до рассмотрения данных. При использовании по отношению к байесовскому коэффициенту это соотношение называется априорным шансом в нашем H_1 и записывается как $O(H_1)$. Это представление полезно, потому что позволяет заметить, насколько сильно (или слабо) мы верим в гипотезу, которую проверяем. Когда это число больше 1, это означает, что априорные шансы подтверждают гипотезу, а когда оно меньше 1, это означает, что они противоречат гипотезе. Например, $O(H_1) = 100$ означает, что без какой-либо другой информации мы считаем, что H_1 в 100 раз более вероятна, чем альтернативная гипотеза. С другой стороны, когда $O(H_1) = 1/100$, альтернативная гипотеза в 100 раз более вероятна, чем наша.

Апостериорные шансы

Если собрать коэффициент Байеса и предыдущие шансы, то получаются апостериорные шансы:

$$\text{апостериорные шансы} = O(H_1)\frac{P(D|H_1)}{P(D|H_2)}.$$

Апостериорные шансы вычисляют, во сколько раз наша гипотеза лучше объясняет данные, чем гипотеза противника.

В табл. 16.1 приведены рекомендации по оценке различных значений апостериорных шансов.

Таблица 16.1. Рекомендации по оценке апостериорных шансов

Апостериорные шансы	Сила доказательств
1 к 3	Интересно, но ничего неопровержимого
3 к 20	Похоже, мы к чему-то движемся
20 к 150	Сильные доказательства в пользу H_1
> 150	Неопровержимые доказательства

По соотношению этих шансов можно понять, стоит ли поменять мнение.

Хотя эти значения могут служить полезным руководством, байесовские рассуждения все еще являются формой рассуждений, это означает, что нужно использовать некоторые суждения. Если вы не согласны с другом, апостериорных шансов со значением 2 может быть достаточно, чтобы почувствовать себя уверенно. Если вы пытаетесь выяснить, пьете ли вы яд, апостериорная вероятность 100 все равно не поможет.

Далее рассмотрим два примера, в которых используется коэффициент Байеса для определения силы убеждений.

Проверка утяжеленной игральной кости

Коэффициент Байеса и апостериорные шансы можно использовать как форму проверки гипотезы, в которой каждый тест является соревнованием двух идеей. Предположим, у вашего друга в сумке лежат три шестигранных кубика. Один кубик утяжеленный — в половине случаев при подбрасывании выпадает шестерка. Два других кубика — традиционные игральные кости, где вероятность выпадения шестерки равна 1/6. Друг достает наугад кубик и бросает 10 раз со следующими результатами:

$$6, 1, 3, 6, 4, 5, 6, 1, 2, 6.$$

Нужно выяснить, является ли кубик утяжеленным. Утяжеленный кубик назовем H_1, а обычный — H_2.

Начнем с определения коэффициента Байеса:

$$\frac{P(D|H_1)}{P(D|H_2)}.$$

Первый шаг — вычисление $P(D|H)$, или правдоподобности H_1 и H_2, учитывая наблюдаемые данные. В этом примере у друга выпало четыре шестерки и шесть не шестерок. Мы знаем, что если кубик утяжеленный, вероятность выпадения шестерки равна 1/2, а вероятность выпадения любой цифры, кроме шестерки, также равна 1/2. Это означает, что вероятность увидеть эти данные при использовании утяжеленного кубика равна:

$$P(D|H_1)=\left(\frac{1}{2}\right)^4\times\left(\frac{1}{2}\right)^6=0{,}00098.$$

В случае честного кубика вероятность выпадения шестерки равна 1/6, тогда как вероятность выпадения чего-либо еще — 5/6. Таким образом, правдоподобность появления этих данные для H_2, при гипотезе о том, что кубик честный, такова:

$$P(D|H_2)=\left(\frac{1}{6}\right)^4\times\left(\frac{5}{6}\right)^6=0{,}00026.$$

Теперь вычислим коэффициент Байеса, который скажет нам, насколько H_1 лучше, чем H_2, если предположить, что каждая гипотеза была в одинаковой степени вероятна (это означает, что предыдущее отношение шансов равно 1):

$$\frac{P(D|H_1)}{P(D|H_2)}=\frac{0{,}00098}{0{,}00026}=3{,}77.$$

Это означает, что H_1 (кубик нечестный) объясняет наблюдаемые данные почти в четыре раза лучше, чем H_2.

Но это верно только в том случае, если H_1 и H_2 одинаково вероятны. Мы знаем, что у друга есть два честных кубика и только один утяжеленный, это означает, что обе гипотезы не одинаково вероятны. Основываясь на распределении игральных костей в сумке, мы знаем, что априорные вероятности для каждой гипотезы таковы:

$$P(H_1)=\frac{1}{3};\ P(H_2)=\frac{2}{3}.$$

Исходя из этого, рассчитаем априорные шансы для H_1:

$$\text{априорные шансы} = O(H_1) = \frac{P(H_1)}{P(H_2)} = \frac{\frac{1}{3}}{\frac{2}{3}} = \frac{1}{2}.$$

Поскольку в сумке есть только один утяжеленный кубик и два честных, то шансов вытащить честный кубик вдвое больше. С априорными шансами для H_1 вычислим полные апостериорные шансы:

$$\text{апостериорные шансы} = O(H_1) = \frac{P(D \mid H_1)}{P(D \mid H_2)} = \frac{1}{2} \times 3{,}77 = 1{,}89.$$

Хотя начальное отношение правдоподобия показало, что H_1 объясняет данные почти в четыре раза лучше, чем H_2, апостериорные шансы показывают, что, поскольку вероятность H_1 в два раза меньше вероятности H_2, объяснение H_1 только вдвое сильнее, чем H_2.

Если вам очень нужно сделать вывод о том, утяжелен ли кубик или нет, лучше всего сказать, что он действительно утяжелен. Но апостериорные шансы менее 2 — не особенно убедительные доказательства в пользу H_1. Если вы действительно хотите узнать, был ли утяжелен кубик, нужно будет бросить его еще несколько раз, пока доказательства в пользу одной или другой гипотезы не станут достаточно велики, чтобы можно было принять более верное решение.

Рассмотрим второй пример использования коэффициента Байеса для определения силы наших убеждений.

Самодиагностика по интернету

Многие попадались в эту ловушку: гуглили свои симптомы поздно ночью, а затем в ужасе утыкались в экран с мыслью, что стали жертвой ужасной неизлечимой болезни. К сожалению, редко кто подключает байесовские рассуждения, чтобы избавиться от ненужной тревоги. Давайте предположим, что вы допустили ошибку при поиске симптомов и нашли два возможных заболевания, которые им соответствуют. Вы не поддадитесь панике, а используете апостериорные шансы, чтобы оценить вероятность каждого заболевания.

Предположим, вы проснулись и обнаружили, что у вас звенит в ушах и плохо со слухом. Весь день вас это беспокоит, и вечером вы решаете,

что надо поискать в интернете потенциальные причины таких симптомов. Беспокойство нарастает, и вы приходите к двум возможным гипотезам:

Ушная сера. В одном ухе слишком много ушной серы. Визит к врачу облегчит это состояние.

Вестибулярная шваннома. Это опухоль головного мозга, растущая на миелиновой оболочке вестибулярного нерва, вызывающая необратимую потерю слуха и, возможно, требующая операции на головном мозге.

Из двух вариантов наличие вестибулярной шванномы является наиболее тревожным. Конечно, может, это и просто ушная сера, но что, если нет? Что, если у вас *опухоль* мозга? Так как возможность опухоли головного мозга беспокоит больше всего, то эта гипотеза будет H_1. Гипотеза H_2 — у вас слишком много ушной серы в ухе.

Посмотрим, могут ли апостериорные шансы успокоить вас.

Как и в последнем примере, мы начнем с рассмотрения вероятности наблюдения этих симптомов, если каждая гипотеза верна, и вычислим коэффициент Байеса. Нужно вычислить $P(D \mid H)$. Вы наблюдали два симптома: потеря слуха и шум в ушах.

Для вестибулярной шванномы вероятность потери слуха составляет 94 %, а вероятность возникновения шума в ушах (тиннитус) — 83 %. Это означает, что вероятность потери слуха и шума в ушах при вестибулярной шванноме составляет:

$$P(D \mid H_1) = 0{,}94 \times 0{,}89 = 0{,}78.$$

Сделаем то же самое для H_2. В случае скопления ушной серы вероятность потери слуха составляет 63 %, а вероятность шума в ушах — 55 %. Правдоподобность появления симптомов при воздействии ушной серы:

$$P(D \mid H_2) = 0{,}63 \times 0{,}55 = 0{,}35.$$

Теперь имеется достаточно информации, чтобы взглянуть на коэффициент Байеса:

$$\frac{P(D \mid H_1)}{P(D \mid H_2)} = \frac{0{,}78}{0{,}35} = 2{,}23.$$

Вот дела! Только один коэффициент Байеса мало помогает в решении проблемы. Принимая во внимание только отношение правдоподобия, кажется,

что шансов на появление вестибулярной шванномы в два раза больше, чем на скопление ушной серы! К счастью, мы еще не закончили анализ.

Следующим шагом является определение априорных шансов каждой гипотезы. Если не считать симптомов, насколько вероятно, что кто-то столкнется с одной проблемой, а не с другой? Найдем эпидемиологические данные для каждого из этих заболеваний. Оказывается, вестибулярная шваннома является редким заболеванием. Только 11 людям из 1 000 000 в год ставят подобный диагноз. Априорные шансы выглядят так:

$$P(H_1) = \frac{11}{1\ 000\ 000}.$$

Неудивительно, что воздействие ушной серы встречается гораздо чаще, с 37 000 случаев на 1 000 000 человек в год:

$$P(H_2) = \frac{37\ 000}{1\ 000\ 000}.$$

Чтобы получить априорные шансы для H_1, нужно посмотреть на соотношение этих двух априорных вероятностей:

$$O(H_1) = \frac{P(H_1)}{P(H_2)} = \frac{\frac{11}{1\ 000\ 000}}{\frac{37\ 000}{1\ 000\ 000}} = \frac{11}{37\ 000}.$$

Основываясь только на априорной информации, у конкретного человека вероятность возникновения серной пробки в 3700 раз выше вероятности возникновения вестибулярной шванномы. Но прежде чем окончательно успокоиться, вычислим все шансы на победу. Умножим коэффициент Байеса на априорные шансы:

$$O(H_1) \times \frac{P(D|H_1)}{P(D|H_2)} = \frac{11}{37\ 000} \times 2{,}23 = \frac{223}{370\ 000}.$$

Этот результат показывает, что гипотеза H_2 примерно в 1659 раз более вероятна, чем H_1. Ну вот, теперь можно расслабиться — утренний визит к врачу для чистки ушей, скорее всего, избавит вас от симптомов.

В повседневных рассуждениях легко переоценить вероятность страшных ситуаций, но используя байесовские рассуждения, можно разделить реальные риски и посмотреть, насколько они вероятны на самом деле.

Заключение

В этой главе вы узнали, как использовать коэффициент Байеса и апостериорные шансы для сравнения двух гипотез. Коэффициент Байеса не фокусируется на предоставлении данных в поддержку наших убеждений, а проверяет, насколько хорошо наши убеждения поддерживают наблюдаемые данные. В результате получается соотношение, которое отражает, во сколько раз одна гипотеза объясняет данные лучше, чем другая. Мы можем использовать его для укрепления своих априорных убеждений, если они объясняют данные лучше, чем альтернативные убеждения. С другой стороны, когда результат незначителен, можно подумать о смене мнения.

Упражнения

Чтобы убедиться, что вы понимаете коэффициент Байеса и апостериорные шансы, попробуйте ответить на эти вопросы.

1. Возвращаясь к задаче с игральными костями, предположим, что ваш друг допустил ошибку и внезапно осознал, что на самом деле было две нечестные кости и только одна честная. Как это изменит априорный и, следовательно, апостериорный шансы этой задачи? Вы более склонны верить, что бросаемая кость нечестная?
2. Вернемся к примеру с редкими заболеваниями. Предположим, вы обратились к врачу и после чистки ушей заметили, что симптомы не исчезли. Еще хуже, появился новый симптом: головокружение. Врач предлагает другое возможное объяснение, лабиринтит — вирусную инфекцию внутреннего уха, при которой в 98 % случаев возникает головокружение. Однако потеря слуха и шум в ушах менее распространены при этом заболевании; потеря слуха происходит только в 30 % случаев, а шум в ушах — только в 28 %. Головокружение также является возможным симптомом вестибулярной шванномы, но встречается только в 49 % случаев. В общей численности населения 35 человек на миллион заболевают лабиринтитом ежегодно. Каковы апостериорные шансы гипотезы, что у вас лабиринтит, по сравнению с гипотезой о вестибулярной шванноме?

17

Байесовские рассуждения в «Сумеречной зоне»

В главе 16 мы применили коэффициент Байеса и апостериорные шансы, чтобы выяснить, во сколько раз одна гипотеза лучше другой. Но эти инструменты байесовского рассуждения могут сделать даже больше, чем просто сравнить идеи. В этой главе мы будем использовать коэффициент Байеса и апостериорные шансы для количественной оценки того, сколько доказательств понадобится, чтобы убедить кого-то в гипотезе. Мы увидим, как оценить силу чьих-то убеждений в определенную гипотезу, и все это на примере эпизода «Сумеречной зоны»[1].

Байесовские рассуждения в «Сумеречной зоне»

Один из моих любимых эпизодов «Сумеречной зоны» называется «Вовремя». В этом эпизоде молодожены Дон и Пэт сидят в закусочной маленького городка и ждут, пока отремонтируют их машину. На столике в кафе они видят машину-гадалку «Мистический предсказатель», которая принимает

[1] «Сумеречная зона» — американский телевизионный сериал-антология, просуществовал пять сезонов на канале CBS с 1959 по 1964 год. Каждый эпизод представляет собой отдельную историю, в которой персонажи сталкиваются часто с тревожными или необычными событиями, явлениями или переживаниями, которые являются опытом вхождения в «Сумеречную зону». — *Примеч. ред.*

вопросы «да» или «нет» и за монетки выкладывает карточки с ответами на каждый вопрос.

Суеверный Дон задает Мистическому предсказателю ряд вопросов. Когда машина отвечает правильно, он начинает верить в ее сверхъестественные способности. Тем не менее Пэт скептически относится к возможностям машины, хотя Предсказатель продолжает давать правильные ответы.

Хотя Дон и Пэт наблюдают одни и те же данные, они приходят к разным выводам. Как объяснить, почему они по-разному рассуждают, когда видят одно и то же? Используем коэффициент Байеса, чтобы лучше понять, как оба персонажа думают о данных.

Коэффициента Байеса и Мистический предсказатель

В этом эпизоде мы столкнулись с двумя конкурирующими гипотезами. Давайте назовем их H и $\overline{H}$ (или «не H»), поскольку одна гипотеза является отрицанием другой:

H — Мистический предсказатель действительно может предсказать будущее.

$\overline{H}$ — Мистическому предсказателю просто повезло.

Наши данные, в этом случае D, — последовательность из n правильных ответов, которые дает Мистический предсказатель. Чем больше n, тем сильнее доказательства в пользу H. Основное предположение заключается в том, что Мистический предсказатель всегда прав, поэтому возникает вопрос: является ли этот результат сверхъестественным или это просто совпадение? Наши данные D всегда представляют последовательность из n правильных ответов. Оценим правдоподобность или вероятность получения наших данных с учетом каждой гипотезы.

$P(D \mid H)$ — это вероятность получить n правильных ответов подряд, учитывая, что Мистический предсказатель может предсказать будущее. Эта вероятность всегда будет равна 1, независимо от количества задаваемых вопросов. Если Мистический предсказатель имеет сверхъестественные свойства, то всегда найдет правильный ответ, независимо от того, задан один вопрос или тысяча. Это также означает, что если Мистический предсказатель выдаст один неверный ответ, вероятность этой гипотезы упадет до 0, так как экстрасенсорная машина никогда не будет угадывать неправильно.

В этом случае может потребоваться выдвинуть более слабую гипотезу, например, что Мистический предсказатель прав в 90 % случаев (аналогичную задачу рассмотрим в главе 19).

$P(D|\overline{H})$ — это вероятность получения n правильных ответов подряд, если Мистический предсказатель выдает ответы случайно. Здесь $P(D|\overline{H})$ составляет $0{,}5^n$. Другими словами, если машина просто угадывает, то каждый ответ с вероятностью 0,5 может быть правильным.

Чтобы сравнить эти гипотезы, посмотрим на отношение двух вероятностей:

$$\frac{P(D|H)}{P(D|\overline{H})}.$$

Напомню, что отношение измеряет, во сколько раз вероятнее данные с учетом H, в отличие от $\overline{H}$, если предполагается, что обе гипотезы одинаково вероятны. Теперь посмотрим, как сравнить их.

Измерение коэффициента Байеса

Как и в предыдущей главе, временно проигнорируем соотношение априорных шансов и сконцентрируемся на сравнении отношения правдоподобия, или коэффициента Байеса. Мы предполагаем (на данный момент), что Мистический предсказатель может как обладать сверхъестественными способностями, так и просто угадывать ответы.

В этом примере числитель $P(D|H)$ всегда равен 1, поэтому для любого значения n мы имеем:

$$BF = \frac{P(D_n|H)}{P(D_n|\overline{H})} = \frac{1}{0{,}5^n}.$$

Представим, что Мистический предсказатель дал три правильных ответа: $P(D_3|H) = 1$ и $P(D|H) = 0{,}53 = 0{,}125$. Очевидно, что H объясняет данные лучше, но никого, даже суеверного Дона, не убедить только тремя правильными догадками. Предполагая, что априорные шансы одинаковы, вычислим коэффициент Байеса для трех вопросов:

$$BF = \frac{1}{0{,}125} = 8.$$

Мы можем использовать те же принципы, которые использовали для оценки апостериорных шансов в табл. 16.1, чтобы оценить здесь коэффициенты

Байеса (если мы предположим, что каждая гипотеза одинаково вероятна). Из табл. 17.1 видно, что коэффициент Байеса (КБ), равный 8, далеко не окончательный.

Таблица 17.1. Руководство по оценке коэффициентов Байеса

КБ	Сила доказательств
1 к 3	Интересно, но ничего неопровержимого
3 к 20	Похоже, мы к чему-то движемся
20 к 150	Сильные доказательства в пользу H_1
> 150	Подавляющие доказательства в пользу H_1

Итак, принимая во внимание три вопроса, на которые дан правильный ответ и КБ = 8, мы должны как минимум заинтересоваться силой Мистического предсказателя, хотя пока не должны быть полностью уверены.

К этому моменту Дон, похоже, уже уверен, что Мистический предсказатель — экстрасенс. Ему достаточно только четырех правильных ответов, чтобы убедиться в этом. С другой стороны, Пэт требуется 14 вопросов, чтобы *только начать всерьез рассматривать* эту возможность, в результате чего коэффициент Байеса составляет 16 384 — гораздо больше доказательств, чем ей нужно.

Однако вычисление коэффициента Байеса не объясняет, почему Дон и Пэт формируют разные убеждения относительно доказательств. Что же происходит?

Учитываем априорные убеждения

Отсутствующий элемент — это априорное убеждение каждого персонажа в гипотезах. Помните, что Дон чрезвычайно суеверен, а Пэт скептик. Очевидно, что Дон и Пэт используют дополнительную информацию в своих ментальных моделях, потому что каждый из них приходит к выводу о разной силе и в разное время. Такое довольно часто встречается в повседневных рассуждениях: два человека по-разному реагируют на одни и те же факты.

Смоделируем это явление, просто представив начальные шансы $P(H)$ и $P(\overline{H})$ без дополнительной информации. Назовем его *отношением априорных шансов*, как было показано в главе 16:

$$\text{априорные шансы} = O(H) = \frac{P(H)}{P(\overline{H})}.$$

Концепция априорных убеждений в отношении коэффициента Байеса на самом деле интуитивна. Скажем, мы идем в закусочную из «Сумеречной зоны», и я спрашиваю вас: «Каковы шансы, что Мистический предсказатель — экстрасенс?» Вы можете ответить: «Ну, один на миллион! Не может быть, чтобы эта штука была сверхъестественной». Математически мы можем выразить это следующим образом:

$$O(H) = \frac{1}{1\ 000\ 000}.$$

Теперь объединим это априорное убеждение с нашими данными. Для этого умножим априорные шансы на результаты отношения правдоподобия, чтобы получить апостериорные шансы для гипотезы, учитывая наблюдаемые данные:

$$\text{апостериорные шансы} = O(H \mid D) = O(H) \times \frac{P(D \mid H)}{P(D \mid \overline{H})}.$$

Предполагая, что у Мистического предсказателя есть только один шанс на миллион иметь экстрасенсорные способности без учета любых доказательств — довольно сильный скептицизм. Байесовский подход достаточно хорошо его отражает. Если вы думаете, что гипотеза о том, что Мистический предсказатель сверхъестествен, крайне маловероятна, то потребуется значительно больше данных, чтобы убедиться в обратном. Предположим, Мистический предсказатель выдает пять правильных ответов. Тогда коэффициент Байеса становится следующим:

$$\text{КБ} = \frac{1}{0{,}5^5} = 32.$$

Коэффициент Байеса, равный 32, — это достаточно сильное убеждение, что Мистический предсказатель действительно сверхъестествен. Но если добавить весьма скептические априорные шансы для расчета апостериорных шансов, то мы получим следующие результаты:

$$\text{апостериорные шансы} = O(H \mid D_5) \times \frac{P(D_5 \mid H)}{P(D_5 \mid \overline{H})} = \frac{1}{1\ 000\ 000} \times \frac{1}{0{,}5^5} = 0{,}000032.$$

Теперь апостериорные шансы указывают, что экстрасенсорность машины крайне маловероятна. Этот результат вполне соответствует интуитивному представлению. Опять же, если вы действительно не верите в гипотезу с самого начала, потребуется много доказательств, чтобы убедить вас в обратном.

При работе в обратном направлении апостериорные шансы могут помочь выяснить, сколько доказательств понадобится, чтобы заставить вас поверить в H. При апостериорных шансах, равных 2, вы начинаете просто рассматривать сверхъестественную гипотезу. Таким образом, если провести расчет для шансов, превышающих 2, то можно определить, что потребуется, чтобы убедить вас.

$$\frac{1}{1\ 000\ 000} \times \frac{1}{0{,}5^n} > 2\,.$$

Если мы проведем расчет для n до ближайшего целого числа, получим:

$$n > 21.$$

При 21-м правильном ответе подряд даже сильный скептик должен задуматься о том, что Мистический предсказатель на самом деле может быть экстрасенсом.

Таким образом, априорные шансы могут сделать гораздо больше, чем просто сказать, как сильно мы верим чему-то, учитывая наш опыт. Они также помогут точно определить, сколько доказательств нужно, чтобы убедиться в гипотезе. Верно и обратное: если после 21 правильного ответа подряд вы сильно верите в H, то, возможно, захотите ослабить априорные шансы.

Развитие собственных экстрасенсорных способностей

Мы научились сравнивать гипотезы и рассчитывать, сколько положительных доказательств потребуется, чтобы убедить нас в H, учитывая наше априорное убеждение в H. Теперь рассмотрим еще один прием: количественную оценку априорных убеждений Дона и Пэт на основе их реакции на доказательства.

Мы не знаем точно, насколько сильно Дон и Пэт верят в возможность того, что Мистический предсказатель — экстрасенс, когда они впервые заходят в закусочную. Но мы *знаем*, что Дону нужно около семи правильных ответов, чтобы убедиться в сверхъестественных способностях Мистического предсказателя. По оценкам, на данный момент апостериорные шансы Дона составляют 150 — порог для *очень сильных* убеждений, согласно табл. 17.1. Теперь выпишем все, что знаем, за исключением $O(H)$, который нужно будет вычислить:

$$150 = O(H) \times \frac{P(D_7 \mid H)}{P(D_7 \mid \overline{H})} = O(H) \times \frac{1}{0{,}5^7}\,.$$

Решение уравнения для $O(H)$ дает результат:

$$O(H)_{\text{Дон}} = 1{,}17.$$

Теперь у нас есть количественная модель для верований Дона. Поскольку его начальное соотношение шансов больше 1, Дон входит в забегаловку с чуть большей готовностью полагать, что Мистический предсказатель сверхъестествен, еще до сбора каких-либо данных. Это имеет смысл, если принять во внимание его суеверный характер.

Теперь Пэт. При 14 правильных ответах Пэт нервничает, называя Мистический предсказатель глупым куском мусора. Хотя она начала подозревать, что Мистический предсказатель может быть экстрасенсом, она не так уверена, как Дон. Я бы оценил, что ее апостериорные шансы равны 5 — с этого момента она может начать думать: «Может быть, Мистический предсказатель *мог бы* обладать экстрасенсорными способностями...» Рассчитаем последующие шансы для убеждений Пэт таким же образом:

$$5 = O(H) \times \frac{P(D_{14} \mid H)}{P(D_{14} \mid \overline{H})} = O(H) \times \frac{1}{0{,}5^{14}}.$$

При решении уравнения для $O(H)$ смоделируем скептицизм Пэт как:

$$O(H)_{\text{Пэт}} = 0{,}0003.$$

Другими словами, Пэт, входя в закусочную, сказала бы, что у Мистического предсказателя есть 1 шанс на 3000 быть сверхъестественным. Это тоже интуитивно; Пэт начинает с очень сильного убеждения, что машина-гадалка — не более чем забавная игрушка, которой они с Доном могут занять себя.

То, что мы сделали здесь, замечательно. Мы использовали наши правила вероятности, чтобы составить количественное утверждение о том, кто во что верит. По сути, мы стали телепатами!

Заключение

В этой главе мы изучили три способа использования коэффициентов Байеса и апостериорных шансов для оценки вероятностных рассуждений. Мы узнали в предыдущей главе, что можно использовать апостериорные шансы для сравнения двух идей. Затем увидели, что если мы знаем нашу априорную веру в шансы одной гипотезы против другой, то можем точно

рассчитать, сколько доказательств потребуется, чтобы убедить нас в том, что стоит изменить убеждения. Наконец, мы использовали апостериорные шансы, чтобы присвоить ценность предыдущим убеждениям каждого человека, посмотрев, сколько нужно доказательств, чтобы убедить его. В конце концов, апостериорные шансы — гораздо больше чем просто способ проверить идеи. Они служат основой в рассуждениях в условиях неопределенности.

Теперь вы можете использовать свои собственные «мистические» способности байесовских рассуждений, чтобы выполнить приведенные ниже упражнения.

Упражнения

Чтобы убедиться, что вы понимаете количественную оценку числа доказательств, которые необходимо предоставить, чтобы убедить кого-либо в гипотезе и оценить силу чужого априорного убеждения, попробуйте ответить на эти вопросы.

1. Каждый раз, когда вы и ваш друг встречаетесь, чтобы посмотреть фильм, вы подбрасываете монетку, чтобы определить, кто выберет фильм. Друг всегда выбирает орла, и каждую пятницу в течение 10 недель выпадает орел. Вы выдвигаете гипотезу, что у монетки два орла, а не орел и решка. Вычислите коэффициент Байеса для гипотезы о том, что монетка с подвохом, в отношении к гипотезе о том, что монетка честная. Что одно только это соотношение говорит о том, обманывает ли ваш друг или нет.
2. Теперь представьте три случая: ваш друг немного шутник, ваш друг большую часть времени честен, но иногда может схитрить, и ваш друг очень надежный. В каждом случае оцените некоторые априорные коэффициенты шансов для вашей гипотезы и вычислите апостериорные шансы.
3. Предположим, вы очень доверяете другу. Задайте априорные шансы обмана равными 1/10 000. Сколько раз должен выпасть орел, прежде чем вы начнете сомневаться в невиновности друга — скажем, с апостериорными шансами 1?
4. Другой ваш друг также общается с вышеописанным другом, и после лишь четырех недель выпадения орла он твердо решил, что вас обманывают. Такая уверенность подразумевает апостериорные шансы около 100. Какую ценность вы бы присвоили априорному убеждению этого друга, что первый друг — мошенник?

18

Когда данные не убеждают

В предыдущей главе мы использовали байесовские рассуждения, чтобы обосновать две гипотезы из эпизода «Сумеречной зоны»:

H — Мистический предсказатель действительно может предсказать будущее.

$\overline{H}$ — Мистическому предсказателю просто повезло.

Мы также узнали, как учесть скептицизм, изменив соотношение априорных шансов. Например, если вы, как и я, считаете, что Мистический предсказатель определенно не экстрасенс, тогда установите априорные шансы крайне низкими — например, 1/1 000 000.

Но в зависимости от своего уровня скептицизма вы можете почувствовать, что даже соотношения шансов 1/1 000 000 будет недостаточно, чтобы убедить вас в силе предсказателя.

Может быть, даже получив 1000 правильных ответов от предсказателя, который, несмотря на ваши очень скептические предыдущие убеждения, подсказал бы вам, что вы астрономически настроены на то, чтобы поверить, что он экстрасенс, вы все равно не купитесь. Конечно, можно сделать наши априорные шансы еще более экстремальными, но лично я не считаю это решение удовлетворительным, потому что никакие данные не убедили бы меня в том, что Мистический предсказатель на самом деле экстрасенс.

В этой главе мы более подробно рассмотрим ситуации, когда данные не убеждают людей так, как мы ожидаем. В реальном мире такое довольно распространено. Любой, кто спорил за праздничным ужином с родственником, должно быть, заметил, что чем чаще приводить доказательства обратного, тем больше люди начинают настаивать на своей правоте! Чтобы понять байесовские рассуждения, нужно понимать, почему возникают подобные ситуации, с математической точки зрения. Это поможет определить и избежать проблем в статистическом анализе.

Друг-экстрасенс бросает кости

Предположим, ваш друг говорит, что он может предсказать исход броска шестигранного кубика с точностью до 90 %, потому что он экстрасенс. В это сложно поверить, и вы решили проверить гипотезу, используя коэффициент Байеса. Как и в примере с Мистическим предсказателем, у вас есть две гипотезы, которые нужно сравнить:

$$H_1 : P(\text{верная}) = \frac{1}{6} \quad H_2 : P(\text{верная}) = \frac{9}{10}.$$

Первая гипотеза, H_1, отражает веру в то, что кость честная, а ваш друг не экстрасенс. Если кость честная, шанс угадать результат равен 1 к 6. Вторая гипотеза, H_2, представляет убеждение вашего друга в том, что он на самом деле может предсказать результат броска кости в 90 % случаев и поэтому получает соотношение 9/10. Далее потребуются данные, чтобы начать проверку гипотез. Друг бросает кость 10 раз и правильно угадывает результат броска в 9 случаях.

Сравнение правдоподобия

По традиции начнем с рассмотрения коэффициента Байеса, предполагая, что априорные шансы для каждой гипотезы равны. Сформулируем соотношение правдоподобия следующим образом:

$$\frac{P(D \mid H_2)}{P(D \mid H_1)}.$$

Результаты укажут, во сколько раз лучше (или хуже) утверждение вашего друга о том, что он экстрасенс, объясняет данные, чем ваша гипотеза. Для этого примера в уравнениях для краткости используем переменную КБ для

обозначения коэффициента Байеса. Вот результат, учитывающий, что ваш друг правильно предсказал 9 из 10 бросков:

$$\text{КБ} = \frac{P(D_{10} \mid H_2)}{P(D_{10} \mid H_1)} = \frac{\left(\frac{9}{10}\right)^9 \times \left(1 - \frac{9}{10}\right)^1}{\left(\frac{1}{6}\right)^9 \times \left(1 - \frac{1}{6}\right)^1} = 468\ 517.$$

Отношение правдоподобия показывает, что гипотеза друга-экстрасенса объясняет данные в 468 517 раз лучше, чем гипотеза, что другу просто везет. Это заслуживает внимания. Согласно таблице коэффициентов Байеса из предыдущих глав, это означает, что мы должны быть практически уверены, что H_2 истинна, а ваш друг — экстрасенс. Если только вы не поверили в возможность мистических сил, что-то здесь не так.

Добавление априорных шансов

В большинстве рассматриваемых здесь примеров, где одна только вероятность дает странные результаты, мы можем решить проблему, добавив априорные вероятности. Ясно, что мы не верим в гипотезу нашего друга почти так же сильно, как верим в собственную, поэтому имеет смысл создать сильные априорные шансы в пользу нашей гипотезы. Начнем с того, что установим отношение шансов достаточно высоким, чтобы оно нейтрализовало экстремальный результат коэффициента Байеса, и посмотрим, решит ли это нашу проблему:

$$O(H_2) = \frac{1}{468\ 517}.$$

Теперь при вычислении полных апостериорных шансов мы обнаруживаем, что снова не убеждены в том, что друг — экстрасенс:

$$\text{апостериорный шанс} = O(H_2) \times \frac{P(D_{10} \mid H_2)}{P(D_{10} \mid H_1)} = 1.$$

Похоже, априорные шансы снова спасли нас от затруднения, которое возникло при учете только одного коэффициента Байеса.

Но предположим, что друг бросает кость еще пять раз и успешно предсказывает все пять результатов. Теперь у нас есть новый набор данных, D_{15}, который представляет 15 бросков кости, 14 из которых ваш друг угадал.

Теперь при вычислении апостериорных шансов мы видим, что даже наш экстремальный априорный шанс мало помогает:

$$\text{апостериорный шанс} = O(H_2) \times \frac{P(D_{15} \mid H_2)}{P(D_{15} \mid H_1)} = \frac{1}{468\ 517} \times \frac{\left(\frac{9}{10}\right)^{14} \times \left(1 - \frac{9}{10}\right)^{1}}{\left(\frac{1}{6}\right)^{14} \times \left(1 - \frac{1}{6}\right)^{1}} = 4592$$

Используя существующий априорный шанс и имея всего пять бросков кости, мы получаем апостериорные шансы, равные 4592. Это означает, что мы вернулись к почти полной уверенности, что друг — действительно экстрасенс!

В большинстве предыдущих примеров мы исправили неинтуитивные апостериорные результаты, добавив адекватный априорный. Мы добавили довольно экстремальный априорный шанс против того, что ваш друг экстрасенс, но шансы по-прежнему сильно поддерживают обратную гипотезу.

Это серьезная проблема — байесовские рассуждения должны соответствовать логике. Понятно, что 15 бросков кости с 14 удачными догадками — это необычно, но вряд ли большинство убедится, что у оппонента есть экстрасенсорные способности. Но если мы не можем объяснить происходящее с помощью проверки гипотез, это означает, что мы действительно не можем полагаться на нее для решения повседневных статистических задач.

Учитываем альтернативные гипотезы

Проблема вот в чем: *мы не хотим верить, что друг является экстрасенсом.* Если бы вы оказались в такой ситуации в реальной жизни, то, скорее всего, быстро пришли бы к какому-то альтернативному выводу. Например, решили бы, что друг использует нечестную кость, которая выбрасывает определенное значение, например, в 90 % случаев. Это *третья* гипотеза. Наш коэффициент Байеса рассматривает только две возможные гипотезы: H_1, гипотезу о том, что кость честная, и H_2, гипотезу о том, что ваш друг — экстрасенс.

Коэффициент Байеса поддерживает гипотезу, что наш друг экстрасенс, а не то, что он правильно угадывает броски честной кости. Когда в таких терминах мы думаем о выводе, то это имеет больше смысла: с такими результатами очень маловероятно, что кость честная. Альтернатива H_2 не

вызывает комфорт, потому что наши представления о мире не поддерживают идею, что H_2 является реалистичным объяснением.

Важно понимать, что проверка гипотез сравнивает только два объяснения события, но зачастую бывает множество возможных объяснений. Если победившая гипотеза не убеждает вас, всегда можно рассмотреть третью.

Посмотрим, что происходит при сравнении H_2, победившей гипотезы, с новой гипотезой, H_3, — что кость нечестная и дает определенный результат в 90 % случаев.

Начнем с новых априорных шансов относительно H_2, которые назовем $O(H_2)'$ (галочка — это стандартное обозначение в математике, означающее «похоже, но не то же самое»). Это выражение будет представлять шансы H_2/H_3. Пока мы считаем, что вероятность того, что ваш друг использует нечестную кость, в 1000 раз выше, чем того, что он действительно экстрасенс (хотя реальный приоритет может быть гораздо более экстремальным). Это означает, что априорные шансы друга быть экстрасенсом составляют 1/1000. Если пересмотреть наши новые апостериорные шансы, то получается следующий интересный результат:

$$\text{КБ} = O(H_2)' \times \frac{P(D_{15} \mid H_2)}{P(D_{15} \mid H_3)} = \frac{1}{1000} \times \frac{\left(\frac{9}{10}\right)^{14} \times \left(1-\frac{9}{10}\right)^{1}}{\left(\frac{9}{10}\right)^{14} \times \left(1-\frac{9}{10}\right)^{1}} = \frac{1}{1000}.$$

Согласно этому вычислению, апостериорные шансы такие же, как и априорные, — $O(H_2)'$. Это происходит потому, что две вероятности одинаковы. Другими словами, $P(D_{15}|H_2) = P(D_{15}|H_3)$. Для обеих гипотез вероятность того, что ваш друг правильно угадал результат броска кости, одинакова с вероятностью использования нечестной кости, потому что вероятность, которую каждый присваивает успеху, одинакова. Это означает, что коэффициент Байеса всегда будет равен 1.

Эти результаты вполне интуитивны; в конце концов, без учета априорных шансов каждая гипотеза объясняет данные, которые мы видели, одинаково хорошо. Это означает, что если до рассмотрения данных мы считаем, что одно объяснение гораздо более вероятно, чем другое, то никакие новые доказательства не изменят наше мнение. Таким образом, проблем с наблюдаемыми данными больше нет; просто нашлось лучшее объяснение этому.

В этом сценарии никакое количество данных не изменит наше мнение о том, что H_3 превосходит H_2, потому что оба объясняют то, что мы наблюдали

одинаково хорошо, и мы уже думаем, что H_3 является гораздо более вероятным объяснением, чем H_2. Здесь интересно то, что мы можем оказаться в такой ситуации, даже если наши прежние убеждения совершенно иррациональны. Может быть, вы очень верите в мистические явления и думаете, что ваш друг — самый честный человек на свете. В этом случае можно задать априорные шансы как $O(H_2)' = 1000$. Если вы верите этому, никакие данные не смогут убедить вас, что друг жульничает.

В таких случаях важно понимать, что если вы хотите найти решение, то должны быть готовы изменить свои априорные убеждения. Если вы не хотите отпустить неоправданные априорные убеждения, то хотя бы должны признать, что больше не рассуждаете в байесовском, или логическом, смысле. Все мы придерживаемся иррациональных убеждений, и это совершенно нормально, если не пытаться использовать байесовские рассуждения для их оправдания.

Споры с родственниками и теории заговора

Любой, кто когда-либо спорил с родственниками за семейными посиделками о политике, изменении климата или своих любимых фильмах, лично столкнулся с ситуацией, в которой они сравнивают две гипотезы, обе одинаково хорошо объясняющие данные (тому, кто спорит), и в итоге остаются только априорные убеждения. Как мы можем изменить чужие (или собственные) убеждения, даже если добавление данных ничего не меняет?

Ели сравнить убеждение, что ваш друг бросает нечестный кубик, и убеждение, что он экстрасенс, большее количество данных ничего не изменит и не повлияет на ваши убеждения. Это потому, что и ваша гипотеза, и гипотеза вашего друга объясняют данные одинаково хорошо. Чтобы друг убедил вас в том, что он экстрасенс, он должен изменить ваши прежние убеждения. Например, поскольку вы подозреваете, что кости могут быть нечестные, друг может предложить вам выбрать кость, которую бросит. Если вы купили новую кость, дали ее своему другу и он продолжает точно предсказывать броски, то вы начнете сомневаться по поводу старых убеждений. Та же самая логика сохраняется всякий раз, когда две гипотезы одинаково объясняют данные. В этих случаях нужно посмотреть, есть ли что-то, что можно изменить в своих априорных убеждениях.

Предположим, что после того как вы купили новую кость для своего друга и он продолжает добиваться успеха, вы *все равно* не поверите ему; теперь

вы утверждаете, что у друга должен быть секретный способ броска. В ответ друг позволяет вам бросить кость самостоятельно и продолжает успешно предсказывать броски — но вы *все еще* не верите ему. В этом сценарии происходит нечто иное, помимо скрытой гипотезы. Теперь у вас есть H_4 — ваш друг жульничает, — и вы не передумали. Это означает, что для любого D_n, $P(D_n|H_4) = 1$. Ясно, что мы находимся вне байесовской территории, так как вы фактически признали, что не передумали, но давайте посмотрим, что происходит с математической точки зрения, если ваш друг настаивает на попытках убедить вас.

Посмотрим, как эти два объяснения, H_2 и H_4, дополнятся с использованием данных D_{10} с 9 верными предсказаниями и 1 неверным. Это объясняется коэффициентом Байеса:

$$\text{КБ} = \frac{P(D_{10}\,|\,H_2)}{P(D_{10}\,|\,H_4)} = \frac{\left(\frac{9}{10}\right)^{14} \times \left(1-\frac{9}{10}\right)^{1}}{1} = \frac{1}{26}.$$

Поскольку вы отказываетесь верить во что-либо, кроме того что друг жульничает, вероятность того, что вы наблюдаете, равна (и всегда будет равна) 1. Даже если данные в точности соответствуют ожиданиям того, что ваш друг — экстрасенс, мы считаем, что наши убеждения также объясняют данные в 26 случаях. Друг, решительно настроенный сломить ваше упрямство, упорствует и бросает кость 100 раз, получая 90 правильных догадок и 10 неправильных. Коэффициент Байеса меж тем демонстрирует нечто очень странное:

$$\text{КБ} = \frac{P(D_{100}\,|\,H_2)}{P(D_{100}\,|\,H_4)} = \frac{\left(\frac{9}{10}\right)^{90} \times \left(1-\frac{9}{10}\right)^{10}}{1} = \frac{1}{131\ 272\ 619\ 177\ 803}.$$

Даже несмотря на то что данные явно подтверждают гипотезу друга, вы отказываетесь сдвинуться с места в своих убеждениях, теперь вы еще сильнее убеждены в своей правоте! Когда мы вообще не позволяем изменить свое мнение, большее число данных еще сильнее убеждает нас в том, что мы правы.

Это поведение знакомо любому, кто спорил с радикально настроенными родственниками или кем-то, кто непреклонно верит в теорию заговора. В байесовских рассуждениях жизненно важно, чтобы убеждения были

как минимум опровержимыми. В традиционной науке *опровержимость* означает, что что-то можно опровергнуть, но в нашем случае это просто означает, что должен быть какой-то способ уменьшить нашу веру в гипотезу.

Опасность неопровержимых убеждений в байесовских рассуждениях заключается не только в том, что нельзя доказать, что они неверны, но и в том, что они даже подкрепляются доказательствами, которые, кажется, противоречат им. Вместо того чтобы настойчиво пытаться убедить вас, друг должен был сначала спросить: «Что я могу показать тебе, чтобы ты передумал?» Если бы вы ответили, что *ничто* не может изменить ваше мнение, тогда другу лучше даже не пытаться дать вам больше доказательств.

Итак, в следующий раз, когда вы будете спорить с родственником по поводу политики или теории заговора, спросите его: «Какие доказательства изменили бы твое мнение?» Если ответа на этот вопрос нет, то отстаивать свои взгляды, приводя еще больше доказательств, смысла нет, поскольку это только повысит уверенность оппонента в своей правоте.

Заключение

Из этой главы вы узнали о нескольких способах проверки гипотез. Хотя коэффициент Байеса — это конкуренция двух идей, вполне возможно, что есть и другие, не менее обоснованные гипотезы, которые стоит проверить.

В других случаях мы обнаруживаем, что две гипотезы одинаково хорошо объясняют данные; вы с равной вероятностью будете наблюдать правильные предсказания вашего друга, если они обусловлены экстрасенсорными способностями или хитрым трюком с кубиком. В этом случае имеет значение только отношение априорных шансов для каждой гипотезы. Это также означает, что получение большего количества данных в таких ситуациях никогда не изменит наших убеждений, потому что это никогда не даст ни одной гипотезе преимущества над другой. В этих случаях лучше всего подумать, как вы можете изменить априорные убеждения, которые влияют на результаты.

В более экстремальных случаях может быть гипотеза, которую просто не хотят опровергать. Похоже на теорию заговора в отношении данных. В этом случае большее количество данных не только никогда не убедит нас изменить убеждения, но и фактически возымеет противоположный эффект. Если гипотеза не может быть опровергнута, дополнительные данные только сильнее убедят в ней.

Упражнения

Чтобы убедиться, что вы понимаете, как справляться с крайними случаями в байесовских рассуждениях, попробуйте ответить на эти вопросы.

1. Когда две гипотезы одинаково хорошо объясняют данные, один из способов изменить мнение — посмотреть, можно ли воздействовать на априорную вероятность. Какие факторы могут повысить вашу априорную веру в экстрасенсорные способности друга?
2. Эксперимент утверждает, что когда люди слышат слово «Флорида»[1], они думают о пенсионерах и это влияет на их скорость ходьбы. Чтобы проверить это, мы собрали две группы из 15 студентов, которые идут по комнате; одна группа слышит слово «Флорида», а другая — нет. Предположим, что члены группы H_1 двигаются с одинаковой скоростью, а группы H_2 двигаются медленнее, потому что слышат слово «Флорида». Также предположим:

$$\text{КБ} = \frac{P(D|H_2)}{P(D|H_1)}.$$

   Эксперимент показывает, что H_2 имеет коэффициент Байеса, равный 19. Предположим, что кто-то не убежден в этом эксперименте, потому что у H_2 были более низкие шансы на выигрыш. Какие априорные шансы объяснили бы, что кого-то не убедили, и каким должен быть КБ, чтобы довести апостериорные шансы до 50 для этого неубежденного человека?

   Теперь предположим, что априорные шансы не изменили мнение скептика. Подумайте об альтернативной H_3, которая объясняет наблюдение, что группа «Флорида» двигается медленнее. Помните, если H_2 и H_3 объясняют данные одинаково хорошо, только априорные шансы в пользу H_3 заставят кого-то утверждать, что H_3 вернее H_2, поэтому нужно переосмыслить эксперимент, чтобы уменьшить эти шансы. Придумайте эксперимент, который может изменить априорные шансы H_3 по сравнению с H_2.

[1] В штате Флорида самая высокая концентрация пенсионеров в США. — *Примеч. ред.*

19

От проверки гипотез к оценке параметров

До сих пор мы использовали апостериорные шансы для сравнения только двух гипотез, что подходит для простых задач; даже если у нас есть три или четыре гипотезы, их можно проанализировать, проведя несколько проверок гипотез. Но иногда нужно найти действительно большое пространство возможных гипотез, чтобы объяснить данные. Например, вы можете угадать количество драже в банке, высоту какого-либо здания или точное количество минут, которое потребуется для прибытия рейса. Во всех этих случаях существует множество всевозможных гипотез, и их слишком много, чтобы привести все.

К счастью, есть способ для обработки этого сценария. В главе 15 мы узнали, как превратить задачу оценки параметров в проверку гипотез. В этой главе мы собираемся сделать обратное: рассматривая практически непрерывный диапазон возможных гипотез, мы можем использовать коэффициент Байеса и апостериорные шансы (проверка гипотезы) в качестве формы оценки параметров! Этот подход позволяет оценивать более двух гипотез и предоставляет простую структуру для оценки любого параметра.

Честна ли ярмарочная игра?

Предположим, вы находитесь на праздничной ярмарке. Прогуливаясь, вы замечаете, что кто-то спорит с работником ярмарки возле бассейна

с маленькими резиновыми уточками. Любопытствуя, вы подходите ближе и слышите, как человек кричит: «Вы жулики! Вы сказали, что шанс получить приз 1 к 2. Я выловил 20 уток и получил только один приз! Получается, что шанс на приз всего лишь 1 к 20!»

Теперь, когда вы хорошо понимаете вероятности, то разрешаете этот спор самостоятельно. Вы объясняете присутствующему и сердитому клиенту, что если сегодня увидите еще несколько игр, то сможете использовать коэффициент Байеса, чтобы определить, кто прав. Вы создаете две гипотезы. H_1 — утверждение работника, что вероятность выигрыша равна $1/2$, и H_2 — утверждение сердитого клиента, что вероятность выигрыша составляет всего $1/20$:

$$H_1 : P(\text{выигрыша}) = \frac{1}{2};$$

$$H_2 : P(\text{выигрыша}) = \frac{1}{20}.$$

Работник утверждает, что поскольку он не смотрел, как клиент вылавливал уточек, то не считает, что следует использовать его сообщенные данные, так как никто не может их проверить. Звучит справедливо. Вы решаете посмотреть следующие 100 игр и использовать их в качестве данных. После того как клиент подобрал 100 уток, вы заметили, что 24 из них получили призы.

Теперь о коэффициенте Байеса. Поскольку у нас нет четкого мнения ни о претензии клиента, ни о заявлениях работника, мы не будем беспокоиться об априорных шансах или вычислении полных апостериорных шансов. Чтобы получить коэффициент Байеса, нужно вычислить $P(D|H)$ для каждой гипотезы:

$$P(D|H_1) = (0{,}5)^{24} \times (1 - 0{,}5)^{76};$$

$$P(D|H_2) = (0{,}05)^{24} \times (1 - 0{,}05)^{76}.$$

По отдельности обе эти вероятности довольно малы, но все, что нас интересует, — их соотношение. Мы рассмотрим соотношение с точки зрения H_2/H_1, чтобы результат сообщал нам, во сколько раз гипотеза клиента лучше объясняет данные, чем гипотеза работника:

$$\frac{P(D|H_2)}{P(D|H_1)} = \frac{1}{653}.$$

Коэффициент Байеса указывает, что H_1, гипотеза работника, объясняет данные в 653 раза лучше, чем H_2; это означает, что гипотеза работника (вероятность получить приз при вылавливании уточки составляет 0,5) является более вероятной.

Это сразу должно насторожить. Очевидно, что вероятность получить только 24 приза, когда было выловлено 100 уточек, кажется маловероятной, если истинная вероятность выигрыша равна 0,5. Мы можем использовать функцию `pbinom()` в R (см. главу 13) для вычисления биномиального распределения, которое сообщит вероятность получить 24 или меньше призов, предполагая, что вероятность получения приза действительно равна 0,5:

```
> pbinom(24,100,0.5)
9.050013e-08
```

Как видите, вероятность получения 24 или менее призов при истинной вероятности выигрыша 0,5 чрезвычайно мала; расширив ее до полного десятичного значения, мы получим вероятность 0,00000009050013! Что-то определенно не так с H_1. Хотя мы не верим гипотезе работника, она все же объясняет данные гораздо лучше, чем данные клиента.

Чего же не хватает? Мы уже сталкивались с тем, что априорная вероятность обычно имеет большое значение, когда только один коэффициент Байеса не дает осмысленного ответа. Но в главе 18 мы видели, что бывают случаи, когда априорная вероятность не является основной причиной. В этом случае использование следующего уравнения кажется разумным, поскольку в любом случае единого мнения нет:

$$O\left(\frac{H_2}{H_1}\right) = 1.$$

Возможно, проблема здесь в том, что у вас уже есть недоверие к ярмарочным играм. Поскольку результат коэффициента Байеса так сильно поддерживает гипотезу работника, априорные шансы должны быть не менее 653, чтобы поддержать гипотезу клиента:

$$O\left(\frac{H_2}{H_1}\right) = 653.$$

Сильное недоверие к честности игры! Здесь должно быть еще что-то, кроме априорных вероятностей.

Рассматриваем множественные гипотезы

Очевидная проблема заключается в том, что хотя интуитивно и понятно, что работник ошибается в своей гипотезе, альтернативная гипотеза клиента слишком экстремальна, чтобы быть верной, поэтому в наличии две неверные гипотезы. Что, если клиент подумал, что вероятность выигрыша равна 0,2, а не 0,05? Мы назовем эту гипотезу H_3. Проверка H_3 против гипотезы работника радикально меняет результаты нашего отношения правдоподобия:

$$bf = \frac{P(D \mid H_3)}{P(D \mid H_1)} = \frac{(0,2)^{24} \times (1-0,2)^{76}}{(0,5)^{24} \times (1-0,5)^{76}} = 917\ 399.$$

H_3 объясняет данные значительно лучше, чем H_1. С коэффициентом Байеса 917 399 мы можем быть уверены, что H_1 — далеко не лучшая гипотеза для объяснения наблюдаемых данных, потому что H_3 разбивает ее в пух и прах. Проблема, с которой мы столкнулись при первой проверке гипотезы, заключалась в том, что убеждения клиента были гораздо худшим описанием события, чем убеждения работника. Но как мы видим, это не значит, что работник был прав. Когда мы выдвинули альтернативную гипотезу, то увидели, что она было гораздо лучшей, чем догадка работника или клиента. Но задачу мы не решили. Что, если есть еще лучшая гипотеза?

Поиск дополнительных гипотез с помощью R

Нам нужно более общее решение, которое ищет все возможные гипотезы и выбирает лучшую. Для этого можно использовать функцию `seq()` в R, чтобы создать последовательность гипотез, которые нужно сравнить с H_1.

Рассмотрим каждый шаг в 0,01 между 0 и 1 как возможную гипотезу — то есть 0,01, 0,02, 0,03 и т. д. Величину 0,01, на которую мы увеличиваем каждую гипотезу, мы назовем `dx` (принятое обозначение из высшей математики, представляющее «наименьшее изменение») и используем ее для определения переменной `hypotheses`, которая представляет все возможные гипотезы, которые нужно рассмотреть. Применим функцию `seq()` в R для генерации диапазона значений для каждой гипотезы от 0 до 1, увеличивая значения на `dx`:

```
dx <- 0.01
hypotheses <- seq(0,1,by=dx)
```

Потребуется функция, которая может вычислить отношение правдоподобия для любых двух гипотез. Функция `bayes.factor()` будет принимать два

аргумента: `h_top`, который обозначает вероятность получения выигрыша за гипотезу клиенты (числитель), и `h_bottom`, который обозначает гипотезу работника. Выглядеть это будет так:

```
bayes.factor <- function(h_top,h_bottom){
  ((h_top)^24*(1-h_top)^76)/((h_bottom)^24*(1-h_bottom)^76)
}
```

Наконец, вычисляем отношение правдоподобия для всех этих возможных гипотез:

```
bfs <- bayes.factor(hypotheses,0.5)
```

Используем базовый функционал построения графиков в R, чтобы увидеть, как выглядят эти отношения правдоподобия:

```
plot(hypotheses,bfs, type='l')
```

На рис. 19.1 показан результирующий график.

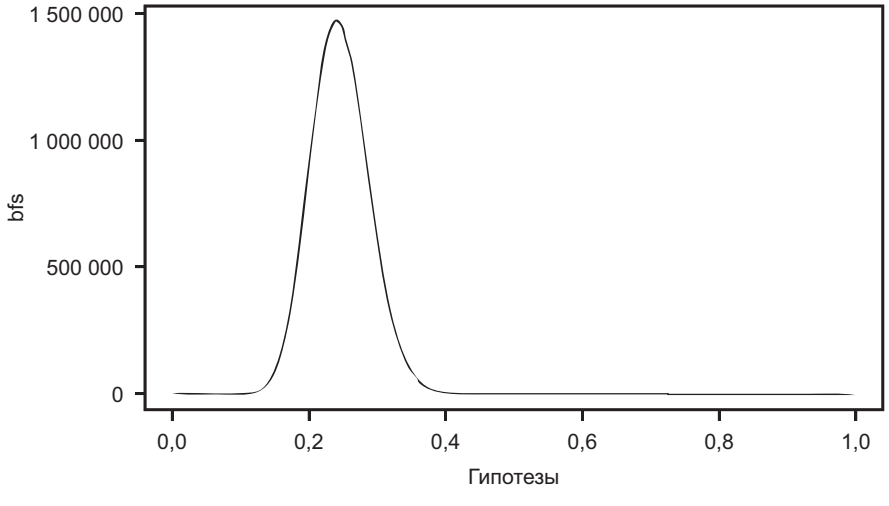


Рис. 19.1. Построение графика коэффициента Байеса для каждой из гипотез

Видно четкое распределение различных объяснений для наблюдаемых данных. Используя R, мы можем рассмотреть широкий диапазон возможных гипотез, где каждая точка на линии представляет коэффициент Байеса для соответствующей гипотезы на оси *X*.

Мы также можем увидеть, насколько велик самый большой коэффициент Байеса, используя функцию `max()` с нашим вектором `bfs`:

```
> max(bfs)
1.47877610^{6}
```

Можно проверить, какая гипотеза соответствует наибольшему отношению правдоподобия, говоря нам, в какие гипотезы стоит верить больше всего. Для этого введите:

```
> hypotheses[which.max(bfs)]
0.24
```

Теперь мы знаем, что вероятность 0,24 является лучшим предположением, так как эта гипотеза дает самое высокое отношение правдоподобия по сравнению с гипотезой работника. Из главы 10 вы узнали, что использование среднего или ожидаемого значения данных часто является хорошим способом оценки параметров. Здесь мы просто выбрали гипотезу, которая объясняет данные наилучшим образом, потому что сейчас нет способа взвесить оценки по вероятности их появления.

Добавление априорных вероятностей к коэффициентам правдоподобия

Вы показываете результаты клиенту и работнику. Оба согласны с тем, что ваши выводы довольно убедительны, но тут подходит другой человек и говорит вам: «Раньше я создавал такие игры и могу сказать вам, что по какой-то причине люди, которые разрабатывают игры с уточками, никогда не устанавливают призовую ставку от 0,2 до 0,3. Держу пари, что шансы 1000 к 1 и что реальный выигрыш не находится в этом диапазоне. Ничего, кроме этого, сказать не могу».

Теперь у нас есть некоторые априорные шансы, которые нужно использовать. Поскольку бывший создатель игр дал нам серьезные шансы относительно своих априорных убеждений в вероятности получения приза, можно попытаться умножить это значение на наш текущий список коэффициентов Байеса и вычислить апостериорные шансы. Для этого создадим список коэффициентов априорных шансов для каждой имеющейся гипотезы. Как сказал бывший создатель игр, отношение шансов для всех вероятностей от 0,2 до 0,3 должно составлять 1/1000. Поскольку он не знает ничего о других гипотезах, отношение шансов для них будет равно 1. Используем простой

оператор `ifelsestate` и вектор `hypotheses` для создания вектора коэффициентов шансов:

```
priors <- ifelse(hypotheses >= 0.2 & hypotheses <= 0.3, 1/1000,1)
```

Затем еще раз применим `plot()`, чтобы отобразить это распределение априорных вероятностей:

```
plot(hypotheses,priors,type='l')
```

На рис. 19.2 показано наше распределение априорных шансов.

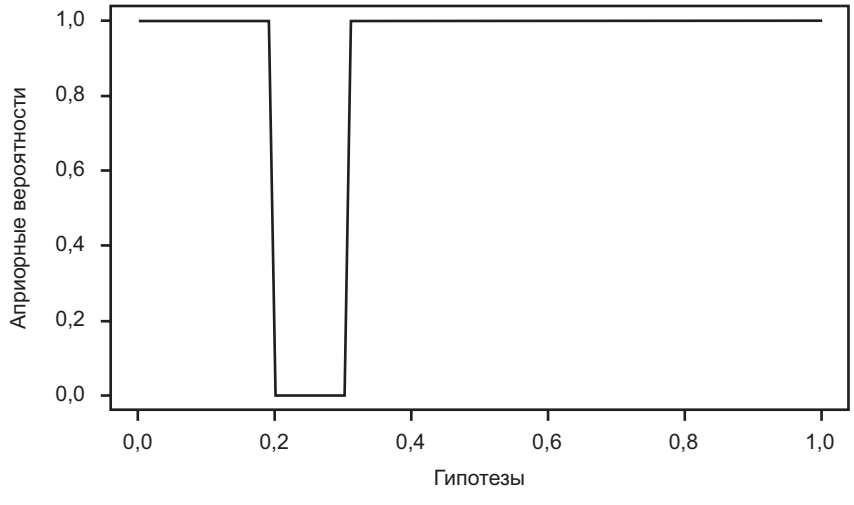


Рис. 19.2. Визуализация коэффициентов априорных шансов

Поскольку R является векторным языком (подробнее об этом см. в приложении А), можно просто умножить `priors` на `bfs` и получить новый вектор исходных данных, представляющих коэффициенты Байеса:

```
posteriors <- priors*bfs
```

Наконец, можно построить график вероятности повторения каждой из наших многочисленных гипотез:

```
plot(hypotheses,posteriors,type='l')
```

График показан на рис. 19.3.

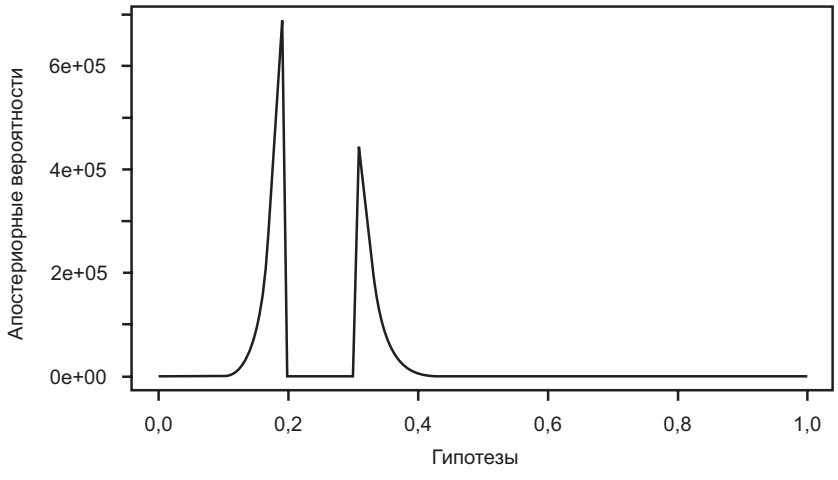


Рис. 19.3. Построение графика распределения коэффициентов Байеса

В итоге получается очень странное распределение возможных убеждений. У нас есть достаточная уверенность в значениях от 0,15 до 0,2 и от 0,3 до 0,35, но мы находим диапазон между 0,2 и 0,3 крайне маловероятным. Но это распределение является честным представлением о силе веры в каждую гипотезу, учитывая то, что мы узнали о процессе производства игр с уточками. Хотя эта визуализация полезна, мы действительно хотим иметь возможность обрабатывать эти данные как истинное распределение вероятностей. Таким образом, можно задавать вопросы о том, насколько мы верим в диапазоны возможных гипотез, и рассчитывать ожидания распределения, чтобы получить единственную оценку гипотезы.

Построение распределения вероятностей

Истинное распределение вероятностей — это такое распределение, где сумма всех возможных убеждений равна 1. Наличие распределения вероятностей позволило бы нам рассчитать ожидание (или среднее значение) данных, чтобы сделать более точную оценку истинной вероятности получения приза. Это также позволило бы легко суммировать диапазоны значений, чтобы получить доверительные интервалы и другие подобные оценки.

Но если сложить все апостериорные шансы для гипотез, они не будут равны 1, как показано в этом расчете:

```
> sum(posteriors)
3.140687510^{6}
```

Значит, нужно нормализовать апостериорные шансы так, чтобы они давали в сумме 1. Для этого разделим каждое значение в векторе апостериорных вероятностей на сумму всех значений:

```
p.posteriors <- posteriors/sum(posteriors)
```

Теперь значения `p.posteriors` складываются, давая в итоге 1:

```
> sum(p.posteriors)
1
```

Наконец, построим новый `p.posteriors`:

```
plot(hypotheses,p.posteriors,type='l')
```

График показан на рис. 19.4.

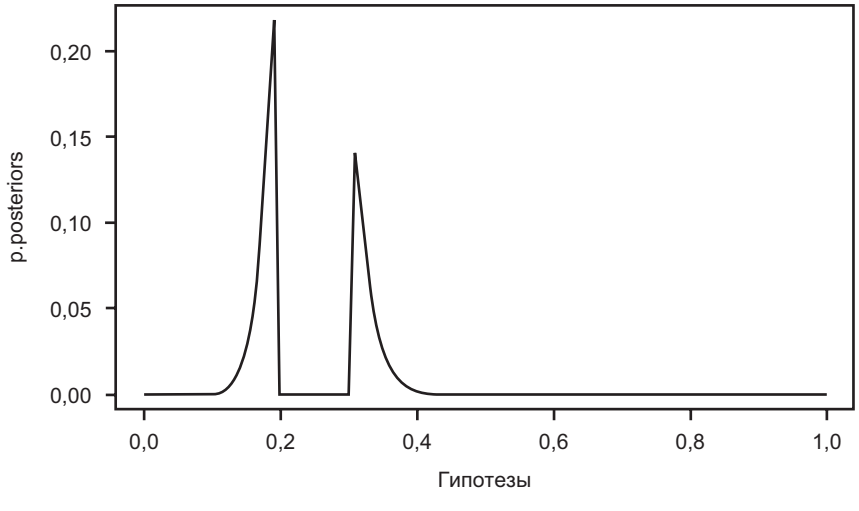


Рис. 19.4. Нормализованные апостериорные шансы (обратите внимание на шкалу оси Y)

Можно использовать `p.posteriors`, чтобы ответить на некоторые общие вопросы о наших данных. Теперь можно рассчитать вероятность того, что истинный показатель получения приза будет меньше, чем утверждают участники. Сложим все вероятности для значений меньше 0,5:

```
sum(p.posteriors[which(hypotheses < 0.5)])
> 0.9999995
```

Как мы видим, вероятность того, что призовая ставка ниже, чем утверждает работник, составляет почти 1. Почти наверняка работник завышает истинную призовую ставку.

Рассчитаем ожидание распределения и используем этот результат в качестве оценки истинной вероятности. Напомню, что ожидание — это сумма оценок, взвешенных по их значению:

```
> sum(p.posteriors*hypotheses)
0.2402704
```

Полученное распределение несколько нетипично, с большим разрывом в середине, поэтому можно выбрать наиболее *вероятную* оценку следующим образом:

```
> hypotheses[which.max(p.posteriors)]
0.19
```

Теперь мы использовали коэффициент Байеса, чтобы получить диапазон вероятностных оценок для истинно возможного показателя выигрыша приза в игре с уточками. Это означает, что мы использовали коэффициент Байеса как форму оценки параметров!

От коэффициента Байеса к оценке параметров

Давайте еще раз взглянем на коэффициенты вероятности. Если мы не использовали априорную вероятность для какой-либо из гипотез, вы, скорее всего, понимали, что это был хороший подход к решению задачи без учета коэффициента Байеса. Мы наблюдали 24 вытащенные уточки с призами и 76 вытащенных уточек без призов. Разве нельзя использовать старое доброе бета-распределение? Как мы уже много раз обсуждали, начиная с главы 5, если нужно оценить частоту какого-либо события, мы всегда можем использовать бета-распределение. На рис. 19.5 показан график бета-распределения, где альфа 24 и бета 76.

За исключением масштаба оси *Y*, график выглядит практически идентично исходному графику наших коэффициентов правдоподобности. Но если мы проделаем несколько простых трюков, то сможем добиться идеального совпадения этих двух графиков. Если масштабировать бета-распределение по размеру `dx` и нормализовать `bfs`, то мы увидим, что эти два распределения достаточно близки (рис. 19.6).

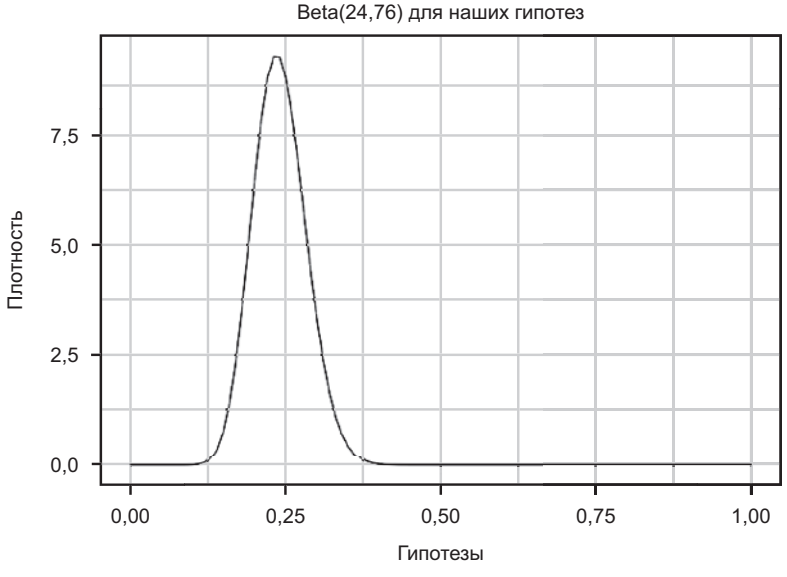


Рис. 19.5. Бета-распределение с альфа 24 и бета 76

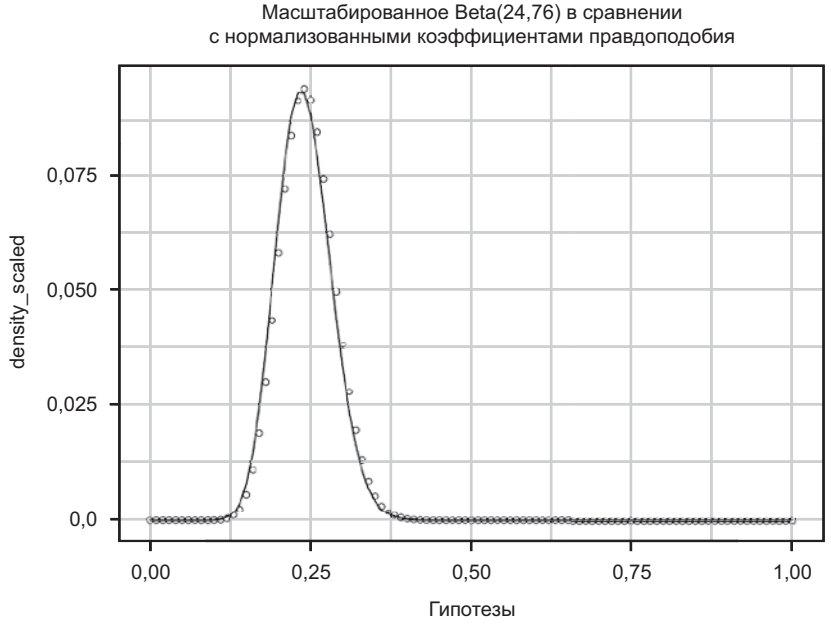


Рис. 19.6. Первоначальное распределение коэффициентов правдоподобия довольно близко соответствует Beta(24,76)

Кажется, сейчас есть только небольшая разница. Это можно исправить, используя самую слабую априорную вероятность, которая указывает на то, что получение приза и неполучение приза одинаково вероятны, то есть путем добавления 1 к параметрам альфа и бета (рис. 19.7).

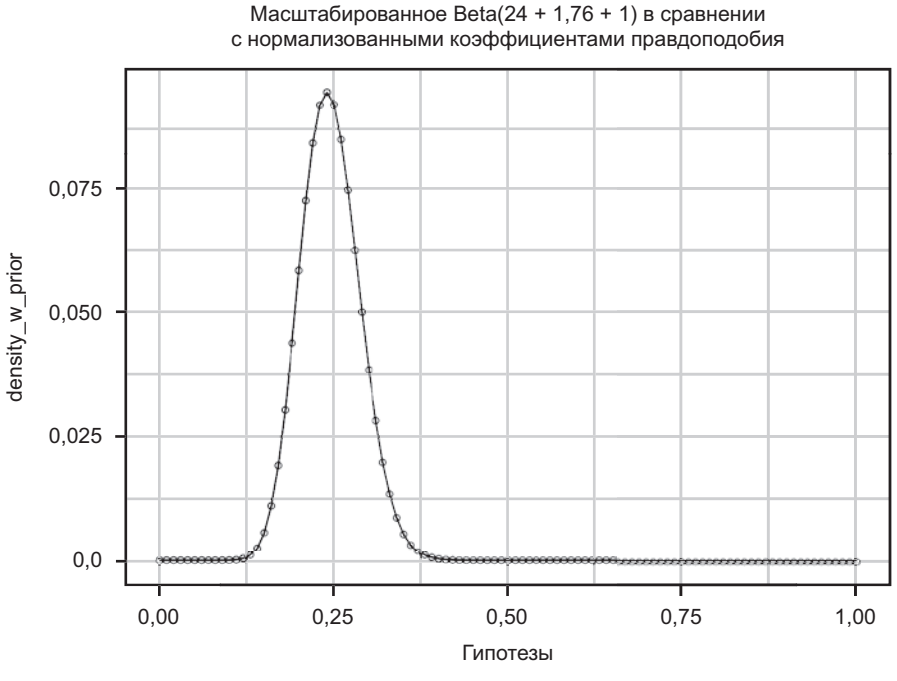


Рис. 19.7. Наши отношения правдоподобия идеально соответствуют распределению Beta(24 + 1,76 + 1)

Теперь эти два распределения идеально выровнены. В главе 5 упоминалось, что бета-распределение было трудно вывести из наших основных правил вероятности. Но с коэффициентом Байеса мы смогли эмпирически воссоздать его модифицированную версию, которая предполагает априорное распределение Beta(1,1). И сделали мы это без всякой заумной математики! Нужно было всего лишь:

1) определить вероятность доказательства, выдвинутого гипотезой;
2) рассмотреть все возможные гипотезы;
3) нормализовать эти значения, чтобы создать распределение вероятностей.

Каждый раз, когда в книге использовалось бета-распределение, это было бета-распределение априорной вероятности. Оно облегчило вычисления, поскольку прийти к апостериорной вероятности можно, комбинируя альфа- и бета-параметры из правдоподобности и априорных бета-распределений. Другими словами:

$$\text{Beta}(\alpha_{\text{апостериорное}}, \beta_{\text{апостериорное}}) =$$

$$= \text{Beta}(\alpha_{\text{априорное}} + \alpha_{\text{правдоподобия}}, \beta_{\text{априорное}} + \beta_{\text{правдоподобия}}).$$

Но построив распределение на основе коэффициента Байеса, можно легко использовать уникальное априорное распределение. Коэффициент Байеса не только является отличным инструментом для настройки проверок гипотез. С его помощью также создается любое распределение вероятностей, которое можно использовать для решения проблемы, будь то проверка гипотез или оценка параметров. Достаточно уметь определить базовое сравнение между двумя гипотезами, и мы уже на месте.

При создании A/B-теста в главе 15 мы выяснили, как свести многие проверки гипотез к проблеме оценки параметров. Теперь вы увидели, как наиболее распространенная форма проверки гипотез также может использоваться для оценки параметров. Учитывая эти две взаимосвязанные идеи, мы не имеем ограничений на тип вероятностных проблем, которые можно решить, используя только самые основные правила вероятности.

Заключение

Мы закончили путешествие по байесовской статистике, и теперь вы можете оценить истинную красоту того, что изучили. Из основных правил вероятности мы можем вывести теорему Байеса, которая позволяет преобразовывать доказательства в утверждение, выражающее силу наших убеждений. Из теоремы Байеса можно вывести коэффициент Байеса — инструмент для сравнения того, насколько хорошо две гипотезы объясняют наблюдаемые данные. Итерируя возможные гипотезы и нормализуя результаты, можно использовать коэффициент Байеса, чтобы создать оценку параметров для неизвестного значения. Это, в свою очередь, позволяет выполнять бесчисленные другие проверки гипотез, сравнивая оценки. И все, что нужно сделать, чтобы выпустить всю эту мощь, — использовать основные правила вероятности, чтобы определить правдоподобность, $P(D \mid H)$!

Упражнения

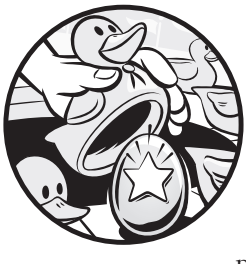
Чтобы убедиться, что вы понимаете использование коэффициента Байеса и априорных шансов для оценки параметров, попробуйте ответить на эти вопросы.

1. Коэффициент Байеса предполагал, что мы рассматриваем H_1: $P(\text{приз}) = 0{,}5$. Это позволило нам получить версию бета-распределения со значением альфа 1 и бета 1. Будет ли иметь значение выбор другой вероятности для H_1? Предположим, что H_1: $P(\text{приз}) = 0{,}24$, а затем посмотрим, отличается ли результирующее распределение, однажды нормализованное до суммы 1, от исходной гипотезы.
2. Напишите априорную вероятность для распределения, в котором каждая гипотеза в 1,05 раза более вероятна, чем предыдущая (предположим, что `dx` остается неизменным).
3. Предположим, вы наблюдали еще одну игру с уточками, где было 34 уточки с призами и 66 уточек без призов. Какую проверку вы бы сделали, чтобы ответить на вопрос: какова вероятность того, что шансов выиграть приз в этой игре больше, чем в той игре, которая приводилась в нашем примере? Реализация этой проверки намного сложнее, чем то, что было показано в этой книге, но наверняка вы сможете изучить все самостоятельно и отправиться в собственное приключение по миру более продвинутой байесовской статистики!

ПРИЛОЖЕНИЯ

A

Краткое введение в язык R



В этой книге для выполнения сложной вычислительной работы используется язык R, который специализируется на статистике и науке о данных. Если у вас нет опыта работы с R или вообще с программированием, не беспокойтесь — это приложение поможет вам начать работу.

R и RStudio

Для запуска примеров кода в этой книге необходимо установить R на компьютер. Для этого перейдите по ссылке https://cran.rstudio.com/ и следуйте инструкциям по установке для используемой операционной системы.

После установки R также нужно установить RStudio, интегрированную среду разработки (IDE), которая делает запуск проектов R чрезвычайно простым. Загрузите и установите RStudio с сайта www.rstudio.com/products/rstudio/download/.

При открытии RStudio вас должны встретить несколько панелей (рис. A.1).

Самая важная панель — большая в середине, называемая *консолью*. В консоли можно ввести любой из примеров кода из книги и запустить его, просто нажав клавишу Enter. Консоль сразу запускает весь код, который вы вводите, что затрудняет отслеживание написанного кода.

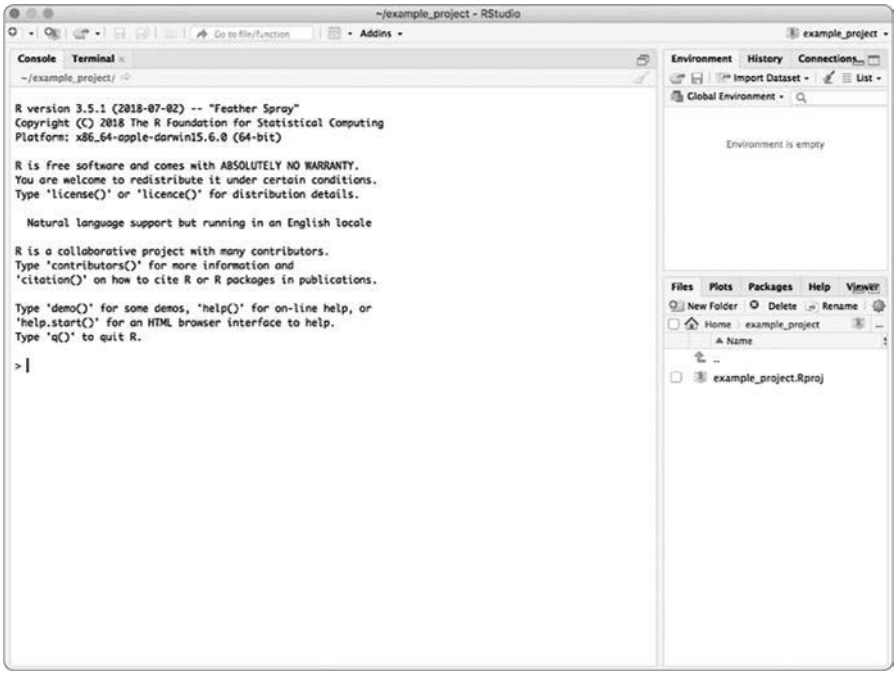


Рис. А.1. Просмотр консоли в RStudio

Чтобы писать программы, которые можно сохранить и вернуться к ним, нужно поместить свой код в сценарий R, который представляет собой текстовый файл. Его можно загрузить в консоль позже. R — чрезвычайно интерактивный язык программирования, поэтому вместо того чтобы думать о консоли как о месте, где можно тестировать код, следует представлять сценарии R как способ быстрой загрузки инструментов, которые можно использовать в консоли.

Создание сценария в R

Чтобы создать сценарий в R, перейдите к File ▸ New File ▸ R Script в RStudio. Появится новая пустая панель в левом верхнем углу (рис. А.2).

На этой панели вы можете ввести код и сохранить его в виде файла. Чтобы запустить код, просто нажмите кнопку Source в правом верхнем углу панели или запустите отдельные строки, нажав кнопку Run. Кнопка Source автоматически загрузит ваш файл в консоль, как если бы вы его там напечатали.

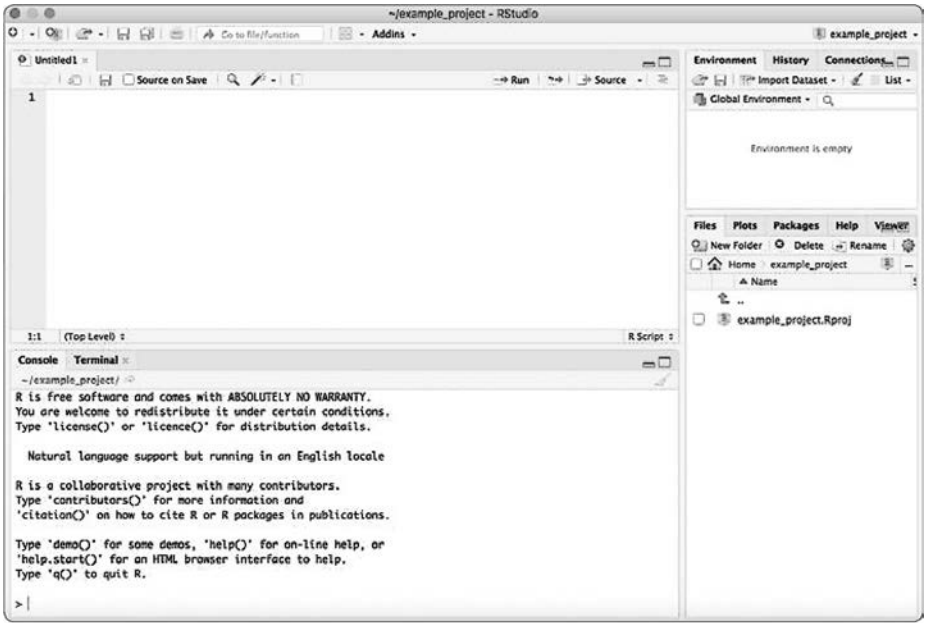


Рис. А.2. Создание сценария в R

Основные понятия R

Мы будем использовать R в качестве продвинутого калькулятора, а это значит, что требуется понять только основы, чтобы разобраться с проблемами и самостоятельно расширить примеры из этой книги.

Типы данных

Все языки программирования имеют различные типы данных, которые можно использовать для разных целей и по-разному манипулировать ими. R включает в себя большое разнообразие типов и структур данных, но в этой книге мы будем использовать только небольшое их количество.

Числа с плавающей точкой

Числа, которые мы используем в R, имеют тип *double* (сокращение от «числа с плавающей точкой двойной точности», *double-precision floating-point*,

которое является наиболее распространенным способом представления десятичных чисел в компьютере). Число с плавающей точкой является типом по умолчанию для представления десятичных чисел. Если не указано иное, все числа, которые вы вводите в консоль, имеют тип *double*.

Мы можем манипулировать такими числами с помощью стандартных математических операций. Например, можно сложить два числа с помощью оператора `+`. Попробуйте воспроизвести это в консоли:

```
> 5 + 2
[1] 7
```

С помощью оператора `/` можно разделить любые числа, что даст десятичный результат:

```
> 5/2
[1] 2.5
```

Можно умножить значения с помощью оператора `*`:

```
> 5 * 2
[1] 10
```

И возвести значение в степень, используя оператор `^`. Например, 5^2 это:

```
> 5^2
[1] 25
```

Также можно добавить перед числом знак «минус», чтобы сделать его отрицательным:

```
> 5 - -2
[1] 7
```

И еще можно использовать экспоненциальную запись с `e+`. Таким образом, 5×10^2 — это просто:

```
> 5e+2
[1] 500
```

Если мы используем `e-`, то получаем тот же результат, что и 5×10^{-2}:

```
> 5e-2
[1] 0.05
```

Это полезно знать, потому что иногда R возвращает результат в экспоненциальной записи, если он слишком большой:

```
> 5*10^20
[1] 5e+20
```

Строки

Еще один важный тип в R — это *строка (string)*, представляющая собой группу используемых символов для отображения текста. В R мы заключаем строку в кавычки, например:

```
> "hello"
[1] "hello"
```

Обратите внимание, что если вы помещаете число в строку, это число нельзя использовать в обычных числовых операциях, поскольку строки и числа — это разные типы. Например:

```
> "2" + 2
Error in "2" + 2 : non-numeric argument to binary operator
```

Мы не будем широко использовать строки в этой книге. В первую очередь они потребуются для передачи аргументов функциям и для обозначения графиков. Но важно помнить о них, если вы используете текст.

Логические типы

Логические, или *бинарные*, типы представляют истинные или ложные значения, выраженные кодами `TRUE` и `FALSE`. Обратите внимание, что `TRUE` и `FALSE` не являются строками — они не заключены в кавычки и пишутся заглавными буквами. (R также позволяет вам просто использовать `T` или `F` вместо записи полных слов.)

Логические типы можно комбинировать с символами `&` («и») и `|` («или») для выполнения основных логических операций. Например, если мы хотим узнать, может ли что-то быть одновременно истинным *и* ложным, то можем ввести:

```
> TRUE & FALSE
```

R вернет:

```
[1] FALSE
```

сообщая нам, что значение не может быть одновременно истиной и ложью. Но как насчет истины *или* лжи?

```
> TRUE | FALSE
[1] TRUE
```

Как и строки, логические значения будут в основном использоваться для предоставления аргументов функциям, которые мы будем использовать, или в качестве результатов сравнения двух разных значений.

Отсутствующие значения

В практической статистике и data science в данных часто отсутствуют некоторые значения. Например, есть данные о температуре для утра и полдня каждого дня в течение месяца, но однажды что-то дает сбой и вы обнаруживаете, что не хватает утренней температуры. Поскольку отсутствующие значения встречаются очень часто, R имеет особый способ их представления: значение `NA`. Важно иметь способ обрабатывать отсутствующие значения, потому что они могут означать очень разные вещи в разных контекстах. Например, при измерении количества осадков отсутствующее значение может означать, что в датчике не было дождя, или это может означать, что было много дождей, но в ту ночь температура была ниже нуля, что привело к поломке датчика и утечке воды. В первом случае мы могли бы считать, что отсутствующие значения означают 0, но во втором случае неясно, каким должно быть значение. Хранение отсутствующих значений отдельно от других значений заставляет нас учитывать эти различия. Чтобы подсказать нам, каковы наши отсутствующие значения каждый раз, когда мы пытаемся их использовать, R будет выводить `NA` для любой операции, используя отсутствующее значение:

```
> NA + 2
[1] NA
```

Как мы увидим чуть позже, разные функции в R могут обрабатывать отсутствующие значения различными способами, но не нужно беспокоиться об отсутствующих значениях для примеров R, используемых в этой книге.

Векторы

Почти каждый язык программирования содержит определенные функции, которые делают его уникальным и подходящим для решения задач в своей области. Особенность R в том, что это *векторный язык*. Вектор — это список

значений, и все, что делает R, — совершает операции над векторами. Мы используем код с `(...)` для определения векторов (но даже если мы введем только одно значение, R сделает это самостоятельно).

Чтобы понять, как работают векторы, рассмотрим пример. Введите следующий код в сценарии, а не в консоли. Сначала создадим новый вектор, присвоив переменную `x` вектору `c(1,2,3)` с помощью оператора присваивания `<-`:

```
x <- c(1,2,3)
```

Теперь, когда у нас есть вектор, мы можем использовать его в расчетах. При выполнении простой операции, например добавления `3` к `x` и ввода результата в консоли, мы получаем довольно неожиданный вывод (особенно если вы привыкли к другому языку программирования):

```
> x + 3
[1] 4 5 6
```

Результат говорит нам, что произойдет, если мы добавим `3` к каждому значению в нашем векторе `x`. (Во многих других языках программирования нужно было бы использовать цикл `for` или какой-нибудь другой итератор для выполнения такой операции.)

Также можно складывать векторы друг с другом. Здесь мы создадим новый вектор, содержащий три элемента, каждый со значением `2`. Назовем этот вектор `y`, а затем прибавим `y` к `x`:

```
> y <- c(2,2,2)
> x + y
[1] 3 4 5
```

Как видите, эта операция добавила каждый элемент в `x` к соответствующему элементу в `y`. А что, если умножить эти два вектора?

```
> x * y
[1] 2 4 6
```

Каждое значение в `x` умножается на соответствующее значение в `y`. Если бы списки не были одинакового размера или кратны одинаковому размеру, мы получили бы ошибку. Если вектор кратен одинаковому размеру, R будет просто многократно применять меньший вектор к большему. Но в этой книге эта операция не будет использоваться.

Мы можем довольно легко комбинировать векторы в *R*, определяя новый вектор на основе существующих. Здесь мы создадим вектор `z` путем объединения `x` и `y`:

```
> z <- c(x,y)
> z
[1] 1 2 3 2 2 2
```

Обратите внимание, что эта операция не вернула вектор векторов; вместо этого мы получили один вектор, который содержит значения обоих в том порядке, в котором `x` и `y` были заданы при определении `z`.

Эффективное использование векторов в R может поначалу показаться сложным. Как ни странно, программисты, которые имеют опыт работы с языками, не основанными на векторах, часто испытывают наибольшие трудности. Не беспокойтесь: в этой книге мы будем использовать векторы, чтобы упростить чтение кода.

Функции

Функции — это блоки кода, которые выполняют определенную операцию со значением, и мы будем использовать их в R для решения задач.

В R и RStudio все функции снабжены документацией. Если вы введете `?`, за которым последует имя функции в консоли, то получите полную документацию по этой функции. Например, при вводе `?sum` в консоли вы должны увидеть документацию в правом нижнем углу экрана (рис. А.3).

В документации дается определение функции `sum()` и некоторые варианты ее применения. Функция `sum()` принимает значения вектора и складывает их. В документации сказано, что она принимает `...` в качестве аргумента, это означает, что она может принимать любое количество значений. Обычно эти значения представляют собой вектор чисел, но они также могут состоять из нескольких векторов.

В документации также указан *необязательный аргумент*: `na.rm = FALSE`. Необязательные аргументы — это аргументы, которые не нужно передавать в функцию, чтобы она работала; если вы не передадите необязательный аргумент, R будет использовать значение аргумента по умолчанию. В случае с `na.rm`, который автоматически удаляет все пропущенные значения,

значением по умолчанию после знака равенства является FALSE. Это означает, что по умолчанию sum() не удалит пропущенные значения.

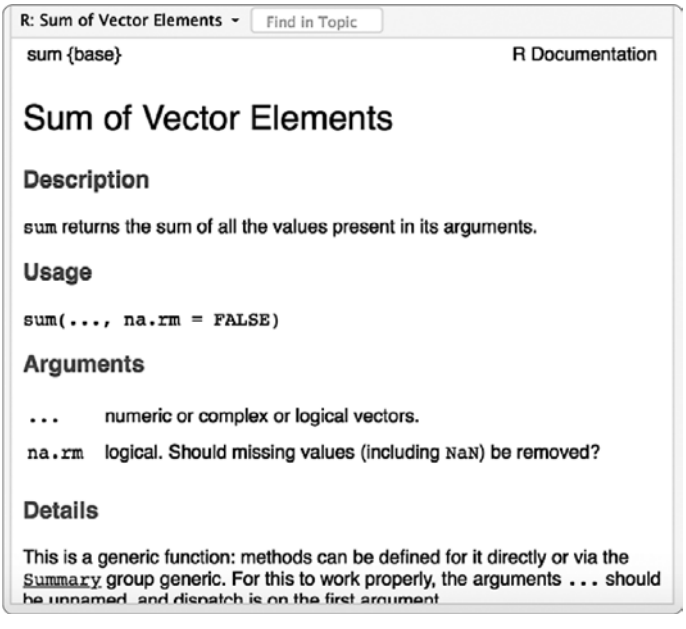


Рис. А.3. Просмотр документации для функции sum()

Основные функции

Вот некоторые из наиболее важных функций *R*.

Функции length() и nchar()

Функция length() возвращает длину вектора:

```
> length(c(1,2,3))
[1] 3
```

Поскольку в этом векторе три элемента, функция length() возвращает 3.

Поскольку все в *R* является вектором, можно использовать функцию length(), чтобы найти длину чего угодно — даже строки, например, doggies:

```
> length("doggies")
[1] 1
```

R говорит, что "doggies" — это вектор, содержащий одну строку.

Теперь, если бы у нас было две строки, "doggies" и "cats", мы бы получили:

```
> length(c("doggies", "cats"))
[1] 2
```

Чтобы найти количество символов в строке, используйте функцию nchar():

```
> nchar("doggies")
[1] 7
```

Обратите внимание, что если мы используем nchar() для вектора c("doggies","cats"), R вернет новый вектор, содержащий количество символов в каждой строке:

```
> nchar(c("doggies","cats"))
[1] 7 4
```

Функции sum(), cumsum() и diff()

Функция sum() принимает вектор чисел и складывает все эти числа:

```
> sum(c(1,1,1,1,1))
[1] 5
```

Из документации предыдущего раздела мы знаем, что sum() принимает ... в качестве аргумента, а это означает, что она может принять любое количество значений:

```
> sum(2,3,1)
[1] 6
> sum(c(2,3),1)
[1] 6
> sum(c(2,3,1))
[1] 6
```

Независимо от того, сколько векторов мы даем, sum() складывает их, как если бы они были одним вектором целых чисел. Если нужно суммировать несколько векторов, вызовите sum() для каждого из них по отдельности.

Помните также, что функция sum() принимает необязательный аргумент na.rm, который по умолчанию имеет значение FALSE. Аргумент na.rm определяет, удаляет ли sum() значения NA или нет.

Если мы оставим для `na.rm` значение `FALSE`, а затем попытаемся использовать `sum()` для вектора с отсутствующим значением, произойдет вот что:

```
> sum(c(1,NA,3))
[1] NA
```

Как мы видели, добавление любого значения к значению `NA` приводит к получению `NA`. Если нужно, чтобы R вместо предыдущего ответа возвращал число, можем дать команду `sum()` удалить значения `NA`, установив `na.rm = TRUE`:

```
> sum(c(1,NA,3),na.rm = TRUE)
[1] 4
```

Функция `cumsum()` принимает вектор и вычисляет его *общую сумму* — вектор той же длины, что и входной, который заменяет каждое число суммой чисел, предшествующих ему (включая это число). Вот пример кода:

```
> cumsum(c(1,1,1,1,1))
[1] 1 2 3 4 5
> cumsum(c(2,10,20))
[1] 2 12 32
```

Функция `diff()` принимает вектор и вычитает каждое число из числа, предшествующего ему в векторе:

```
> diff(c(1,2,3,4,5))
[1] 1 1 1 1
> diff(c(2,10,3))
[1] 8 -7
```

Обратите внимание, что результат функции `diff()` содержит на один элемент меньше, чем исходный вектор. Это потому, что из первого значения в векторе ничего не вычитается.

Оператор : и функция seq()

Чтобы не перечислять вручную каждый элемент вектора, можно генерировать векторы автоматически. Чтобы автоматически создать вектор целых чисел в определенном диапазоне, используется оператор `:`, чтобы отделить начало и конец диапазона. R может даже выяснить, нужно ли считать по направлению вверх или вниз (оборачивание этого оператора в `c()` не обязательно):

```
> c(1:5)
[1] 1 2 3 4 5
> c(5:1)
[1] 5 4 3 2 1
```

При использовании : R будет проводить подсчет от первого значения до последнего.

Иногда нужно посчитать что-то иным способом, кроме приращения на единицу. Функция `seq()` позволяет создавать векторы последовательности значений, которые увеличиваются на определенную величину. Аргументы `seq()` расположены по порядку:

1) начало последовательности;
2) конец последовательности;
3) величина, на которую нужно увеличить последовательность.

Вот несколько примеров использования `seq()`:

```
> seq(1,1.1,0.05)
[1] 1.00 1.05 1.10
> seq(0,15,5)
[1] 0 5 10 15
> seq(1,2,0.3)
[1] 1.0 1.3 1.6 1.9
```

Если нужно отсчитать до определенного значения с помощью функции `seq()`, используйте отрицательное значение в качестве приращения, например:

```
> seq(10,5,-1)
[1] 10 9 8 7 6 5
```

Функция ifelse()

Функция `ifelse()` дает команду R выполнить одно из двух действий на основе некоторого условия. Эта функция может немного сбить с толку, если вы привыкли к обычной структуре управления `if...else` в других языках. В R она принимает следующие три аргумента (по порядку):

1) утверждение о векторе, которое может быть истинным или ложным по отношению к его значениям;
2) то, что происходит в случае, если утверждение истинно;
3) то, что происходит в случае, если утверждение ложно.

Функция `ifelse()` работает сразу с целыми векторами. Когда речь идет о векторах, содержащих одно значение, его использование интуитивно понятно:

```
> ifelse(2 < 3,"small","too big")
[1] "small"
```

Здесь утверждение состоит в том, что 2 меньше 3, и мы просим R вывести `"small"` («маленькое»), если это так, и `"too big"` («слишком большое»), если это не так.

Предположим, у нас есть вектор `x`, который содержит несколько значений:

```
> x <- c(1,2,3)
```

Функция `ifelse()` вернет значение для каждого элемента вектора:

```
> ifelse(x < 3,"small","too big")
[1] "small" "small" "too big"
```

Мы также можем использовать векторы в аргументах результатов для `ifelse()`. Предположим, что в дополнение к нашему вектору `x` был еще один вектор `y`:

```
y <- c(2,1,6)
```

Нужно создать новый список, который содержит наибольшее значение из `x` и `y` для каждого элемента в векторе. Можно использовать `ifelse()` в качестве очень простого решения:

```
> ifelse(x > y,x,y)
[1] 2 2 6
```

R сравнил значения в `x` с соответствующими значениями в `y` и вывел наибольшее из двух для каждого элемента.

Случайные выборки

Мы будем часто использовать R для случайной выборки значений. Это позволяет компьютеру выбрать случайное число или значение за нас. Мы используем этот пример для имитации таких действий, как подбрасывание монетки, игра в «камень, ножницы, бумага» или выбор числа от 1 до 100.

Функция runif()

Одним из способов случайной выборки значений является функция `runif()`, сокращение для «случайной последовательности», которая принимает требуемый аргумент `n` и возвращает это же число выборок в диапазоне от 0 до 1:

```
> runif(5)
[1] 0.8688236 0.1078877 0.6814762 0.9152730 0.8702736
```

Мы можем использовать эту функцию с `ifelse()` для генерации значения `A` в 20 % случаев. При этом мы будем использовать `runif(5)` для создания пяти случайных значений от 0 до 1. Затем, если значение меньше 0,2, мы вернем `A`; в противном случае вернем `B`:

```
> ifelse(runif(5) < 0.2,"A","B")
[1] "B" "B" "B" "B" "A"
```

Так как числа, которые мы генерируем, случайные, каждый раз при запуске функции `ifelse()` будут разные результаты. Вот некоторые возможные:

```
> ifelse(runif(5) < 0.2,"A","B")
[1] "B" "B" "B" "B" "B"
> ifelse(runif(5) < 0.2,"A","B")
[1] "A" "A" "B" "B" "B"
```

Функция `runif()` может принимать необязательные второй и третий аргументы, которые являются минимальным и максимальным значениями диапазона для случайной последовательности чисел. По умолчанию функция использует диапазон от 0 до 1 включительно, но вы можете установить любой диапазон:

```
> runif(5,0,2)
[1] 1.4875132 0.9368703 0.4759267 1.8924910 1.6925406
```

Функция rnorm()

Можно произвести выборку из нормального распределения, используя функцию `rnorm()`, которая подробно описана в главе 12:

```
> rnorm(3)
[1] 0.28352476 0.03482336 -0.20195303
```

По умолчанию `rnorm()` выбирает нормальное распределение со средним значением 0 и стандартным отклонением 1, как в этом примере. Это означает, что образцы будут иметь «колоколообразное» распределение около 0, при этом большинство образцов близко к 0, а очень мало — меньше –3 или больше 3.

Функция `rnorm()` имеет два необязательных аргумента, `mean` и `sd`, которые позволяют установить другое среднее значение и стандартное отклонение соответственно:

```
> rnorm(4,mean=2,sd=10)
[1] -12.801407 -9.648737 1.707625 -8.232063
```

В статистике выборка из нормального распределения часто более распространена, чем выборка из равномерного распределения, поэтому `rnorm()` очень удобна.

Функция sample()

Иногда нужно выбрать что-то еще, помимо хорошо изученного распределения. Предположим, у вас есть ящик с носками разных цветов:

```
socks <- c("red","grey","white","red","black")
```

Если нужно смоделировать случайный выбор любых двух носков, вы можете использовать функцию `sample()`, которая принимает в качестве аргументов вектор значений и количество элементов для выборки:

```
> sample(socks,2)
[1] "grey" "red"
```

Функция `sample()` ведет себя так, как если бы мы выбрали два случайных носка из ящика, не кладя их обратно. Если мы выберем пять носков, мы получим все носки, которые были в ящике:

```
> sample(socks,5)
[1] "grey" "red" "red" "black" "white"
```

Это означает, что если мы попытаемся взять шесть носков из ящика, где есть только пять, то получим ошибку:

```
> sample(socks,6)
Error in sample.int(length(x), size, replace, prob) :
cannot take a sample larger than the population when 'replace = FALSE'
```

Если нужно выполнить выборку и «положить носки обратно», мы можем установить необязательный аргумент `replace` в значение `TRUE`. Теперь каждый раз мы достаем носок и кладем его обратно в ящик. Это позволяет доставать больше носков, чем есть в ящике. Это также означает, что распределение носков в ящике никогда не меняется.

```
> sample(socks,6,replace=TRUE)
[1] "black" "red" "black" "red" "black" "black"
```

С помощью этих простых инструментов выборки можно запускать удивительно сложные симуляции в R, которые избавят вас от множества математических вычислений.

Использование set.seed() для предсказуемых случайных результатов

Случайные числа, сгенерированные R, не являются действительно случайными числами. Как и во всех языках программирования, случайные числа генерируются *генератором псевдослучайных чисел*, который принимает начальное значение и использует его для создания последовательности чисел, достаточно случайных для большинства целей. Начальное значение устанавливает начальное состояние генератора случайных чисел и определяет, какие числа будут выбраны в последовательности. В R мы можем вручную установить это начальное значение с помощью функции `set.seed()`. Установка начального значения чрезвычайно полезна в случаях, когда нужно снова использовать те же случайные результаты:

```
> set.seed(1337)
> ifelse(runif(5) < 0.2, "A","B")
[1] "B" "B" "A" "B" "B"
> set.seed(1337)
> ifelse(runif(5) < 0.2, "A","B")
[1] "B" "B" "A" "B" "B"
```

Когда мы дважды использовали один и тот же начальный фрагмент с функцией `runif()`, он сгенерировал один и тот же набор предположительно случайных значений. Основным преимуществом использования `set.seed()` является воспроизводимость результатов. Это может значительно упростить отслеживание ошибок в программах, связанных с выборкой, поскольку результаты не меняются при каждом запуске программы.

Определение собственных функций

Иногда полезно написать собственные функции для определенных операций, которые придется выполнять неоднократно. В R можно определять функции, используя ключевое слово `function` (*«ключевое слово»* в языке программирования — это просто специальное слово, зарезервированное для конкретного использования).

Вот определение функции, принимающей один аргумент, `val`, обозначающий значение, которое пользователь введет в функцию, а затем удваивает значение `val` и возводит его в куб.

```
double_then_cube <- function(val){
  (val*2)^3
}
```

После того как функция определена, ее можно использовать как встроенную функцию R. Вот наша функция `double_then_cube()`, примененная к числу 8:

```
> double_then_cube(8)
[1] 4096
```

Поскольку все, что мы делали для определения нашей функции, *векторизовано* (то есть все значения работают с векторами значений), функция будет работать и для векторов, и для отдельных значений:

```
> double_then_cube(c(1,2,3))
[1] 8 64 216
```

Мы также можем определить функции, которые принимают более одного аргумента. Определенная здесь функция `sum_then_square()` складывает два аргумента вместе, а затем возводит результат в квадрат:

```
sum_then_square <- function(x,y){
  (x+y)^2
}
```

Добавляя два аргумента (`x, y`) в функцию, в определении R мы указываем, что функция `sum_then_square()` ожидает два аргумента. Теперь мы можем использовать нашу новую функцию, например, следующим образом:

```
> sum_then_square(2,3)
[1] 25
> sum_then_square(c(1,2),c(5,3))
[1] 36 25
```

Также можно определить функции, которым требуется несколько строк. В R при вызове функции та всегда возвращает результат вычисления в последней строке определения. Это означает, что мы могли бы переписать `sum_then_square()` следующим образом:

```
sum_then_square <- function(x,y){
  sum_of_args <- x+y
  square_of_result <- sum_of_args^2
  square_of_result
}
```

Как правило, при написании функций вы создаете их в файле сценария R, чтобы можно было сохранить их и использовать позже.

Создание основных графиков в R

В R мы можем очень быстро создавать графики данных. Хотя R имеет необычную библиотеку графиков `ggplot2`, которая содержит множество полезных функций для генерации красивых графиков, мы пока ограничимся базовыми функциями построения графиков, которые сами по себе очень полезны. Чтобы показать, как работает построение графиков, создадим два вектора значений, `xs` и `ys`:

```
> xs <- c(1,2,3,4,5)
> ys <- c(2,3,2,4,6)
```

Затем мы можем использовать эти векторы в качестве аргументов функции `plot()`, которая построит для нас данные. Функция `plot()` принимает два аргумента: значения точек графика на оси *X* и значения этих точек на оси *Y* в следующем порядке:

```
> plot(xs,ys)
```

Эта функция должна генерировать график, показанный на рис. А.4 в левом нижнем окне RStudio.

Этот график показывает взаимосвязь между нашими значениями `xs` и соответствующими им значениями `ys`. Если мы вернемся к функции, то сможем присвоить этому графику заголовок, используя необязательный аргумент `main`. Мы также можем изменить метки осей *X* и *Y* с помощью аргументов `xlab` и `ylab`, например:

```
plot(xs,ys,
     main="example plot",
     xlab="x values",
     ylab="y values"
     )
```

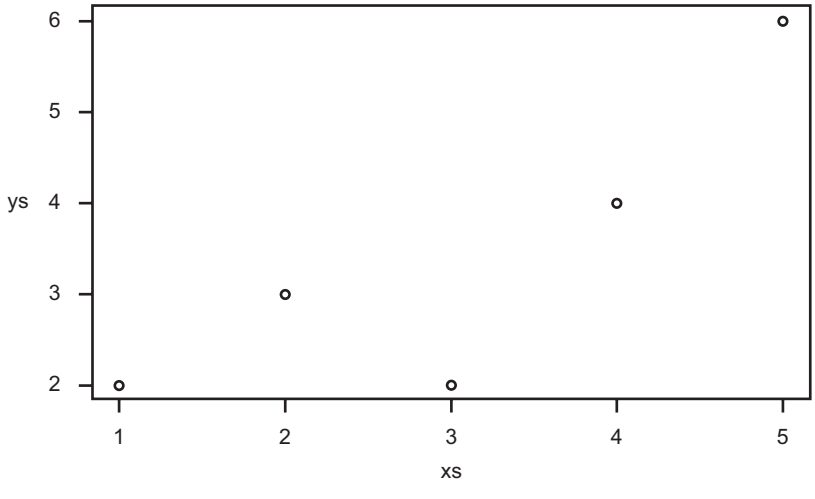


Рис. А.4. Простой график, созданный с помощью функции plot() в R

Новые метки должны отображаться так, как показано на рис. А.5.

Можно также изменить тип графика, используя аргумент `type`. Первый тип графика, который мы сгенерировали, называется *точечным графиком*, но если нужно создать линейный график, который рисует линию через каждое значение, следует установить `type = "l"`:

```
plot(xs,ys,
     type="l",
     main="example plot",
     xlab="x values",
     ylab="y values"
     )
```

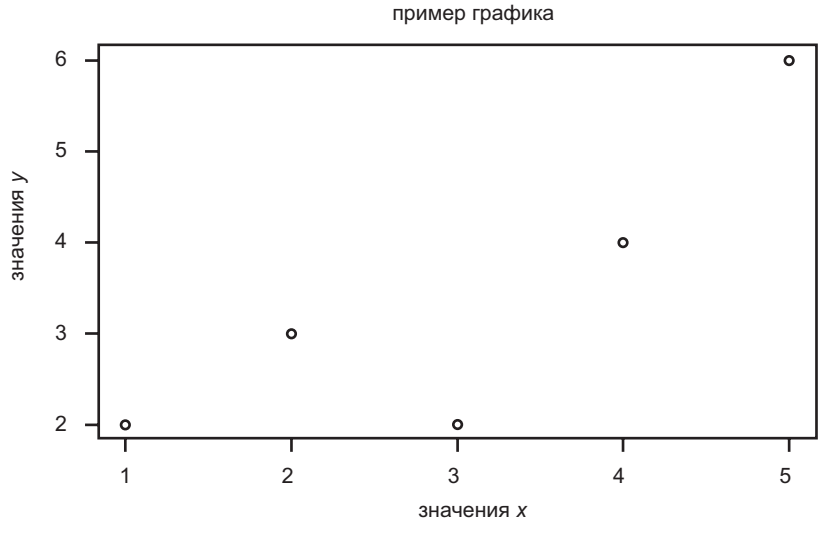


Рис. А.5. Изменение заголовка и меток графика с помощью функции plot()

Тогда это будет выглядеть так, как на рис. А.6.

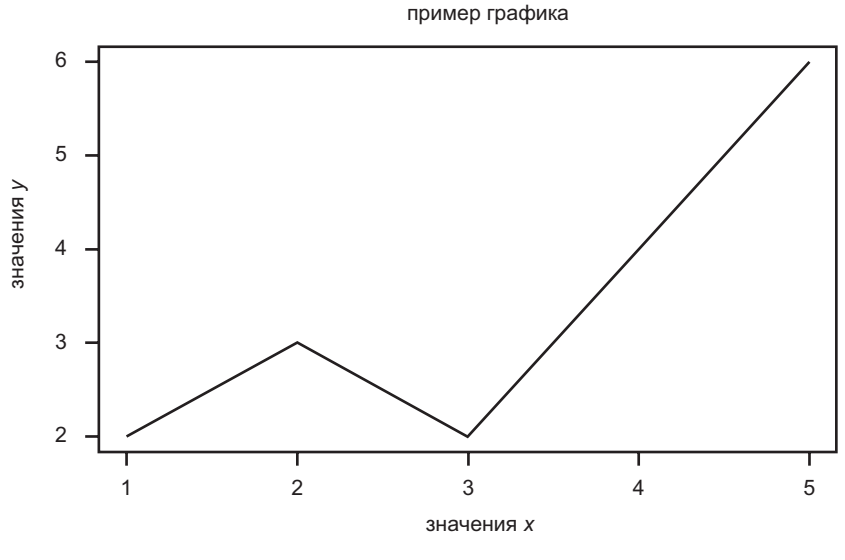


Рис. А.6. Линейный график, сгенерированный с помощью функции plot() в R

Или мы можем сделать и то и другое! Функция R под названием `lines()` может добавлять линии к существующему графику. Он принимает большинство тех же аргументов, что и `plot()`:

```
plot(xs,ys,
     main="example plot",
     xlab="x values",
     ylab="y values"
     )
lines(xs,ys)
```

На рис. А.7 показан график, который будет сгенерирован этой функцией.

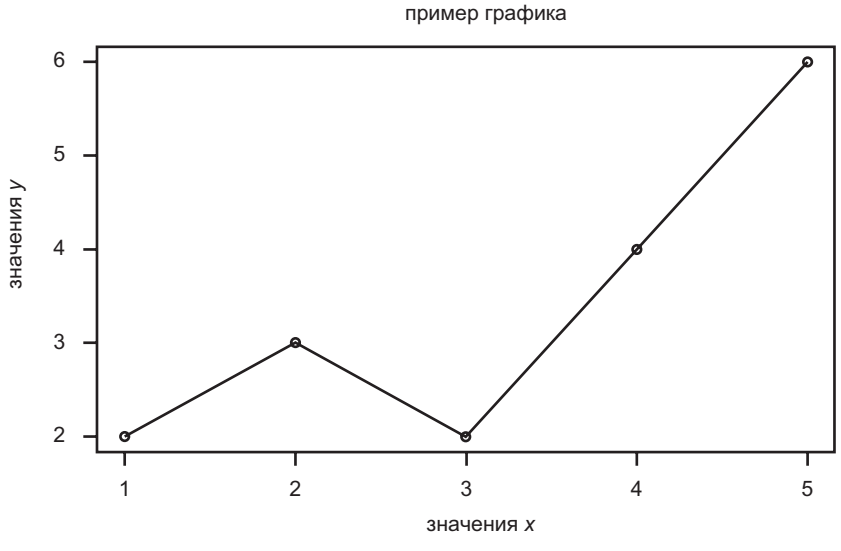


Рис. А.7. Добавление линий к существующему графику с помощью функции lines() в R

Существует множество более удивительных способов использования основных графиков в *R*, и вы можете обратиться к `?plot` для получения дополнительной информации о них. Но если вы хотите создавать действительно красивые графики в R, стоит изучить библиотеку *ggplot2* (https://ggplot2.tidyverse.org/).

Упражнение: моделирование цен на бирже

Теперь давайте применим все свои навыки для создания имитации биржевого тикера! Люди часто моделируют цены на акции, используя общую сумму нормально распределенных случайных значений. Для начала мы будем моделировать движение запасов в течение определенного периода времени, генерируя последовательность значений от 1 до 20, увеличивая ее на 1 каждый раз с помощью функции `seq()`. Мы назовем вектор, представляющий период времени `t.vals`.

```
t.vals <- seq(1,20,by=1)
```

Рис. А.8. График, сгенерированный для моделируемого биржевого тикера

Теперь `t.vals` — это вектор, содержащий последовательность чисел от 1 до 20, увеличивающихся на 1. Далее создадим несколько смоделированных цен, взяв общую сумму нормально распределенного значения для каждого момента времени в `t.vals`. Для этого мы будем использовать `rnorm()` для выборки количества значений, равного длине `t.vals`. Затем используем

cumsum() для вычисления общей суммы этого вектора значений, что будет представлять идею движения цены вверх или вниз на основе случайного сдвига; менее экстремальные сдвиги встречаются чаще, чем более экстремальные.

```
price.vals <- cumsum(rnorm(length(t.vals),mean=5,sd=10))
```

Наконец, можно построить график для всех этих значений, чтобы посмотреть, как они выглядят! Используем функции plot() и lines() и пометим оси в соответствии с тем, что они представляют.

```
plot(t.vals,price.vals,
     main="Simulated stock ticker",
     xlab="time",
     ylab="price")
lines(t.vals,price.vals)
```

Функции plot() и lines() должны сгенерировать график, показанный на рис. А.8.

Заключение

Приложение охватывает основы языка R в достаточном объеме, чтобы вы могли понять примеры из этой книги. Рекомендую следовать примерам из книги, а затем самостоятельно поэкспериментировать с примерами кода, чтобы закрепить знания. У языка R есть отличная онлайн-документация, что поможет вам в дальнейшем обучении (https://cran.r-project.org/manuals.html).

Б

Математический минимум

В этой книге мы иногда будем использовать идеи из высшей математики, хотя никакого реального решения задач не потребуется! Что *потребуется*, так это понимание некоторых основ, таких как производная и (особенно) интеграл. Это приложение ни в коем случае не является попыткой глубоко изучить эти концепции или показать вам, как их решать. Оно предлагает краткий обзор этих идей и того, как они представлены в математической записи.

Функции

Функция — это просто математическая «машина», которая принимает одно значение, что-то делает с ним и возвращает другое значение. Это очень похоже на работу функций в языке R (см. приложение А): они принимают значение и возвращают результат. Например, в высшей математике может быть функция f, определенная следующим образом:

$$f(x) = x^2.$$

В этом примере f принимает значение x и возводит его в квадрат. Например, если мы введем значение 3 в f, то получим:

$$f(3) = 9.$$

Это немного отличается от алгебраических задач средней школы, где обычно есть значение y и некоторое уравнение с x.

$$y = x^2.$$

Одна из причин важности функций заключается в том, что они позволяют абстрагироваться от реальных вычислений, которые мы делаем. Это означает, что мы можем сказать что-то вроде $y = f(x)$ и просто заниматься абстрактным поведением самой функции, а не тем, как она определена. Это подход, который мы будем использовать для этого приложения.

К примеру, вы готовитесь к забегу на 5 километров и используете умные часы, чтобы отслеживать расстояние, скорость, время и другие факторы. Сегодня вы вышли на пробежку и пробежали полчаса. Однако умные часы работали со сбоями и записывали только скорость в милях в час в течение получасового пробега. На рис. Б.1 показаны данные, которые удалось восстановить. Представьте, что скорость бега создана функцией s, которая принимает аргумент t, время в часах. Функция обычно пишется в терминах аргумента, который она принимает, поэтому мы будем писать $s(t)$, что приводит к значению, представляющему текущую скорость в момент

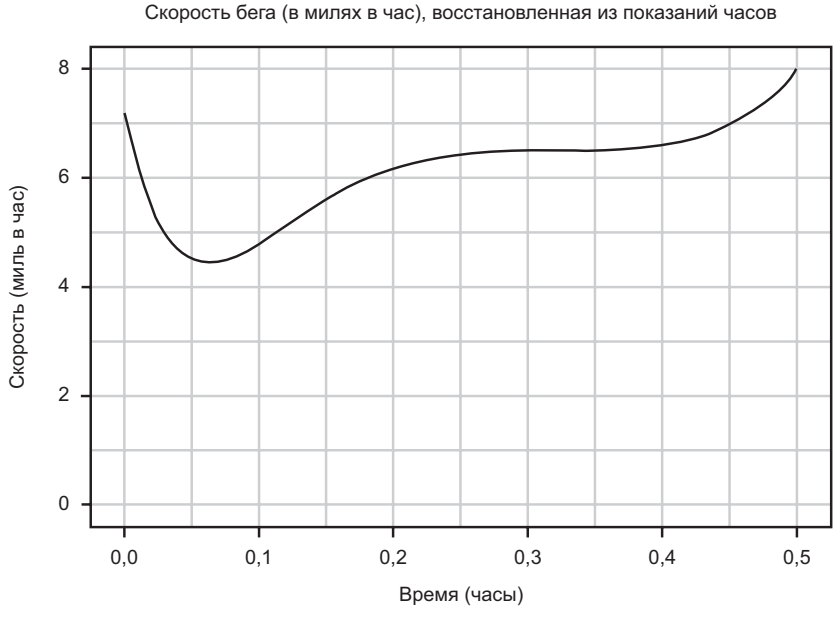


Рис. Б.1. Скорость в течение заданного времени забега

времени t. Вы можете представлять функцию s как машину, которая принимает текущее время и возвращает вашу скорость в это время. В высшей математике обычно используется конкретное определение $s(t)$, такое как $s(t) = t^2 + 3t + 2$, но здесь мы просто говорим об общих понятиях, поэтому не будем беспокоиться о точном определении s.

ПРИМЕЧАНИЕ

В книге мы будем использовать R для решения задач высшей математики, поэтому очень важно, чтобы вы понимали фундаментальные идеи, а не механически заучивали решения.

Только благодаря этой функции мы можем узнать несколько фактов. Понятно, что ваш темп был немного неравномерным во время этого пробега, поднимаясь и опускаясь с максимума около 8 миль в час в конце и минимума чуть менее 4,5 миль в час вначале.

Тем не менее есть еще много интересных вопросов, на которые вы, возможно, захотите ответить, например:

- Как далеко вы пробежали?
- Когда вы потеряли наибольшую скорость?
- Когда вы набрали наибольшую скорость?
- В какое время ваша скорость была относительно постоянной?

Мы можем сделать довольно точную оценку последнего вопроса из этого графика, но на другие, кажется, невозможно ответить, учитывая то, что мы имеем. Однако оказывается, что можно ответить на все эти вопросы с помощью высшей математики! Посмотрим, как именно.

Определение того, как далеко вы пробежали

До этого наш график показывал только скорость бега в определенное время, так как мы узнаем, как далеко вы пробежали?

Теоретически это не кажется слишком сложным. Предположим, например, что вы пробегали 5 миль в час последовательно за весь пробег. В этом случае вы пробежали 5 миль в час за 0,5 часа, поэтому общее расстояние составило 2,5 мили. Это интуитивно понятно, поскольку вы бегали со скоростью 5 миль в час, но пробежали всего полчаса, то есть вы пробежали половину пути, который пробежали бы за час.

Но наша проблема связана с разной скоростью почти в каждый момент, когда вы бежали. Давайте посмотрим на проблему по-другому. На рис. Б.2 показаны нанесенные данные для постоянной скорости движения.

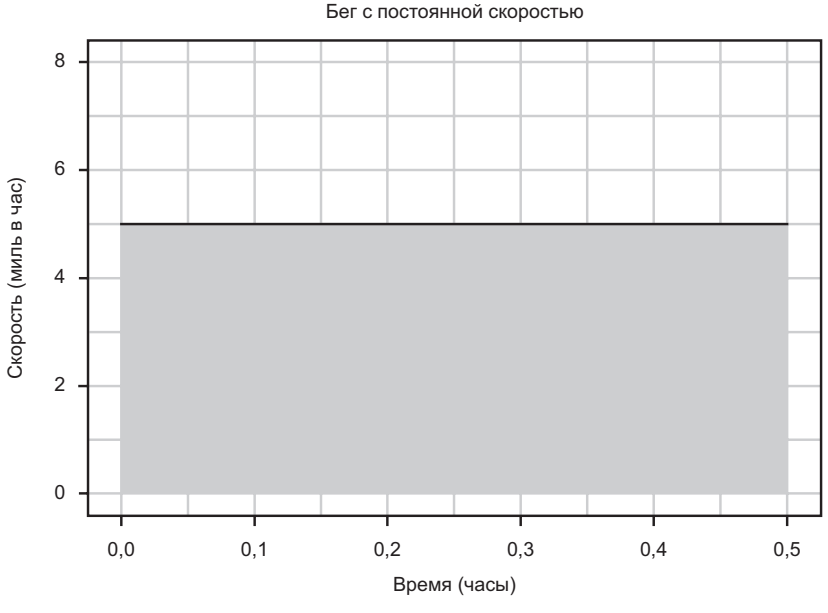


Рис. Б.2. Визуализация расстояния как области графика скорости/времени

Вы можете видеть, что эти данные создают прямую линию. Если мы проанализируем область под этой линией, то увидим, что это большой блок, который фактически отражает пройденное вами расстояние! Блок имеет высоту 5 и длину 0,5, поэтому площадь его составляет $5 \times 0{,}5 = 2{,}5$, что дает нам результат в 2,5 мили!

Теперь давайте посмотрим на упрощенную проблему с изменяющимися скоростями, когда вы бежали со скоростью 4,5 мили в час от 0,0 до 0,3 часа, со скоростью 6 миль в час от 0,3 до 0,4 часа и 3 мили в час до конца 0,5 мили. Если мы представим эти результаты в виде блоков, или *башен*, как на рис. Б.3, то сможем решить проблему таким же образом.

Первая башня составляет $4{,}5 \times 0{,}3$, вторая — $6 \times 0{,}1$, а третья — $3 \times 0{,}1$, так что:

$$4{,}5 \times 0{,}3 + 6 \times 0{,}1 + 3 \times 0{,}1 = 2{,}25.$$

Затем, посмотрев на область под башней, мы получаем общее пройденное расстояние: 2,25 мили.

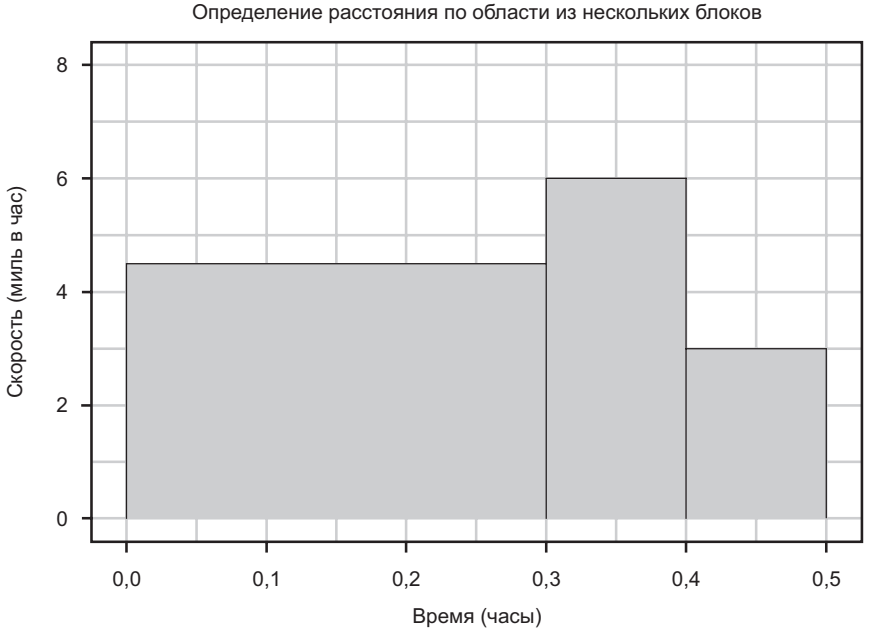


Рис. Б.3. Мы можем легко рассчитать общее пройденное расстояние, сложив эти башни

Измерение площади под кривой: интеграл

Вы уже видели, что мы можем определить область под линией, чтобы понять, как далеко вы продвинулись. К сожалению, линия для наших исходных данных изогнута, что делает проблему еще сложнее: как можно вычислить площадь башни нашей кривой линией?

Мы можем начать этот процесс, представив несколько больших башен, которые достаточно близки к нашей кривой. Если мы начнем с трех башен, как мы видим на рис. Б.4, это будет неплохой оценкой.

Рассчитав площадь под каждой из этих башен, мы получим значение 3,055 мили для приблизительного количества пройденных миль. Но можно было бы сделать лучше, добавив больше башен меньшего размера, как показано на рис. Б.5.

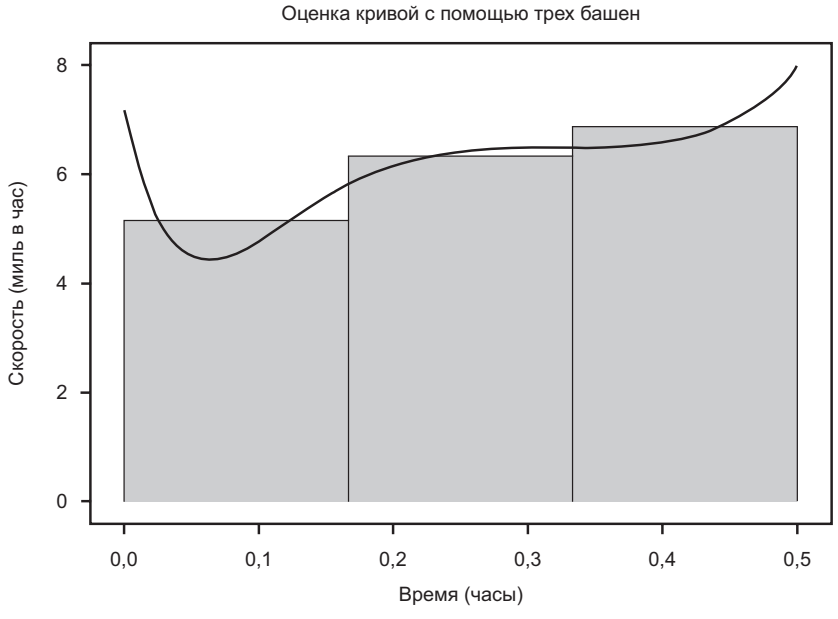


Рис. Б.4. Аппроксимация кривой тремя башнями

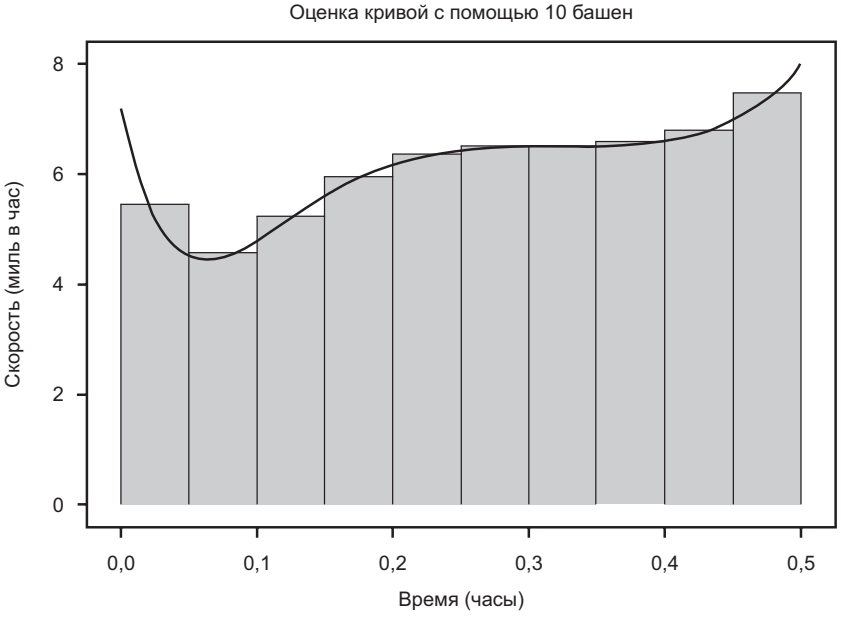


Рис. Б.5. Лучшее приближение к кривой с использованием 10 башен вместо трех

Суммируя площади этих башен, мы получаем 3,054 мили, что является более точной оценкой. Если представить, что мы повторим этот процесс бесконечно, используя более тонкие башни, в конечном итоге мы получим полную площадь под кривой, как на рис. Б.6.

Рис. Б.6. Полное получение области под кривой

Это точная площадь, пройденная за полчаса пробега. Если бы мы могли сложить бесконечно много башен, мы бы получили 3,053 мили. Наши оценки были довольно близки, и, поскольку мы используем все больше башен меньшего размера, наша оценка становится ближе к действительному результату. Сила высшей математики в том, что она позволяет нам вычислить эту точную площадь под кривой или интегралом. В высшей математике мы представили бы *интеграл* для нашей $s(t)$ от 0 до 0,5 в математической записи в виде:

$$\int_0^{0,5} s(t)\,dt,$$

где $\int$ — это просто причудливая S, означающая сумму (или общее число) площади всех маленьких башен в $s(t)$. Запись dt напоминает, что мы говорим

о маленьких кусочках переменной t; d — математический способ обращения к этим маленьким башням. Конечно, в этой части записи есть только одна переменная t, поэтому мы вряд ли запутаемся. Аналогично в этой книге мы обычно отбрасываем dt (или его эквивалент для используемой переменной), поскольку это очевидно в примерах.

В последней записи мы устанавливаем начало и конец интеграла, это означает, что мы можем найти расстояние не только для всего пробега, но и для его части. Предположим, нужно узнать, как далеко вы пробежали от 0,1 до 0,2 часа. Мы бы отметили это следующим образом:

$$\int_{0,1}^{0,2} s(t)\,dt.$$

Можно визуализировать этот интеграл, как показано на рис. Б.7.

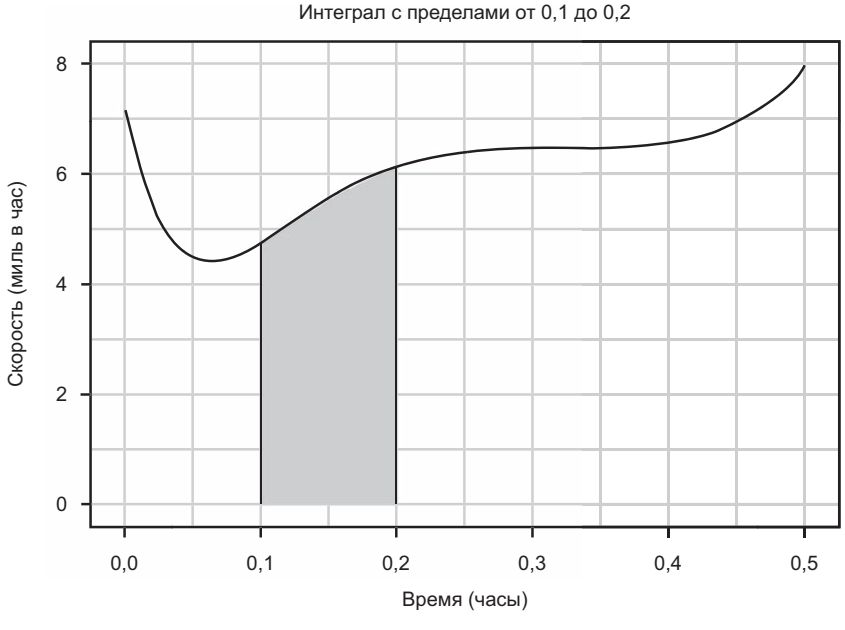


Рис. Б.7. Визуализация области под кривой от 0,1 до 0,2

Площадь только этой заштрихованной области составляет 0,556 мили.

Мы можем даже представить интеграл нашей функции как другую функцию. Предположим, мы определили новую функцию $dist(T)$, где T — это наше общее время пробега:

$$\text{dist}(T)\int_0^T s(t)\,dt.$$

Это приводит к функции, которая сообщает *расстояние*, пройденное за время *T*. Мы также можем понять, почему стоит использовать *dt* — потому что мы можем видеть, что наш интеграл применяется к аргументу в нижнем регистре *t*, а не к заглавному аргументу *T*. На рис. Б.8 показано общее расстояние, которое вы пробежали в любой момент времени T во время пробега.

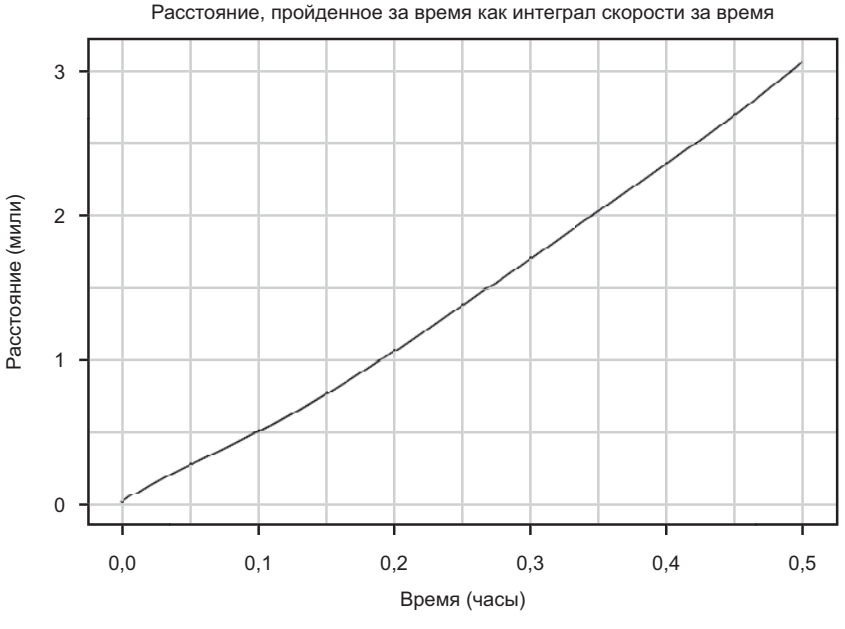


Рис. Б.8. Построение интеграла преобразует график времени и скорости в график времени и расстояния

Таким образом, интеграл преобразовал нашу функцию *s*, которая была скоростью за время, в функцию *dist*, расстояние, пройденное за время. Как показано ранее, интеграл функции между двумя точками представляет собой расстояние, пройденное между двумя разными точками на временной шкале. Теперь мы смотрим на общее расстояние, пройденное в любой данный момент времени *t* от начала времени 0. Интеграл важен, потому что он позволяет вычислять площадь под кривыми, что гораздо сложнее вычислить, чем если бы у нас были прямые линии. В этой книге мы будем использовать концепцию интеграла для определения вероятностей того, что события находятся между двумя диапазонами значений.

Измерение быстроты изменения: производная

Вы видели, как можно использовать интеграл для определения пройденного расстояния, когда все, что у нас есть, — это запись вашей скорости в разное время. Но с учетом измерения различной скорости мы также можем быть заинтересованы в определении *быстроты изменения* вашей скорости в разное время. Когда мы говорим о быстроте изменения скорости, мы имеем в виду *ускорение*. На нашем графике есть несколько интересных моментов относительно быстроты изменения: точки, когда вы теряете скорость быстрее всего, когда вы набираете скорость быстрее всего и когда скорость наиболее устойчива (то есть быстрота изменения примерно равна 0).

Как и в случае с интегрированием, основная проблема определения ускорения заключается в том, что оно, кажется, всегда меняется. Если бы у нас была постоянная быстрота изменения, вычисление ускорения не было бы таким сложным, как показано на рис. Б.9.

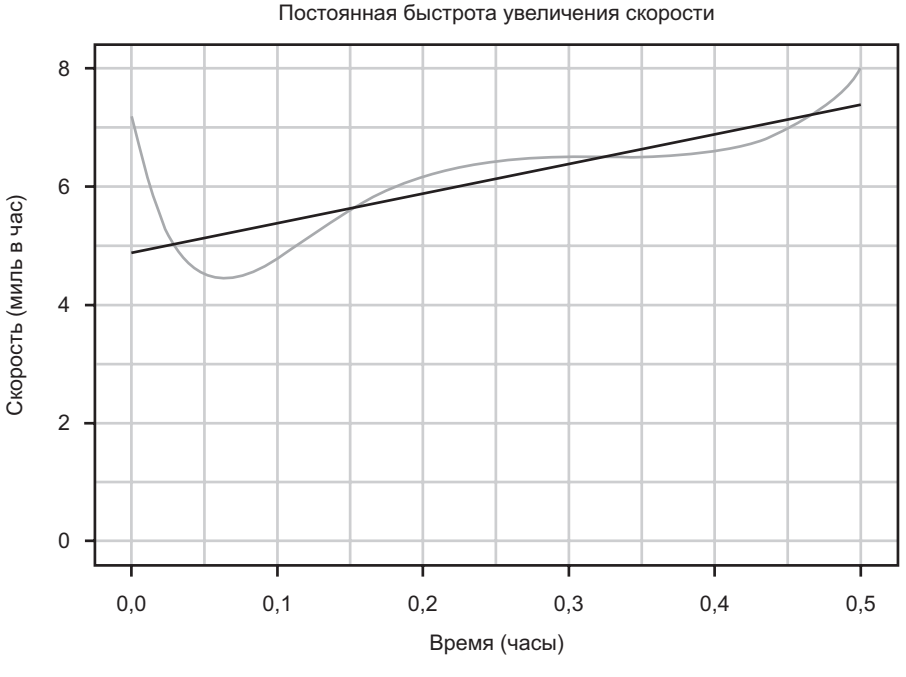


Рис. Б.9. Визуализация постоянной быстроты изменения (по сравнению с вашей фактической быстротой изменения)

Вы можете помнить из базовой алгебры, что можно нарисовать любую линию, используя эту формулу:

$$y = mx + b,$$

где b — точка, в которой линия пересекается, ось Y и m — наклон линии. *Наклон* представляет собой быстроту изменения прямой линии. Для линии на рис. Б.9 полная формула имеет вид:

$$y = 5x + 4{,}8.$$

Наклон со значением 5 означает, что для каждого раза, когда x увеличивается на 1, y увеличивается на 5; 4,8 — точка, в которой линия пересекает ось X. В этом примере мы интерпретируем эту формулу как $s(t) = 5t + 4{,}8$, это означает, что для каждой пройденной мили вы ускоряетесь на 5 миль в час и что вы начинаете отсчет с 4,8 мили в час. Поскольку вы пробежали полмили, используя эту простую формулу, мы можем выяснить:

$$s(t) = 5 \times 0{,}5 + 4{,}8 = 7{,}3,$$

что означает, что в конце пробега вы будете бежать со скоростью 7,3 мили в час. Мы могли бы точно так же определить вашу точную скорость в любой точке бега, если бы ускорение было постоянным!

Для реальных данных, поскольку линия извилистая, определить уклон в определенный момент времени нелегко. Вместо этого мы можем выяснить уклоны частей линии. Если мы разделим наши данные на три отрезка, то сможем нарисовать линии между каждой частью, как на рис. Б.10.

Теперь очевидно, что эти линии не идеально подходят к нашей кривой, но они позволяют увидеть участки, где вы ускорились быстрее всего, замедлились сильнее всего и двигались с постоянной скоростью. Если разделить нашу функцию на еще больше частей, мы получим лучшие оценки, как на рис. Б.11.

Здесь имеется картина, аналогичная той, что была при нахождении интеграла, где область под кривой была разделена на все меньшие и меньшие башни, пока не было сложено бесконечно много маленьких башен. Теперь мы хотим разбить нашу линию на бесконечно много маленьких отрезков. В конце концов, вместо одного m, представляющего уклон, мы имеем новую функцию, представляющую быстроту изменения в каждой точке нашей исходной функции. Это называется производной, представленной в математической нотации следующим образом:

$$\frac{d}{dx} f(x).$$

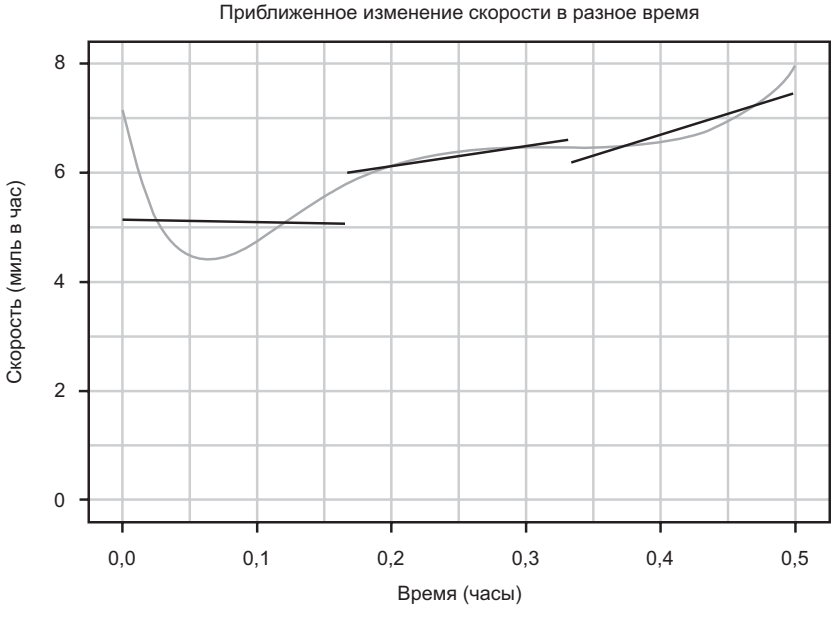


Рис. Б.10. Использование нескольких наклонов для получения более точной оценки скорости изменения

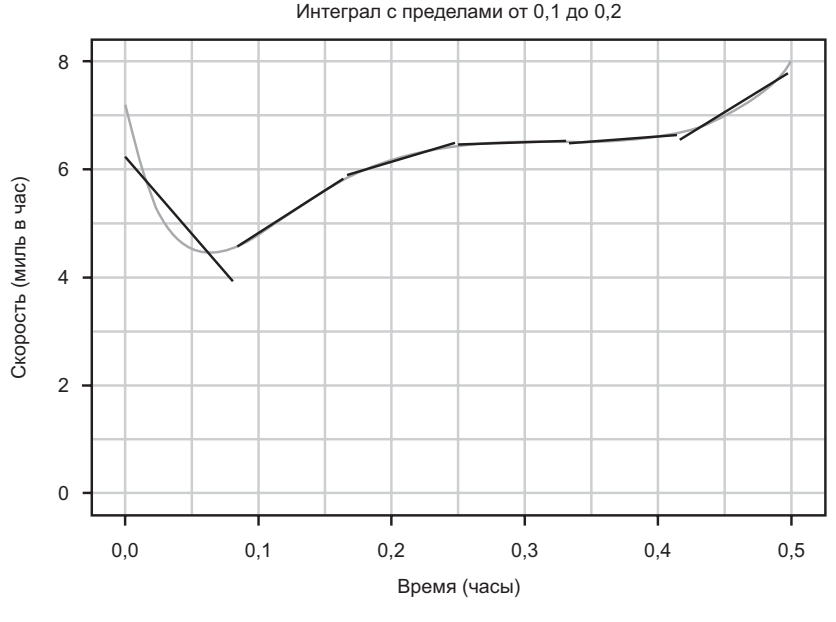


Рис. Б.11. Добавление большего количества наклонов позволяет лучше приблизиться к кривой

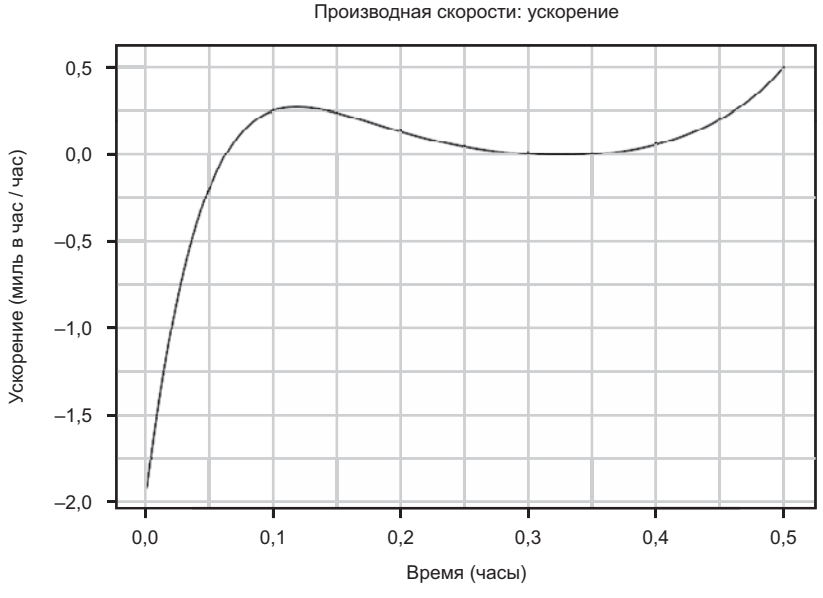


Рис. Б.12. Производная это еще одна функция, которая описывает наклон s(x) в каждой точке

Опять же *dx* просто напоминает нам, что мы рассматриваем очень маленькие части аргумента *x*. На рис. Б.12 показан график производной для функции $s(t)$, которая позволяет видеть точный темп изменения скорости в каждый момент пробега. Другими словами, это график ускорения во время пробега. Глядя на ось *Y*, вы видите, что вначале быстро потеряли скорость и примерно через 0,3 часа у вас был период ускорения 0, что означает, что темп не изменился (а это хорошо при подготовке к гонке!). Мы также можем точно видеть, когда вы набрали наибольшую скорость. Глядя на исходный график, мы не могли с легкостью определить, набираете ли вы скорость быстрее примерно через 0,1 часа (сразу после первого ускорения) или в конце пробега. С производной, однако, ясно, что последний всплеск скорости в конце действительно был быстрее, чем вначале.

Производная работает так же, как наклон прямой линии, только она указывает, насколько изогнутая линия наклонена в определенной точке.

Основная теорема анализа

Мы рассмотрим последнюю действительно замечательную концепцию высшей математики. Между интегралом и производной есть очень интересная

связь. (Доказательство этого отношения выходит далеко за рамки этой книги, поэтому мы сосредоточимся здесь только на самом отношении.) Предположим, у нас есть функция $F(x)$ с прописной буквой *F*. Что делает эту функцию особенной, так это то, что ее *производная* — $f(x)$. Например, производная нашей функции *dist* является функцией *s*; то есть изменение расстояния в каждый момент времени — это скорость. Производная скорости — ускорение. Мы можем описать это математически как:

$$\frac{d}{dx}F(x) = f(x).$$

В терминах дифференциального исчисления мы называем *F первообразной f*, потому что *f* является производной от *F*. В нашем примере первообразной ускорения будет скорость, а первообразной скорости будет расстояние. Теперь предположим, что для любого значения *f* нужно взять его интеграл от 10 до 50; то есть необходимо:

$$\int_{10}^{50} f(x)dx\,.$$

Мы можем получить это, просто вычитая $F(10)$ из $F(50)$, так что:

$$\int_{10}^{50} f(x)dx = F(50) - F(10).$$

Соотношение между интегралом и производной называется *основной теоремой анализа, или формулой Ньютона — Лейбница*. Это довольно удивительный инструмент, потому что позволяет математически решать интегралы, что зачастую намного сложнее, чем поиск производных. Используя основную теорему анализа, если мы можем найти первообразную функции, для которой нужно найти интеграл, то можем также легко выполнить интегрирование. Понимание этого — основа интегрирования вручную.

На курсе по дифференциальному исчислению обычно подробно изучают интегралы и производные. Но в этой книге мы используем высшую математику только изредка, а помогал нам в этом R. Тем не менее полезно иметь общее представление о том, что такое дифференциальное исчисление и что значат все эти загадочные символы!

В

Ответы к упражнениям

Здесь вы найдете все упражнения и ответы к ним. Для некоторых упражнений есть несколько способов решения, поэтому я дам как минимум один вариант.

Часть I. Введение в теорию вероятностей

Глава 1. Байесовские рассуждения в обычной жизни

У1. Перепишите утверждения ниже, используя математическую нотацию из этой главы:

- **вероятность дождя низкая;**
- **вероятность дождя при условии облачности высокая;**
- **вероятность, что вы с зонтом при условии дождя выше, чем просто вероятность, что вы с зонтом.**

О1.

$$P(\text{дождь}) = \text{низкая}$$
$$P(\text{дождь} \mid \text{облачно}) = \text{высокая}$$
$$P(\text{зонтик} \mid \text{дождь}) \gg P(\text{зонтик})$$

У2. Запишите, используя математические обозначения из этой главы, данные из такой истории. Придумайте гипотезу, объясняющую эти данные.

Вы приходите домой с работы и замечаете, что дверь открыта, а окно разбито. Войдя, вы видите, что вашего ноутбука нет на месте.

О2. Сначала необходимо описать наши данные с помощью переменной:

D = дверь открыта, окно разбито, ноутбук отсутствует.

Эти данные представляют три факта, которые вы наблюдали по прибытии домой. Непосредственным объяснением этих данных является то, что вас ограбили! Математически это можно выразить так:

$$H_1 = \text{вас ограбили!}$$

Теперь мы можем выразить это в виде «Вероятность увидеть все это при условии, что вас ограбили», следующим образом:

$$P(D \mid H_1).$$

У3. Дополним историю выше новыми данными. Покажите, как новая информация меняет ваши представления, и придумайте новую гипотезу для объяснения данных. Используйте обозначения из этой главы!

К вам подбегает соседский ребенок и долго извиняется, что случайно попал камнем в ваше окно. Он говорит, что заметил ноутбук и испугался, что его украдут. Открыв дверь, он унес его к себе до вашего прихода.

О3. Теперь у нас есть другая гипотеза о наблюдаемых данных:

H_2 = ребенок случайно разбил окно и взял ноутбук на хранение.

Мы можем выразить это следующим образом:

$$P(D \mid H_2) \gg P(D \mid H_1).$$

И ожидаем получить:

$$\frac{P(D \mid H_2)}{P(D \mid H_1)} = \text{большое число}.$$

Конечно, вы можете подумать, что ребенок тот еще шутник и все в округе знают его как хулигана. Это может изменить ваше мнение о том, насколько правдоподобно его объяснение, и привести к гипотезе о том, что он украл

ноутбук. В процессе чтения этой книги вы узнаете больше о том, как это можно отразить математически.

Глава 2. Измеряем неопределенность

У1. Какова вероятность бросить два шестигранных кубика и получить в сумме больше 7?

О1. Существует 36 возможных способов бросить два кубика (если считать, что единица и шестерка отличаются от шестерки и единицы). Можно выписать все на бумаге (или найти способ сделать это с помощью кода, что будет быстрее). Пятнадцать пар из этих 36 *больше* 7. Таким образом, вероятность получения значения больше 7 составляет 15/36.

У2. Какова вероятность бросить три шестигранных кубика и получить в сумме больше 7?

О2. При трех бросках существует 216 различных возможных результатов. Вы можете записать их на листе бумаги, и это допустимо, но займет довольно много времени. Вы можете понять, почему изучение основ написания кода полезно, поскольку есть масса программ (даже запутанных), которые можно написать для решения этой проблемы. Например, мы можем найти ответ с помощью этого простого набора циклов `for` в R:

```
count <- 0
for(roll1 in c(1:6)){
  for(roll2 in c(1:6)){
    for(roll3 in c(1:6)){
      count <- count + ifelse(roll1+roll2+roll3 > 7,1,0)
    }
  }
}
```

Счетчик равен 181, поэтому вероятность того, что выпадет больше 7, равна 181/216. Но как уже отмечалось, существует множество способов вычислить это значение. Альтернативой является эта единственная (трудно читаемая!) строка R-кода, которая выполняет то же самое, что и цикл `for`:

```
sum(apply(expand.grid(c(1:6),c(1:6),c(1:6)),1,sum) > 7)
```

При обучении программированию вы должны сосредоточиться на получении правильного ответа, а не на использовании определенного подхода для его получения.

У3. Играют команды «Янки» и «Ред Сокс». Вы — преданный фанат «Соксов» и заключаете с другом пари на их выигрыш. Если «Соксы» проиграют — вы платите другу 30 долларов, если выиграют — друг платит вам 5 долларов. Какую вероятность вы присвоили гипотезе, что «Ред Сокс» выиграет?

О3. Мы видим, что предполагаемые шансы победы «Ред Сокс» составляют:

$$O(\text{«Ред Сокс» выиграет}) = \frac{30}{5} = 6.$$

Вспоминая нашу формулу для преобразования шансов в вероятности, мы можем преобразовать шансы в вероятность того, что выиграет «Ред Сокс»:

$$P(\text{«Ред Сокс» выиграет}) = \frac{O(\text{«Ред Сокс» выиграет})}{1 + O(\text{«Ред Сокс» выиграет})} = \frac{6}{7}.$$

Итак, основываясь на сделанной ставке, вы можете сказать, что вероятность того, что «Ред Сокс» выиграет, составляет около 86 %.

Глава 3. Логика неопределенности

У1. Какова вероятность выкинуть 20 на 20-гранной игральной кости три раза подряд?

О1. Вероятность выпадения 20 составляет 1/20, и чтобы определить вероятность выпадения трех 20 подряд, мы должны использовать правило произведения вероятностей:

$$P(\text{три раза по } 20) = \frac{1}{20} \times \frac{1}{20} \times \frac{1}{20} = \frac{1}{8000}.$$

У2. Прогноз погоды сообщает, что завтра с 10 %-ной вероятностью пойдет дождь. Вы забываете зонтик дома в половине случаев. Какова вероятность, что завтра вы окажетесь под дождем без зонта?

О2. Опять же мы можем использовать правило произведения вероятностей, чтобы решить эту проблему. Мы знаем, что $P(\text{дождя}) = 0{,}1$ и $P(\text{забывания зонта}) = 0{,}5$, поэтому:

$$P(\text{дождя, забывания зонта}) = P(\text{дождя}) \times P(\text{забывания зонта}) = 0{,}05.$$

Как видите, вероятность того, что вы окажетесь под дождем без зонта, составляет всего 5 %.

У3. Сырые яйца с вероятностью 1/20 000 заражены сальмонеллой. Вы съели два сырых яйца, какова вероятность, что вы съели яйцо с сальмонеллой?

О3. Для решения этой задачи нужно использовать правило суммы вероятностей, потому что если *какое-либо* яйцо содержит сальмонеллу, вы заболеете:

$$P(\text{яйцо}_1) + P(\text{яйцо}_2) - P(\text{яйцо}_1) \times P(\text{яйцо}_2) =$$

$$= \frac{1}{20\,000} + \frac{1}{20\,000} - \frac{1}{20\,000} \times \frac{1}{20\,000} = \frac{39\,999}{400\,000\,000}$$

...что составляет всего лишь менее 1/10 000.

У4. Какова вероятность выкинуть два орла за два броска монеты или три шестерки за три броска шестигранного кубика?

О4. Для этого упражнения нам нужно объединить правило произведения и правило суммы. Сначала давайте посчитаем P(двух орлов) и P(трех шестерок) отдельно. Для каждой вероятности используется правило произведения:

$$P(\text{два орла}) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

$$P(\text{три шестерки}) = \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} = \frac{1}{216}$$

Теперь нужно использовать правило суммы, чтобы вычислить вероятность того, что произойдет одно из этих событий, P(двух орлов или трех шестерок):

$$P(\text{два орла}) + P(\text{три шестерки}) - P(\text{два орла}) \times P(\text{три шестерки}) =$$

$$= \frac{1}{4} + \frac{1}{216} - \frac{1}{4} \times \frac{1}{216} = \frac{73}{288},$$

...что составляет чуть больше 25 %.

Глава 4. Как получить биноминальное распределение

У1. Каковы параметры биномиального распределения для вероятности выкинуть 1 или 20 на 20-гранной кости, если бросить кость 12 раз?

О1. Мы рассматриваем событие, происходящее 1 раз из 12 попыток, поэтому $n = 12$ и $k = 1$. Мы имеем 20 сторон, и важны две из них, поэтому $p = 2/20 = 1/10$.

У2. В колоде из 52 карт 4 туза. Вы вытягиваете карту, возвращаете ее обратно, тасуете колоду и снова вытягиваете карту. Сколькими способами можно вытянуть только одного туза за пять попыток?

О2. Для этого нам даже не нужна комбинаторика. Если представить, что A означает «туз», а x — что-то еще, остается пять возможных случаев:

- Axxxx
- xAxxx
- xxAxx
- xxxAx
- xxxxA

Можно было бы просто назвать это $\binom{5}{1}$ или, как в R, `choose(5,1)`. В любом случае ответ — 5.

У3. Продолжая предыдущую задачу: какова вероятность вытянуть пять тузов за десять попыток (помните, что карта возвращается в колоду)?

О3. Это то же самое, что и $B(5; 10, 1/13)$. Как и ожидалось, вероятность этого крайне мала: около 1/32 000.

У4. При поиске новой работы полезно иметь больше одного предложения — это открывает возможность поторговаться. Пусть вероятность получить после собеседования предложение о работе равна 1/5, и за месяц вы проходите семь собеседований. Какова вероятность, что к концу месяца вы получите хотя бы два предложения?

О4. Мы можем использовать следующий код R для вычисления ответа:

```
> pbinom(1,7,1/5,lower.tail = FALSE)
0.4232832
```

Как видите, вероятность получения двух *или более* предложений о работе составляет около 42 %, если вы были на интервью в семи компаниях.

У5. Вы получили немало писем от рекрутеров и обнаружили, что в следующем месяце у вас 25 собеседований. Ох, это утомительно, а вероятность получить предложение о работе, когда проходишь собеседование

усталым, падает до 1/10. Вы готовы пройти 25 собеседований, только если это в два раза повысит вероятность получить хотя бы два предложения. Надо ли проходить 25 собеседований или остановиться на 7?

О5. Давайте напишем еще немного кода на R, чтобы разобраться в этом:

```
p.two.or.more.7 <- pbinom(1,7,1/5,lower.tail = FALSE)
p.two.or.more.25 <- pbinom(1,25,1/10,lower.tail = FALSE)
```

Даже с уменьшенной вероятностью предложения ваша вероятность получить по крайней мере два предложения за 25 собеседований составляет 73 %. Однако у вас получится это сделать только при наличии *вдвое* больше шансов.

Как мы видим в R:

```
> p.two.or.more.25/p.two.or.more.7
[1] 1.721765
```

...у вас всего в 1,72 раза больше шансов получить два или более предложений, так что хлопоты того не стоят.

Глава 5. Бета-распределение

У1. Вы хотите использовать бета-распределение, чтобы определить, честная ли монетка — то есть равны ли для нее вероятности выкинуть орел и решку. Вы подбрасываете монетку 10 раз и получаете 4 орла и 6 решек. Используя бета-распределение, найдите вероятность того, что орел выпадает в более чем 60 % бросков.

О1. Мы бы смоделировали это как Beta(4,6). Нужно вычислить интеграл от 0,6 до 1, что мы можем сделать в R следующим образом:

```
integrate(function(x) dbeta(x,4,6),0.6,1)
```

Это говорит о 10 %-ной вероятности того, что истинная вероятность выпадения орла составляет 60 % или больше.

У2. Вы еще 10 раз подбрасываете монетку и в итоге получаете 9 орлов и 11 решек. Какова вероятность того, что монетка честная, при использовании нашего определения честности плюс-минус 5 %?

О2. Теперь мы имеем распределение Beta(9,11). Но нам нужно знать вероятность того, что монетка честная, то есть шанс выпадения орла составляет

0,5 в любом случае с вероятностью 0,05. Это означает, что нужно интегрировать новое распределение между 0,45 и 0,55. Мы можем сделать это с помощью такой строки кода в R:

```
integrate(function(x) dbeta(x,9,11),0.45,0.55)
```

Теперь мы видим, что существует 30 %-ная вероятность того, что монетка честная, с учетом новых имеющихся данных.

У3. Данные — лучший способ убедиться в верности своих утверждений. Вы еще 200 раз подбрасываете монетку и в итоге получаете 109 орлов и 111 решек. Какова теперь вероятность того, что монетка честная (плюс-минус 5 %)?

О3. Учитывая предыдущее упражнение, ответ довольно прост:

```
integrate(function(x) dbeta(x,109,111),0.45,0.55)
```

Теперь мы на 86 % уверены, что монетка достаточно честная. Обратите внимание, что ключом к большей уверенности было добавление большего количества данных.

Часть II. Байесовские и априорные вероятности

Глава 6. Условная вероятность

У1. Мы хотим использовать теорему Байеса для определения вероятности того, что в 2010 году пациент с синдромом Гийена — Барре был привит от гриппа. Какая информация нам нужна?

О1. Мы хотим выяснить $P(\text{вакцины против гриппа} \mid \text{синдрома Гийена — Барре})$. Можно решить эту задачу, используя теорему Байеса, при условии, что вся эта информация у нас имеется:

$$P(\text{грипп} \mid \text{ синдром Гийена} - \text{Барре}) =$$
$$= \frac{P(\text{грипп}) \times P(\text{синдром Гийена} - \text{Барре} \mid \text{грипп})}{P(\text{синдром Гийена} - \text{Барре})}.$$

Из этих сведений мы не знаем только вероятность получения вакцины от гриппа в первую очередь. Вероятно, мы могли бы получить эту информацию

из Центров по контролю и профилактике заболеваний или другой национальной службы сбора данных.

У2. Какова вероятность того, что случайно выбранный из всей популяции человек — женщина и не дальтоник?

О2. Мы знаем, что $P(\text{женщины}) = 0{,}5$ и что $P(\text{дальтоника} \mid \text{женщины}) = 0{,}005$, но мы хотим знать вероятность того, что кто-то является женщиной и *не* дальтоником, что равно $1 - P(\text{дальтоника} \mid \text{женщины}) = 0{,}995$. Итак:

$$P(\text{женщина, не дальтоник}) = P(\text{женщина}) \times P(\text{не дальтоник} \mid \text{женщина}) = 0{,}5 \times 0{,}995 = 0{,}4975.$$

У3. Какова вероятность того, что мужчина, привитый от гриппа в 2010 году, будет страдать либо от дальтонизма, либо от синдрома Гийена — Барре?

О3. Эта задача поначалу может показаться сложной, но мы можем ее немного упростить. Давайте начнем с того, что просто поработаем над вероятностью получения дальтонизма, если выбранный человек — мужчина, и вероятностью наличия синдрома Гийена — Барре, если он был привит от гриппа. Обратите внимание, что мы идем по сокращенному пути, поскольку наличие мужского пола не зависит от наличия синдрома Гийена — Барре (насколько нам известно), а прививка от гриппа не влияет на дальтонизм. Мы сделаем каждую из них отдельной вероятностью:

$$P(A) = P(\text{дальтоник} \mid \text{женщина}).$$

$$P(B) = P(\text{синдром Гийена — Барре} \mid \text{грипп}).$$

К счастью, мы уже проделали всю эту работу ранее в этой главе, поэтому мы знаем, что $P(A) = 4/1000$ и $P(B) = 3/100\,000$. Теперь можно просто использовать правило суммы вероятностей, чтобы решить эту задачу:

$$P(A \text{ или } B) = P(A) + P(B) - P(A) \times P(B \mid A).$$

И поскольку вероятность наличия дальтонизма, насколько нам известно, не имеет ничего общего с вероятностью наличия синдрома Гийена — Барре, мы знаем, что $P(B \mid A) = P(B)$. Подставляя числа, мы получаем ответ: 100 747/25 000 000, или 0,00403. Это немного больше, чем шанс быть дальтоником и мужчиной, потому что вероятность наличия синдрома Гийена — Барре очень мала.

Глава 7. Теорема Байеса и Lego

У1. Канзас-Сити, вопреки названию, стоит на границе двух штатов, Миссури и Канзаса. Агломерация Канзас-Сити состоит из 15 округов: 9 в штате Миссури и 6 в Канзасе. В штате Канзас всего 105 округов, в Миссури — 114. Используя теорему Байеса, вычислите вероятность, что человек, переехавший в агломерацию Канзас-Сити, окажется в штате Канзас. Используйте P (Канзас), P (Канзас-Сити) и P (Канзас-Сити | Канзас).

О1. Надеюсь, очевидно, что в районе Канзас-Сити 15 округов, и 6 из них находятся в Канзасе, поэтому вероятность оказаться в Канзасе, учитывая, что вы знаете, что кто-то живет в районе Канзас-Сити, должна составлять 6/15, что эквивалентно 2/5. Однако цель этого упражнения — не просто получить ответ, а показать, что теорема Байеса предоставляет инструменты для его решения. При работе над более сложными задачами будет очень полезно укрепить доверие к теореме Байеса.

Итак, чтобы вычислить $P(\text{Канзаса} \mid \text{Канзас-Сити})$, мы можем использовать теорему Байеса следующим образом:

$$P(\text{Канзас} \mid \text{Канзас-Сити}) = \frac{P(\text{Канзас-Сити} \mid \text{Канзас}) \times P(\text{Канзас})}{P(\text{Канзас-Сити})}.$$

Учитывая имеющиеся данные, мы знаем, что из 105 округов Канзаса 6 находятся в районе Канзас-Сити:

$$P(\text{Канзас-Сити} \mid \text{Канзас}) = \frac{6}{105}.$$

И между Миссури и Канзасом находится 219 округов, 105 из которых находятся в Канзасе:

$$P(\text{Канзас}) = \frac{105}{219}.$$

И из этого общего количества 219 округов 15 находятся в районе Канзас-Сити:

$$P(\text{Канзас-Сити}) = \frac{15}{219}.$$

Таким образом, заполнение всех частей теоремы Байеса дает результат:

$$P(\text{Канзас} \mid \text{Канзас-Сити}) = \frac{\frac{6}{105} \times \frac{105}{219}}{\frac{15}{219}} = \frac{2}{5}.$$

У2. В колоде 52 красные и черные карты, в том числе четыре туза: два красных и два черных. Вы вынули из колоды черный туз и перемешали ее. Ваш друг вытянул карту черной масти. Какова вероятность, что это туз?

О2. Как и в предыдущем вопросе, мы можем легко увидеть, что есть 26 черных карт и 2 из них — тузы, и вероятность вытащить туза составляет 2/26, или 1/13, если мы вытянули черную карту. Но, опять же, мы хотим укрепить некоторое доверие к теореме Байеса и не использовать так много математических умственных сокращений. Используя теорему Байеса, мы получаем:

$$P(\text{туз} \mid \text{черная масть}) = \frac{P(\text{черная масть} \mid \text{туз}) \times P(\text{туз})}{P(\text{черная масть})}.$$

В колоде 26 черных карт из имеющейся 51 карты, так как мы убрали 1 красного туза. Если мы знаем, что вытянули туз, вероятность того, что он черный, равна:

$$P(\text{черная масть} \mid \text{туз}) = \frac{2}{3}.$$

В колоде теперь 51 карта, из которых только 3 туза, поэтому мы имеем:

$$P(\text{туз}) = \frac{3}{51}.$$

Наконец, мы знаем, что из оставшейся 51 карты 26 черных, так что:

$$P(\text{черная масть}) = \frac{26}{51}.$$

Теперь у нас есть достаточно информации, чтобы решить задачу:

$$P(\text{туз} \mid \text{черная масть}) = \frac{\frac{2}{3} \times \frac{3}{51}}{\frac{26}{51}} = \frac{1}{13}.$$

Глава 8. Априорная и апостериорная вероятности и правдоподобие в теореме Байеса

У1. Как уже говорилось, вы можете не согласиться с нашей оценкой правдоподобия для первой гипотезы.

Как это повлияет на меру нашей убежденности в превосходстве H_1 над H_2?

$$P(\text{разбитое окно, открытая входная дверь, пропавший ноутбук} \mid \text{ограбление}) = \\ = \frac{3}{10}.$$

О1. Для начала запомните следующее:

$$P(\text{разбитое окно, открытая дверь, пропавший ноутбук} \mid \text{ограбление}) = \\ = P(D \mid H_1).$$

Чтобы увидеть, как это меняет убеждения, все, что нам нужно сделать, это заменить эту часть в соотношении:

$$\frac{P(H_1) \times P(D \mid H_1)}{P(H_2) \times P(D \mid H_2)}.$$

Мы уже знаем, что знаменатель формулы равен 1/21 900 000, а $P(H_1) = 1/1000$, поэтому, чтобы получить ответ, нам просто нужно добавить измененное убеждение в $P(D \mid H_1)$:

$$\frac{\frac{1}{1000} \times \frac{3}{100}}{\frac{1}{21\,900\,000}} = 657.$$

Итак, когда мы считаем, что $(D \mid H_1)$ в 10 раз меньше вероятности, соотношение будет в 10 раз меньше (хотя все еще в значительной степени будет говорить в пользу H_1).

У2. Насколько малой должна быть априорная вероятность ограбления, чтобы гипотезы H_1 и H_2 при имеющихся данных были равновероятны?

О2. В предыдущем ответе уменьшение вероятности $P(D \mid H_1)$ в 10 раз уменьшило соотношение. На этот раз мы хотим изменить $P(H_1)$ так, чтобы

соотношение было равно 1, это означает, что нужно сделать его в 657 раз меньше:

$$\frac{\frac{1}{1000\times 657}\times\frac{3}{100}}{\frac{1}{21\,900\,000}}=1.$$

Итак, новая вероятность $P(H_1)$ должна быть равна 1/657 000, а это является довольно сильным убеждением в маловероятности того, что вас ограбили!

Глава 9. Байесовские априорные вероятности и распределение вероятностей

У1. Друг находит монетку на земле, подбрасывает ее и получает шесть орлов подряд, а затем одну решку. Найдите бета-распределение, которое описывает этот случай. Используйте интегрирование, чтобы определить вероятность того, что истинная вероятность выпадения орла находится в диапазоне от 0,4 до 0,6. Это значит, что монетка является относительно честной.

О1. Мы можем представить это как бета-распределение с $\alpha = 6$ и $\beta = 1$, поскольку мы получили шесть орлов и одну решку. В R можно интегрировать это следующим образом:

```
> integrate(function(x) dbeta(x,6,1),0.4,0.6)
0.04256 with absolute error < 4.7e-16
```

С вероятностью около 4 % эта монетка честная, и исходя только из правдоподобности, мы сочли бы ее нечестной.

У2. Придумайте априорную вероятность того, что монетка честная. Используйте бета-распределение таким образом, чтобы с вероятностью не менее 95 % истинная вероятность выпадения орла составляла от 0,4 до 0,6.

О2. Любое $\alpha_{\text{априорное}} = \beta_{\text{априорное}}$ даст нам честную априорную вероятность, и чем больше эти значения, тем сильнее априорная вероятность. Например, при использовании 10 мы получим:

```
> prior.val <- 10
> integrate(function(x) dbeta(x,6+prior.val,1+prior.val),0.4,0.6)
0.4996537 with absolute error < 5.5e-15
```

Но, конечно, вероятность того, что монетка честная, составляет всего 50 %. Методом проб и ошибок мы можем найти число, которое нам подходит. Используя $\alpha_{априорное} = \beta_{априорное} = 55$, мы находим, что это дает априорную вероятность, которая соответствует цели:

```
> prior.val <- 55
> integrate(function(x) dbeta(x,6+prior.val,1+prior.val),0.4,0.6)
0.9527469 with absolute error < 1.5e-11
```

У3. Теперь посмотрите, сколько еще орлов (без решек) потребуется, чтобы убедить вас в существовании реальной вероятности того, что монетка нечестная. В этом случае наша вера в то, что вероятность монетки составляет от 0,4 до 0,6, падает ниже 0,5.

О3. Опять же, мы можем решить эту проблему просто методом проб и ошибок, пока не получим работающий ответ. Помните, что мы все еще используем Beta(55,55) в качестве априорной вероятности. На этот раз мы хотим увидеть, сколько можно добавить к α, чтобы повысить вероятность получения честной монетки примерно до 50 %. Мы видим, что с еще пятью орлами наша апостериорная вероятность опускается до 90 %.

```
> more.heads <- 5
> integrate(function(x) dbeta(x,6+prior.val+more.heads,1+
            prior.val),0.4,0.6)
0.9046876 with absolute error < 3.2e-11
```

И если бы выпало еще 23 орла, мы бы обнаружили, что вероятность того, что монетка будет честной сейчас, составит около 50 %. Это показывает, что даже сильную априорную вероятность можно преодолеть с помощью большего количества данных.

Часть III. Оценка параметров

Глава 10. Введение в усреднение и оценку параметров

У1. Можно получить ошибки, не в полной мере взаимоисключающие. По шкале Фаренгейта 98,6 градуса — это нормальная температура тела, а 100,4 градуса — типичный порог лихорадки. Скажем, вы ухаживаете за ребенком, которому жарко и который кажется больным, но все

повторные показания термометра находятся между 99,5 и 100,0 градуса: высоковато, но не совсем лихорадка. Вы ставите термометр самому себе и получаете несколько показаний между 97,5 и 98. Что может быть не так с термометром?

О1. Похоже, что термометр может давать *смещенные* измерения, которые, как правило, отклоняются на 1 градус по Фаренгейту. Если добавить 1 градус к результатам, мы увидим, что они располагаются между 98,5 и 99, что кажется правильным для тех, кто обычно имеет температуру тела, равную 98,6 градуса по Фаренгейту.

У2. Учитывая, что вы чувствуете себя здоровым и у вас всегда стабильная нормальная температура, как можно изменить измерения 100, 99,5, 99,6 и 100,2, чтобы оценить, есть ли у ребенка температура?

О2. Если измерения смещены, это означает, что они систематически ошибочны, поэтому никакая выборка не исправит это сама по себе. Чтобы исправить исходные измерения, мы могли бы просто добавить 1 градус к каждому.

Глава 11. Измерение разброса данных

У1. Одним из преимуществ расхождения является то, что возведение в квадрат различий делает штрафы экспоненциальными. Приведите несколько примеров, когда это будет полезно.

О1. Экспоненциальные штрафы очень полезны во многих повседневных ситуациях. Одна из наиболее очевидных — физическое расстояние. Предположим, кто-то изобрел телепорт, который может перенести вас в другое место. Если вы промахнетесь на метр, ничего страшного; 5 километров тоже можно простить; но смещение на 50 километров может быть невероятно опасным. В этом случае нужно, чтобы штраф за удаленность от цели становился намного более суровым по мере ее увеличения.

У2. Рассчитайте среднее значение, расхождение и стандартное отклонение для следующих значений: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.

О2. Среднее значение = 5,5, расхождение = 8,25, стандартное отклонение = 2,87.

Глава 12. Нормальное распределение

О СТАНДАРТНОМ ОТКЛОНЕНИИ

R имеет встроенную функцию sd, которая вычисляет *выборочное* стандартное отклонение, отличающееся от стандартного отклонения, обсуждаемого в этой книге. Идея выборочного стандартного отклонения заключается в том, что вместо n усредняется $n - 1$. Выборочное стандартное отклонение используется в классической статистике для оценки математического ожидания при заданных значениях. В примере функция my.sd вычисляет стандартное отклонение, которое мы использовали в книге:

```
my.sd <- function(val){
  val.mean <- mean(val)
  sqrt(mean((val.mean-val)^2))
}
```

По мере роста количества данных разница между выборочным стандартным отклонением и настоящим стандартным отклонением становится несущественной. Однако при небольшом количестве данных из примеров в этой книге можно ощутить небольшое различие. Для всех примеров из главы 12 используется my.sd, но иногда, удобства ради, я буду использовать sd по умолчанию.

У1. Какова вероятность наблюдения значения на пять сигм большего или меньшего, чем среднее значение?

О1. Мы можем использовать integrate() для нормального распределения со средним значением 0 и стандартным отклонением 1. Затем мы просто интегрируем от 5 до некоторого достаточно большого числа, например 100:

```
> integrate(function(x) dnorm(x,mean=0,sd=1),5,100)
2.88167e-07 with absolute error < 5.6e-07
```

У2. Лихорадка — это любая температура выше 100,4 градуса по шкале Фаренгейта. Учитывая следующие измерения, какова вероятность того, что у пациента жар?

100,0, 99,8, 101,0, 100,5, 99,7

О2. Начнем с определения среднего и стандартного отклонения данных:

```
temp.data <- c(100.0, 99.8, 101.0, 100.5, 99.7)
temp.mean <- mean(temp.data)
temp.sd <- my.sd(temp.data)
```

Затем мы просто используем `integrate()`, чтобы определить вероятность того, что температура превышает 100,4:

```
> integrate(function(x) dnorm(x,mean=temp.mean,sd=temp.sd),100.4,200)
0.3402821 with absolute error < 1.1e-08
```

При таких измерениях вероятность повышения температуры составляет около 34 %.

У3. Предположим, что в главе 11 мы попытались измерить глубину колодца по времени падения монет и получили следующие значения:

2,5, 3, 3,5, 4, 2

Расстояние, на которое падает объект, может быть рассчитано (в метрах) по следующей формуле:

$$\text{расстояние} = 1/2 \times G \times \text{время}^2,$$

где G составляет 9,8 м^2/с. Какова вероятность того, что глубина колодца превышает 500 метров?

О3. Начнем с помещения наших данных о времени в R:

```
time.data <- c(2.5,3,3.5,4,2)
time.data.mean <- mean(time.data)
time.data.sd <- my.sd(time.data)
```

Далее нужно выяснить, сколько времени необходимо, чтобы достичь 500 метров. Вычислим:

$$\frac{1}{2} \times G \times t^2 = 500.$$

Если G равно 9,8, мы можем вычислить, что время (t) составляет около 10,10 секунды (также можно рассчитать это, создав функцию в R и вручную выполняя итерацию, или поискать решение на чем-то вроде Wolfram Alpha). Теперь нужно просто интегрировать нормальное распределение за пределами 10.1:

```
> integrate(function(x)
dnorm(x,mean=time.data.mean,sd=time.data.sd),10.1,200)
2.056582e-24 with absolute error < 4.1e-24
```

Это дает практически нулевую вероятность, поэтому мы можем быть уверены, что глубина скважины *не равна* 500 метрам.

У4. Какова вероятность того, что колодца нет (то есть колодец имеет фактическую глубину 0 метров)? Вы заметите, что вероятность выше, чем можно было бы ожидать, учитывая наблюдения, что колодец есть. Есть два хороших объяснения того, что эта вероятность выше, чем должна быть. Во-первых, нормальное распределение является плохой моделью для измерений; во-вторых, при составлении чисел для примера я выбрал значения, которые вы вряд ли увидите в реальной жизни. Что для вас более вероятно?

А4. Если мы проделаем то же самое интегрирование, но с –1 до 0, то получим:

```
> integrate(function(x)
dnorm(x,mean=time.data.mean,sd=time.data.sd),-1,0)
1.103754e-05 with absolute error < 1.2e-19
```

Это мало, но вероятность того, что *колодца нет*, превышает 1/100 000. Но ведь мы видим колодец! Он прямо перед нами! Таким образом, даже если вероятность мала, она не так уж близка к нулю. Стоит подвергать сомнению модель или данные? Как байесовец вы должны ставить под сомнение модель, но не данные. Например, движение цен на акции обычно сопровождается очень высокими событиями сигма во время финансовых кризисов. Это означает, что нормальное распределение — плохая модель динамики запасов. Однако в этом примере нет причин подвергать сомнению предположения о нормальном распределении, и на самом деле проблема в исходных числах, которые я выбрал для предыдущей главы, пока мой научный редактор не указала, что значения кажутся слишком разбросанными.

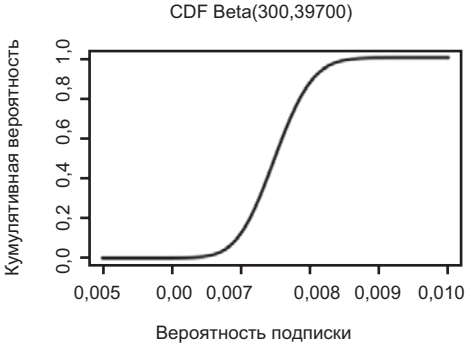
Одно из величайших достоинств статистического анализа — скептицизм. На практике я несколько раз получал плохие данные для работы. Несмотря на то что модели всегда несовершенны, очень важно убедиться, что можно доверять данным. Посмотрите, верны ли ваши предположения о мире, а если нет, то убедитесь, что вы все еще доверяете и модели, и данным.

Глава 13. Инструменты оценки параметров: PDF, CDF и квантильная функция

У1. Используя пример кода для построения PDF на с. 155, постройте функции CDF и квантильную.

О1. Взяв код из главы, просто замените `dbeta()` на `pbeta()` для CDF следующим образом:

```
xs <- seq(0.005,0.01,by=0.00001)
plot(xs,pbeta(xs,300,40000-300),type='l',lwd=3,
     ylab="Cumulative probability",
     xlab="Probability of subscription",
     main="CDF Beta(300,39700)")
```



Для квантильной функции нужно заменить `xs` на фактические квантили:

```
xs <- seq(0.001,0.99,by=0.001)
plot(xs,qbeta(xs,300,40000-300),type='l',lwd=3,
     ylab="Probability of subscription",
     xlab="Quantile",
     main="Quantile of Beta(300,39700)")
```

У2. Возвращаясь к задаче измерения снегопада из главы 10, скажем, у вас есть следующие измерения (в дюймах) снежного покрова:

7,8, 9,4, 10,0, 7,9, 9,4, 7,0, 7,0, 7,1, 8,9, 7,4

Каков 99,9 %-ный доверительный интервал для истинного значения снежного покрова?

О2. Сначала вычислим среднее значение и стандартное отклонение для этих данных:

```
snow.data <- c(7.8, 9.4, 10.0, 7.9, 9.4, 7.0, 7.0, 7.1, 8.9, 7.4)
snow.mean <- mean(snow.data)
snow.sd <- sd(snow.data)
```

Затем используем `qnorm()` для вычисления верхней и нижней границ доверительного интервала 99,9 %.

- Нижний — `qnorm(0,0005,mean=snow.mean,sd=snow.sd)` = 4,46.
- Верхний — `qnorm(0,9995,mean=snow.mean,sd = snow.sd)` = 11,92.

Это означает, что мы очень уверены в том, что выпадает не менее 4,46 дюйма снега и не более 11,92 дюйма.

У3. Девочка продает конфеты. Пока она посетила 30 домов и продала 10 конфет. Сегодня она посетит еще 40 домов. Каков 95 %-ный доверительный интервал для проданных за остаток дня конфет?

О3. Сначала нужно рассчитать 95 % доверительный интервал для вероятности продажи конфеты. Мы можем смоделировать это как Beta(10,20), а затем использовать `qbeta()`, чтобы вычислить значения:

- Нижнее значение — `qbeta(0,025,10,20)` = 0,18.
- Верхнее значение — `qbeta(0,975,10,20)` = 0,51.

Учитывая, что осталось посетить 40 домов, мы можем ожидать, что она продаст от 40 × 0,18 = 7,2 до 40 × 0,51 = 20,4 конфеты. Конечно, девочка может продать только целые конфеты, поэтому мы уверены, что она продаст от 7 до 20 конфет.

Если нужна конкретика, мы могли бы фактически вычислить квантильную функцию для биномиального распределения на каждом экстремуме ее уровня продаж с помощью `qbinom()`! Я оставлю *это* в качестве упражнения, которое вы можете изучить самостоятельно.

Глава 14. Оценка параметров с априорными вероятностями

У1. Предположим, вы играете в аэрохоккей с друзьями и подбрасываете монетку, чтобы узнать, кто будет подавать шайбу. Проиграв 12 раз, вы понимаете, что друг, который приносит монету, почти всегда идет первым: 9 из 12 раз. Некоторые из ваших друзей начинают что-то подозревать. Определите априорное распределение вероятностей для следующих убеждений:

- **убеждения человека, который слабо верит, что друг обманывает, и реальная скорость выпадения орла ближе к 70 %;**
- **убеждения человека, который очень сильно верит, что монетка честная, и дает 50 %-ную вероятность выпадения орла;**
- **убеждения человека, который твердо верит, что монета склонна к выпадению орла в 70 % случаев.**

О1. Выбор этих априорных значений немного субъективен, но вот несколько примеров, которые соответствуют каждому убеждению:

- Beta(7,3) — достаточно слабое априорное значение, представляющее уверенность в том, что показатель приближается к 70 %.
- Beta(1000,1000) — очень сильное убеждение в том, что монетка честная.
- Beta(70,30) — это гораздо более сильное убеждение в том, что монетка смещена до 70 %-ной вероятности выпадения орла.

У2. Чтобы проверить монету, вы подбрасываете ее еще 20 раз и получаете 9 орлов и 11 решек. Используя априорные вероятности, которые вы рассчитали в предыдущем вопросе, определите обновленные апостериорные убеждения в истинной вероятности выпадения орла с точки зрения 95 %-ного доверительного интервала.

О2. Теперь у нас есть обновленный набор данных с 32 наблюдениями, которые включают 18 орлов и 14 решек. Используя `qbeta()` в R и априорные вероятности из предыдущих задач, мы можем получить 95 %-ный доверительный интервал для этих различных убеждений.

Просто рассмотрим код для Beta(7,3), поскольку другие примеры идентичны этому.

- Нижняя граница для 95 %-ного интервала — `qbeta (0,025,18 + 7,14 + 3)` = 0,445, а верхняя граница — `qbeta (0,975,18 + 7,14 + 3)` = 0,737.
- Для Beta(1000,1000) имеем: 0,479 – 0,523.
- И для Beta(70,30) имеем: 0,5843 – 0,744.

Итак, как вы можете видеть, слабая априорная вероятность обеспечивает самый широкий диапазон возможностей, очень сильная априорная вероятность честности монетки остается вполне правдоподобной, а сильная априорная вероятность в 70 % по-прежнему склоняется к более высокому диапазону возможных значений истинной честности монетки.

Часть IV. Проверка гипотез: сердце статистики

Глава 15. От оценки параметров к проверке гипотез: создание байесовских A/B-тестов

У1. Предположим, опытный директор по маркетингу говорит вам о своей уверенности в том, что вариант без картинок (B) не будет работать иначе, чем исходный вариант. Как это объяснить в нашей модели? Внедрите это изменение и посмотрите, как изменятся окончательные выводы.

О1. Вы можете объяснить это, увеличив силу априорной вероятности. Например:

```
prior.alpha <- 300
prior.beta <- 700
```

Это потребует гораздо больше доказательств для изменения убеждений. Чтобы увидеть, как это меняет выводы, мы можем перезапустить код:

```
a.samples <- rbeta(n.trials,36+prior.alpha,114+prior.beta)
b.samples <- rbeta(n.trials,50+prior.alpha,100+prior.beta)
p.b_superior <- sum(b.samples > a.samples)/n.trials
```

Тогда новое значение `p.b_superior` равно 0,74, что намного ниже, чем исходное значение 0,96.

У2. Ведущий дизайнер видит ваши результаты и настаивает на том, что вариант B без картинок не будет работать лучше. Она считает, что вы должны принять коэффициент конверсии для варианта B, близкий

к 20 %, а не к 30 %. Реализуйте решение для этого и снова просмотрите результаты анализа.

О2. Вместо того чтобы использовать одну априорную вероятность для изменения убеждений, стоит использовать две — одну, отражающую исходную априорную вероятность, для A, и другую, отражающую веру ведущего дизайнера, для B. Вместо того чтобы использовать слабую априорную вероятность, мы будем использовать немного более сильную:

```
a.prior.alpha <- 30
a.prior.beta <- 70

b.prior.alpha <- 20
b.prior.beta <- 80
```

При запуске этого моделирования нужно использовать два отдельных априорных значения:

```
a.samples <- rbeta(n.trials,36+a.prior.alpha,114+a.prior.beta)
b.samples <- rbeta(n.trials,50+b.prior.alpha,100+b.prior.beta)
p.b_superior <- sum(b.samples > a.samples)/n.trials
```

На этот раз `p.b_superior` составляет 0,66, что ниже, чем раньше, но это все же дает небольшую уверенность в том, что B может быть лучшим вариантом.

У3. Предположим, что 95 %-ная уверенность означает, что вы более или менее убеждены в правильности гипотезы. Также предположим, что больше нет ограничений на количество писем, которые можно отправить в тесте. Если истинное преобразование для A составляет 0,25, а для B — 0,3, изучите, сколько выборок потребуется, чтобы убедить директора по маркетингу в том, что B на самом деле лучше. Изучите то же самое для ведущего дизайнера. Можно сгенерировать образцы конверсий с помощью следующего фрагмента R:

```
true.rate <- 0.25
number.of.samples <- 100
results <- runif(number.of.samples) <= true.rate
```

О3. Вот основной код для решения этой задачи в случае директора по маркетингу (для ведущего дизайнера вам нужно добавить отдельные априорные вероятности). Можно использовать цикл `while` в R для перебора примеров (или просто вручную пробовать новые значения).

```
a.true.rate <- 0.25
b.true.rate <- 0.3
prior.alpha <- 300
prior.beta <- 700

number.of.samples <- 0
#using this as an initial value so that the loop starts
p.b_superior <- -1
while(p.b_superior < 0.95){
  number.of.samples <- number.of.samples + 100
  a.results <- runif(number.of.samples/2) <= a.true.rate
  b.results <- runif(number.of.samples/2) <= b.true.rate
  a.samples <- rbeta(n.trials,
                     sum(a.results==TRUE)+prior.alpha,
                     sum(a.results==FALSE)+prior.beta)
  b.samples <- rbeta(n.trials,
                     sum(b.results==TRUE)+prior.alpha,
                     sum(b.results==FALSE)+prior.beta)
  p.b_superior <- sum(b.samples > a.samples)/n.trials
}
```

Обратите внимание, что поскольку этот код сам по себе является моделированием, вы будете получать разные результаты при каждом запуске, поэтому запустите его несколько раз (или создайте более сложный пример, который запускается еще несколько раз).

Чтобы убедить директора по маркетингу, нужно около 1200 образцов. Ведущему дизайнеру должно хватить около 1000 образцов. Обратите внимание, что хотя ведущий дизайнер считает, что *B* хуже, в нашем примере он имеет также более слабые априорные вероятности, поэтому требуется меньше доказательств, чтобы изменить его мнение.

Глава 16. Введение в коэффициент Байеса и апостериорные шансы: конкуренция идей

У1. Возвращаясь к задаче с игральными костями, предположим, что ваш друг допустил ошибку и внезапно осознал, что на самом деле было две нечестные кости и только одна честная. Как это изменит априорный и, следовательно, апостериорный шансы этой задачи? Вы более склонны верить, что бросаемая кость нечестная?

О1. Первоначальные априорные шансы были следующими:

$$\frac{\frac{1}{3}}{\frac{2}{3}} = \frac{1}{2}.$$

И коэффициент Байеса составил 3,77, что дало нам апостериорные шансы со значением 1,89. Наши новые априорные шансы следующие:

$$\frac{\frac{2}{3}}{\frac{1}{3}} = 2.$$

Итак, апостериорные шансы составляют 2 × 3,77 = 7,54. Сейчас мы определенно более склонны верить, что бросаемая кость нечестная, но апостериорные шансы все еще не очень высоки. Мы бы хотели собрать больше доказательств, прежде чем полностью сдаться.

У2. Вернемся к примеру с редкими заболеваниями. Предположим, вы обратились к врачу и после чистки ушей заметили, что симптомы не исчезли. Еще хуже, появился новый симптом: головокружение. Врач предлагает другое возможное объяснение, лабиринтит — вирусную инфекцию внутреннего уха, при которой в 98 % случаев возникает головокружение. Однако потеря слуха и шум в ушах менее распространены при этом заболевании; потеря слуха происходит только в 30 % случаев, а шум в ушах — только в 28 %. Головокружение также является возможным симптомом вестибулярной шванномы, но встречается только в 49 % случаев. В общей численности населения 35 человек на миллион заболевают лабиринтитом ежегодно. Каковы апостериорные шансы гипотезы, что у вас лабиринтит, по сравнению с гипотезой о вестибулярной шванноме?

А2. Мы немного запутаем ситуацию и сделаем H_1 вероятностью лабиринтита, а H_2 — вероятностью вестибулярной шванномы, поскольку мы уже видели, насколько маловероятна вестибулярная шваннома. Необходимо пересчитать каждую часть апостериорных шансов, потому что мы рассматриваем новую часть данных, «имеет головокружение», а также совершенно новую гипотезу.

Начнем с коэффициента Байеса. Для H_1 мы имеем:

$$P(D|H_1) = 0{,}98 \times 0{,}30 \times 0{,}28 = 0{,}082.$$

Новая вероятность для H_2:

$$P(D|H_2) = 0{,}63 \times 0{,}55 \times 0{,}49 = 0{,}170.$$

Таким образом, коэффициент Байеса для новой гипотезы:

$$\frac{P(D|H_1)}{P(D|H_2)} = 0{,}48.$$

Это означает, что, учитывая только коэффициент Байеса, наличие вестибулярной шванномы является примерно в два раза лучшим объяснением, чем лабиринтит. Теперь нужно посмотреть на соотношение шансов:

$$O(H_1) = \frac{P(H_1)}{P(H_2)} = \frac{\frac{35}{1\,000\,000}}{\frac{11}{1\,000\,000}} = 3{,}18.$$

Лабиринтит встречается гораздо реже, чем избыток ушной серы, и лишь примерно в три раза чаще, чем вестибулярная шваннома. При сложении апостериорных шансов мы можем увидеть:

$$O(H_1) \times \frac{P(D|H_1)}{P(D|H_2)} = 3{,}18 \times 0{,}48 = 1{,}53.$$

Конечным результатом является то, что лабиринтит — лишь немного лучшее объяснение, чем вестибулярная шваннома.

Глава 17. Байесовские рассуждения в «Сумеречной зоне»

У1. Каждый раз, когда вы с другом встречаетесь, чтобы посмотреть фильм, вы подбрасываете монетку, чтобы определить, кто выберет фильм. Друг всегда выбирает орла, и каждую пятницу в течение 10 недель выпадает орел. Вы выдвигаете гипотезу, что у монетки два орла, а не орел и решка. Вычислите коэффициент Байеса для гипотезы о том, что монетка нечестная, в отношении к гипотезе о том, что монетка честная. Что одно только это соотношение говорит о том, обманывает ваш друг или нет.

А1. Допустим, H_1 — это гипотеза о том, что монетка на самом деле с подвохом, а H_2 — гипотеза о том, что монетка честная. Если монетка действительно имеет подвох, вероятность выпадения 10 орлов подряд равна 1, поэтому мы знаем, что:

$$P(D|H_1)=1.$$

И если монетка честная, то вероятность выбросить 10 орлов составляет $0{,}5^{10} = 1/1{,}024$. Итак, мы знаем, что:

$$P(D|H_2)=\frac{1}{1024}.$$

Коэффициент Байеса утверждает, что:

$$\frac{P(D|H_1)}{P(D|H_2)}=\frac{1}{\frac{1}{1024}}=1024.$$

Это означает, что, учитывая только коэффициент Байеса, мы в 1024 раза более уверены, что монетка с подвохом.

У2. Теперь представьте три случая: ваш друг немного шутник, ваш друг большую часть времени честен, но иногда может схитрить и ваш друг очень надежный. В каждом случае оцените некоторые априорные коэффициенты шансов для вашей гипотезы и вычислите апостериорные шансы.

О2. Это немного субъективно, но давайте сделаем некоторые оценки. Нужно найти три различных отношения априорных шансов. Для каждого случая мы просто умножаем априорные шансы на коэффициент Байеса из предыдущей задачи, чтобы получить апостериорную вероятность.

Если ваш друг — шутник, то он, скорее всего, обманет вас, поэтому мы установим $O(H_1) = 10$. Тогда апостериорные шансы станут равны $10 \times 1024 = 10\,240$.

Если ваш друг в основном честен, но может схитрить, вы не удивитесь, если он вас обманет, но не стоит ожидать этого, поэтому мы сделаем априорные шансы равными $O(H_1) = 1/4$, это означает, что апостериорные шансы равны 240.

Если вы действительно доверяете своему другу, вы, возможно, захотите задать априорные шансы для обмана. Априорные шансы здесь могут быть равны $O(H_1) = 1/10\,000$, что дает апостериорные шансы со значением примерно 1/10, то есть вы по-прежнему в 10 раз более уверены в том, что монетка честная, чем в том, что ваш друг обманывает.

У3. Предположим, вы очень доверяете другу. Задайте априорные шансы обмана равными 1/10 000. Сколько раз должен выпасть орел, прежде чем вы начнете сомневаться в невиновности друга — скажем, с апостериорными шансами 1?

О3. При 14 подбрасываниях монеты коэффициент Байеса будет следующим:

$$\frac{1}{\frac{2}{0{,}5^{14}}} = 16\,384\,.$$

Апостериорные шансы будут равны 16 384/10 000 = 1,64. В этот момент вы начинаете чувствовать неуверенность в невиновности вашего друга. Но, бросив монетку менее 14 раз, вы все равно можете поддержать идею о том, что монетка честная.

У4. Другой ваш друг также общается с вышеописанным другом, и после лишь четырех недель выпадения орла он твердо решил, что вас обманывают. Такая уверенность подразумевает апостериорные шансы около 100. Какую ценность вы бы присвоили априорному убеждению этого друга, что первый друг — мошенник?

О4. Мы можем решить эту задачу, заполнив пробелы. Исходя из того что мы знаем, коэффициент Байеса будет равен 16:

$$P(D \mid H_2) = 0{,}5^4 = \frac{1}{16}.$$

Нужно просто найти значение, которое мы затем умножим на 16, что равно 100.

$$100 = O(H_1) \times 16\,,$$
$$O(H_1) = \frac{100}{16} = 6\frac{1}{4}.$$

И теперь мы присвоили точное значение априорным шансам вашего подозревающего друга!

Глава 18. Когда данные не убеждают

У1. Когда две гипотезы одинаково хорошо объясняют данные, один из способов изменить мнение — посмотреть, можно ли воздействовать на

априорную вероятность. Какие факторы могут повысить вашу априорную веру в экстрасенсорные способности друга?

О1. Поскольку мы говорим об априорных убеждениях, ответы на этот вопрос, вероятно, будут немного разными для всех. Мне кажется, что просто предсказать результат броска кубика особенно легко. Хотелось бы увидеть, как этот друг продемонстрирует экстрасенсорные способности в эксперименте по моему выбору. Например, он мог бы угадать последнюю цифру номера долларовых купюр в моем кошельке — так ему будет намного труднее обмануть меня.

У22. Эксперимент утверждает, что, когда люди слышат слово «Флорида», они думают о пенсионерах и это влияет на скорость их ходьбы. Чтобы проверить это, мы собрали две группы из 15 студентов, которые идут по комнате; одна группа слышит слово «Флорида», а другая — нет. Предположим, что H_1 = группы не двигаются с разной скоростью, а H_2 = группа «Флорида» двигаются медленнее, потому что слышит слово «Флорида». Также предположим:

$$КБ = \frac{P(D|H_2)}{P(D|H_1)}.$$

Эксперимент показывает, что H_2 имеет коэффициент Байеса, равный 19. Предположим, что кто-то не убежден в этом эксперименте, потому что у H_2 были более низкие шансы на выигрыш. Какие априорные шансы объяснили бы, что кого-то не убедили, и каким должен быть КБ, чтобы довести апостериорные шансы до 50 для этого неубежденного человека?

О2. Этот вопрос взят из существующей статьи «Автоматичность социального поведения»[1].

Если эксперимент кажется вам сомнительным, значит, вы не одиноки. Известно, что результаты исследования было сложно воспроизвести.

Если вас это не убедило, мы скажем, что априорные шансы должны быть равны около 1/19, чтобы отрицать результаты. Чтобы иметь апостериорные шансы, равные 50, потребуется:

$$50 = \frac{1}{19} \times 950.$$

[1] Джон А. Барг (John A. Bargh), Марк Чен (Mark Chen) и Лара Берроуз (Lara Burrows), «Automaticity of Social Behavior: Direct Effects of Trait Construct and Stereotype Activation on Action», *Journal of Personality and Social Psychology* 71, номер 2 (1996).

Теперь предположим, что априорные шансы не изменили мнение скептика. Подумайте об альтернативной H_3, которая объясняет наблюдение, что группа «Флорида» двигается медленнее. Помните, если H_2 и H_3 объясняют данные одинаково хорошо, только априорные шансы в пользу H_3 заставят кого-то утверждать, что H_3 вернее H_2, поэтому нужно переосмыслить эксперимент, чтобы уменьшить эти шансы. Придумайте эксперимент, который может изменить априорные шансы H_3 по сравнению с H_2.

Вполне возможно, что вторая группа двигалась в среднем медленнее. Нетрудно представить, что в группе из всего 15 человек, услышавшей слово «Флорида», было больше людей невысокого роста, которые могли пройти небольшое расстояние за более длительное время. Чтобы убедиться в этом, мне нужно как минимум увидеть, как этот эксперимент многократно воспроизводится с множеством разных групп людей, чтобы убедиться, что не случайность привела к тому, что группа, услышавшая слово «Флорида», замедлила движение.

Глава 19. От проверки гипотез к оценке параметров

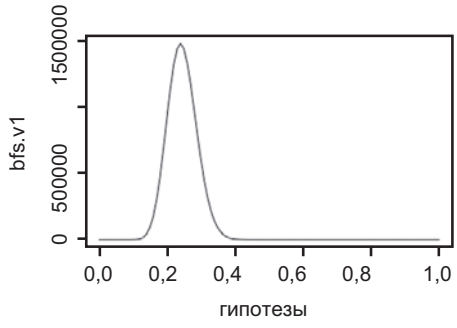
У1. Коэффициент Байеса предполагал, что мы рассматриваем H_1:P(приз) = 0,5. Это позволило нам получить версию бета-распределения со значением альфа 1 и бета 1. Будет ли иметь значение выбор другой вероятности для H_1? Предположим, что H_1:P(приз) = 0,24, а затем посмотрим, отличается ли результирующее распределение, однажды нормализованное до суммы 1, от исходной гипотезы.

О1. Мы можем повторно запустить весь код, но на этот раз создадим одну группу `bfs` для версии 0,5, а другую — для версии 0,24:

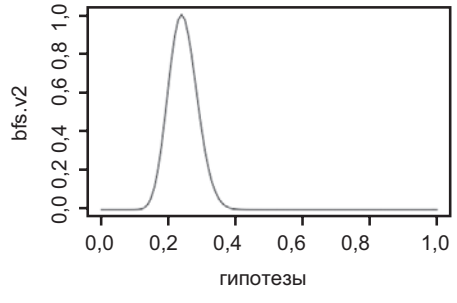
```
dx <- 0.01
hypotheses <- seq(0,1,by=0.01)
bayes.factor <- function(h_top,h_bottom){
  ((h_top)^24*(1-h_top)^76)/((h_bottom)^24*(1-h_bottom)^76)
}
bfs.v1 <- bayes.factor(hypotheses,0.5)
bfs.v2 <- bayes.factor(hypotheses,0.24)
```

Затем построим каждую из них отдельно:

```
plot(hypotheses,bfs.v1,type='l')
```

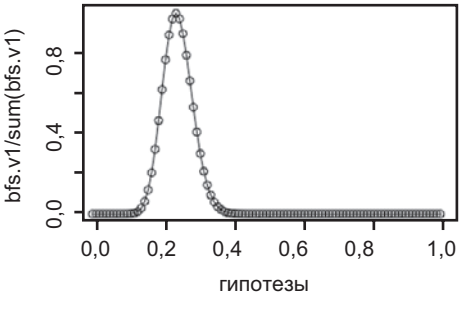



```
plot(hypotheses,bfs.v2,type='l')
```



Здесь мы видим единственное отличие — ось *Y*. Выбор более слабой или более сильной гипотезы меняет только масштаб распределения, но не его форму. Если мы нормализуем и построим эти два графика вместе, то увидим, что они идентичны:

```
plot(hypotheses,bfs.v1/sum(bfs.v1),type='l')
points(hypotheses,bfs.v2/sum(bfs.v2))
```



У2. Напишите априорную вероятность для распределения, в котором каждая гипотеза в 1,05 раза более вероятна, чем предыдущая (предположим, что `dx` остается неизменным).

О2. Давайте воссоздадим `bfs` из оригинала (см. код в предыдущем ответе для первой части вычислений):

```
bfs <- bayes.factor(hypotheses,0.5)
```

Далее новые априорные значения будут начинаться с 1 (поскольку предыдущей гипотезы не существует), затем 1,05, 1,05*1,05, 1,05*1,05*1,05 и т. д. Есть несколько способов сделать это, но мы просто начнем с вектора 1,05 на единицу меньше длины нашей гипотезы (поскольку первая равна 1), используя функцию `replicate()` в R:

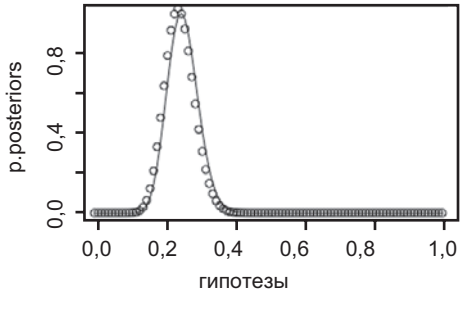
```
vals <- replicate(length(hypotheses)-1,1.05)
```

Затем добавим 1 в этот список и сможем использовать функцию `cumprod()` (которая похожа на `cumsum()`, но для умножения) для создания априорных вероятностей:

```
vals <- c(1,vals)
priors <- cumprod(vals)
```

Наконец, мы просто вычисляем апостериорные вероятности и нормализуем их, а затем можем визуализировать новое распределение:

```
posteriors <- bfs*priors
p.posteriors <- posteriors/sum(posteriors)
plot(hypotheses,p.posteriors,type='l')
#add the bfs alone for comparison
points(hypotheses,bfs/sum(bfs))
```



Обратите внимание, что это не сильно меняет наше окончательное распределение. Несмотря на то что оно дает гораздо более сильные априорные шансы для последней гипотезы — она примерно в 125 раз вероятнее остальных, — коэффициент Байеса настолько низок, что в конечном итоге не имеет большого значения.

У3. Предположим, вы наблюдали еще одну игру с уточками, где было 34 уточки с призами и 66 уточек без призов. Какую проверку вы бы сделали, чтобы ответить на вопрос: какова вероятность того, что шансов выиграть приз в этой игре больше, чем в той игре, которая приводилась в нашем примере? Реализация этой проверки намного сложнее, чем то, что было показано в этой книге, но наверняка вы сможете изучить все самостоятельно и отправиться в собственное приключение по миру более продвинутой байесовской статистики!

О3. Очевидно, что для решения этой задачи нам нужно настроить A/B-тест, как в главе 15. Мы можем легко придумать два распределения для примера «34 утки с призами, 66 без призов», просто повторяя процесс, который уже рассматривался в этой главе. Сложная часть — это выборка из априорной вероятности, которую мы создали сами. В прошлом для выборки из известного распределения мы использовали встроенные функции, такие как `rbeta()`, но для этого случая нет эквивалентной функции. Чтобы решить эту задачу, следует использовать продвинутую технику выборки, например выборку отбраковки или даже алгоритм Метрополиса — Хастингса. Если вы хотите решить эту задачу, самое время приступить к изучению более продвинутой книги по байесовскому анализу. Гордитесь собой, так как вы хорошо разобрались в основах байесовской статистики!

Уилл Курт

Байесовская статистика: Star Wars®, LEGO®, резиновые уточки и многое другое

Перевел с английского *А. Павлов*

Заведующая редакцией	*Ю. Сергиенко*
Ведущий редактор	*К. Тульцева*
Литературный редактор	*А. Павлов*
Художественный редактор	*В. Мостипан*
Корректоры	*С. Беляева, Г. Шкатова*
Верстка	*Л. Егорова*

Изготовлено в России. Изготовитель: ООО «Прогресс книга».
Место нахождения и фактический адрес: 194044, Россия, г. Санкт-Петербург,
Б. Сампсониевский пр., д. 29А, пом. 52. Тел.: +78127037373.

Дата изготовления: 04.2021. Наименование: книжная продукция. Срок годности: не ограничен.

Налоговая льгота — общероссийский классификатор продукции ОК 034-2014, 58.11.12 — Книги печатные профессиональные, технические и научные.

Импортер в Беларусь: ООО «ПИТЕР М», 220020, РБ, г. Минск, ул. Тимирязева, д. 121/3, к. 214, тел./факс: 208 80 01.

Подписано в печать 02.04.21. Формат 70×100/16. Бумага офсетная. Усл. п. л. 24,510. Тираж 700. Заказ 0000.