

Сергей
Самойленко

ВЕРОЯТНОСТИ И НЕПРИЯТНОСТИ

МАТЕМАТИКА ПОВСЕДНЕВНОЙ ЖИЗНИ



ЭВОЛЮЦИЯ
МИО

Серия «Наука для всех»

Сергей Самойленко

Вероятности и неприятности

Математика повседневной жизни

Москва
«Манн, Иванов и Фербер»
2022

УДК 519.2
ББК 22.17
С17

Научный редактор Евгений Поникаров

Самойленко, Сергей Борисович

С17 Вероятности и неприятности. Математика повседневной жизни / Сергей Борисович Самойленко. — Москва : Манн, Иванов и Фербер, 2022. — 256 с. — (Наука для всех).

ISBN 978-5-00169-565-3

Книга познакомит вас с повседневными приложениями теории вероятностей и математической статистики, мягко вводя в мир нешкольной математики. Лейтмотивом изложения станут широко известные «законы Мёрфи», или «законы подлости», — несерьезные досадные закономерности, наблюдаемые каждый день, но имеющие, однако, объективное математическое обоснование. Кроме разнообразных примеров из области теории вероятностей, в книге немало говорится и о смежных разделах: теории мер, марковских цепях, стохастических процессах, теории очередей, динамическом хаосе и т. п.

Эта книга подойдет и школьнику, которому не терпится попасть в университет, и студенту, недоумевающему: «Куда я попал?», — и преподавателю, которому нужны оригинальные живые примеры, а также просто любопытному читателю, желающему развить навыки математического мышления, чтобы научиться отсеивать информационный шум и мусор в потоке новостей.

УДК 519.2
ББК 22.17

Все права защищены. Никакая часть данной книги не может быть воспроизведена в какой бы то ни было форме без письменного разрешения владельцев авторских прав.

ISBN 978-5-00169-565-3

© Сергей Борисович Самойленко, 2022
© Оформление. ООО «Манн, Иванов и Фербер», 2022

Оглавление

Введение	7
Глава 1. Знакомимся с неприятностями	12
Разновидности неприятностей	12
А при чем тут математика?	15
Закон велосипедиста	19
Измеряем уровень подлости	20
От закона велосипедиста к парадоксу инспекции	25
Глава 2. Знакомимся со случайностями и вероятностями	31
Что мы имеем в виду, говоря о вероятности?	33
Возможность невероятного	41
О коварстве географических карт	45
Проверяем честность реальной монеты	47
Откуда же берется случайность?	48
От монеток к бабочкам и самой судьбе	53
Глава 3. Головокружительный полет бутерброда с маслом	57
Айда кидать бутерброды в Монте-Карло!	58
Как правильно говорить о случайных величинах	60
Как правильно задавать вопрос природе?	67
Еще немного анализа размерностей	72
Виновато ли масло?	77
Глава 4. Статистика как научный способ чего-либо не знать	83
Слово в защиту статистики	84
Как возможность ошибиться делает науку наукой	87
Запутываем статистикой и помогаем распутаться	92
Где заканчивается свобода в математике?	95
Измеряем нашу доверчивость	99
Так правда ли, что дожди предпочитают выходные дни?	106
Беспорядок внутри самих чисел	111
Глава 5. Закон арбузной корки и нормальность ненормальности	114
Начнем с многомерного арбуза	116
Мне одному кажется, что я нормальный?	119
В погоне за Нормой	122

Тот самый закон подлости	124
Счастье — это найти друзей с тем же диагнозом, что и у тебя	126
Этот странный закольцованный мир	132
Сравниваем и ищем с помощью вероятности	135
Глава 6. Почему уж не везет так не везет?	139
Синтезируем злодейку-судьбу	140
Ценность релаксации	144
О марковских цепях и пессимистах с оптимистами	150
«Лила» и игра с бесконечностью	154
Почему автобуса все нет?!	164
Глава 7. Прелести чужой очереди	168
Еще раз про пуассоновский процесс	168
Теория для заскучавших в коридоре	171
Совсем немного о случайных функциях	178
Мне только спросить!	181
Стационарный бардак	184
Лучшее — враг хорошего	190
Глава 8. Проклятие режиссера и проклятые принтеры	195
Стратегия балбеса	196
О методе пристального всматривания	202
Быстрее, еще быстрее!	208
Мостим дорогу благими намерениями	212
Ну вот! Еще и принтер сломался!	215
Глава 9. Термодинамика классового неравенства	217
Как говорить об экономике?	218
Подходите, всем хватит!	220
Новая экономическая политика	224
Люди — молекулы	227
Измеряем температуру у рынка	230
Постигаем Дао энтропии	233
Игры с энтропией	240
Экономика должна быть экономной	244
Заключение	251
Рекомендуемая литература	253
Об авторе	254

Введение

В далеком 1977 году в свет вышла книга, которую быстро начали разбирать на цитаты все кому не лень — от журналистов до ученых. Выдержки из нее превратились в «народную мудрость», стали появляться в заголовках газет и журналов и даже упоминаться в серьезных научных трудах. Однако сама по себе она ничему не учила, в ней не предлагалось новаторских методик, она не раскрывала глаза на какую-то «правду». В ней можно было найти только то, что хорошо известно всем на свете, и именно этим она подкупила читателя. Книга называлась «Закон Мёрфи и другие причины, почему все идет не так», а написал ее американский публицист Артур Блох*. Почему же некие «законы» пришлось по душе широкой публике? Потому что они относятся к повседневным неприятностям, досадным совпадениям, надоевшему несовершенству нашего мира. А мы, люди, очень любим жаловаться. Особенно когда жалобы «объективны». Иначе говоря, виноваты в неприятностях могут быть какие угодно обстоятельства, случайности или закономерности, но только не тот, кто жалуется, и не тот, кто его выслушивает.

В этой книге речь тоже пойдет о различных неприятностях. Привычных, ожидаемых и настолько предсказуемых, что они получили статус «законов». Их в книге Блоха и нами самими сформулировано великое множество, это и закон падающего бутерброда,

* Блох А. Закон Мёрфи. — Мн. : Попурри, 2005.

и закон Мёрфи*: *«Если какая-нибудь неприятность может произойти, она случится»*, — и законы Чизхолма, утверждающие: *«Когда все идет хорошо, что-то должно случиться в самом ближайшем будущем»*, и наблюдение Этторе: *«Соседняя очередь всегда движется быстрее»*. Большая их часть вполне тривиальна, но, согласно закону Муира, *«Когда мы пытаемся вытащить что-нибудь одно, оказывается, что оно связано со всем остальным»*. Наша задача — найти рациональное зерно этих закономерностей. Не для того, чтобы с ними бороться, а для удовольствия. И поскольку при этом мы будем использовать математику, удовольствие будет своеобразным и полезным, в отличие от самого результата. Ну а если рассуждения заведут нас слишком далеко, можно взять на вооружение постулат Персига: *«Число разумных гипотез, объясняющих любое данное явление, бесконечно»*. Со всеми этими глубокомысленными фразами и законами мы и станем разбираться, опираясь на язык математики и по возможности строгие выкладки.

Современная математика — огромная страна со сложным «ландшафтом». В ней есть и цветущие долины, и древние памятники, развлекательные центры и пряничные городки, даже супермаркеты с готовыми решениями на все случаи жизни. Все это связано хорошо оборудованными дорогами с указателями и путеводителями. Но есть в математической стране и глухие участки с густыми непроходимыми лесами, горами и топкими болотами, через которые проходят внезапно исчезающие тропинки с шаткими мостиками гипотез и предположений. Наконец, она окружена неизведанными землями, куда если и осмеливался ступить человек,

* Первоисточником закона Мёрфи была не книга Блоха. В ней собраны многочисленные его следствия, но сам закон появился раньше. Блох приписывает его Эдварду Мёрфи, инженеру Лаборатории реактивного движения, так сформулировавшему закономерность в 1949 году: *«Если что-то можно сделать неправильно, он так и сделает»* (If there is any way to do it wrong, he will). В книге Анны Роу 1952 года формулировка «закона Мёрфи, или четвертого начала термодинамики» звучит так: *«Если что-то может пойти не так, это пойдет не так»* (If anything can go wrong it will), и она приписывается безымянному физики. Как позже установлено, это был физик Ховард Перси Робертсон, который дал интервью Роу в 1949 году. Однако близкие по смыслу формулировки существовали намного раньше. Например, в 1877 году британский инженер Альфред Холт писал: *«Установлено, что, если что-нибудь может в море пойти неправильно, это рано или поздно пойдет неправильно»* (It is found that anything that can go wrong at sea generally does go wrong sooner or later).

то лишь очень отважный и часто одинокий в своих поисках. Я не случайно так увлекся этой аллегорией. Она гораздо ближе к пониманию того, что такое наука, чем кажется на первый взгляд. Ведь в любом городе можно ходить по-разному от одной площади до другой, от одного здания к другому. Наконец, в любом городе по-разному можно жить.

Выходя на улицы родного города ребенком, вы изучаете правила перехода улиц, назначение тротуаров и магазинов, узнаете первые надежные тропинки. Если уже взрослым вы впервые попадаете в новый интересный для вас город, то, скорее всего, выберете для ознакомления экскурсионный маршрут, который уже отработан годами и представляет собой своеобразное произведение искусства. Так за какие-нибудь пару часов вы получите яркие впечатления о городе, которые останутся с вами на всю жизнь. Но вы не сможете сказать, что узнали его по-настоящему. Быть может, вас туда занесет по работе — скажем, случится более или менее длинная командировка. Тогда неплохо удастся изучить основные полезные маршруты, и у вас появятся навыки мастерски пользоваться общественным транспортом, перемещаясь быстро, эффективно и удобно. Но и после нескольких недель такой жизни город может остаться незнакомым вам.

Наконец, порой случается так, что город становится вашим по-настоящему. Возможно, вы полюбите его и будете бродить по его улочкам бесцельно, получая удовольствие от самих прогулок. Вы станете отыскивать новые проходы от одной площади к другой через закоулки и дворы, удивительным путем попадать парками и тропинками в нужную точку. Эти дороги могут оказаться на удивление короткими, а способны завести бог знает куда. Но это не страшно: вы знаете этот город и никогда в нем не заблудитесь.

Общий школьный курс похож на освоение элементарных правил жизни в городе. Университетский курс математики уже ближе к экскурсии. Вам покажут главные древние памятники и знаменитые площади, к которым ведут большие проспекты. Глубокое погружение в ту или иную прикладную задачу напоминает командировку: тут не до блужданий, важно четко понять, на какую

линию садиться и на какой остановке пересаживаться каждый день, чтобы не терять драгоценных сил и времени. Но с математикой у вас может случиться и настоящая любовь. И тогда вы уже не остаетесь в рамках лишь практической пользы или удобства, вам становится важно понять, почувствовать, что математика как большой город — это не только дома и площади, даже не линии метро и трамвайных маршрутов. Это единая система, соединяющая всё, что в ней есть, не только взаимным расположением, но и смыслами, контекстами, историями.

Эта книга не совсем о математике. Я приглашаю вас на прогулку по некоторым ее местечкам, хорошо известным и имеющим большую практическую пользу. Но двигаться мы будем несколько необычным маршрутом. Не прямым, как в учебнике, и не сложным и запутанным, как в научной работе, а легким, как бесцельное шатание в хорошей компании под интересный разговор. То и дело мы будем оказываться на развилках и площадях с четко обозначенными названиями, соответствующими разделам математики. Оглянувшись, мы отправимся дальше, но читатель может отметить про себя, что пересеченный нами проспект или бульвар — целое направление, куда можно углубиться самостоятельно, будь на то интерес или необходимость.

В стране математики говорят на своем языке, и не все указатели и надписи легко перевести на русский. Иногда я буду приводить цитаты на языке аборигенов. Иначе говоря, в книге есть формулы. Но это вовсе не единственный алфавит языка математики. Формулы можно выразить графически, и я всегда буду сопровождать уравнения иллюстрациями, которые можно понять интуитивно. Почему же я не отказался от формул, как многие авторы научно-популярных книг? В нашей математической стране не принято верить каждому встречному, не принято сильно полагаться на интуицию, чутье и даже на опыт. Да, опыт, в отличие от физики или психологии, здесь имеет сравнительно невысокую цену. В ходу только доказательство — самая твердая валюта, которой неведомы ни девальвация, ни инфляция, ни мода, ни конъюнктура. Она не обесценивается тысячелетиями (и это не фигура речи, мы используем доказательства тысячелетней давности каждый день). Таким

образом, все, что я вам здесь наговорю, не должно приниматься на веру. Любое мое утверждение, вывод, даже самый неожиданный, можно проверить строгими доказательствами. Именно поэтому везде, где уместно, есть ключи-заметки в виде формул, которыми я руководствовался. Это, впрочем, не лишает читателя возможности любоваться непонятными значками, воспринимая их как орнамент, а автор оставляет за собой право давать математическим закономерностям не очень серьезные и даже фривольные житейские интерпретации. Ведь так гораздо интереснее!

Глава 1

Знакомимся с неприятностями

Разновидности неприятностей

Какие-то наши неприятности *детерминированы*: случайности не играют в их возникновении ключевой роли. Например, если вам понизили зарплату на 10%, а потом извинились и увеличили на 10%, в итоге этих махинаций вы останетесь в убытке, поскольку:

$$x(1 - 0,1)(1 + 0,1) = x(1 - 0,01) < x.$$

Более того, если зарплату сначала повысят, а потом, не извинившись даже, понизят на те же 10%, результат выйдет таким же. Ведь совершенно неважно, в каком порядке перемножаются коэффициенты. Это очень просто, немножко обидно, но к удаче отношения не имеет.

Примером случайной, хоть и весьма вероятной неприятности может быть волшебство, происходящее в наших карманах с наушниками: кладем их аккуратно смотанными в карман, а через полчаса там происходит чудо и вынимаем мы дикий узел проводов. В 2007 году вышла серьезная научная статья двух ученых из солнечного и безмятежного Сан-Диего под заголовком «Спонтанное

образование узлов на возбуждаемой нити»*. В этой работе детально анализируется и моделируется запутывание наушников в кармане. Авторы, основываясь на теории узлов, теории вероятностей и физических экспериментах, убедительно показывают, что при стандартном способе сматывания наушники действительно должны запутываться, причем спустя всего лишь несколько секунд тряски. Впрочем, это мы и так наблюдаем почти каждый день. Сюрпризом здесь может оказаться только ожидаемая скорость запутывания.

Основной причиной образования узлов оказались, во-первых, петли, которые мы сами создаем, наматывая провод на руку, а во-вторых, три конца наушников: штекер и два динамика. В процессе случайного перемещения по карману они попадают в петли, что само по себе не страшно. Проблема возникает тогда, когда мы, пытаясь наушники распутать, тянем за эти концы. Тут-то и выясняется, что узлы на петле из веревки в трехмерном мире, раз появившись, не могут исчезнуть без нарушения целостности веревки. Пока мы трясем их, не выпуская концы наушников из рук, узлы, образованные этими концами и петлями, никуда не исчезают, а только затягиваются. Причем сами по себе петли ни в чем не виноваты. Если бы концов у наушников не было и они представляли собой замкнутое кольцо, то неразвязываемый узел образоваться бы не смог. Ведь узлы не только не исчезают, раз появившись, но и не возникают сами, если их изначально не было.

С этой неприятностью вполне можно бороться математическим способом: нужно либо исключить концы, что в случае наушников неинтересно, либо убрать петли. А это можно сделать с помощью операции сложения. Но не той, что мы изучали в школе, а той, что применяется к петлям на веревках и лентах. Как и числа, петли бывают разных знаков, причем для каждой «положительной» можно построить такую «отрицательную», что в сумме они дадут «ноль»: прямую веревку. Примеры таких петель показаны на рис. 1.1.

* Raymer D. M., Smith D. E. Spontaneous knotting of an agitated string // PNAS. October 16, 2007. Vol. 104. No. 42. Pp. 162–167.

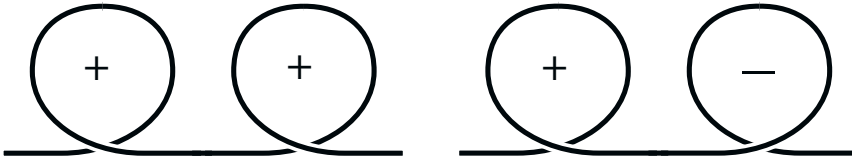


Рис. 1.1. Примеры сложения петель разных знаков

Попробуйте мысленно нанизать на шнурок несколько таких петель разных знаков и вычислите результат и его знак. Чтобы наушники не запутывались, число положительных и отрицательных петель должно оказаться равным. Таких способов сложения проводов несколько, один из них показан на рис. 1.2. Здесь петли разных «знаков» появляются сразу парами и взаимно уничтожают друг друга, не формируя узлов. Уже много лет я складываю наушники именно так, чувствуя себя крутым топологом, и всякий раз радуюсь как фокусу, когда они сами собой полностью разматываются от одного небрежного встряхивания рукой.

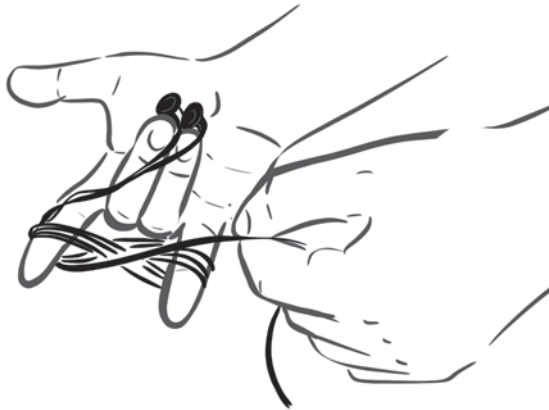


Рис. 1.2. Один из способов складывания проводов, не приводящий к их запутыванию. Он хорош еще и тем, что попутно вы складываете пальцы в мудру любви

Но и среди стохастических по природе законов не все одинаково интересно. Например, закон Бука («Ключи всегда находишь в последнем кармане») не имеет рационального основания. Простой

подсчет показывает, что при равной вероятности отыскать ключи для всех карманов последний ничем не отличается от прочих. Впрочем, этот закон можно трактовать разве что как забавный трюизм: утверждение Бука верно всегда, поскольку тот карман, в котором ключи будут обнаружены, окажется завершающим в процессе поиска и, следовательно, последним. Однако и здесь есть о чем поговорить. В процессе перебора карманов так называемая условная вероятность того, что ключи лежат в последнем из них, действительно повышается. Но это уже нельзя трактовать как вероятность того, что ключи находятся в последнем кармане, тут уже другая задача. Мы вернемся к этому примеру в главе 5.

Нас будут интересовать законы парадоксальные и поучительные, те, которые выглядят злым роком, выбирающим из множества вариантов самые досадные и неприятные, наперекор интуиции, подсказывающей, что этот вариант не должен быть самым вероятным. И, прежде чем приступить к детальным и точным рассуждениям о случайностях и вероятностях, предположим, что какая-то интуиция в отношении случайных процессов и вероятностей у нас уже есть. Это вполне допустимо даже в математической книге — до какого-то момента использовать интуитивное представление о предмете, а потом дать строгое определение. Тем самым, во-первых, мы определяем границы применимости нашей интуиции, а во-вторых, расширяем их в правильном с научной точки зрения направлении. Но не будем забывать о законе Вертерна: *«Предположение — мать любой неразберихи»*, и все наши гипотезы и даже строгие выводы постараемся, где возможно, проверять с помощью имитационного моделирования.

А при чем тут математика?

Петли, наушники, законы подлости, неприятности... при чем же тут математика? Почему вообще имеет смысл рассуждать о законах подлости не так, как Артур Блох, когда он просто посмеялся и нашел меткий афоризм?

С математикой знакомы все, но мало кто готов ответить на вопрос: что делают математики? Считают и вычисляют? Рисуют

треугольники и круги на бумаге в клеточку? Передвигают туда-сюда буквы в уравнениях? Придумывают странные значки и закрючки, чтобы потом писать непонятные тексты? Решают задачи, вычисляя что-то по заказу инженеров, медиков, химиков и других практиков?

Если вы никогда этого не делали, загляните в какой-нибудь математический журнал — просто из любопытства. Сейчас это легко сделать не выходя из дома: поищите в Сети что-то на тему «гомологическая теория типов» или «топология». Вы поразитесь тому, насколько то, что вы там обнаружите, не похоже на школьный образ математики. Но вот что важно: эта колоссальная разница не говорит о том, что есть одна, «простая» математика и другая, «сложная». Математику часто называют языком. Как на любом живом человеческом языке можно писать анекдоты и незамысловатые детские стишки или неуловимо тонкую поэзию, тяжеловесный роман или многостраничный договор, так и с помощью математики можно рассуждать о числах и отрезках, а можно — о петлях и поверхностях, многомерных пространствах и даже основах самой этой науки. Не нужно думать, что числа и отрезки — самое простое, с чем работают математики! Современные теория чисел и геометрия — огромные и во многом неизведанные области, в которых ведутся очень интенсивные исследования.

Но что же все-таки изучают математики? Для чего им этот язык? Чаще всего речь идет о тех или иных *моделях*. Например, что может быть моделью количества? Число, скажете вы. Но любое ли число годится для этого? Младшие школьники, впервые сталкиваясь с отрицательными числами, испытывают замешательство, ведь модель числа оказывается шире привычного им понятия количества. Переход от количества к шагам помогает понять, что числа годятся для моделирования движений на прямой. Тогда отрицательные числа обретают наконец смысл. А чем можно моделировать скорость? Тоже числом. Но если я скажу вам, что двигаюсь со скоростью 60, будет ли этого достаточно для описания того, что со мной происходит? Точно нет! Остается неясно ни куда я двигаюсь, ни, собственно, с какой скоростью:

60 может означать как 60 км/ч, так и 60 мм/год. Отсюда можно заключить, что для моделирования скорости только числа недостаточно. А если, желая объяснить вам, как я перемещаюсь, я изобразю стрелку, станет ли понятнее? Стрелка — ориентированный отрезок — в качестве модели скорости лучше. Она показывает направление, а сравнив ее с какой-то эталонной стрелкой, принятой за единицу, можно определить ее масштаб. Более того, стрелки можно складывать и умножать на числа, получая новые корректные стрелки! И, главное, если мне удастся придумать, как однозначно сопоставлять скорости предметов стрелкам на бумаге, причем окажется, что если v_1 соответствует стрелка a , а скорости v_2 — стрелка b , сумме скоростей $3v_1 + v_2$ будет соответствовать стрелка $3a + b$ и никакая иная, — то это уже будет свойством, позволяющим мне не бегать по двору, изучая скорости, а, сидя в кресле, рисовать стрелки на бумаге.

А можно ли чем-то моделировать стрелки? Абстрактной моделью в этом случае способен стать упорядоченный набор чисел с определенными правилами сложения и умножения на число, который называется *вектором*. Так математики пришли к мысли о линейных векторных пространствах, элементами которых являются векторы. Изучая свойства этих пространств (*изучая*, а не *придумывая*, разницу мы обсудим позже), математики выработали единый язык, который называется *линейной алгеброй*, для разговора о таких разных вещах, как, например, цвета, вращения предметов в пространстве, спектры звуковых сигналов. Пользуясь этим языком, уже можно найти оптимальную стратегию в экономической игре или научить компьютер распознавать нашу речь, рукописные буквы либо лицо человека в толпе.

Математики работают с *математическими структурами* — универсальными моделями всего, с чем имеет дело человеческий разум. *Группы, поля, решетки, графы, петли, косы, языки и бесконечномерные пространства...* Все это структуры с четко определенными свойствами и, если угодно, поведением. Вот уже тысячи лет математики исследуют взаимосвязи между ними, обнаруживают в реальном и математическом мире, что еще можно с их помощью моделировать и при каких условиях.

Я не случайно называл манипуляции с петлями на проводе наушников «сложением», а сами петли «положительными» и «отрицательными». Такая терминология оправдана тем, что петли на струне образуют структуру, называемую *группой*. Для ее построения нужно иметь множество* A и некую операцию $+$, которая будет удовлетворять следующим четырем свойствам.

1. *Замкнутость*: для любых двух элементов из множества A результат операции $+$ всегда будет элементом этого же множества.

2. *Ассоциативность*: для любых a, b, c из множества A верно, что $(a + b) + c = a + (b + c)$.

3. *Существование нейтрального элемента*: в A есть единственный элемент 0 , такой, что $0 + a = a + 0 = a$ для любого a из A .

4. *Обратимость*: для каждого элемента a в A существует единственный обратный ему элемент $(-a)$, такой, что $a + (-a) = 0$.

Группа — общая модель для обратимого ассоциативного комбинирования действий или объектов. Ее образуют числа с операцией сложения, и они же формируют группу с операцией умножения. Несложно убедиться, что аксиомам группы удовлетворяют и петли на веревке или ленте. Понятие группы настолько важно в математике, что, хотя они сами нам в этой книге и не понадобятся, нелишним будет о них рассказать тем, кто с таким подходом еще не знаком, или напомнить тем, кто о группах уже слышал, но не связал свою жизнь с их изучением.

Мы в основном будем иметь дело с двумя структурами: *случайными величинами* и *случайными функциями*. Но, знакомясь с ними, мы встретим многие другие понятия и модели и обозначим некоторые связи между ними.

А начнем мы с простого инструментария, который будет полезен на протяжении всего рассказа. И для этого нам потребуется... велосипед!

* Полагаю, читатель знаком с понятием множества, а также с отношениями и операциями над множествами: пересечением, объединением и пр. Для понимания книги это не обязательно, но для понимания современной математики строго необходимо. Так что любопытного неопита я отсылаю к списку литературы в самом конце книги, а еще лучше — к преподавателю. Поверьте, если школьного учителя попросить растолковать вам, что такое множества и что с ними можно делать, вы оба получите удовольствие!

Закон велосипедиста

Я большой энтузиаст любительского велосипедного спорта. Многие задачи, вошедшие в эту книгу, я обмозговывал в седле, вертя их мысленно и так и эдак, пытаюсь найти наиболее наглядный и простой подход к их объяснению. Что может быть лучше, чем мчаться по трассе ранним утром, по холодку, скачиваясь с легкого склона... Это ощущение стоит того, чтобы ради него преодолевать бесконечные подъемы или сопротивление встречному ветру! Правда, порой кажется, что подъемов больше, чем спусков, а ветер норовит быть встречным, куда ни поверни. В книгах по мерфологии в связи с этим приводится **закон велосипедиста**:

**Независимо от того, куда вы едете, —
это в гору и против ветра.**

Живу я на Камчатке. В Петропавловске много горок — катаюсь по городу, их не миновать. Однако меня должна успокаивать такая мысль: начиная свой путь из дома, я возвращаюсь снова туда, а это значит, что суммарный спуск должен быть равен суммарному подъему. Особенно честным будет маршрут, в котором прямой и обратный пути совпадают.

Представим себе 2-километровую трассу, которая состоит из одной симметричной горки: километр вверх, километр вниз. Вверх по склону я могу достаточно долго ехать со скоростью 10 км/ч, а на спуске стараюсь держать скорость 40 км/ч (я осторожный велосипедист). Исходя из этих условий, на подъем я буду тратить в четыре раза больше времени, чем на спуск, и общая картина получится такой: 4/5 времени путешествия уйдет на тягучий подъем и лишь 1/5 — на приятный спуск. Обидно — 80% времени прогулки займет сложный участок пути! Этот результат не зависит от длины горок, а определяется лишь соотношением скоростей. Если я выкачусь из нашего холмистого города в сторону океана или в долину реки Авачи, горок почти не будет, но в моем распоряжении остаются встречный и попутный ветер или участки с плохой

дорогой, которые также способны отнять значительную часть времени путешествия.

Взглянем на закон велосипедиста несколько иначе. Если я сделаю множество селфи на протяжении своей велопробулки в случайные моменты, а потом займусь их подсчетом и классификацией, то обнаружу, что большинство картинок показывает мне согбенную фигуру в оранжевом шлеме, упорно ползущую вверх по склону либо сопротивляющуюся встречному ветру. Доля снимков с летящим и сияющим велосипедистом, как на рекламной картинке, увы, составит лишь около 20%. А что скажет статистика? Если мы выпустим на холмистую трассу большую толпу велосипедистов, пождем немного и понаблюдаем за их плотностью, то увидим, что большая часть спортсменов толпится на трудных участках, а доля безмятежно улыбающихся лиц не так уж и велика!

Измеряем уровень подлости

Давайте, как когда-то в школе, покажем на графике зависимость перемещения велосипедиста от времени при движении по симметричной треугольной горке. Только сделаем всё «по-взрослому», в так называемых *собственных масштабах задачи**: расстояние станем измерять не в километрах, а в долях общего пути. Так же поступим и со временем путешествия. Первую половину пути велосипедист двигался медленно и долго — $4/5$ всего времени, — а вторую преодолел быстро — за $1/5$ времени (рис. 1.3).

Что же нам показывает полученный график? Во-первых, мы можем сравнить скорости на разных участках (наклоны) со *средней скоростью*, которая соответствует диагональной линии. Во-вторых, становится наглядным соотношение $80/50$ — 80% времени путешествия заняла трудная половина маршрута. Кроме того, из графика можно заключить, что за первую половину расчетного времени путешествия велосипедист успеет преодолеть лишь треть пути. Пока все предельно просто и понятно.

* Подробнее о собственных масштабах и обезразмеривании задачи мы поговорим в главе 2, когда речь пойдет о бутербодах.

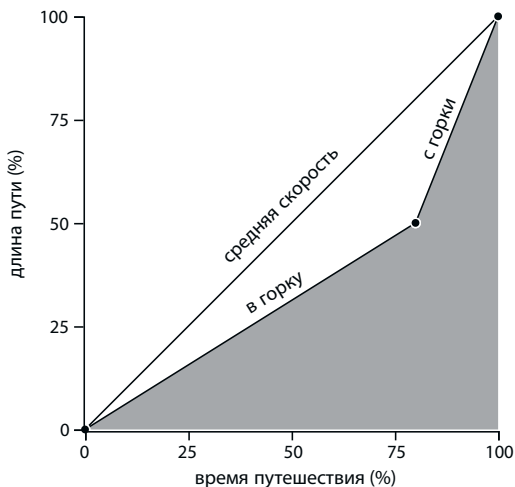


Рис. 1.3. Диаграмма перемещения велосипедиста в долях от общего пути и времени

А что, если маршрут велосипедиста усложнится и перестанет быть симметричным? Что, если участков с подъемами и спусками окажется несколько, и все они будут разными по сложности? Можно изобразить путешествие и на этот раз — например, так, как показано на рис. 1.4.

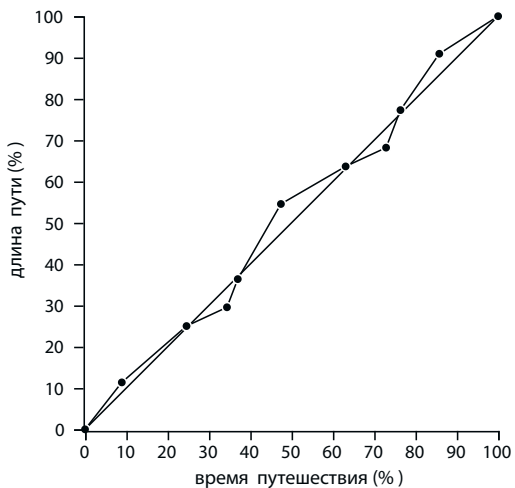


Рис. 1.4. Диаграмма перемещения велосипедиста для более сложного маршрута

Диаграмма хорошо отражает характер пути, но не дает представления об общем соотношении легких и трудных участков; иными словами, она ничего не говорит о *распределении* скоростей. О том, какой смысл мы вкладываем в слово «распределение», речь пойдет в следующей главе; пока же доверимся интуиции и тому, что мы используем его достаточно часто и порой не вкладываем в него точный математический смысл. Чтобы увидеть это распределение, упорядочим отрезки пути по скорости от самых медленных до самых быстрых, после чего вновь нанесем их на диаграмму (рис. 1.5).

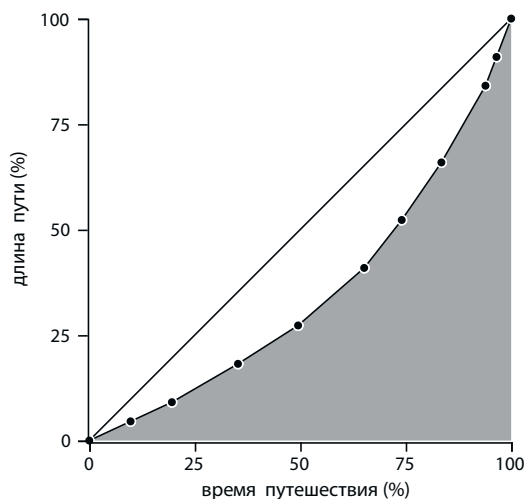


Рис. 1.5. Диаграмма перемещения велосипедиста для распределения скоростей

Мы потеряем при этом информацию о последовательности участков, зато получим обобщающую картину, отражающую то, что можно было бы условно назвать «справедливостью» распределения. Более того, если вместо одного велосипедиста мы взглянем на группу спортсменов, едущих по этому маршруту в произвольном направлении, то наша диаграмма практически не изменится, разве что несколько сгладится из-за разброса скоростей. Ее смысл останется прежним: она покажет, насколько этот маршрут отклоняется от самого справедливого, на котором время преодоления участка не зависит от его «трудности», а определяется только его длиной.

Пора пояснить, откуда взялась такая странная терминология. С начала XX века у эконометристов, демографов, экологов и маркетологов появились вполне универсальные способы суждения о несправедливости этого мира — *кривая Лоренца* и связанный с ней *индекс Джини*.

Для известного распределения в некоторой популяции чего-нибудь ценного, например денег, можно, отсортировав элементы множества по возрастанию уровня богатства, построить *кумулятивную кривую*. Она строится путем последовательного суммирования вкладов каждого члена группы и показывает, как по мере добавления новых членов растет общее благосостояние популяции. Далее нужно поделить все значения, отмеченные по оси X , на численность популяции, а по оси Y — на общее ее благосостояние, перейдя от конкретных чисел к долям или процентам. Получится кривая, носящая имя американского экономиста Макса Отто Лоренца. Когда мы строили график перемещения велосипедиста по простой треугольной горке, мы, по существу, создали кривую Лоренца для распределения скоростей по отрезкам пути, состоящего всего из двух столбцов, как показано на рис. 1.6.

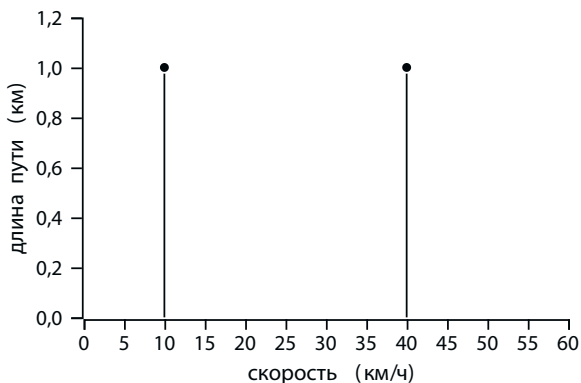


Рис. 1.6. Распределение скорости велосипедиста по пройденному пути

Конечно, не всякий график перемещения можно воспринимать как кривую Лоренца. Для начала нужно отсортировать периоды путешествия по возрастанию скорости, после чего приступить

к построению. Можно построить *гистограмму скоростей*, сгруппировав известные нам данные по принадлежности к известным интервалам значений, после чего последовательно суммировать вклады всех данных гистограммы, начиная с малых значений и заканчивая самыми большими. Результатом должна стать всюду вогнутая кривая, которая проходит ниже диагонали, — настоящая кривая Лоренца. Упомянутая диагональ называется *кривой равенства*, она в нашем случае соответствует постоянной (средней) скорости на всем пути или гистограмме с единственным столбиком (такое распределение называется *вырожденным*). В экономическом контексте кривая равенства отражает всеобщее равенство благосостояния в обществе. Чем больше кривая Лоренца отклоняется от кривой равенства, тем менее «справедливым» можно считать распределение. И, раз уж мы изучаем законы подлости и несправедливости нашего мира, разумно использовать терминологию и инструменты, созданные именно для исследования справедливости.

Площадь под кривой Лоренца для любого невырожденного распределения будет меньше площади под кривой равенства. Их разница может служить формальной характеристикой неравенства или «несправедливости» распределения. Эту роль на себя берет *индекс Джини*. Он вычисляется как удвоенная площадь замкнутой фигуры, образуемой кривой равенства и кривой Лоренца (ее мы показали заливкой на рис. 1.5), и лежит в диапазоне от 0 до 1. Для кривой равенства, идеального вырожденного мира, индекс Джини равен 0, а в самом кошмарном варианте, когда все богатство группы принадлежит одному ее члену, он равен 1. В рассмотренном нами примере он составляет 0,35. Это неплохой показатель. Скажем, распределение богатства среди населения в России сейчас имеет индекс Джини 0,39, в США — 0,49, в Австрии и Швеции не превышает 0,3, а для всего мира он в 2017 году составил 0,66. Так что приведенная нами в качестве примера ситуация с велосипедистами, конечно, несправедлива, но вполне терпима.

Обратите внимание на то, что с помощью некоторого формального индекса мы стали сопоставлять совершенно разные

и несравнимые вещи. Это одновременно и заманчиво, и опасно. Нужно отдавать себе отчет в том, что формальные индексы и числовые показатели *всегда чему-то равны*, независимо от того, есть в этом какой-либо смысл или нет. Мы сравниваем распределение богатства среди населения стран и распределение времени, затрачиваемого на преодоление пути, с точки зрения отличия от некоторого варианта, который сочли бы справедливым. Пока мы ведем фривольные и подчас хулиганские разговоры о законах подлости, пожалуй, это оправданное сравнение; но в науке так, конечно, делать нельзя. Кривую Лоренца и индекс Джини можно формально рассчитать и для гистограммы яркости пикселей на картинке или для частотности слов в живой речи. Но к справедливости это не будет иметь никакого отношения, да и смысла останется совсем немного, поэтому, имея в виду индекс Джини для чего попало, мы будем его называть *индексом подлости*, чтобы не вводить читателя в заблуждение наукообразностью терминов.

Кривые Лоренца и индекс подлости позволяют нам смело сравнивать возмутительно разные вещи. Математика — точная наука, но никто не запрещает математикам хулиганить. В своем, конечно, кругу и без драк.

От закона велосипедиста к парадоксу инспекции

Вывод, который делает велосипедист, пыхтя на пониженной передаче: «Мир несправедлив, бóльшую часть сил отнимает самая дурацкая часть работы», — часто именуют **принципом Парето** или **принципом «80/20»**: «80% усилий дают 20% результатов». Это абсолютная эмпирика: принцип Парето никто не доказывал, но его так часто цитируют, что он уже производит впечатление истины. Его используют и как оправдание неудачам, и даже как инструкцию, обнаруживают в самых разных проявлениях. Иногда это работает: например, принципу «80/20» соответствует индекс подлости около 0,6, как для распределения богатства в мире.

У принципа Парето есть полезное для понимания более строгое обобщение. Закон подлости, названный Артуром Блохом в честь безымянного велосипедиста, имеет официальное научное звание: **парадокс инспекции**. Это хорошо известное явление встречается в разных исследованиях, связанных с социологическими опросами, тестированием, и в *теории отказов* (разделе прикладной математики, занимающемся надежностью сложных систем), неявно, но систематически смещая наблюдаемые результаты в сторону наиболее часто наблюдаемых явлений.

Приведем классический пример, связанный с неудовольствием пассажиров общественного транспорта. На линии в некоем городе работает множество автобусов. В относительно короткий час пик они переполняются, всё же остальное время ходят почти пустыми. Если мы станем опрашивать *пассажиров*, то выясним, что бóльшая их доля оказалась невезучей и ехала в переполненном транспорте (по той простой причине, что в переполненном автобусе было больше людей), и получим выражение общего недовольства. Если же мы опросим *водителей*, то они тоже начнут жаловаться, но, как ни странно, на незаполненность большинства маршрутов и неразумность руководства, гоняющего их попусту. Гибкий график сгладит ситуацию, но в любом случае кривая Лоренца будет отклоняться от кривой равенства, соответствующей невероятной ситуации всегда одинакового числа пассажиров во всех автобусах.

В учебниках по теории вероятностей часто встречается специальный непрозрачный мешок, в который математики складывают разнообразные объекты, а потом наугад вытаскивают их, делая подчас весьма глубокомысленные выводы. Разрешение нашего парадокса в том, что, анализируя систему пассажиропотока в целом, мы кладем в мешок автобусы, а проводя опрос, достаем из него наугад пассажиров и по их данным пытаемся делать выводы об автобусах. Рисунок 1.7 показывает, в чем тут разница.

Рассмотрим эту ситуацию подробнее, построив кривую Лоренца (на этот раз настоящую) для числа пассажиров в автобусах, показанных на рис. 1.7.

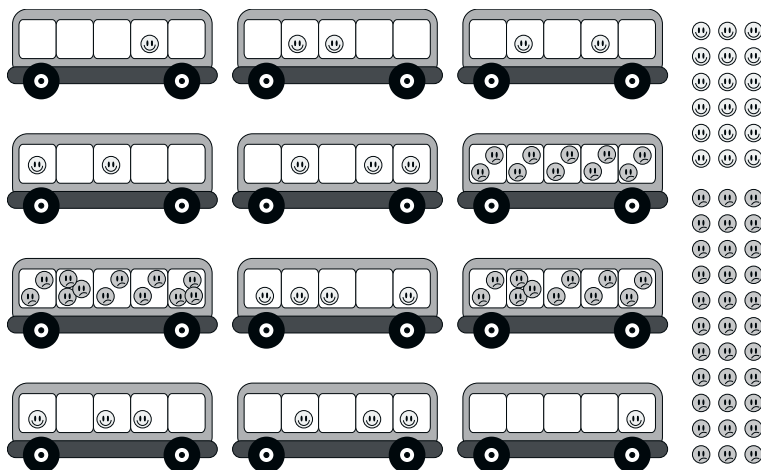


Рис. 1.7. Статистика по автобусам говорит, что в 75% машин есть свободные места, то есть они ходят не в полной мере эффективно. А опрос пассажиров обнаружит, что 61% людей, воспользовавшихся автобусом в этот день, оказались в переполненном транспорте и остались недовольны

Для этого нужно отсортировать машины по числу пассажиров и последовательно суммировать вклад каждого в общий пассажиропоток.

Число пассажиров в автобусе	1	1	2	2	2	3	3	3	4	10	11	12
Кумулятивная сумма пассажиров	1	2	4	6	8	11	14	17	21	31	42	54
Кумулятивная сумма автобусов	1	2	3	4	5	6	7	8	9	10	11	12

Полученные кумулятивные суммы следует разделить на их максимальные значения, чтобы получить доли, например, в процентах, после чего их можно нанести на диаграмму (рис. 1.8).

Кривая Лоренца в данном случае показывает, как распределение числа элементов в некоторых группах (горизонтальная ось) смещается при анализе распределения элементов по принадлежности к группам (вертикальная ось). В этом, собственно, и состоит парадокс инспекции: картинка, которую наблюдает инспектор,

оказывается искаженной. Ведь он анализирует не группы, а их элементы, и при этом наблюдаемые значения смещаются в сторону более «весомой» части распределения.

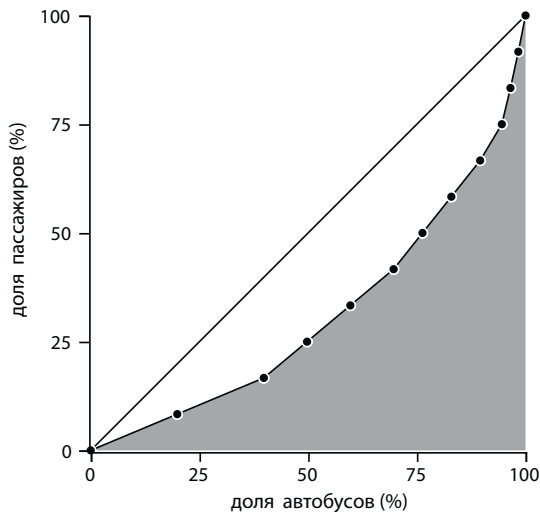


Рис. 1.8. Кривая Лоренца хорошо иллюстрирует несправедливость ситуации с автобусами: половина возит лишь четверть всего пассажиропотока, а на 25% перегруженных машин приходится половина пассажиров

Сам по себе закон велосипедиста очень прост, но он то и дело будет усугублять другие законы подлости, прибавляя им угрюмой эмоциональной окраски. Размышляя об этом, мне нравится представлять, как искажается восприятие мира инспектором, становясь контрастнее. В растровых графических редакторах есть инструмент «Кривые». Он позволяет дизайнеру или фотографу тонко менять контраст картинки, манипулируя распределением числа пикселей по яркости. Вот, например, как меняет восприятие реальности кривая Лоренца, полученная нами для автобусов. Картина мира становится мрачнее, как мы и ожидаем (рис. 1.9).

Крайнее проявление парадокса инспекции возникает, если в группах, помещенных в наш теоретический мешок, есть не просто редкие элементы, а элементы, не наблюдаемые вовсе. Тогда мы

получаем то, что статистики, демографы и публицисты называют *систематической ошибкой выжившего*.

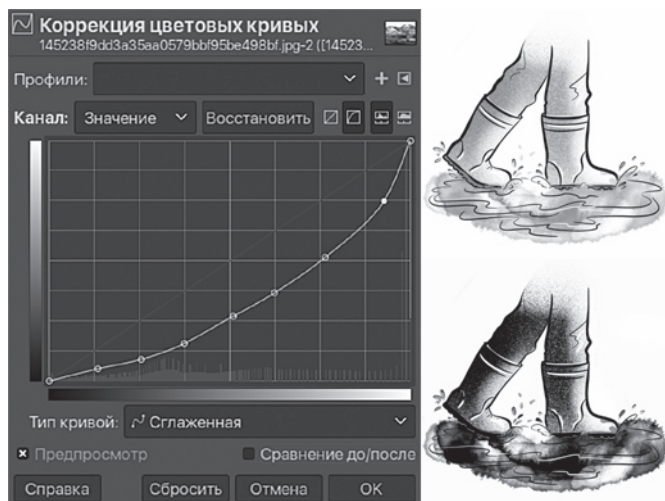


Рис. 1.9. Кривая Лоренца, примененная в качестве фильтра «Цветовая кривая» в растровом графическом редакторе, делает любую картину мрачнее

Часто ее демонстрируют на примере с дельфинами, которые спасают людей, оказавшихся волею несчастного случая в открытом море. Дельфины обнаруживают на поверхности моря любопытный несъедобный объект (человека) и играют с ним, подталкивая носом. При этом они необязательно толкают его в сторону ближайшего берега — часть людей они уводят в открытое море, поскольку разумно предположить, что для дельфина берег, да еще и населенный людьми, опасен. Однако, если всё же дельфины толкают потерпевшего именно к берегу, в сторону спасения, и он благодаря этому выживает, весь мир облетает новость: дельфины спасли человека! О поведении дельфинов во всех прочих печальных случаях, увы, мы не узнаем ничего. Эти элементы из мешка мы не достанем и в статистику они не попадут, так что мы получим явно искаженную картину.

Об этом явлении часто рассказывают в различных демотивирующих статьях для начинающих бизнесменов, уверяя их в том, что успешный путь, описываемый в мотивационных книгах, скорее

всего, не для них: «неудачники книг не пишут». Впрочем, к законам подлости это отношения не имеет, тут мы касаемся психологии. Парадокс инспектора и ошибка выжившего действительно способны исказить восприятие действительности, омрачая ее либо придавая излишне радужную окраску. Но с научной точки зрения это методические ошибки при получении и обработке данных. К сожалению, они приводят к расхожему мнению о статистике как нечестном манипулировании фактическими данными среди людей, весьма далеких от этих методик. О таких ошибках знать полезно, чтобы избегать их в своей работе и критически относиться к новостям, слухам и недобросовестным исследованиям. Этой теме посвящена относительно недавняя книга Джордана Элленберга «Как не ошибаться»*, содержащая множество ярких примеров того, как статистические данные и числа могут быть до забавного неверно поданы и интерпретированы.

Мы встретимся с парадоксом инспекции и его влиянием еще не раз: стоя в очереди или на автобусной остановке, рассуждая о судьбе. Поняв, что это не козни рока, а простейшая математика, с которой бороться смысла нет, можно научиться получать удовольствие и от затяжных подъемов, и от нудных, но неизбежных этапов работы — хотя бы решая в уме задачи или медитируя. Даосы стремились жить вечно и верно рассудили, что вместе с работой над телом для достижения их цели требуется подготовка ума. Ведь для вечной жизни нужно не только умение отпускать привязанности, но и терпение, а также способность получать удовольствие от затяжных участков.

* Издана на русском языке: Элленберг Дж. Как не ошибаться. Сила математического мышления. М. : Манн, Иванов и Фербер, 2017.

Глава 2

Знакомимся со случайностями и вероятностями

Разговор о законах подлости как источнике житейских неурядиц часто начинается со знаменитого **закона бутерброда**. Он просто формулируется, легко проверяется и широко известен:

Бутерброд всегда падает маслом вниз.

Понятно, что «всегда» здесь явное преувеличение: легко представить себе условия, в которых бутерброд упадет, но при этом намазанная маслом сторона останется в сохранности. Что же люди понимают под этим законом? Скорее всего, что бутерброд падает маслом вниз достаточно часто, чтобы это было заметно. Но чаще ли происходит неблагоприятный исход, чем благоприятный? Бутерброды бывают разные, падают при различных обстоятельствах и с разной высоты. Параметров столько, что говорить о закономерностях в такой задаче, возможно, нет смысла. По-всякому бывает. Иногда маслом вниз — тогда становится обидно, мы вспоминаем закон и закрепляем его в памяти. А если бутерброд падает неинтересно — маслом кверху — или кусок хлеба вовсе без масла, и говорить не о чем: понятно же, что закон шуточный!

В принципе бутерброд подобен монетке, которую математики используют для получения случайных величин с двумя возможными значениями: «орел» и «решка». Если монетка «честная», то ей неважно, какой стороной падать. По идее, с бутербродами дела должны обстоять так же.

Мы вернемся к ним и посвятим им целую главу, в которой очень внимательно изучим их падение, но пока присмотримся к самой, наверное, простой вероятностной системе: монетке. Ее в книгах о теории вероятностей подбрасывают каким-то особым магическим образом — так, чтобы выбор начального положения, начальной скорости и скорости закручивания при подбрасывании никак не влиял на вероятность исхода. Но очевидно же, что это невозможно! Монетка представляет собой *механическую систему* и подчиняется законам механики, а они не содержат случайных величин. Будущее в законах движения такого простого тела, как монетка, однозначно определяется его прошлым состоянием. Если монетку будет подбрасывать робот или демон Лапласа — мифическое существо, обладающее полной информацией о координатах и скоростях любой механической системы, — то при неизменных начальных данных будут получаться идентичные результаты. Более того, такому демону можно было бы заказать ту или иную сторону при сколь угодно хитром закручивании монеты. Когда я смотрю выступления цирковых жонглеров, которые невероятно ловко и точно управляют с десятком разнообразных предметов, в голову приходит мысль, что демоны Лапласа существуют и живут среди нас. Вот для кого, кажется, нет никакой случайности: ведь часто акробатические номера выполняются под куполом цирка или на весьма неустойчивой башне из всякой всячины. Случайность в этом случае может обернуться трагедией, так что ее необходимо исключить!

Мы с вами, конечно, не роботы и не демоны, а большинство не умеют жонглировать и тремя апельсинами. Но неужели люди подбрасывают монетки настолько неряшливо и непредсказуемо, что законы механики могут приводить к случайностям? Да и откуда вообще берется случайность в мире, описываемом строгими и предсказуемыми законами механики? Существует ли она в реальном мире? Многие мои знакомые, в том числе искусственные в науке, уверены, что

настоящих случайностей не бывает, есть лишь нехватка информации, неточные расчеты, глубинное непонимание человеком механики физического мира. Однако «Бог не играет в кости с Вселенной». Эта фраза, неоднократно повторенная Альбертом Эйнштейном, стала девизом механистической картины мира, которая в XXI веке вынуждена уживаться с квантовой механикой, ее неустранимой, как нам сейчас кажется, стохастичностью (случайностью).

Но в чем же разница между истинно хаотическими или стохастическими системами, принципиально непредсказуемыми, и теми, где трудно угадать поведение, рассчитать которое все же возможно? Когда стоит переходить на язык вероятностей и о чем он позволяет говорить, что невозможно выразить иначе, не прибегая к этому языку?

Что мы имеем в виду, говоря о вероятности?

Начнем разбираться с простенькой монеткой и посмотрим, каким может быть источник неопределенности в эксперименте с подбрасыванием. Задача подробно рассматривалась в 1986 году Джозефом Келлером*, и здесь мы приведем простое объяснение возникновению неопределенности в этом нехитром процессе, основанное на рассуждениях из его статьи. В самом первом приближении то, какой стороной упадет монета, зависит от времени ее полета t и угловой скорости ω . Если измерять последнюю в оборотах за единицу времени, то число оборотов, совершаемое монетой, выражается предельно просто: $n = t\omega$. Эта зависимость задает линии равного числа оборотов в координатах (t, ω) , а они, в свою очередь, ограничивают области, соответствующие четному и нечетному числу оборотов: тому, сменится ли сторона монетки после подбрасывания или нет. Пример такой диаграммы показан на рис. 2.1.

С помощью этой полосатой диаграммы можно выяснить, каким будет результат подбрасывания монетки, закрученной на известное число оборотов в секунду и пойманной через известное время после

* Keller J. B. The probability of heads // American Mathematical Monthly. 1986. Vol. 93. Pp. 191–197.

броска. Если попадаем в белую полосу, выпадет та же сторона, что была сверху при броске; если в серую — обратная. Линии равного числа оборотов представляют собой гиперболы; видно, что по мере увеличения числа оборотов чередование областей становится все более частым, а сами области оказываются тоньше. Человеческая рука несовершенна, и очень небольшой разброс начальных значений перекрывает сразу много областей, делая исход непредсказуемым. В диапазоне действия руки (прямоугольник на диаграмме) смещения на 5% достаточно для того, чтобы перескочить с белой полосы на серую. Остается вопрос: как из этого построения следует «честность» настоящей монеты? Как из такой диаграммы получить *вероятность* выпадения орла или решки?

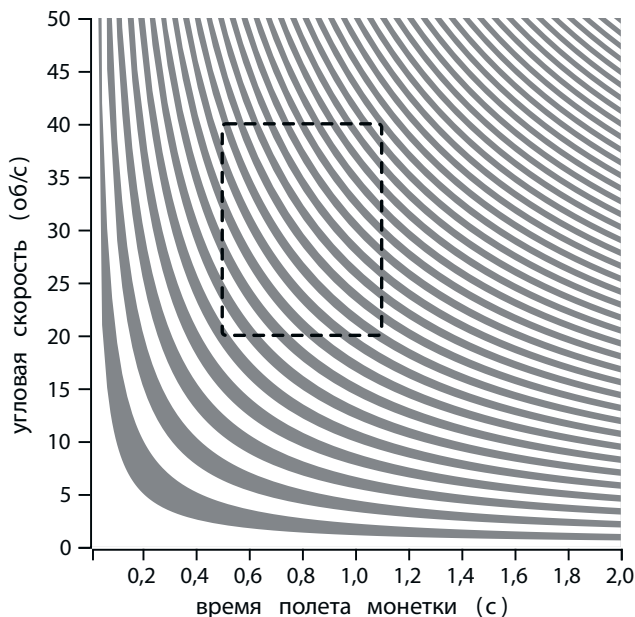


Рис. 2.1. Диаграмма, показывающая четность количества оборотов монеты в полете. Прямоугольником показана область, в которой чаще всего происходит процесс гадания на монетке при подбрасывании рукой

Чтобы перевести наши рассуждения на язык вероятностей, окунемся в математику, которую не проходят в школе. И хотя от нее

ожидают чего-то сложного, сейчас она упростит дело и поможет лучше понять, о чем мы рассуждаем.

Во введении я говорил, что математики изучают не числа или геометрические фигуры, как может показаться после изучения школьного курса. Они работают со сложными структурами (абстрактными алгебрами, полукольцами, полями, моноидами, топологическими пространствами и прочей абстрактной всячиной), описывают их, вроде бы совершенно не привязываясь к практике, корректно определяют, изучают их свойства, доказывают теоремы. А потом они оттачивают мастерство в поиске подобных структур в самых разных явлениях природы и областях человеческих знаний, совершая удивительно полезные прорывы, в том числе в чисто прикладных областях. Сейчас мы рассмотрим, как строится базис теории вероятностей, основанный на достаточно абстрактном понятии *меры*.

Мы описали механику монетки и получили области, описывающие множества решений с определенными свойствами. Области — плоские фигуры. Как правильно перейти от них к вероятностям? Нужно измерять наши области, и мы естественным путем приходим к их площади. Площадь — *мера* плоской фигуры. Это точный математический термин, обозначающий *функцию*, которая множеству ставит в соответствие некую неотрицательную числовую величину.

В математике есть целый раздел, который называется *теорией меры*. Она родилась на рубеже XIX–XX веков (у ее истоков стояли французы Эмиль Борель и Анри Леон Лебег) и открыла математикам широкие возможности для анализа очень сложно устроенных объектов: канторовых и фрактальных множеств. Теория меры легла в основу функционального анализа и современной теории вероятностей. Определение вероятности как меры позволяет увидеть все ее основные свойства как для дискретных, так и для непрерывных множеств.

Хотя наша книга не учебник, на этом стоит остановиться, чтобы взглянуть на понятия теории вероятностей как бы с «высоты птичьего полета» и почувствовать вкус «большой» математики. Я прошу читателя не пугаться, если что-то в приводимых ниже определениях покажется непонятным. Если язык математики вам незнаком,

воспринимайте это как отрывок текста «в оригинале» на незнакомом вам языке. Он может быть не полностью понятен, но в нем нет искажений «переводчика» и не нарушена целостность. При изучении истории, литературы или иностранных языков необходимо работать или хотя бы знакомиться с оригинальными текстами и полными цитатами. Язык математики тоже требует знакомства с «оригиналом», поскольку в текстах определений и теорем ничего ни прибавить, ни убавить без потерь не получится. Попытки сократить текст «для ясности» порой приводят к серьезным неточностям и вовсе к ошибкам. Итак, вот как звучит определение меры.

Пусть имеется множество X .

Набор его подмножеств F называется алгеброй, если для F верно:

- 1) пустое множество принадлежит F : $\emptyset \in F$;
- 2) если множество $A \in F$, то и его дополнение $X \setminus A \in F$;
- 3) если A и $B \in F$, то их объединение $A \cup B \in F$.

Из этого определения следует, что пересечение множеств A и B принадлежит F , а также то, что объединение или пересечение любого конечного числа множеств принадлежит F . Говорят, что алгебра замкнута относительно конечного объединения и пересечения.

Набор подмножеств F называется сигма-алгеброй, если вместо 3) потребовать более сильное условие: чтобы объединение счетного числа множеств A_i принадлежало F : если $A_i \in F$, то $\bigcup_i A_i \in F$.

Из этого определения следует, что и пересечение счетного числа множеств принадлежит F . Иными словами, сигма-алгебра замкнута относительно счетного объединения и пересечения.

Пусть F — алгебра множеств. Функция μ , сопоставляющая любому множеству $A \in F$ какое-нибудь неотрицательное число, называется мерой, если:

- 1) мера пустого множества равна 0: $\mu(\emptyset) = 0$;
- 2) для любых непересекающихся множеств $A, B \in F$, то есть $A \cap B = \emptyset$, верно $\mu(A \cup B) = \mu(A) + \mu(B)$. Такое свойство называется аддитивностью.

Если же взять F — сигма-алгебру, а во втором условии взять счетное количество непересекающихся множеств, то получится

более сильное условие $\mu(\cup_i A_i) = \sum_i \mu(A_i)$, которое называется сигма-аддитивностью. Такая мера называется сигма-аддитивной.

Из определения меры следуют такие свойства:

1) если A включается в B , то мера A не больше, чем у B : если $A \subseteq B$, то $\mu(A) \leq \mu(B)$;

2) если A включается в B , то мера разности множеств равна разности мер: если $A \subseteq B$, то $\mu(B \setminus A) = \mu(B) - \mu(A)$;

3) для любых A и B верно $\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$.

Знакомые каждому примеры мер — количества (количество яблок в мешке, например), а также длины, площади, объемы фигур.

Количество элементов — так называемая считающая мера. Каждому подмножеству A поставим в соответствие количество элементов в нем: для конечных A положим $\mu(A) = |A|$, а для бесконечных — $\mu(A) = \infty$.

Длина на прямой, площадь на плоскости, объем в пространстве — тоже мера. Во всех случаях условие аддитивности выполняется.

Всякая ли неотрицательная числовая функция может быть мерой? Совсем нет. Например, возраст ставит человеку в соответствие вполне определенное положительное число. Но он не подходит под определение меры. Предположение о том, что возраст может быть таковой, приводит к забавным парадоксам. Представьте себе кошку, которой пять лет. Естественно, что и правой, и левой половине животного тоже по пять лет, ведь они возникли одновременно. Если бы возраст был мерой, как, например, кошачий вес, то, согласно свойству аддитивности, кошке как сумме ее половинок должно быть уже десять лет. Подобное деление, впрочем, можно продолжить и достичь сколь угодно большого возраста. С другой стороны, мера части не может превосходить меры целого. Иначе говоря, хвост должен быть строго моложе кошки, а шерстинки на хвосте, соответственно, еще моложе. Так мы приходим к выводу, что мельчайшие клетки, из которых состоит пятилетняя кошка, должны были появиться на свет практически только что. Подобные рассуждения можно применить к таким измеримым величинам, как температура или скорость, которые не являются мерами. Два человека бегут не вдвое быстрее одного.

По этому поводу в книге Артура Блоха был сформулирован **закон новшества**.

Если вы хотите, чтобы команда выиграла прыжки в высоту, найдите одного человека, который может прыгнуть на семь футов, а не семь человек, прыгающих на один фут.

В свою очередь, импульс (количество движения) или энергия уже обладают свойствами меры. Вес, количество денег, объем знаний, громкость (амплитуда) крика — хоть и не всегда легко измеримы, но тоже могут служить мерой на множестве людей.

Но вернемся к вероятностям. На интуитивном уровне с этим понятием знакомы сейчас практически все. Ее оценивают политологи и журналисты на ток-шоу, ее обсуждают, говоря о глобальном потеплении или завтрашнем дожде, о ней рассказывают анекдоты: *«Какова вероятность встретить на Тверском бульваре живого динозавра? — Одна вторая: или встречу, или нет»*.

Широко распространено понимание вероятности как частоты, с которой могут происходить события при многократных испытаниях или наблюдениях. Это представление согласуется с нашим повседневным опытом, но оставляет ряд сложных вопросов. Например, когда байесовский спам-фильтр выдает следующий результат: *«Вероятность того, что сообщение “Заработать в интернете может любой! Жми! Узнай как!” — спам, составляет 82%»*, с частотой чего это можно связать? Если протестировать сообщение несколько раз, ничего не изменится; можете переставить слова, но результат останется тем же, а при изменении текста сообщения мы переходим к другой задаче. О какой же вероятности речь? Другой пример. Камчатские сейсмологи каждый год публикуют прогноз сейсмической опасности — вероятности сильного землетрясения в ближайшее время. Однако и здесь неясно, можно ли дать частотное толкование такого прогноза. В главе 6 мы разберемся с этим примером, а сейчас приведем определение вероятности, данное замечательным русским математиком Андреем Николаевичем Колмогоровым в 1930-е. Оно может показаться далеким от интуитивного представления

и чересчур сложным. Но интуиция — неважный помощник в рассуждениях на такую абстрактную тему, как вероятность. Сформулированное Колмогоровым определение — надежный и универсальный инструмент, применимый к очень широкому кругу задач. В следующих главах мы будем неоднократно обращаться к нему, вырабатывая правильную интуицию у читателя.

Современная теория вероятностей базируется на понятии вероятностного пространства. Его определение потребует ввести несколько новых терминов.

Элементарное событие — результат какого-либо эксперимента или наблюдения за системой, имеющей случайное поведение. При этом один эксперимент порождает ровно одно событие. Например: «выпадение тройки при бросании игральной кости», «наблюдение интервала в 7 минут между автомобилями в дорожном потоке».

Множество всех таких событий называют пространством элементарных событий. Ну что же, мы теперь готовы познакомиться с тем, как в математике определяется вероятность.

Вероятностным пространством называется тройка, включающая пространство элементарных событий Ω , сигма-алгебру его подмножеств F и функцию P , называемую *вероятностью*, которая каждому элементу из F ставит в соответствие неотрицательное число, причем:

- 1) $P(\emptyset) = 0$;
- 2) $P(\Omega) = 1$;
- 3) функция P сигма-аддитивна, то есть вероятность счетного объединения непересекающихся событий равна сумме их вероятностей: $P(\cup_i A_i) = \sum_i P(A_i)$.

Как видите, вероятность — сигма-аддитивная мера на пространстве элементарных событий, имеющем меру 1. Соответственно, описанные выше свойства меры на языке вероятностей примут следующий вид.

Если из события A следует событие B , то вероятность A не больше, чем вероятность B : если $A \subseteq B$, то $P(A) \leq P(B)$.

Если из события A следует событие B , то вероятность того, что наступит B , но не наступит A , равна разности вероятностей: если $A \subseteq B$, то $P(B \setminus A) = P(B) - P(A)$. В частности, если $B = \Omega$, то получаем

формулу для вероятности противоположного события. Если событие, означающее, что событие A не произошло, обозначить \bar{A} , то $P(\bar{A}) = 1 - P(A)$.

Для любых A и B верно $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Рассмотрим простейший пример вероятностного пространства. Пусть мы бросаем монету, то есть в нашем эксперименте возможны всего два исхода, и $\Omega = \{O \text{ (орел)}, P \text{ (решка)}\}$. Сигма-алгебра — множество всех подмножеств Ω , и в ней всего четыре элемента: $\{\emptyset, \{O\}, \{P\}, \{O, P\}\}$. Она включает невозможное событие — отсутствие какого-либо результата (\emptyset), а также тривиальное — получение какого-либо из возможных результатов $\{O, P\}$, то есть все множество элементарных событий.

Если монета честная, то зададим такую вероятность: $P(O) = 50\%$, $P(P) = 50\%$. Кроме того, $P(\emptyset) = 0$, $P(O, P) = 100\%$. Очевидно, что свойство сигма-аддитивности (которая в данном случае сводится к аддитивности) выполняется. Именно поэтому у нас получилось вероятностное пространство.

Дискретным случайным величинам соответствуют конечные или счетные множества, в них естественной (считающей) мерой оказывается обыкновенный подсчет количества элементов. Соответственно, вероятность в дискретном вероятностном пространстве получают с помощью комбинаторного подсчета вариантов, знакомого каждому студенту или интересующемуся математикой школьнику. Для непрерывных случайных величин вероятность как мера больше похожа на длину или площадь. Точное определение *случайной величины* мы дадим в следующей главе, пока же положимся на ее интуитивное понимание как величины, которую можно измерить или наблюдать. Но повторные измерения могут привести к иным результатам, заранее не известным.

Для полноценной работы со случайными событиями и вероятностями вводится одно важнейшее понятие, которое нехарактерно для других мер: независимость событий. С ней и связанной с нею условной вероятностью мы познакомимся в главе 4 и разберемся, что же имеет в виду байесовский спам-фильтр. Впрочем, если читателю уже приходилось решать задачи, в которых появляются независимые события (например, выпадение двух «орлов» при двух

подбрасываниях монеты), то он знает, что вероятность пересечения для независимых событий вычисляется как произведение их вероятностей.

Если заменить в обсуждаемых определениях и свойствах вероятности сумму на «максимум», а произведение на «минимум», можно построить альтернативную теорию. Она называется *теорией возможностей*. Это характерный подход для математики в целом. Начинаем с абстрактных рассуждений: числа образуют определенную структуру с операциями сложения и умножения; замечаем, что на ограниченном числовом интервале можно построить такую же числовую структуру, но с другими операциями: минимум и максимум. Строим понятие меры на новой структуре и выясняем, что она открывает новый взгляд на мир! В отличие от теории вероятностей, здесь можно построить две согласованные меры — *возможность* и *необходимость*. Это направление, созданное американцем азербайджанского происхождения Лотфи Заде, служит основанием для *нечеткой логики* и используется в системах автоматического распознавания образов и принятия решений.

Возможность невероятного

Первое свойство мер: «Мера пустого множества равна нулю», — кажется тривиальным, но оно интересно своей асимметричностью. Если мера подмножества равна нулю, из этого не следует, что оно пусто! Например, линия — это, очевидно, непустое подмножество точек плоскости (и точек в ней бесконечно много), но ее мера на плоскости, то есть площадь, не просто исчезающе мала, а в точности равна нулю. Бывают и более экзотические примеры — канторовы и фрактальные множества, имеющие сложную структуру, содержащие бесконечное число точек, зримо «занимающие» некоторую площадь или объем, но тем не менее имеющие нулевую меру.

С появлением вычислительной техники множества с необычными свойствами сошли со страниц математических книг и журналов в область, понятную широкой публике. Они вызывают интерес

не заложенной в них математикой, а своеобразной гармоничностью, красотой и завораживающей глубиной, которой обладают их визуализации. Треугольник Серпинского, множество Мандельброта и тесно связанные с ним множества Жулиа, как и многие другие математические объекты, стали визуальным символом века компьютерной графики, прежде недоступной человеку (рис. 2.2).

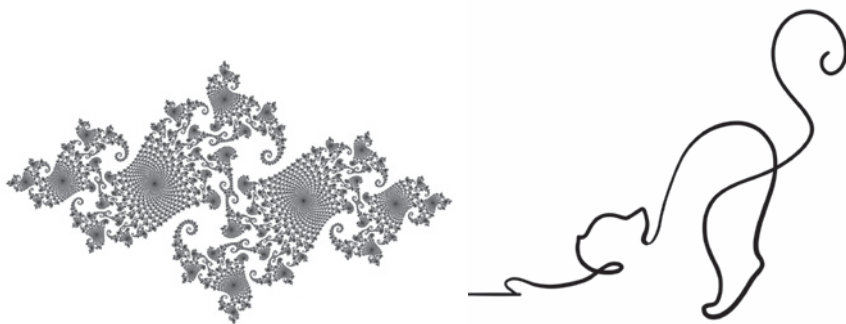


Рис. 2.2. Некоторые красивые объекты нулевой меры: линия на плоскости, спорадическое множество Жулиа

Готовя эту иллюстрацию, я нашел замечательное изображение несвязного множества Жулиа на прозрачном фоне с высоким разрешением. Вставив его в векторный редактор, я столкнулся с забавной трудностью: было очень нелегко попасть курсором в это изображение, чтобы выделить его. Оно такое «рыхлое», что вероятность попадания в закрашенную точку на экране была заметно меньше, чем в прозрачный фон. В вероятностном пространстве тоже могут существовать подмножества нулевой меры, но это не означает, что события из этих подмножеств невозможны. С четвертой-пятой попытки я смог выделить изображение, поскольку точки на экране все-таки имеют конечный размер. Но что было бы, попади в мое распоряжение настоящее несвязное множество Жулиа с бесконечным разрешением?

Представьте себе, что вы пользуетесь программным генератором случайных чисел, который выдает произвольное вещественное число от 0 до 1. Какова вероятность выпадения 0? А 1/2 или

e/π ? Во всех этих случаях ответ — ноль! Вернее, самое маленькое доступное компьютеру положительное число, так называемый машинный эпсилон, ведь машина оперирует конечным числом знаков после запятой. «Подождите, — скажете вы, — в каком смысле ноль? Эти же числа невозможные». Проведем эксперимент. В результате мы получим какое-то конкретное число. Тогда «по построению» вероятность его появления не может быть нулевой. Все верно, но прежде чем выпадет конкретное число, нам придется перебрать бесконечное число случайных чисел! Дело в том, что отдельное число, как точка на отрезке, имеет нулевую меру и честную нулевую вероятность. Отлична от нуля лишь мера сплошного отрезка, пусть даже очень маленького. Именно поэтому мы говорим не о вероятности получить некоторое значение случайной величины, а о *плотности вероятности*, которая при умножении на конечную меру подмножества в вероятностном пространстве даст конечную величину — вероятность попасть в это подмножество.

Любопытно, но, окажись у нас идеальный генератор случайных чисел с бесконечной точностью, вероятность получить с его помощью какое-либо рациональное число* (не какое-то конкретное, а вообще любое) тоже будет равна нулю. Драматизма этому факту придает то обстоятельство, что множество рациональных чисел не просто бесконечно, оно *всюду плотно*. Это значит, что в любой сколь угодно малой окрестности выбранной рациональной точки можно обнаруживать новые и новые рациональные точки. Если мы захотим изобразить это множество графически на числовой оси, то можем брать карандаш и смело рисовать сплошную прямую на ней. Однако и это множество имеет нулевую меру на множестве вещественных чисел! Доказательство того, что рациональные числа образуют плотное подмножество нулевой меры множества вещественных чисел, наделало шума в конце XIX века. В таких случаях математики говорят: *случайно выбранное вещественное число почти наверняка будет иррациональным*. Как бы странно ни звучало, но «почти наверняка» — точный математический термин,

* Напомним, что рациональными называют дроби вида p/q , где p и q — целые числа.

означающий, что событие — дополнение подмножества вероятностного пространства нулевой меры.

Если бы пифагорейцам удалось заглянуть в науку будущего, они пришли бы в недоумение, обнаружив, что верные и понятные рациональные числа — как им казалось, единственно возможные, на которых строилась вся их математика, — практически не встречаются на числовой оси! Вот уж точно — закон подлости! И если в быту мы чаще всего встречаем целые числа или несложные дроби, то даже в повседневной физике или геометрии «работает» большое количество иррациональных зависимостей (корни различных степеней) и трансцендентных функций (синусы, логарифмы и т. п.), делающее рациональные и целые решения редкостью. Среди фундаментальных физических констант нет «фундаментально» рациональных чисел. Некоторые из них — такие как скорость света, заряд электрона, постоянные Планка и Больцмана* — приняты рациональными или целыми по соглашению. Просто единицы измерения подобраны так, чтобы фиксировать количество значимых цифр в этих константах, поэтому в таблицах такие величины указаны «точно», но эта точность в известном смысле искусственная, принятая для удобства.

Если кто-то терпеливо проведет тысячу экспериментов с монеткой и радостно скажет вам, что у него получилось столько же выпадений «орлов», сколько и «решек», можете смело выразить сомнение или поздравить его с редкой удачей. Хоть бросание монетки — дискретный случайный процесс, по мере накопления статистики мощность вероятностного пространства будет расти, а мера события «число “орлов” совпадает с числом “решек”» станет уменьшаться. Можно показать, что вероятность этого «самого вероятного» события уменьшается с ростом числа испытаний как $\sqrt{2/\pi n}$. Для сотни бросаний это около 8%, для десяти тысяч — в десять раз меньше.

Мы еще вернемся к этим рассуждениям в одной из следующих глав, когда зададимся вопросом о том, насколько каждый из нас может считать себя нормальным.

* Эти значения приняты с 20 мая 2019 года.

О коварстве географических карт

Я хочу вернуться к толкованию вероятности и продемонстрировать эквивалентность ее колмогоровского и частотного определений. Мы раскроем загадку одного закона подлости, который не вошел в классические книги по мерфологии, но хорошо известен туристам, геологам и всем, кто пользуется топографическими картами:

То место, куда направляется турист, чаще всего оказывается либо на сгибе карты, либо на краю листа.

Раскроем карту, чтобы найти на ней какой-нибудь объект. Предположим, нас одинаково часто интересуют объекты, расположенные на всех участках карты. Причем не объекты сами по себе как точки. Весь смысл использования карты состоит в обозрении *окрестностей* объекта, некой конечной площади. Пусть нам достаточно будет некоторой малой доли α от площади карты S , чтобы понять, как попасть туда, куда нужно. Если то, что мы ищем, окажется недалеко от сгиба или края карты, скажем ближе какого-то критического расстояния d , мы можем счесть, что закон туриста сработал. Доля таких пограничных площадей в общей площади карты даст нам вероятность испытать этот закон подлости на себе. Вот как выглядят неприятные участки карты при $\alpha = 0,5\%$ и всего одном сгибе (рис. 2.3).

Для окрестности в форме квадрата $a = \sqrt{\alpha S}$. Неприятные полосы будут иметь площадь $d\sqrt{S} = S\sqrt{\alpha}$. Четыре полосы: две вертикальные и две горизонтальные — расположатся у края; любой дополнительный изгиб, горизонтальный или вертикальный, добавит еще одну полосу. А теперь воспользуемся свойством аддитивности мер и вычислим меру объединения всех полосок как сумму их площадей, за вычетом площади пересечений. При этом следует заметить, что пересекающиеся полосы формируют квадратики площадью $d^2 = \alpha S$.

Сложив карту так, чтобы получилось n горизонтальных и m вертикальных изгибов, мы получим суммарную площадь неприятной зоны, равную $(n+2)\sqrt{\alpha}S + (m+2)\sqrt{\alpha}S - (n+2)(m+2)\alpha S$. Разделив ее на площадь всей карты S , получим неприятную долю общей

площади, выраженную только через количество сгибов и α . Отсюда получаем вероятность оказаться в этой доле при случайном выборе объекта:

$$p = (n + m + 4)\sqrt{\alpha} - (n + 2)(m + 2)\alpha.$$



Рис. 2.3. Серым выделены «нехорошие» участки. Отдельно показан участок с полупроцентной площадью для карты размерами 40×50 см, она имеет размер, слегка превышающий 3 см

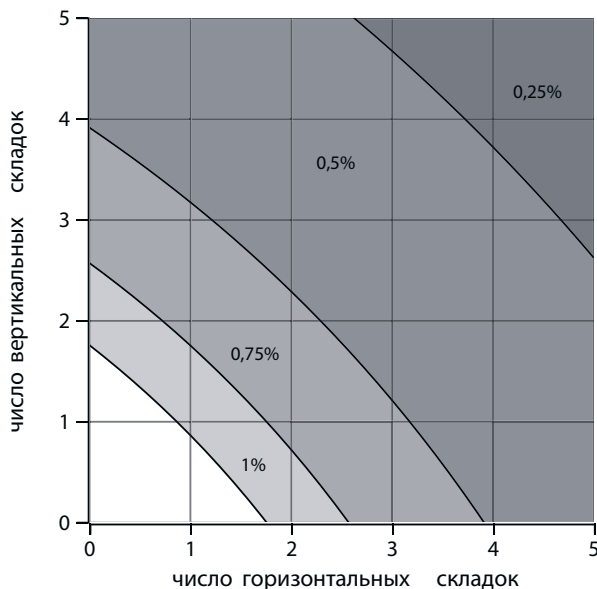


Рис. 2.4. Зоны, в которых вероятность оказаться на сгибе карты или на ее краю, превышают 50%. Числами отмечены значения α

На рисунке 2.4 заливкой показаны области, в которых эта доля превышает 50% для различных значений α . Например, приняв $\alpha = 0,75\%$ и сложив карту вдвое в одном направлении (одна складка) и вчетверо — в другом (три складки), мы найдем, что вероятность попасть в неудобное место превысит 50%.

Чаще всего карты имеют по три вертикальные и три горизонтальные складки, что дает вероятность выполнения закона подлости около 60% при весьма незначительном $\alpha = 0,5\%$.

Проверяем честность реальной монеты

Теперь мы можем вернуться к вопросу, с которого начался наш разговор: насколько может быть честна реальная монетка? Колмогоровское определение вероятности дополнило ее частотное определение и свело его к геометрическому (как к доле «объема» события в общем «объеме» возможностей). Таким образом, доля площади белых полосок на рис. 2.1 отражает вероятность того, что монетка в результате эксперимента не поменяет исходной ориентации, а доля серых — вероятность получить обратную ориентацию. Монетку мы можем считать честным генератором двух этих равновероятных исходов, только если сможем показать, что общая площадь белых полосок равна общей площади закрашенных.

Но вот беда! Если добросовестно рассматривать всю четверть координатной плоскости, то площадь каждой отдельной полоски на диаграмме окажется бесконечной. Более того, и полосок бесконечное число! Как же сравнивать бесконечные суммы бесконечных значений? Нам опять поможет понятие меры. Аддитивное свойство позволит нам аккуратно показать, что бесконечность не мешает площадям серых и белых областей быть одинаковыми. В явном виде уравнения для наших кривых имеют вид $\omega = n/t$. Если площадь под кривой $\omega = 1/t$ равна S , то благодаря свойству аддитивности площадь под кривой $\omega = n/t$ будет равна $S_n = nS$. В свою очередь, для отдельных полосок получаем: $S_n - S_{n-1} = nS - (n-1)S = S$, а это значит, что разница площадей не зависит от «номера» гиперболы. Это

не особенность именно гипербол, тот же вывод можно сделать для любой кривой вида $y = nf(x)$. А раз так, попадания в белую или серую часть диаграммы равновероятны для всей области определения, как и ожидается для «честной» монетки.

Рассуждения, которые мы сейчас привели, кажутся достаточно простыми, но дают весьма общий результат, применимый к любым аддитивным величинам. Абстрактное понятие меры позволило нам сравнивать бесконечные величины, оставаясь в рамках логики и здравого смысла.

Абстракции — это хорошо, но можно возразить, что в реальности мы подбрасываем монетки не со всеми возможными параметрами. Как показали эксперименты со скоростной камерой, при бросании монеты рукой угловые скорости попадают в диапазон от 20 до 40 оборотов в секунду, а длительность полета — от половины до одной секунды. Эта область на рис. 2.1 выделена прямоугольником. В ней суммарная площадь белых полосок чуть больше, чем серых, и можно сделать вывод, что вероятность выпадения той же стороны, что была вверху при броске, составит 50,6%.

В 2007 году Перси Диаконис и соавторы опубликовали статью, в которой дается развернутый анализ процесса подбрасывания монетки. Детальное описание механики летящего и вращающегося диска, который не просто крутится, а еще и прецессирует (его ось вращения сама поворачивается в полете, описывая коническую поверхность), показывает, что при ручном подбрасывании из позиции «орел сверху» вероятность выпадения «орла» составляет 51%. К смыслу этого результата мы еще вернемся.

Откуда же берется случайность?

В сувенирных лавках можно найти магнитные маятники для «выбора желаний». Это тоже механические генераторы случайности, и их иногда ошибочно называют «хаотическими маятниками». Начав движение с каких-то начальных позиции и скорости, маятник совершает ряд «непредсказуемых» колебаний и наконец останавливается в одном из секторов. Однако колебания и здесь не непредсказуемы, просто они очень чувствительны к начальным

условиям. Для каждого сектора, в котором может остановиться маятник, существует *область притяжения* в пространстве координат-скорости. Это множество таких начальных условий, при которых маятник обязательно притянется к определенной точке в указанном секторе. Точка остановки маятника называется *аттрактором* — притягивающей точкой. В случае маятника с рис. 2.5 пространство координат и скоростей четырехмерно, и так просто области притяжения показать не удастся. Но если ограничиться двумя секторами и свести задачу к одномерной (такой маятник называется *осциллятором Дюффинга*), то пространство начальных значений превратится в плоскость, так что области притяжения можно будет увидеть. Они выглядят как замысловатая фигура, напоминающая древний символ «инь-ян» и быстро превращающаяся в узкие полоски, которые разделяют области притяжения.

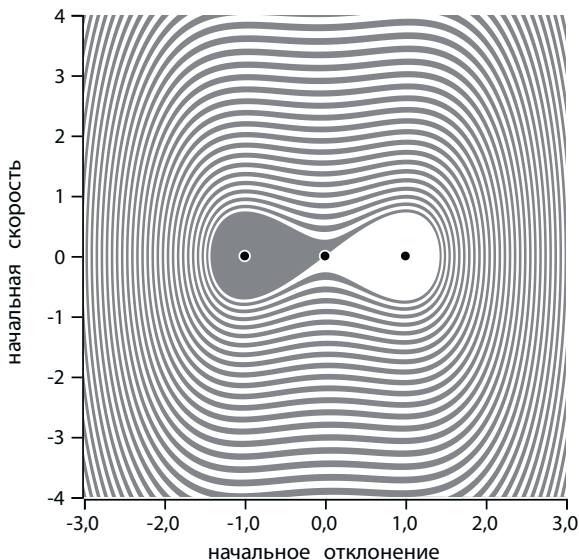


Рис. 2.5. Области притяжения аттракторов для одномерного маятника желаний — осциллятора Дюффинга

Как и в случае с монетой, немного смещая начальные условия, мы попадаем от одного аттрактора к другому. Так же действует и игральная кость, и рулетка, но они не могут считаться сами по

себе генераторами случайности. Это не истинно хаотические системы, и их поведение теоретически можно рассчитать точно. Иначе говоря, вероятностные методы применительно к таким системам помогают восполнить *наше* незнание о них, но не соответствуют неотъемлемым свойствам самих систем.

Но существует ли настоящая случайность, глубинная, невычислимая в принципе, описываемая только на языке вероятностей? Да, причем такие системы можно разделить на два типа: стохастические и хаотические.

Хороший пример истинно стохастической системы — появление автомобилей на дороге. Люди не договариваются, не согласовывают свои планы, каждый элемент ансамбля за пределами дороги действует независимо. И хотя в поведении людей есть определенные закономерности — часы пик утром и вечером, пустые дороги ночью и т. д., — мы не обладаем и никогда не будем обладать достаточной информацией о каждом участнике движения, чтобы предсказать появление любого из них. Можно взлететь над дорогой на вертолете и посмотреть, какие машины мы скоро увидим, расширив наше знание о ней, но и это не будет исчерпывающим описанием системы. Надо еще «взлететь» над временем, чтобы увидеть все прошлое и все причинно-следственные связи между элементами. Однако и этого недостаточно. Нужно заглянуть каждому участнику движения в мозг и выяснить, что он намерен делать и что станет делать, если другие участники изменят его планы. Таким образом, наряду с макроскопическим описанием системы в игру вступает скрытое от нас внутреннее состояние ее элементов, и оно порой выходит на первый план. Другой яркий пример стохастической системы — механика элементарных частиц на квантовом уровне, распад нестабильных атомов, изменения в генетическом коде, а также, видимо, землетрясения и котировки ценных бумаг на бирже. Единственное, что остается исследователю, — рассматривать их как истинные случайные величины и описывать в терминах теории вероятностей.

Есть и другой источник случайностей — *динамический хаос*. Хаотические системы отличаются от стохастических тем, что описываются небольшим числом точных уравнений и параметров, в которых не содержится случайности или скрытой внутренней

структуры. Однако их поведение не просто сложно, а хаотично и истинно непредсказуемо. Если мы начнем раскачивать маятник желаний, пусть очень аккуратно, с предельно точно контролируемой частотой и амплитудой, то обнаружим, что его плавные движения невозможно просчитать надолго. Никакими алгоритмами на сколь угодно точных вычислительных машинах нам не удастся рассчитать точное поведение маятника на произвольно далекое будущее. Он не остановится на каком-либо секторе, а будет совершать свои движения, но никогда не вернется в одну и ту же точку в пространстве координат-скорости дважды. Еще один пример предельно простой хаотической системы — идеальный шарик, подпрыгивающий в поле тяжести на идеальном столике с пружинкой. Сравнительно простые уравнения Лоренца показали, что мы никогда не сможем предсказывать погоду больше чем на пару-тройку недель: это тоже хаотическая система.



Рис. 2.6. Сечение плоскостью странного аттрактора для осциллятора Дюффинга

В XX веке *теории динамического хаоса* удалось объяснить природу такой непредсказуемости. Простой одномерный маятник желаний, который мы рассматривали, имел две устойчивые стационарные точки — два аттрактора, — и одну неустойчивую, от которой система старается уйти; она показана белым кружком на рисунке 2.5. В хаотическом режиме вместо набора аттракторов в системе

появляется бесконечное множество неустойчивых стационарных траекторий. Это множество бесконечно, но имеет нулевую меру и представляет собой очень сложно устроенную несвязную структуру. Попав на одну из таких траекторий, в принципе невозможно ей следовать, используя какие-либо конечные алгоритмы. И вот что самое удивительное — оказалось, это бесконечное множество неустойчивых траекторий само по себе притягивающее! Хаотическая система непрерывно перескакивает от окрестности одной неустойчивой траектории к другой, все время оставаясь в пределах этого странного аттрактора. Так эти множества и называются: *странные аттракторы*. Вот как завораживающе красиво выглядит сечение плоскостью странного аттрактора для одномерного маятника желаний (осциллятора Дюффинга), подверженного гармоническим колебаниям (рис. 2.6). Этот объект можно описать в трехмерном пространстве (отклонение \times скорость \times фаза вынужденного колебания). Если расечь аттрактор в нем плоскостью, то можно увидеть его структуру — это называется *сечением Пуанкаре*. Каждая точка здесь — след траектории, а оттенок точек отражает относительную скорость, с которой траектории разбегаются друг от друга. Вот еще пара красивых странных аттракторов (рис. 2.7).



Рис. 2.7. Слева: сечение Пуанкаре для траектории шарика, подпрыгивающего на подпружиненном столике. Множество точек принадлежит поверхности сферы, соответствующей закону сохранения энергии. Справа: объемная область, которая включает в себе странный аттрактор, рождающийся при вынужденных колебаниях толстой пластины

Гладкость хаотической траектории позволяет немного заглянуть в будущее хаотической системы. Это объясняет одно досадное наблюдение: с одной стороны, синоптики порой не могут уверенно предсказать погоду на неделю, а с другой, если вы скажете, что завтра будет такая же погода, как и сегодня, то не ошибетесь примерно в трех случаях из четырех. Вообще же анекдоты о синоптиках несправедливы, и нужно отдать должное человеческой мысли и упорству, которые позволили предсказывать погоду на современном уровне!

Динамический хаос очень сложен и красив как теория, он порождает изумительные по элегантности образы, но может быть и полезен. Например, алгоритмы, с помощью которых генерируются случайные числа в компьютерах, тоже детерминированы. Для всех примеров в этой книге я применял генератор псевдослучайных чисел, который не использовал какой-нибудь реальный стохастический процесс (альфа-распад или подсчет машин на дороге), а вычислял следующее «случайное» число на базе предыдущих, полученных им ранее.

От монеток к бабочкам и самой судьбе

Наблюдения за тем, как малые отклонения вырастают в глобальные изменения системы, приводят к мысли об «эффекте бабочки». Напомню, что под ним подразумевается цепочка далеко идущих драматичных последствий от некоторого незначительного, на первый взгляд, события. Раздавленная исследователями прошлого бабочка в рассказе Рэя Брэдбери «И грянул гром» привела к кардинальной перестройке будущего. А одну из своих лекций Эдвард Лоренц, создатель теории динамического хаоса, озаглавил так: «*Может ли взмах крыльев бабочки в Бразилии вызвать торнадо в Техасе?*».

На этот эффект мы неявно ссылаемся, сетуя: «*Не поверни я за угол, все было бы иначе!*», «*Не сел бы он в этот поезд, с ним не случилось бы катастрофы!*» или «*Из-за такой мелочи разругались и разошлись!*» Но мы видим, что сосуществуют истинно стохастический

квантовый мир и сверхточные атомные часы, устойчивые гамильтоновы системы в мире звезд и галактик и хаос колец Сатурна или пояса Койпера, тепловое движение молекул и удивительная точность работы биологических систем или механизмов автомобиля. Нет, взмах крыла бабочки не рождает ураганов, а бесследно исчезает, порождая цепочку вихрей, передающих энергию и информацию все более и более мелким вихрям, пока и энергия, и информация не исчезнут в хаосе флуктуаций. Надо четко понимать, что малые отклонения приводят к кардинальной перестройке системы только в случае, если она неустойчива либо находится на пороге *бифуркации*, или *катастрофы*, — так на языке математики называются глобальные перестройки в поведении системы при малых непрерывных изменениях параметров. Бифуркации всегда образуют множества нулевой меры в пространстве параметров — это точки или границы. Малые возмущения не приводят к катастрофам *почти всюду* (это тоже точный термин, означающий «везде, кроме множества нулевой меры»), а неустойчивые состояния в природе наблюдаются редко, не проходя «проверку временем».

Если пара молодых людей распалась «из-за ерунды», ей суждено было разойтись в любом случае, она была неустойчивой. Устойчивые пары проходят войны и голод, а потом, бывает, и распадаются, но не из-за мелочей, а в результате глубоких перемен, которые могут произойти с личностью в течение жизни. В цепочке событий, приведших к катастрофе поезда, нелегко однозначно выделить ключевое, конкретную ошибку или роковую случайность. Скорее всего, ключевым будет не событие, а систематическое нарушение правил, приводящее систему к неустойчивому состоянию. Если в системе множество параметров и ряд из них случаен (а наша жизнь устроена именно так), то информация в ней имеет свойство теряться и уже никак не удастся восстановить, в какой именно момент в нашей жизни «все пошло не так». Мы поговорим о роли памяти в случайных процессах через две главы. Не терзайте себя сожалениями о случившемся, а присмотритесь к происходящему сейчас, чтобы не пропустить настоящей точки бифуркации.

В связи с этим можно вспомнить один из законов мерфологии, который некий Дрейзен назвал **законом восстановления**:

Время улучшения ситуации обратно пропорционально времени ее ухудшения.

В качестве примера приводится следующее наблюдение: *на склеивание вазы уходит больше времени, чем на то, чтобы ее разбить*. Этот закон удивительно точно описывает соотношение между характерными скоростями для процесса *релаксации* устойчивой системы, которую можно описать убывающим экспоненциальным законом $e^{-\lambda t}$ и скоростью развития *катастрофического* процесса в неустойчивой системе, в линейном приближении — экспоненциального роста малого возмущения $e^{\lambda t}$. Эти скорости действительно обратно пропорциональны друг другу.

В примере с вазой процесс склеивания — не релаксация, не переход к наиболее вероятному состоянию. Он ближе к другому процессу — *самоорганизации*, — который в первом приближении описывается *логистическим* законом и ближе по скорости к релаксации, чем к катастрофе (рис. 2.8).

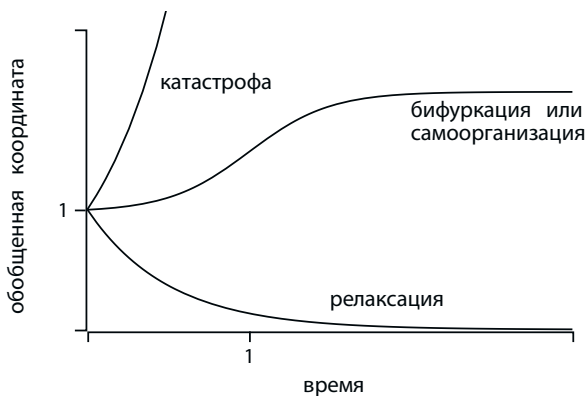


Рис. 2.8. Типичные нестационарные процессы: катастрофа, релаксация и самоорганизация, — имеющие одинаковое характерное время

Иногда, гуляя в снегопад, я удивляюсь тому, что снежинка падает мне на нос. Удивляюсь оттого, что вероятность этого события ничтожно мала. Если рассудить, снежинка родилась высоко в небе над

Тихим океаном, кружилась в беспорядочных турбулентных потоках в облаке, падала, непрерывно меняя направление движения... чтобы попасть на кончик моего носа на Камчатке! А какой ошеломительный путь прошли фотоны от далекой звезды?! Десятки тысяч лет они неслись сквозь Вселенную, их не поглотила пыль, им не встретился астероид! Родились они в бушующем квантовом мире далекой звезды, а закончили свой путь в квантовом мире белка опсина на сетчатке в моем глазу. Даже считать вероятность этого события нет смысла, она исчезающе мала. Но событие случается, и я вижу мерцающий свет звезды. Теперь понятно: это все потому, что площадь моего носа и даже молекулы опсина имеют ненулевую меру. Но все равно удивительно: то, что почти наверняка не должно было произойти, все же происходит!

О роли предопределенности или случайности в нашей судьбе, об истинности или призрачности нашего знания о природе пусть спорят философы. Я же призываю читателя взглянуть на мир с высоты математических абстракций и восхититься его красотой и согласованностью.

Глава 3

Головокружительный полет бутерброда с маслом

Тема падающих бутербродов не дает покоя ни широкой публике, ни исследователям. Десятки лет проводятся эксперименты, снимается кино, пишутся статьи, падающий бутерброд обрастает легендами и неправильными выводами. Мало какая столь же бесполезная задача привлекала к себе такое внимание. И если вы думаете, что это баловство, то имейте в виду, что за ее решение даже премии дают — правда, тоже несерьезные. В 1996 году Роберт Мэтьюз получил Шнобелевскую премию за работу «Падающий бутерброд, закон Мёрфи и фундаментальные константы»*, опубликованную в *European Journal of Physics*. Несмотря на шуточную тему и соответствующую реакцию научного сообщества, это небезынттересная статья, в которой проводится тщательный анализ процесса соскальзывания и делается далеко идущий вывод: на какой бы планете ни возникли антропоморфные существа, живущие в атмосфере, они будут обречены на закон бутерброда. После такого триумфа бесполезных исследований можно бы тему и закрыть, но зачем упускать возможность рассмотреть

* Matthews R. A. Tumbling toast, Murphy's Law and the fundamental constants // *European Journal of Physics*. 1995. Vol. 16. No. 4. Pp. 172–176.

на примере занятой задачки интересные и объективно полезные методы!

Айда кидать бутерброды в Монте-Карло!

Мы редко подбрасываем бутерброды, как монетку, — по крайней мере, когда становимся старше двух лет. Чаще всего мы невольно повторяем примерно один и тот же эксперимент: бутерброд, изначально расположенный маслом вверх, выскальзывает из рук или съезжает со стола. В процессе соскальзывания он закручивается, летит в воздухе и наконец шлепается на стол или на пол. На начальный этап падения влияет ряд параметров: трение о пальцы или поверхность стола, начальное положение бутерброда и его начальная скорость, высота падения — наконец, размеры бутерброда. Налицо динамическая система с несколькими входными параметрами и одним выходным — положением бутерброда на полу. Внутри системы, как и в случае с монеткой, работают механические законы, которые описываются дифференциальными уравнениями, и они *детерминистические*. Это значит, что в них нет никаких случайностей. Результат зависит только от входных данных, и при точном повторении параметров мы должны получать идентичные результаты. Это относится к модели бутерброда, представленной в виде системы дифференциальных уравнений. А что насчет настоящих бутербродов, шероховатых и неповторимых, роняемых настоящими людьми в ресторанах, на улице или на диване? Изменчивость реального мира можно описать, подавая на вход детерминистической системы случайные параметры.

Однако даже алгебра случайных величин, включающая в себя лишь сложение и умножение, — дело непростое, а у нас дифференциальные уравнения! Мы не полезем в эти увлекательные дебри, а используем отработанную во многих областях технику — *метод Монте-Карло*. Он состоит в определении свойств некой сложной системы в результате многократных испытаний с различными случайными параметрами. Подчеркну еще раз: исследуемая система не

стохастична и не хаотична, и на случайные входные данные она реагирует предсказуемо. В методе Монте-Карло случайность нужна лишь для того, чтобы эффективно перебрать как можно больше вариантов и заглянуть во все реалистичные «углы», получив представление о поведении системы. Это универсальный метод, применяемый в самых разнообразных задачах. Обычно студенты впервые знакомятся с методом Монте-Карло, изучая численное интегрирование, например вычисляя площадь какой-либо сложной фигуры, задаваемой системой неравенств, которая не имеет приличного аналитического представления. То обстоятельство, что вероятность — мера, позволяет использовать метод Монте-Карло для вычисления мер (площадей и объемов) геометрических фигур.

Особенность предстоящего эксперимента с бутербродом состоит в том, что нас интересует зависимость вероятности того или иного его исхода от параметров задачи. Мы будем искать ответ на вопрос: при каких обстоятельствах выполняется закон бутерброда? Станем подавать на вход нашей динамической системы различные конкретные параметры и набирать статистику по падениям маслом вверх и маслом вниз. И результатом ряда экспериментов будет число — вероятность падения маслом вниз.

Я убежден, что намеренно ронять на пол настоящие бутерброды из хлеба и масла неправильно, поэтому воспользуемся математическим моделированием. Для решения задачи я взял один из доступных симуляторов физического мира, которые используют для создания онлайн-игр. Он легко позволил создать виртуальные стол и пол, а также два бутерброда. Один оказывался на краю стола, а второй «выскальзывал из пальцев», то есть соскальзывал с точечной опоры (рис. 3.1).

В моих силах задать все параметры задачи: начальные позицию и угол бутерброда, горизонтальную скорость для случая смахивания со стола, коэффициенты трения, размеры бутерброда и высоту падения. В момент, когда бутерброд касается пола, фиксируется угол бутерброда, вернее угол вектора, нормального к нему. О том, с какой стороны оказалось масло, нам скажет знак синуса этого угла: положительному значению соответствует удачный случай, а отрицательному — положение маслом вниз. Результат заносится

в таблицу, и новый виртуальный бутерброд готов к падению. Задачу мы поставим такую: оценить вероятность приземления бутерброда маслом вниз при его падении с заданной высоты.

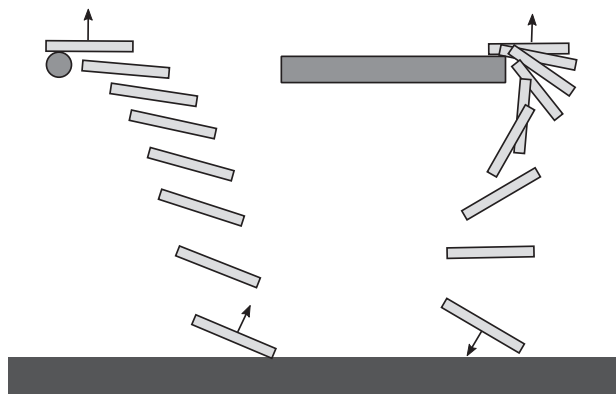


Рис. 3.1. Математические эксперименты с бутербродами

При этом мы ничего пока не будем говорить о масле. Но обещаю, что ему будет посвящен отдельный разговор, где мы подробно рассмотрим его роль в этом законе.

Как правильно говорить о случайных величинах

Метод Монте-Карло подразумевает, что в качестве параметров используются случайные переменные. И здесь наконец пора разобраться с тем, что же такое *случайная величина*.

Вернемся к математическим структурам. Какой структурой можно моделировать результаты выпадения числа на игральной кости или уровень воды в реке, ведь там постоянное волнение? Как работать с числом автомобилей, проезжающих перекресток в течение часа? Какой структурой можно описать состояние электрона в атоме водорода? С одной стороны, это конкретные числа из вполне определенного множества значений: для кости, например, из множества $\{1, 2, 3, 4, 5, 6\}$, — и какое-нибудь значение легко получить, проведя эксперимент. Однако повторный опыт даст иной

результат — это явно не просто число: сегодня оно одно, завтра другое. Может даже возникнуть философский вопрос: а имеет ли смысл говорить о каком-то точном значении «уровня воды в реке» или числе автомобилей, ведь эти величины невозможно «поймать» и зафиксировать? Возможно ли в каком-либо смысле *точное* знание о случайной переменной?

Часто, говоря о таких случайных величинах, ограничиваются одним средним значением, и мы говорим о «средней скорости в час пик» или об «орбите электрона». Но это отличный способ запутаться или даже намеренно запутать. Если фраза «средняя скорость в час пик равна 15 км/ч» дает неплохое представление о ситуации на улице в целом, то переучивать студентов-физиков от мышления орбитами к оперированию волновыми функциями уже весьма непросто. Ну и, наконец, какой смысл в среднем значении числа, выпадающего на игральной кости? Посчитать-то его можно, любой с этим справится: $(1 + 2 + 3 + 4 + 5 + 6) / 2 = 3,5$. Но это число не говорит ровным счетом ничего о рассматриваемой случайной величине. Его даже нет на гранях кубика.

Может быть, нужно указать два числа: среднее и дисперсию? Это уже лучше, но опять же пример с игральной костью показывает, что это явно не вся информация об интересующем нас объекте. А что, если случайные величины — не числа, а множества? Скажем, уровень воды в реке можно попытаться описать интервалом возможных значений с учетом волнения, а для примера с машинами сказать, что за час проезжает от 1 до 100 автомобилей и т. д. Но легко увидеть, что и множества возможных значений тоже недостаточно: например, при многократном повторении измерения количества автомобилей на улице какие-то числа будут встречаться чаще, а каких-то мы не дождемся вовсе.

В предыдущей главе, определяя вероятность, мы ввели меру как функцию на вероятностном пространстве. Для случайной величины элементарными событиями этого пространства будут элементы области ее определения, а мерой задается *распределение вероятностей* для этой величины. И вот это уже исчерпывающая и точная информация. Итак, подводим итог: случайная величина однозначно и полностью характеризуется своим распределением. Распределение, в свою

очередь, представляет собой функцию. Ее область определения — множество возможных значений случайной величины, а область значений этой функции — вероятности для этих значений.

Для уровня воды в реке или скорости машин распределение может быть выражено в виде гладкой колоколообразной кривой. Количество машин, зафиксированных на дороге в единицу времени, должно быть натуральным числом, и его распределение можно представить в виде дискретной функции, определенной только на натуральных числах, или точной формулы. Наконец, моделью игральной кости может быть таблица, показывающая вероятность выпадения каждого из возможных чисел (рис. 3.2).

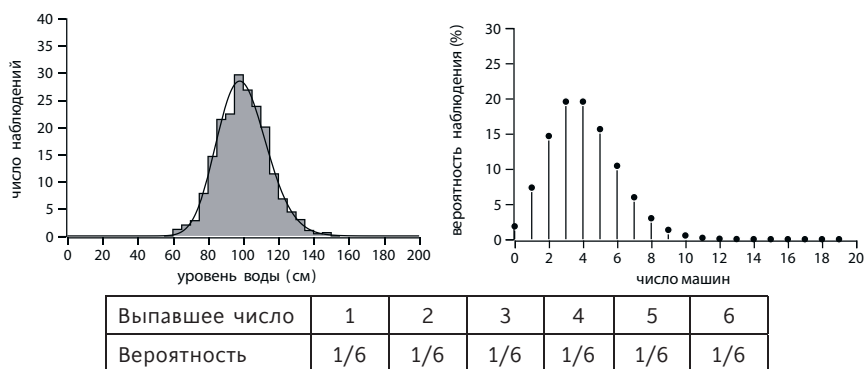


Рис. 3.2. Примеры представления распределений различных случайных величин

Функции можно представлять аналитически или в виде приближения другой функцией, таблицы, гистограммы либо графика. Все эти представления — модели одного и того же объекта, случайной величины. Самое важное тут — не столько конкретный вид представления, сколько математические свойства этой функции. Для распределений вероятностей свойства бывают разными: *количество параметров, количество мод, энтропия, бесконечная делимость, аддитивность, устойчивость, интегрируемость* и т. д. Изучением распределений и их свойств занимается теория вероятностей. Но на практике часто встречается иная задача: необходимо найти модель для случайной величины, если мы не имеем полной информации о ней, но значения которой можем наблюдать, проводя эксперименты.

Из огромного арсенала известных распределений с точно определенными свойствами исследователь выбирает не столько «самую похожую» функцию, сколько функцию, наиболее совпадающую по свойствам с наблюдаемой случайной величиной. Это составляет суть *статистического анализа*, который знаком каждому студенту, прикоснувшемуся к математической статистике.

Сейчас нам нужно задать параметры бутерброда случайными числами, не имея статистических данных, а руководствуясь лишь нужными нам свойствами этих величин. Это важная и интересная часть метода Монте-Карло, от которой зависят и решение, и его корректность.

Размеры бутерброда. Какими они могут быть? Разумной величины канане имеет сантиметра три в ширину, а студенческий добрый «лапоть» может быть сантиметров пятнадцать. При этом вероятность встретить бутерброд миллиметровой или метровой ширины в практическом смысле равна нулю. Больше про бутерброды я ничего сказать не могу и приму их размеры *равномерно распределенными* в указанном диапазоне (рис. 3.3). Запишем это так:

$$l \sim \text{Uniform}([3,15]).$$

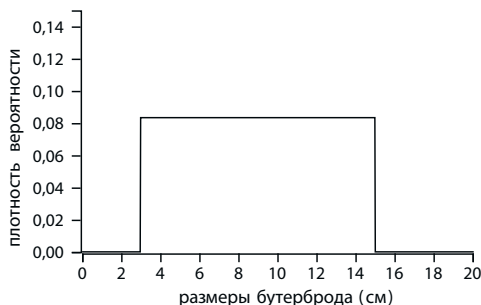


Рис. 3.3. Возможное распределение для размеров падающего бутерброда

В случае равномерного распределения на некотором отрезке $[a, b]$ случайная величина имеет всюду одинаковую плотность, равную $1/(b-a)$. В этом случае плотность распределения принимает вид прямоугольника, а вероятность попасть в какой-нибудь отрезок

пропорциональна его длине. Такой выбор не идеален: всё же средние бутерброды мы встречаем чаще крошечных или гигантских. Но позже мы увидим, что это слабое место можно изящно обойти.

Начальное положение. Тут мы, не мудрствуя, зададим равномерное распределение для смещения бутерброда за край стола, лишь бы он вообще упал:

$$dl \sim \text{Uniform}([l/2, l]).$$

Коэффициент трения. Это неотрицательная безразмерная величина, зависящая только от материала. Столы и скатерти бывают разные, пальцы сжимают бутерброд с разной силой. Диапазон коэффициента от 0,01 до 0,9, при этом крайние значения маловероятны, в среднем можно ожидать около 0,3. Для моделирования неизвестного коэффициента трения нам поможет любое колоколообразное несимметричное распределение неотрицательной величины (рис. 3.4), например *гамма-распределение*:

$$\mu \sim \text{Gamma}(8, 25).$$

Оно будет часто встречаться в этой книге. Почему? Об этом вы узнаете в самом конце.

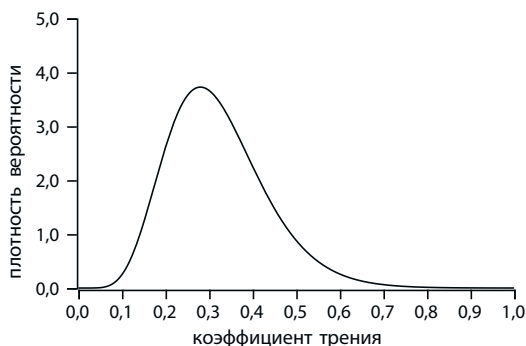


Рис. 3.4. Возможное распределение для коэффициента трения между бутербродом и поверхностью стола

Начальная скорость. Мы редко запускаем бутерброды с большой скоростью, чаще всего не кидаем их вовсе, но всё же иногда

смахиваем. Про величину скорости известно лишь то, что она положительна. Можно предположить, что при смахивании в среднем мы движемся так же, как в среднем руки, — со скоростью около 0,5 м/с. Если про случайную величину известно только это, то ее разумно описать экспоненциальным распределением (рис. 3.5):

$$v_0 \sim \text{Exp}(2).$$

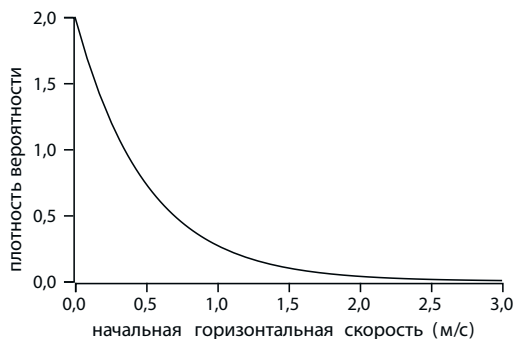


Рис. 3.5. Возможное распределение для скорости, с которой бутерброд смахивается со стола

Его мода (положение максимума на графике) равна нулю, так что доля бутербродов, упавших без большой начальной скорости, будет вполне приличной. В тонком «хвосте» окажутся бутерброды, нечаянно запускаемые в полет при смахивании крошек со стола. Тут стоит обратить внимание на то, что экспоненциальное распределение, вообще говоря, отлично от нуля на всей положительной полуоси; а это значит, что ненулевую вероятность имеют и сверхзвуковая, и сверхсветовая скорости. Однако вероятность наблюдать их при указанном параметре чрезвычайно мала: для скорости, превышающей 10 м/с, она равна одной миллиардной, так что этой опасностью вполне можно пренебречь.

Эксперимент строился так: я «ронял» со стола фиксированной высоты сотню бутербродов, подсчитывал среди них долю тех, что упали маслом вниз, и, используя частотное определение вероятности, отражал на графике зависимость вероятности такого исхода от высоты стола. Вот что у меня получилось (рис. 3.6).

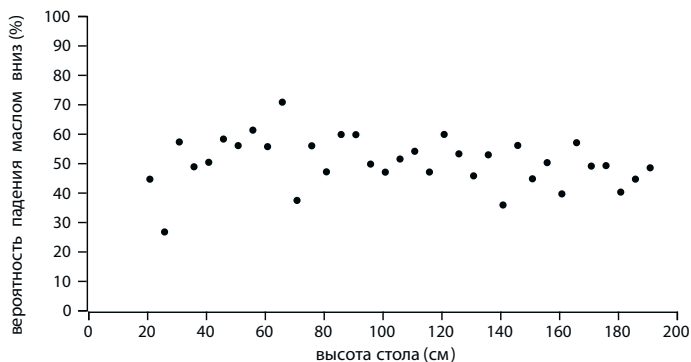


Рис. 3.6. Вероятность приземления маслом вниз разных бутербродов с разными условиями в зависимости от высоты падения. Для каждой высоты проводилось 100 испытаний

Какая-то тенденция видна, но в глаза не бросается. При усреднении получается, что искомая вероятность от высоты стола почти не зависит и едва превышает 50%. Можно ли доверять такому эксперименту? Опровергает ли он закон бутерброда? Может, мы недостаточно много бросали бутербродов — вон какие шумные получились данные!* Увеличим число бросаний и посмотрим, что получится (рис. 3.7).

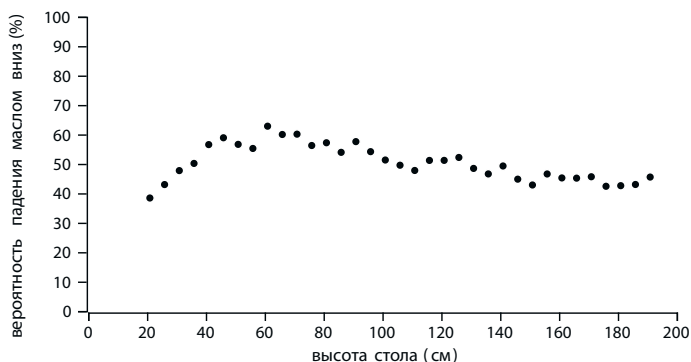


Рис. 3.7. Вероятность приземления маслом вниз разных бутербродов, посчитанная для большего числа испытаний (по 500 на каждую высоту)

* Действительно, 100 бросаний — это мало. А почему и по сравнению с чем мало, мы обсудим в следующей главе.

Выбросов стало меньше, но еще отчетливее видно, что закон бутерброда какой-то невыразительный. Отклонения от 50% не настолько значительны, чтобы стоило говорить о каком-то «законе». Что же, мы готовы его развенчать?

Метод Монте-Карло выглядит заманчиво простым: знай себе подставляй какие попало данные и смотри, что получается. Математика — честная штука: на какой попало вопрос она готова дать какой попало ответ. А вот имеет ли смысл этот ответ, сильно зависит от вопроса. Правильно ли мы проводили наши эксперименты?

Как правильно задавать вопрос природе?

Перед тем как приступить к экспериментам, не таким игрушечным, как у нас, а настоящим и дорогостоящим, использующим орбитальный спутник, ускоритель элементарных частиц или тысячу настоящих бутербродов с маслом, необходимо провести подготовительную работу. И один из мощных и красивых методов, позволяющих понять, как верно и оптимально провести эксперимент, — *анализ размерностей задачи*.

Механику бутерброда мы рассчитывали, пользуясь импульсами и силами — физическими величинами, которые, в свою очередь, связаны уравнениями аналитической механики. И вновь это не просто числа. В физике количественные величины, которые мы измеряем и подставляем в уравнения, не «умещаются» в поле чисел. Они оснащены дополнительной структурой, которая называется *размерностью*. Не все корректные математические выражения имеют смысл, если в них участвуют размерные величины. Скажем, нет смысла складывать скорость и массу, невозможно сравнить силу и расстояние. Однако можно рассмотреть произведение скорости и массы, получив новую размерную величину — количество движения, или импульс; можно возвести скорость в квадрат и поделить на расстояние, получив таким образом величину, имеющую размерность ускорения.

Анализ размерности и теория подобия родились давно. Со времен лорда Рэля они используются в механике, электродинамике, астрофизике и космологии, позволяя с пугающей изящностью подходить к решению очень сложных задач. Однако исследования в этой области не завершены, и строгое определение структуры, образуемой количественными (размерными) величинами, было дано лишь в 2016 году испанским математиком Альваро Рапозо*.

Ограничения, накладываемые размерностями на физические формулы, часто воспринимаются учениками и студентами как лишняя морока, за которой нужно следить. Но логически согласованные ограничения чрезвычайно полезны! Они отсеивают неверные выражения, позволяют «предвидеть» структуру решения физической задачи до ее детального разбора, это мощный инструмент при планировании и анализе экспериментальных данных.

Но вот что важно. Мы рассчитывали падение бутерброда в компьютерной программе, используя не размерные, а обыкновенные числа. Как можно «освободить» физическую величину от размерности и превратить в число? Для этого предназначены хорошо нам знакомые *единицы измерения* физических величин: все эти *метры, фунты, минуты и ньютоны*. Единицы измерения берут на себя размерную часть величины, оставляя нам множитель — вещественное число, с которым уже может иметь дело вычислительная машина. Например, скорость в выбранном направлении величиной 72 км/ч можно представить числом 72. Но тут есть тонкость: от выбора единиц измерения зависит числовое представление. При других единицах (скажем, метрах и секундах) *эта же скорость* будет представлена другим числом: 20. Числа разные, но величина одна, и она не зависит от конкретных единиц.

Возникает вопрос: существует ли в каком-либо смысле «самая лучшая» система единиц? Оказывается, да, но для каждой задачи

* Сами по себе размерности образуют так называемую *свободную абелеву группу*, а размерные величины — *локально тривиальное расслоение*. Я не буду здесь расшифровывать, что означают эти термины: в двух словах это не получится. Пусть для заинтересованного читателя упоминание об алгебраических структурах будет указателем направления, с которого начинается настоящая математика.

она своя. При решении нужно использовать в качестве единиц измерения размерные величины, входящие в задачу.

В этой главе у нас летают бутерброды, в предыдущей — монетки. Приведем еще один пример из области полетов. Как сравнивать летные качества различных птиц? Понятно, что скорости, которые развивают птицы, различны: у голубя — 90 км/ч, у стрижа — 140 км/ч, у журавля, воробья или кряквы — 50 км/ч, у колибри — 80 км/ч. Но все эти птицы существенно различаются по размерам и манере полета. Если длину попугая измерять в попугаях, а время — в периодах взмаха его крыльев, можно получить некую, как говорят, *собственную скорость попугая*. Можно скорости, которые способны развивать эти птицы, разделить на собственные значения и получить безразмерную скорость, показывающую, на сколько длин корпуса может переместиться птица за один взмах крыльев. Вот что получается при таком сравнении.

Птица	Стриж	Голубь	Колибри	Журавль	Кряква	Воробей
Скорость, км/ч	140	90	80	50	50	46
Длина тела, см	18	30	8	100	40	12
Частота взмахов, 1/с	5	5	200	2,5	9	13
Собственная скорость, м/с	0,9	1,5	16	2,5	3,6	1,6
Безразмерная скорость	43	17	1,4	5	3,8	8

Видно, что стриж по праву считается лучшим летуном, а вот колибри неэффективно расходует энергию. Впрочем, у этой птицы нет задачи лететь долго, как у голубя. Одинаковые абсолютные скорости журавля, воробья и утки существенно разнятся при переводе в безразмерные величины. Такого рода расчеты используются, чтобы моделировать настоящий большой самолет, испытывая маленькую модель в аэродинамической трубе. Если все безразмерные параметры этих двух систем близки, они могут считаться *физически подобными*, и моделирование имеет смысл. Мы уже пользовались таким подходом, отражая на диаграммах Лоренца относительные

единицы вместо абсолютных. Это позволяло нам сравнивать различные явления и распределения.

Какой будет самая подходящая система единиц при анализе полета бутерброда? Длину и стола, и бутерброда надо измерять не в сантиметрах или метрах, а в бутербродах. За единицу времени можно взять величину $\sqrt{l/g}$, где l — длина бутерброда, а g — ускорение свободного падения. Легко убедиться, подставив какие-нибудь единицы измерения, что эта величина имеет размерность времени. Получив результат таким путем, мы сразу можем обобщить его как для крошечного канале, так и для солидного «лаптя». Итак, повторим наши вычисления, благо виртуальные бутерброды у нас не закончатся никогда, отражая на графике высоту стола в собственных единицах. Если мы всё сделаем правильно, то для двух разных по размерам бутербродов мы должны получить очень похожие графики. Проверим это (рис. 3.8).

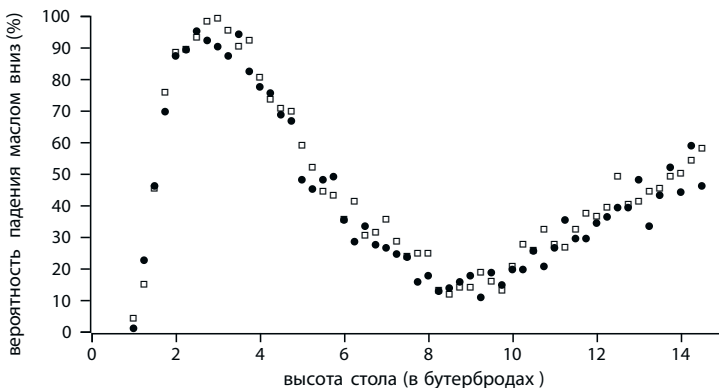


Рис. 3.8. Вероятность приземления маслом вниз бутерброда некой фиксированной величины при различной высоте падения, определенной в собственных единицах задачи. Черные точки соответствуют бутерброду размером 5 см, белые — 10 см

В первоначальной постановке задачи мы, перебирая различные размеры, получали облако результатов, в котором оказалась скрыта интересующая нас зависимость. При увеличении числа испытаний мы это облако усреднили и получили неинтересный ответ, лишенный важных деталей. Чтобы ярче показать, в чем состояла

методическая ошибка, представьте, что мы захотим вычислить вероятность падения бутерброда маслом вниз, перебирая случайным образом и начальные условия, и размеры бутерброда, и высоту. Это равносильно усреднению всех результатов разом. В итоге мы получим уверенную серединку — вероятность, очень близкую к $1/2$, как при подбрасывании монеты! Очень логичный и ожидаемый результат, но он неинтересен. Усредняя множество данных для разных размеров, мы уже приблизились к такому выводу. Но если цель моделирования состоит в выявлении закономерности, то имеет смысл минимизировать число параметров.

Обезразмеренные данные теперь четко говорят в пользу нашего закона, ограничивая его, однако, определенным диапазоном высот: от 2 до 5 размеров бутерброда (от высоты локтя над столом до высоты руки сидящего человека). За пределами этого диапазона у бутерброда повышается шанс повернуться более выгодной для нас стороной перед падением.

А что, если заглянуть дальше и кидать бутерброды из окна? Понятно, что при падении с большой высоты уже неважно, какой стороной приземлилось то, во что превратится бутерброд, и сопротивление воздуха стабилизирует падение, но чисто теоретически: что мы ожидаем увидеть? Наверное, должны наблюдаться некие колебания вероятности по мере увеличения времени полета. Давайте посмотрим (рис. 3.9).

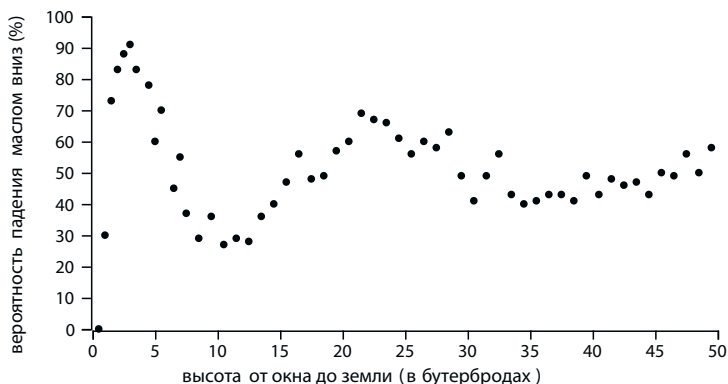


Рис. 3.9. Вероятность приземления маслом вниз бутерброда при падении с большой высоты

В целом форму зависимости мы угадали, но любопытно, что амплитуда колебаний вероятности уменьшается и она сходится к 50%. О чем это может говорить? Тот же ли это эффект, что и в случае с монеткой, когда при увеличении длительности полета становятся более существенными последствия отклонений начальных условий? Оказывается, в данном случае природа выравнивания вероятностей иная.

Еще немного анализа размерностей

Какой бы несерьезной ни была тема нашей книги, мы говорим на языке математики, а он стремится к точным решениям. Можно встретить даже такую фразу: «Если для решения вам понадобился только компьютер, то это еще не математика». Метод Монте-Карло позволил нам получить представление о решении, но это то, что называется грубой силой. Это совсем не так интересно, как хоть какое-то, но аналитическое решение.

Анализ размерностей позволит нам построить теоретический вид зависимости, полученной методом Монте-Карло. Для этого не понадобится решать дифференциальные уравнения; более того, рассуждения не выйдут за пределы вполне примитивных и очевидных соотношений. В том и состоит очарование анализа размерностей — который, впрочем, иногда выглядит фокусничеством. Итак, приступим, ограничиваясь для простоты соскальзыванием бутерброда длины l со стола высоты H с нулевой горизонтальной скоростью.

1. Угол поворота падающего бутерброда зависит от времени и угловой скорости:

$$\varphi = t\omega.$$

2. Угловая скорость равна произведению времени соскальзывания и углового ускорения:

$$\omega = t_0\varepsilon.$$

3. Время соскальзывания можно выразить через ускорение свободного падения и часть длины бутерброда, которая соприкасалась со столом, следующей пропорцией:

$$t_0 \propto \sqrt{l_0 / g},$$

где l_0 — длина части бутерброда, лежавшей на столе. Здесь мы используем отношение пропорциональности, обозначенное знаком \propto . Выражение $y \propto x$ можно переписать как $y = Cx$, где C — некая неизвестная константа. Я очень люблю это отношение. Пропорциональность «вбирает в себя» все сложное, что превращается в константу: и то, что при повороте меняется момент силы тяжести, и то, что при соскальзывании перемещается центр вращения. Все это, конечно, нужно знать для точного расчета, но в результате получится только безразмерный коэффициент, а в нашем анализе он не играет роли. Одним значком мы избавили себя от утомительного интегрирования.

4. Угловое ускорение происходит от ускорения силы тяжести и зависит от плеча, к которому она прилагается:

$$\varepsilon \propto \sqrt{\frac{g}{l-l_0}}.$$

И опять знак \propto позволил нам не вычислять момент инерции пластины для оси, лежащей в ее плоскости, а также изменяющейся проекции силы тяжести (это еще два интеграла).

5. Наконец, время падения зависит от высоты стола и ускорения свободного падения:

$$t \propto \sqrt{H / g}.$$

6. Подставляя все эти выражения в первую формулу, получаем результат:

$$\varphi \propto \sqrt{\frac{l_0 H}{l(l-l_0)}},$$

который, если измерять все длины в бутербродах, превращается в

$$\varphi \propto \sqrt{\frac{xh}{1-x}}.$$

Здесь $l_0 = xl$ и $H = hl$. Что ж, все сходится: угол — величина безразмерная, и зависит она от безразмерных коэффициентов — но не от масштаба времени. Остается чистая геометрия. Знаменатель не опасен — при $x > 0,5$ бутерброд не упадет вовсе (мы рассматриваем нулевую горизонтальную скорость), так что $0 < x < 0,5$.

То, какой стороной упадет бутерброд, определяется знаком синуса угла φ , то есть функцией $\text{sign}(\sin\varphi)$. Она возвращает -1 для случая «маслом вверх» и 1 для «маслом вниз». Мы можем использовать эту функцию для выражения вероятности падения детерминистического бутерброда, если приведем ее к диапазону от 0 до 1:

$$P_{\downarrow}(x, h) = \frac{1}{2}[1 + \text{sign}(\sin\varphi)] = \frac{1}{2}\left\{1 + \text{sign}\left[\sin\left(C\sqrt{\frac{xh}{1-x}}\right)\right]\right\},$$

где стрелочка в индексе символически означает ориентацию масла. Коэффициент C , появившийся в формуле вероятности, выражает все то, что осталось спрятанным, с помощью знака пропорциональности. Это действительно очень хитрый ход, он избавил нас от утомительного интегрирования (и даже трех). Но как же нам теперь узнать, чему равен этот коэффициент? Из эксперимента. Причем достаточно одного испытания с измерением угла в момент падения, чтобы получить оценку этого значения! С помощью симулятора я легко выяснил, что $C = 2,3$.

Мы получили устрашающее двухпараметрическое распределение. Что же с ним теперь делать? Нас интересует вероятность того, что бутерброд упадет маслом вниз, если x будет равно 0,2; или 0,4; или любому числу от 0 до 0,5. Мы использовали союз «или», при этом каждый случай рассматривается как независимый и исключаящий все прочие при проведении конкретного эксперимента. Вспомним, что вероятность — мера вероятностного пространства,

а раз так, то она аддитивна. Это позволяет нам просто сложить вероятности $P_{\downarrow}(x, h)$ для всех значений x , умножив их предварительно на вероятность попадания в конкретный диапазон значений. Разобьем отрезок от 0 до 0,5 на n частей и вычислим оценку вероятности в виде суммы:

$$P_{\downarrow}(h) \approx \frac{2}{n} \sum_{i=0}^n P_{\downarrow}\left(\frac{i}{2n}, h\right)$$

Здесь множитель $2/n$ выражает вероятность для случайной величины x попасть в отрезок ширины $1/n$. Вот как выглядят результаты для значительного числа разбиений ($n = 100$) на фоне серии численных экспериментов с нулевой горизонтальной скоростью (рис. 3.10).

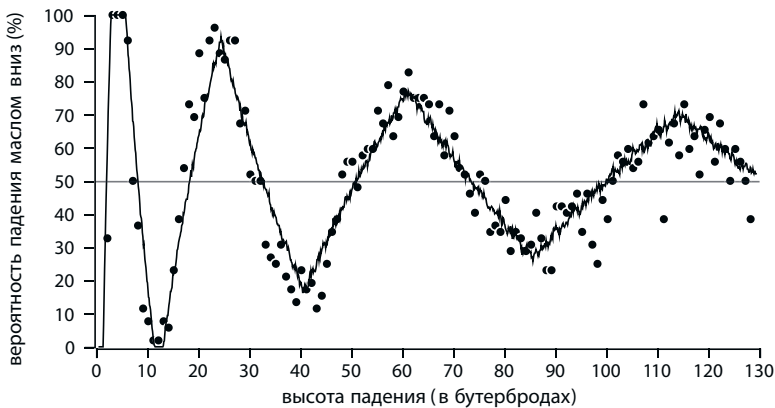


Рис. 3.10. Теоретическая и экспериментальная оценка вероятности приземления бутерброда маслом вниз при падении с большой высоты. Начальная горизонтальная скорость в экспериментах равна нулю

Решение, которое мы приводили до этого, содержит больше случайных параметров, поэтому оно оказалось более сглаженным и приближенным к 50%, но в принципе подобный анализ можно провести и для более общего случая.

Обратите внимание на то, что вероятность P_{\downarrow} при увеличении h стремится к значениям, близким к 50%. И это происходит вовсе

не из-за неопределенности и влияния начальных ошибок. Вычисления показали, что это результат сложения множества *гармоник*, образуемых значениями x при суммировании $P_{\downarrow}(x, h)$. Если мы забудем про несчастный бутерброд и продолжим график P_{\downarrow} , то увидим, что оценка вероятности так и продолжит колебаться вблизи 50%, постепенно стремясь к этому значению.

А можно ли выяснить без прямых вычислений, будет ли вероятность продолжать сходиться к 50% или когда-нибудь снова станет расти? И здесь тоже есть место нетривиальной и глубокой математике. Дело в том, что каждому значению x соответствует определенная частота колебаний*, а весь набор формирует так называемый *спектр* суммарной функции. Если он *дискретный*, то есть состоит из отдельных частот, суммарная функция (она называется *Фурье-образом*) будет периодичной. *Непрерывному* спектру в виде константы на отрезке от 0 до 0,5 будет соответствовать аperiodичная функция, похожая на убывающие колебания. Но это мы заглянули в другой большой и важный раздел математики — *функциональный анализ*. Больше он нам не понадобится, так что если вас напугал этот абзац, не переживайте. Его смысл выразим одной фразой: можно строго показать, что при падении бутерброда с большой высоты вам не удастся угадать, упадет он маслом вверх или вниз.

Великий итальянец Энрико Ферми, «дедушка» метода Монте-Карло (отцом считается польский математик Станислав Улам), приучал своих учеников проводить простые оценочные вычисления, прикидывать на клочке бумаги или на пальцах, что мы ожидаем получить, прежде чем приступить к точному решению задачи. Примечателен такой момент: если оценка окажется верной, станет понятно, что суть проблемы ухвачена; если же нет, то это тем более полезный результат — значит, задача оказалась интереснее, чем кажется!

В нашем случае простой оценки достаточно, задача о бутерброде не стоит более тщательного решения. Метод Монте-Карло

* Колебания в нашей задаче не гармонические и не синусоидальные, но это не мешает нам складывать такие гармоники. Заменой переменных можно привести их к традиционному для преобразований Фурье виду.

продемонстрировал нам только наметки решения, анализ размерности очертил лишь некоторую его общую структуру, но вместе они смогли показать нам, как устроена искомая вероятность. В повседневной работе эрудиция позволяет математику видеть в подобных наметках решения готовые структуры, выбирать подходящие методы и делать далеко идущие предположения и выводы.

Роберт Мэтьюз в своем эпохальном исследовании тоже использовал анализ размерностей, чтобы показать, что закон бутерброда фундаментален. Его вывод основан на том, что предельная высота организма, вставшего на задние конечности с целью передними взять бутерброд с маслом, определяется прочностными свойствами биологических тканей и гравитацией. В свою очередь, характерный размер бутерброда должен соответствовать масштабу существа — и коренастые карлики на какой-нибудь тяжелой планете, и хрупкие дылды на планете с малой гравитацией будут выбирать себе бутерброды по размеру. Тут мы подходим к тому, что в науке называется *спекуляцией*. Это не перепродажа всякого добра втридорога, а сомнительные предположения, лежащие в основание логического построения. В частности, мы предполагаем наличие у существ рук, пропорции которых сходны с нашими, а это более чем спорно.

Виновато ли масло?

В мерфологии известно неправильное цитирование закона Менкена Гроссманом:

**Сложные проблемы всегда имеют простые,
легкие для понимания неправильные решения.**

Очень часто можно услышать, что в законе бутерброда виновато масло, которое плотнее хлеба и потому «перевешивает». И хотя это не относится к предмету нашей книги, я хочу разобрать этот вопрос, чтобы поставить в нем наконец точку. Чтобы

кто угодно потом мог сослаться на то, что «ученые доказали, что наличие масла не влияет на то, какой стороной шлепнется бутерброд»!

В детстве мы забавлялись тем, что подбрасывали высоко вверх голубиное маховое или хвостовое перо, воткнутое в кусочек пластилина диаметром один-два сантиметра. Перо подлетало метра на четыре, после чего красиво и плавно спускалось на авторотации, как вертолет с заглушенным двигателем. Потом мы подросли, и наши забавы стали менее безобидными. Мы раздобывали гайку и вкручивали в нее два болта с противоположных сторон, спрессовывая начинку из крошенных спичечных головок. Оставалось привязать к одному из болтов ленту или кусок веревки, хорошенько раскрутить и запустить в небо метров на пятнадцать. В падении легкая лента стабилизировала вертикальное положение снаряда, обеспечивая качественный удар об асфальт и небольшой взрыв; порой гайка разлеталась на куски. (Будьте осторожны, если решитесь поделить этим опытом со своими детьми!)

В обоих экспериментах мы видим, что легкое перо или лента быстро оказывались над тяжелой частью аппарата и стабилизировали падение. Это, видимо, и приводит к интуитивному мнению, что тяжелое масло и легкий хлеб тоже должны вести себя так. Представим себе воздушный шар: более плотная корзина всегда располагается под менее плотным шаром. Более того, опыт подсказывает, что если взяться двумя пальцами за геометрическую середину предмета с несимметрично распределенной массой, то он кувыркнется так, чтобы тяжелая часть оказалась внизу. Но ни одно из этих явлений не работает в случае падающего бутерброда.

Начнем со второго процесса — «перевешивания». Я не случайно занудно уточнил: «...если взяться за *геометрическую* середину предмета...». Здесь имеется в виду, что точки касания лежат на некой прямой, образующей ось вращения, которая проходит сквозь *геометрическую середину* предмета. В таком случае действительно устойчивым положением будет такое, в котором центр тяжести ниже оси. Но если образуемая пальцами ось вращения проходит

через него, то система окажется в безразличном равновесии и ей будет все равно, как она ориентирована.

Что же заставляет ориентироваться «правильно» перышко с грузиком, мину из гайки или воздушный шар с воздухоплавателями в корзине? Воздух. Он «держит» наши предметы так, что ось вращения проходит выше центра тяжести. Точнее, набегающий поток воздуха, который создает силу, распределенную по площади тела. И условная точка ее приложения будет располагаться вблизи геометрического центра площади фигуры. Чтобы стало яснее, нарисуем силы, действующие на условный воздушный шар как на предмет с неравномерной плотностью (рис. 3.11).

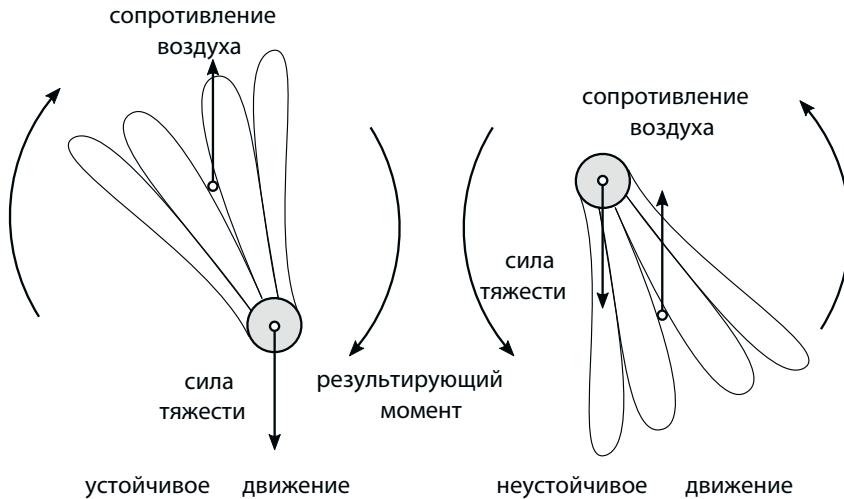


Рис. 3.11. Силы, приводящие воздушный шар в устойчивое положение

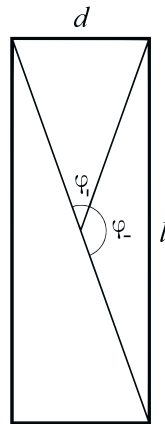
А что бутерброд?

Во-первых, если мы «выключим» воздух, он будет просто падать. В свободном падении тело вращается именно вокруг центра масс, так что у бутерброда нет резона поворачиваться как-то по-особому. Как нам говорили в школе: «В падающем лифте наблюдается невесомость». Масло в бутерброде столь же «невесомо».

Плотное масло может повлиять на процесс соскальзывания, оно эффективно поднимет центр масс над точкой касания и изменит в выражении для углового ускорения l на $\sqrt{l^2 + d^2} = l\sqrt{1 + \delta^2}$, где $\delta = d/l$ — относительная толщина бутерброда. При небольших значениях δ это выражение приближенно равно $l(1 + \delta^2/2)$. Получаем, как говорят, эффект второго порядка. Для бутерброда с соотношением ширины к толщине 5 к 1 эти относительные изменения не превышают 2%. И это максимальная верхняя граница эффекта: мы переместили центр масс на поверхность бутерброда, что соответствует бесконечно плотному маслу!

Теперь «включим» воздух обратно, оставив плотность масла бесконечно превышающей плотность хлеба. Имеем тонкую плотную пластину масла с невесомым, но сопротивляющимся воздуху «парашютом» хлеба. Пока плоскость бутерброда расположена горизонтально или близко к тому, на нее действует момент сил воздушного сопротивления, пропорциональный парусности — площади, с которой взаимодействует поток воздуха: $M_{\perp} \propto l^2$. В вертикальном положении парусность уменьшится и, соответственно, момент будет другим: $M_{\parallel} \propto ld$. Отношение этих моментов: $M_{\parallel}/M_{\perp} \propto \delta$. Я использовал здесь знак пропорциональности, поскольку коэффициенты сопротивления для пластинки, расположенной поперек и вдоль потока, различаются, и мне они неизвестны. Но они и не нужны — уже видно, что влияние воздуха в вертикальном положении (а именно оно делает неравнозначным положение масла) слабее, чем в горизонтальном. Теперь вспомним, что бутерброд вращается, а значит, он подставляется потоку то торцом, то плоскостью. Мы можем ввести меру действия сил сопротивления. Если угловая скорость вращения несущественно изменяется за один период (а для воздуха это так), то имеет смысл в качестве меры взять изменение момента импульса, пропорциональное времени действия силы. В свою очередь, период действия пропорционален углу, «заметаемому» бутербродом в течение этого периода. В итоге меры действия моментов M_{\parallel} и M_{\perp} будут пропорциональны $M_{\parallel}\varphi_{\parallel}$ и $M_{\perp}\varphi_{\perp}$, а углы, которые замечают торец и плоскость, показаны на чертеже ниже. Мы могли бы в качестве меры использовать работу

сил сопротивления и получили бы такое же соотношение. Отношение углов легко вычислить:



$$\frac{\varphi_+}{\varphi_-} = \frac{\varphi_+}{\pi/2 - \varphi_+}.$$

При малых значениях отношения d/l можно воспользоваться приближением: $\varphi_+ \approx \delta$ (используем свойство тангенса малого угла, выполняющегося с 10% точностью при углах меньше 30°), а значит, имеем:

$$\frac{M_{\varphi_+}}{M_{\varphi_-}} \approx \delta \frac{\delta}{\pi/2 - \delta} \propto \delta^2.$$

Опять получается, что влияние несимметричности для плоского бутерброда ограничивается эффектом второго порядка. Обычно плотность масла превышает плотность хлеба примерно вдвое. Таким образом, смещение центра масс не должно превышать трети толщины бутерброда при разумном слое масла (не больше толщины хлеба). Это уменьшит влияние масла до 0,2%.

Если читателю показалось, что мы сейчас палили из пушки по воробьям, то я с ним полностью соглашусь. Но, во-первых, мне не хочется больше слышать о «перевешивающем» масле; во-вторых, я не желаю быть голословным; а в-третьих, я стремился показать, как физик оценивает величины, представляя процесс, но не обладая

полными данными. Конечно, в момент приземления масло может прилипнуть к полу и не дать бутерброду подпрыгнуть и перевернуться вновь, но механику удара, упругой деформации и подскока кусочка хлеба я уж точно разбирать не буду. И так многовато анализа для этой проблемы. И вторую Шнобелевскую уже не дадут.

Не так важна была цель нашего пути: опровержение либо оправдание закона бутерброда, — как сам путь. Он показал, как совмещение разных математических методов позволяет взглянуть на задачу с разных сторон, и дает достаточно точное знание — даже без детального решения задачи. В согласованности различных математических дисциплин, подходов и точек зрения состоит сила и красота математики. Тут уместно вспомнить чудесные слова Марины Цветаевой: «Я не хочу иметь точку зрения, я хочу иметь зрение». Изучение разных областей математики способно дать исследователю настоящее «объемное» многомерное зрение, позволяющее заглянуть в кажущееся закрытым и скрытым пространство знаний.

Глава 4

Статистика как научный способ чего-либо не знать

Цифры обманчивы, особенно когда я сам ими занимаюсь; по этому поводу справедливо высказывание, приписываемое Дизраэли: «Существует три вида лжи: ложь, наглая ложь и статистика».

Марк Твен*

Как часто летом мы намереваемся на свои выходные выехать на природу, прогуляться в парке или устроить пикник, а потом дождь разбивает наши планы, заточив нас в доме! И ладно бы это случилось раз или два за сезон; порой складывается впечатление, что непогода преследует именно выходные дни, раз за разом выпадая на субботу или воскресенье!

Совсем недавно вышла статья австралийских исследователей «Недельные циклы пиковой температуры и интенсивность

* Надо признаться, что эта фраза, ставшая расхожей с легкой руки Марка Твена, не была произнесена Дизраэли, и вообще неясно, кто ее автор.

городских тепловых островов»*. Ее подхватили новостные издания и перепечатали результаты с таким заголовком: «Вам не кажется! Ученые выяснили: погода на выходных действительно хуже, чем в будние дни». В цитируемой работе приводится статистика температуры и осадков за много лет в нескольких городах Австралии, и вправду выявляющая понижение температуры на $0,3^{\circ}\text{C}$ в определенные часы субботы и воскресенья. Там же этому дается объяснение. Оно связывает локальную погоду с уровнем загрязненности воздуха из-за возрастающего транспортного потока. Незадолго до того подобное исследование проводилось в Германии** и привело примерно к тем же выводам.

Согласитесь, доли градуса — весьма тонкий эффект. Сетую на непогоду в долгожданную субботу, мы обсуждаем, был ли день солнечным или дождливым. Такое обстоятельство проще зафиксировать, а позже вспомнить, даже не обладая точными приборами. Мы проведем собственное небольшое исследование на эту тему и получим замечательный результат: можно уверенно утверждать, что мы не знаем, связаны ли на Камчатке день недели и непогода. Исследования с отрицательным результатом обычно не попадают на страницы журналов и в новостные ленты, но нам важно понять, на каком основании мы можем что-либо уверенно заявлять о случайных явлениях. И в этом плане отрицательный результат ничем не хуже положительного.

Слово в защиту статистики

Статистику обвиняют во множестве грехов: и во лжи, и в возможностях манипуляций, и, наконец, в непонятности. Но мне очень хочется реабилитировать эту область знаний, показать, насколько сложна задача, для которой она предназначена, и как непросто понять ответ, который дает статистика.

* Earl N., Simmonds I., N. Tapper N. Weekly cycles in peak time temperatures and urban heat island intensity // Environ. Res. Lett. 2016. Vol. 11.

** Bäumer D., Vogel B. An unexpected pattern of distinct weekly periodicities in climatological variables in Germany // Geophysical Research Letters. 2007. Vol. 34.

Теория вероятностей оперирует точными знаниями о случайных величинах в виде распределений или исчерпывающих комбинаторных подсчетов. Еще раз подчеркну, что иметь точное знание о случайной величине возможно, если мы говорим о распределении. Но что, если это знание нам недоступно, а единственное, чем мы располагаем, — наблюдения? У разработчика нового лекарства есть ограниченное число испытаний, у создателя системы управления транспортным потоком — лишь ряд измерений на реальной дороге, у социолога — результаты опросов. Причем он может быть уверен в том, что, отвечая на какие-то вопросы, респонденты просто соврали.

Понятно, что одно наблюдение не дает ровным счетом ничего. Два — немногим больше. Сколько нужно наблюдений — три, четыре, сто, — чтобы получить какое-то знание о случайной величине, в котором можно быть уверенным в математическом смысле? И что это за знание? Скорее всего, оно будет представлено в виде таблицы или гистограммы, дающей возможность оценить некоторые параметры случайной величины, например область определения, среднее или дисперсия, асимметричность и т. д. Быть может, глядя на гистограмму, удастся угадать точную форму распределения. Это и есть основная задача математической статистики: по наблюдаемым реализациям случайной величины выяснить ее распределение, то есть получить по возможности точное и исчерпывающее ее описание. Но — внимание! — все результаты наблюдений сами будут случайными величинами! Пока мы не владеем точным знанием о распределении, все результаты наблюдений дают нам лишь вероятностное описание случайного процесса. Случайное описание случайного процесса — еще бы здесь не запутаться, а то и захотеть запутать намеренно!

Что же делает математическую статистику точной наукой? Ее методы позволяют заключить наше незнание в четкие рамки и дать вычислимую меру уверенности в том, что в этих рамках наше знание о случайной величине согласуется с фактами. Это язык, на котором можно говорить о случайностях неизвестной природы так, чтобы рассуждения имели смысл. Такой подход очень полезен в философии, психологии и социологии, где очень легко пуститься в пространные рассуждения и дискуссии без надежды на получение настоящего знания и тем более доказательства. Грамотной статистической

обработке данных посвящено множество книг, ведь это абсолютно необходимый инструмент для медиков, социологов, экономистов, физиков, психологов — словом, всех специалистов, научно исследующих «реальный мир», который отличается от идеального математического лишь степенью нашего незнания о нем. Я получил упрек за то, что использовал кавычки вокруг слов «реальный мир», как если бы не верил в его существование. Такое направление в философии действительно есть, оно называется солипсизмом, но я не его сторонник. Кавычками я хочу подчеркнуть, что не разделяю мир на реальный и идеальный, физический и математический. Я не вижу причин считать математические структуры тем, чего нет в мире, в котором мы живем. Это глубокий вопрос и давний спор: математик исследует настоящую Вселенную или изобретает свою, ненастоящую? Я не хочу долго рассуждать на эту тему, поскольку не вижу, как тот или иной ответ может помочь математику или физику в его работе. Но одним из чудес нашего мира по праву считается то, что он описывается на языке математики, доступном человеку.

Теперь еще раз взгляните на эпиграф к этой главе и осознайте, что статистика, которую так пренебрежительно называют третьим видом лжи, — единственное, чем располагают все естественные науки. Это ли не главный закон подлости мироздания! Все физические и наблюдаемые нами экономические законы строятся на математических моделях и их свойствах, но проверяются они статистическими методами в ходе измерений и наблюдений. В повседневности наш разум делает обобщения и подмечает закономерности, выделяет и распознаёт повторяющиеся образы. Это, наверное, лучшее, что умеет человеческий мозг. Именно этому в наши дни учат искусственный интеллект. Но разум экономит силы и склонен делать выводы по единичным наблюдениям, не сильно беспокоясь о точности или обоснованности этих заключений. По этому поводу есть замечательное самосогласованное утверждение из книги Стивена Браста «Исола»*: «Все делают общие выводы из одного примера. По крайней мере, я делаю именно так». И пока речь идет об искусстве, характере домашних любимцев или

* Издана на русском языке: Браст С. Исола. М. : АСТ, 2002.

обсуждении политики, об этом можно сильно не беспокоиться. Однако при строительстве самолета, организации диспетчерской службы аэропорта или тестировании нового лекарства уже нельзя сослаться на то, что «мне так кажется», «интуиция подсказывает» и «в жизни всякое бывает». Тут приходится ограничивать работу своего разума рамками строгих математических методов.

Эта книга не учебник, мы не будем детально исследовать статистические методы и ограничимся лишь одной из техник проверки гипотез. Но мне хотелось бы показать ход рассуждений и форму результатов, характерных для этой области знания. И, возможно, кому-то из читателей, к примеру будущему студенту, не только станет понятно, зачем его мучают матстатистикой, всеми этими QQ-диаграммами, t - и F -распределениями, но и придет в голову другой важный вопрос: а как вообще возможно знать что-нибудь наверняка о случайном явлении? И что именно мы узнаём, используя статистические данные?

Как возможность ошибиться делает науку наукой

Математическая статистика использует методы теории вероятностей, а ее столпы — закон больших чисел и центральная предельная теорема.

Естественное предположение, что наблюдаемые данные отражают реальное неизвестное распределение, оказывается верным. Например, гистограмма наблюдаемых величин приближается к истинной плотности распределения, если число наблюдений стремится к бесконечности.

Как закон больших чисел, так и центральная предельная теорема — не одно утверждение. Каждый из этих результатов представляет собой несколько разных теорем, охватывающих широкий спектр задач и условий. Мы познакомимся с их упрощенными формулировками, дающими хорошее представление об этих важных результатах.

Закон больших чисел — несколько разных теорем, утверждающих, что среднее значение наблюдений случайной величины при

определенных условиях в том или ином смысле стремится к неизвестному математическому ожиданию этой величины. В простейшем случае он выглядит так. Пусть X_1, X_2, \dots, X_n — независимые одинаково распределенные случайные величины с математическим ожиданием a , $S_n = X_1 + X_2 + \dots + X_n$. Тогда

$$\frac{S_n}{n} \rightarrow a, \quad n \rightarrow \infty.$$

Иными словами, среднее значение наблюдений стремится к математическому ожиданию. В частности, из закона больших чисел вытекает, что частота наблюдений какого-либо события стремится к вероятности этого события, то есть он прочно связывает «бытовое» частотное толкование вероятности и теоретическое как меры на вероятностном пространстве.

Центральная предельная теорема говорит о том, что при определенных условиях сумма независимых или слабо зависимых случайных величин, каждая из которых вносит небольшой вклад в общую сумму, имеет распределение, близкое к нормальному (гауссовскому). Теорема получила свое название за универсальность и важность, поскольку ее условия часто реализуются на практике. Например, многие биологические характеристики (рост человека или размах рук) подчиняются нормальному распределению, поскольку на них влияет множество факторов (скажем, действует много разных генов), вносящих по отдельности небольшой вклад. В простейшем случае теорема выглядит так.

Пусть опять X_1, X_2, \dots, X_n — независимые одинаково распределенные случайные величины с математическим ожиданием a и дисперсией σ^2 . Тогда

$$\frac{S_n - na}{\sigma\sqrt{n}} \rightarrow N(0,1), \quad n \rightarrow \infty.$$

Здесь $N(0,1)$ обозначает стандартное нормальное распределение со средним 0 и дисперсией 1.

Иными словами, при больших n сумма S_n близка к гауссовской случайной величине с математическим ожиданием (средним значением) na и дисперсией $n\sigma^2$.

Эту теорему обычно доказывают, применяя методы функционального анализа. Но мы увидим позже, что ее можно понять и даже расширить, введя понятие энтропии как меры вероятности состояния системы: нормальное распределение имеет наибольшую энтропию при наименьшем числе ограничений. В этом смысле оно оптимально при описании неизвестной случайной величины либо случайной величины, являющейся суммой многих других величин, распределение которых тоже неизвестно.

Эти два закона лежат в основе количественных оценок достоверности наших знаний, основанных на наблюдениях. Здесь речь о статистическом подтверждении или опровержении предположения, которое можно сделать из каких-то общих оснований, и математической модели. Это может показаться странным, но сама по себе статистика не производит новых знаний. Набор фактов превращается в знание лишь после построения связей между фактами, образующих определенную структуру. Именно эти структуры и связи позволяют делать предсказания и выдвигать общие предположения, которые основаны на чем-то, выходящем за пределы статистики. Они называются *гипотезами*. Самое время вспомнить один из законов мерфологии — **постулат Персига**:

**Число разумных гипотез, объясняющих
любое данное явление, бесконечно.**

Задача математической статистики — ограничить это бесконечное число, а вернее, свести все гипотезы к одной, причем вовсе не обязательно верной. Итак, у нас есть случайная величина X , распределение P которой неизвестно (иногда совсем, иногда частично). Гипотеза — любое предположение о P . Простая гипотеза — предположение, что P — какое-то конкретное известное распределение. Сложная гипотеза — предположение, что P принадлежит целому классу распределений. Как правило, исследователь проверяет простую гипотезу.

Эта исходная гипотеза обычно называется нулевой. Что может выступить в таком качестве? В определенном смысле — что угодно,

любое утверждение об исследуемой системе. Например, если у нас есть данные о росте призывников, мы можем проверить гипотезу, что неизвестный средний рост равен 1,76 м (или 2,10 м). Если у нас есть данные по количеству аистов и новорожденных, то мы можем проверить гипотезу, что эти две величины независимы. Если у нас есть два больших литературных произведения, мы можем проверить гипотезу, что их написал один автор, построив какую-то математическую модель.

Классическая постановка вопроса при этом такова: позволяют ли наблюдения отвергнуть нулевую гипотезу или нет? Точнее, с какой долей уверенности мы можем утверждать, что наблюдения нельзя получить, исходя из нулевой гипотезы? При этом если мы не смогли доказать, опираясь на статистические данные, что нулевая гипотеза ложна, то она принимается истинной.

Тут можно подумать, что исследователи вынуждены совершать одну из классических логических ошибок, которая носит звучное латинское имя *ad ignorantiam*. Это аргументация истинности некоторого утверждения, основанная на отсутствии доказательства его ложности. Классический пример — слова, сказанные сенатором Джозефом Маккарти, когда его попросили предъявить факты для поддержки выдвинутого им обвинения, что некий человек — коммунист: *«У меня немного информации по этому вопросу, за исключением того общего заявления компетентных органов, что в его досье нет ничего, что бы исключало его связи с коммунистами»*. Или еще ярче: *«Снежный человек существует, поскольку никто не доказал обратного»*. Выявление разницы между научной гипотезой и подобными уловками составляет предмет целой области философии: *методологии научного познания*. Один из ее ярких результатов — *критерий фальсифицируемости*, выдвинутый замечательным философом Карлом Поппером в первой половине XX века. Он призван отделять научное знание от ненаучного и на первый взгляд кажется парадоксальным:

**Теория или гипотеза может считаться научной,
только если существует, пусть даже гипотетически,
способ ее опровергнуть.**

Чем не один из законов мерфологии? Получается, любая научная теория автоматически потенциально неверна, а теория, верная «по определению», не может считаться научной*.

Но всё же: почему мы, если не можем на базе статистических данных отвергнуть гипотезу, вправе считать ее истинной? Дело в том, что статистическая гипотеза берется не из желания исследователя или его предпочтений, она должна вытекать из каких-то общих формальных законов. Например, из центральной предельной теоремы либо принципа максимальной энтропии, о котором мы поговорим в самом конце книги. Эти законы корректно отражают *степень нашего незнания*, не добавляя без необходимости лишних предположений или гипотез. В известном смысле это прямое использование знаменитого философского принципа, известного как *бритва Оккама*:

**Что может быть сделано на основе меньшего
числа предположений, не следует делать,
исходя из большего.**

Вообще с точки зрения принципа фальсифицируемости любое утверждение о существовании чего-либо ненаучно, ведь отсутствие свидетельства ничего не доказывает. В то же время утверждение об отсутствии чего-либо можно легко опровергнуть, предоставив экземпляр, косвенное свидетельство или доказав существование по построению. И в этом смысле статистическая проверка гипотез анализирует утверждения об *отсутствии* искомого эффекта и может предоставить в известном смысле точное опровержение.

Именно этим в полной мере оправдывается термин «нулевая гипотеза»: она содержит необходимый минимум знаний о системе.

* Более того, критерию Поппера не удовлетворяют такие науки, как математика и логика; впрочем, их относят не к естественным наукам, а к формальным. Однако очень важно понимать, что принцип фальсифицируемости говорит не об *истинности* теории, а только о том, научна она или нет. Он помогает определить, дает ли некая теория язык, на котором имеет смысл рассуждать о мире, или нет.

Запутываем статистикой и помогаем распутаться

Очень важно подчеркнуть: если статистические данные говорят о том, что нулевая гипотеза может быть отвергнута, это не значит, что мы тем самым доказали истинность какой-либо альтернативной гипотезы. Вспомним постулат Персига: *«Число разумных гипотез, объясняющих любое данное явление, бесконечно»*. Опровержение нулевой гипотезы не делает все остальные верными. Отвергая ее, мы освобождаем место для нового умозаключения, как в легенде об убийстве деспота-дракона.

Вообще математическая статистика и теория вероятностей рассуждают вовсе не о ложности или истинности каких-либо утверждений. Их следует крайне осторожно смешивать с логикой; здесь кроется масса трудноуловимых ошибок, особенно когда в дело вступят зависимые события. Вот пример такого смешения. Очень маловероятно, что человек может стать папой римским (примерно один к семи миллиардам); следует ли из этого, что папа Иоанн Павел II не был человеком? Утверждение кажется абсурдным.

А вот другой пример: проверка показала, что мобильный тест на содержание алкоголя в крови дает не более 1% как ложноположительных, так и ложноотрицательных результатов. Следовательно, в 98% случаев он верно выявит пьяного водителя. Это правильный вывод, но он вступает в кажущееся противоречие со следующими рассуждениями. Протестируем 1000 водителей, и пусть 100 из них будут действительно пьяны. В результате мы получим $900 \times 1\% = 9$ ложноположительных и $100 \times 1\% = 1$ ложноотрицательный результат: на одного проскочившего пьяницу придется девять невинно обвиненных случайных водителей. Выходит, речь должна идти лишь о 10% правильных ответов, а не о 98%. Чем не закон подлости! Паритет возникнет, только если доля пьяных водителей окажется равна $1/2$ либо если отношение долей ложноположительных и ложноотрицательных результатов будет близким к реальному отношению пьяных водителей к трезвым. Причем чем трезвее обследуемая нация, тем несправедливее будет применение описанного нами прибора!

Здесь мы столкнулись с *зависимыми событиями*. Введем понятие условной вероятности — вероятности наступления одного события, если известно, что произошло другое событие. Для двух событий A и B (причем $P(B) > 0$) она обозначается $P(A|B)$ и вычисляется следующим образом:

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

Пример: мы бросили игральную кость. Пусть событие $A = \{\text{выпала } 1\}$. $P(A) = 1/6$. Пусть теперь известно, что при бросании произошло событие $B = \{\text{выпало нечетное число}\}$. Теперь, очевидно, вместо шести возможных вариантов есть всего три, так что $P(A|B) = 1/3$. Именно это мы и получаем по нашему определению: $A \cap B = \{\text{выпала } 1\}$, $P(A \cap B) = 1/6$, $P(B) = 1/2$, откуда $1/6 : 1/2 = 1/3$.

Если наступление события B не меняет вероятность наступления события A , то должно быть $P(A|B) = P(A)$. В силу определения условной вероятности это значит, что $P(A \cap B) = P(A)P(B)$. Это соотношение оказывается определением важнейшего понятия в теории вероятностей — независимости: события A и B называются независимыми, если $P(A \cap B) = P(A)P(B)$. Определение работает, даже если вероятности событий A или B равны 0.

Из определения условной вероятности можно получить выражение для пересечения произвольных событий:

$$P(A \cap B) = P(A)P(B).$$

Пересечение множеств — операция коммутативная, $A \cap B = B \cap A$. Отсюда немедленно следует, что $P(A \cap B) = P(B \cap A)$, и *теорема Байеса*:

$$P(A|B)P(B) = P(A \cap B),$$

которую можно использовать для вычисления условных вероятностей.

Применим эти новые определения и соотношения, чтобы разобраться в примере с водителями и тестом на алкогольное опьянение. Мы имеем следующие события: A — водитель пьян, B — тест

выдал положительный результат. Вероятности: $P(A) = 10\%$ — для случая, когда остановленный водитель пьян; $P(B|A) = 99\%$ — тест выдаст положительный результат, если известно, что водитель пьян (исключается 1% ложноотрицательных результатов), $P(A|B) = 99\%$ — тестируемый пьян, если тест дал положительный результат (исключается 1% ложноположительных результатов). Вычислим вероятность того, что тест даст верный результат, не обвинит невиновного и не пропустит виноватого. Оба эти варианта независимы и вероятность того, что не случится ни та, ни другая ошибка, равна $P(B|A)P(A|B) = 98,02\%$. Это близко к тому, что ожидалось интуитивно. О чем же мы рассуждали, говоря о несправедливости теста? Мы вычислили $P(B)$ — вероятность получить положительный результат теста на дороге:

$$P(B) = \frac{P(A)P(A|B)}{P(B|A)} = P(A) = 10\%.$$

Понятие условной вероятности позволяет корректно вести логические рассуждения на языке теории вероятностей. Неудивительно, что теорема Байеса нашла широкое применение в теории принятия решений, системах распознавания образов, спам-фильтрах, программах, проверяющих тексты на плагиат, и многих других информационных технологиях. Подобные примеры тщательно разбираются студентами, изучающими медицинские тесты или юридические практики. Но, боюсь, журналистам и политикам не преподают ни математическую статистику, ни теорию вероятностей. Зато они охотно апеллируют к статистическим данным, вольно интерпретируют их и несут полученное «знание» в массы.

Разберем еще один пример ошибочной интерпретации статистических данных. В июне 2011 года был выпущен публичный отчет о росте уровня занятости в США, он составил 18 тысяч новых работников по всей стране. В газетах штата Висконсин об этом была опубликована статья, в которой отмечалось, что более половины роста (9,8 тысячи человек) приходится именно на этот штат. Статья завершалась хвалебным отзывом о плодотворной работе правительства штата и позже с удовольствием цитировалась политиками и чиновниками.

Притом что обе цифры верны и подтасовок в них нет, штат Висконсин никак не может претендовать на доминирующий вклад в общий рост уровня занятости. В том же году в штате Массачусетс появилось 10,4 тысячи новых рабочих мест (58% от общей цифры), а в Калифорнии — 28,8 тысячи (160%). Я полагаю, читатель начинает догадываться, что приводимые тут проценты не имеют большого смысла, поскольку в этом же году в ряде штатов, например в Миссури или Вирджинии, произошло сокращение рабочих мест. Таким образом, 18 тысяч — сумма всех положительных и отрицательных изменений.

Где заканчивается свобода в математике?

Здесь стоит ненадолго остановиться. Мы уже достаточно подкованы в математике, чтобы не просто с умным видом поиздеваться над ошибкой журналистов и доверчивостью чиновников, а разобратся в том, что именно произошло. Речь в статье шла о долях, при этом использовались суммы величин, которые могут быть и отрицательными. Что же здесь не так? Ведь долю, то есть рациональное число, можно вычислить от величины любого знака. Здесь нам опять пригодится понятие меры.

Доли, или удельный вклад, имеет смысл вычислять от величины, относящейся к *мерам* — аддитивной и неотрицательной. Говоря в предыдущей главе о мере как функции над множествами, мы упоминали требование ее неотрицательности, но не заостряли на нем внимание. Само понятие меры появилось как расширение таких категорий, как количество, длина или объем, а эти величины, очевидно, не могут быть отрицательными. Но что случится с нашим определением, если мы разрешим мере быть отрицательной? Может, тем самым мы расширим это понятие и оно станет еще полезнее? Расширили же мы понятие вероятности, введя условную вероятность. Бытует мнение (особенно среди «практиков», инженеров и программистов), что математики изобретают аксиомы и изменяют определения по мере необходимости. Что это вопрос практичности, договоренностей либо даже вкуса. Нет, ребята, математика так не работает.

Приведу два примера, из которых станет ясно, что аксиомы не придумываются. В главе 1, рассматривая петли на наушниках, мы указали, что они образуют группу с операцией сложения, соответствующей нанизыванию их на одну веревку. Для любой группы должны выполняться четыре аксиомы: замкнутость операции группового сложения, ее ассоциативность, наличие единственного нуля (нейтрального элемента), наконец, наличие обратного элемента. Апочему мы ничего не говорим о коммутативности сложения (о том, что $a + b = b + a$)? Легко убедиться в том, что для наших петель, как и для чисел, это свойство выполняется. Кроме того, мы сразу сказали, что ноль — нейтральный элемент, независимо от порядка сложения с ним: $(0 + a = a + 0 = a)$. Раз это должно работать для нуля, почему это не может работать для всех элементов группы?

Дело в том, что коммутативность не вытекает из четырех аксиом группы. Легко найти некоммутативную группу, классическим примером будут движения на плоскости. Если рассмотреть два движения: поворот относительно некоей опорной точки и смещение вдоль какого-то вектора, — то результат будет зависеть от порядка этих движений. Убедиться в этом легко, перемещая лист бумаги по поверхности стола. Почему же сложение с нулем должно быть коммутативно? Это требование ассоциативности, а именно выполнения равенства: $(a + 0) + b = a + (0 + b)$. Если бы сложение с нулем зависело от того, справа или слева он находится, то ассоциативность перестала бы работать для всех элементов группы. Эти два свойства не могут идти по отдельности. В то же время добавление свойства коммутативности согласуется с определением группы и расширяет ее до так называемой абелевой группы. Я помню, как был сначала озадачен, а потом восхищен тем, что коммутативность сложения для чисел не вводится искусственно, а может быть выведена из базового определения операции сложения.

Приведу еще один пример, который, возможно, примирит кого-то с диктатурой в математике. Помните школьное правило: «на ноль делить нельзя»? А почему, кто это запретил? Кроме того, теперь мы достаточно грамотны, чтобы уточнить вопрос: что такое «ноль», на который нельзя делить? Тот ли, который оказывается нейтральным элементом при сложении, или речь о каком-то ином объекте?

Сразу скажу: да, тот самый, поскольку он, по определению группы, единственный*. Более или менее искушенный в математике читатель скажет, что в пределах алгебраической структуры, которая называется *полем чисел* (рациональных или вещественных, именно их мы проходим в школе), не существует делителей нейтрального элемента по сложению, они просто не содержатся во множестве этих чисел. Можно добавить, что при умножении на ноль — как на *поглощающий элемент* для этой операции — мы полностью теряем информацию о втором множителе, подобно тому, как тень на стене не содержит полной информации о форме или цвете трехмерного объекта, отбрасывающего ее. Так что произвести операцию, обратную умножению, то есть деление, у нас в этом случае не получится.

Но можно ведь искусственно дополнить множество чисел специальными элементами — делителями нуля. Дополнили же когда-то множество *рациональных чисел*, привычных нам дробей, *иррациональными*, такими как $\sqrt{2}$, — чтобы можно было рассуждать о длине диагонали единичного квадрата или возведении в рациональные степени. Более того, в шестом классе, когда мы эти корни вводили, нас учили, что квадратный корень из отрицательного числа взять невозможно. Но потом, в десятом классе, множество *вещественных чисел* расширили до *комплексных*, дополнив его мнимой единицей. И вот, пожалуйста, невозможное стало возможным. Так в чем проблема с делением на ноль?

Дело в том, что и рациональные, и вещественные, и комплексные числа построены так, что все они образуют поля, при этом вся арифметика в них согласована. Но если искусственно ввести нетривиальные делители нуля, то получится иная арифметика, своеобразная и не согласующаяся с привычной нам со школы алгеброй полей. Алгебраическая структура, на которой определены сложение и умножение, а также своеобразное деление для всех элементов, включая ноль, называется *колесом***.

* Единственность нуля тоже нетривиальна и интересна. Если кто-то из читателей впервые об этом задумывается, то вот вам пицца для размышлений: сколькими способами можно построить ноль и единицу в рациональных числах? И будут ли все эти способы соответствовать единственному нулю и единице?

** Carlström J. Wheels — On Division by Zero. 2004 // Mathematical Structures in Computer Science. Cambridge University Press, 2011. Vol. 14. No. 1. Pp. 143–184.

нарная операция x/y , обратная умножению, а как унарный оператор $/y$, подобный y^{-1} . Таким образом, деление определяется как произведение $x \cdot /y$. Кроме того, алгебраическая система дополняется символами $/0$ и $0/0$, которые иногда обозначаются как ∞ и \perp . Они имеют особые свойства и не равны ни одному другому элементу системы.

Непротиворечивая система аксиом колеса кроме коммутативности, ассоциативности сложения с умножением содержит следующие правила:

$$\begin{aligned} 0 \cdot 0 &= 0 \\ // x &= x \\ /(xy) &= /y / x \\ xz + yz &= (x + y)z + 0z \\ (x + yz) / y &= x / y + z + 0y \\ (x + 0y)z &= xz + 0y \\ /(x + 0y) &= /x + 0y \\ 0 / 0 + x &= 0 / 0 \end{aligned}$$

Из этих аксиом неизбежно следует, что в общем случае:

$$0x \neq 0, \quad x - x \neq 0, \quad x / x \neq 1.$$

Увы, групповые свойства сложения в такой системе нарушаются, поскольку не для всех элементов x выполняется тождество $x + 0 = x$.

Так что «просто добавить» делители нуля и обратный ему элемент не получится, нужно перестраивать всю систему ради ее непротиворечивости. Подобные трудности возникнут и при попытке искусственно ввести вторую мнимую единицу: согласованную алгебру с двумя единицами создать не получится, а вот с тремя все работает. Так строится *кольцо кватернионов*. Они широко используются для моделирования вращений в трехмерном пространстве, например в компьютерных играх и симуляциях реальности. Увеличивая число дополнительных мнимых единиц, мы в следующий раз получим «хорошую» самосогласованную алгебру, когда их будет семь; она называется *алгеброй октонионов*. На нее возлагаются надежды

как на способ соединить квантовую теорию и гравитацию, получив «священный Грааль» физики: Теорию Всего. А больше можно? Формально да: при 15 дополнительных единицах строится алгебра седенионов. И — о чудо! — в алгебре седенионов уже есть нетривиальные делители нуля, но сама она, похоже, теряет ценность как алгебраическая система! Так что мы не можем просто придумать что-то новое в математике, если оно как-то не согласуется с существующими, повсеместно используемыми понятиями. Допустимо построить непротиворечивую систему, изучить ее свойства и пользоваться ими для моделирования либо реального мира, либо других систем.

Вернемся к мере. Ее неотрицательность необходима, иначе можно нарушить третье из свойств мер, перечисленных на с. 37: «Мера подмножества не превышает меры множества» (вклад штата Калифорния превысил общий рост по всей стране). Кроме того, при этом теряется польза от аддитивности и становится затруднительно вычислить меру для объединения подмножеств; таким образом, само это понятие теряет свою полезность. Число рабочих мест — полноценная мера (как количественная характеристика конечного множества), а вот рост числа рабочих мест — нет, это уже изменение меры.

Может возникнуть вопрос: а каков же на самом деле был вклад правительства штата Висконсин в борьбу с безработицей? Он имеет смысл, поскольку если бы не было этого вклада, то общий результат по стране был бы заметно меньшим. Корректно ответить несложно. Мы можем рассматривать как меру отдельно положительные и отрицательные вклады и таким образом говорить о том, что Висконсин предоставил 27% от общего числа новых рабочих мест (результат простого суммирования всех новых работников по стране). В свою очередь, из всех новых безработных 23% пришлось на жителей штата Миссури.

Измеряем нашу доверчивость

Вернемся к статистике. Из множества разнообразных ее задач мы рассмотрим здесь только одну: проверку статистических гипотез. Для тех, кто уже связал свою жизнь с естественными или

социальными науками, в этих примерах не будет чего-то ошеломительно нового. Но это хорошая задача, показывающая ход математической мысли и не уводящая в дебри технических деталей.

Предположим, мы многократно измеряем случайную величину X , имеющую среднее значение μ и стандартное отклонение σ . Согласно центральной предельной теореме, распределение *наблюдаемого среднего значения* будет близким к нормальному. Из закона больших чисел следует, что его среднее будет стремиться к μ , а из свойств нормального распределения — что после n измерений наблюдаемая дисперсия среднего будет уменьшаться как σ / \sqrt{n} . Стандартное отклонение можно рассматривать как абсолютную погрешность измерения среднего, относительная погрешность при этом будет равна $\delta = \sigma / (\mu \sqrt{n})$. Это общие выводы, не зависящие для достаточно больших значений n от конкретной формы распределения случайной величины X . Из них следуют два полезных правила (не закона).

1. Минимальное число испытаний n должно диктоваться желаемой относительной погрешностью δ . При этом если

$$n \geq \left(\frac{2\sigma}{\mu\delta} \right)^2,$$

то вероятность того, что наблюдаемое среднее останется в пределах заданной погрешности, будет не менее 95%. При μ , близком к нулю, относительную погрешность лучше заменить на абсолютную.

2. Пусть нулевой гипотезой будет предположение, что наблюдаемое среднее значение равно μ . Тогда, если наблюдаемое среднее не выходит за пределы $\mu \pm 2\sigma / \sqrt{n}$, вероятность того, что нулевая гипотеза верна, будет не менее 95%.

При использовании этих правил неизвестное σ можно оценить в первой серии экспериментов при относительно небольших значениях n , после чего уточнить необходимое число экспериментов. Зачастую, если у нас есть предположение о законе распределения, значение σ можно однозначно вывести из значения μ .

Если заменить в этих правилах 2σ на 3σ , степень уверенности вырастет до 99,7%. Это очень сильное правило, которое в физических науках отделяет предположения от экспериментально

установленного факта. В атомной физике критерий истинности — еще более сильное правило 5σ .

Для нас полезно будет рассмотреть приложение этих правил к *распределению Бернулли с параметром*, которое описывает случайную величину, принимающую ровно два значения, условно «успех» и «неудача», с вероятностью успеха p и неудачи $1 - p$. В этом случае $\mu = p$ и $\sigma = p(1 - p)$, так что для необходимого числа экспериментов и доверительного интервала получим такие выражения:

$$n \geq \frac{4}{\delta^2} \frac{1-p}{p} \quad \text{и} \quad np \pm 2n\sqrt{p(1-p)}.$$

В главе 2 мы упомянули результат, опубликованный Перси Диаконисом и говорящий о принципиальной, хоть и небольшой, нечестности процесса подбрасывания монеты. Напомню: вероятность того, что она выпадет той же стороной, которая была сверху при подбрасывании, оказалась равна 51%. Насколько велико такое отклонение? Можно ли его заметить в экспериментах?

Примем скучную нулевую гипотезу: *монета, подбрасываемая человеком, выпадает совершенно случайно, и результат эксперимента не зависит от ее начального положения*. Что нам нужно для того, чтобы опровергнуть это предположение? Нас интересует точность до второго знака после запятой, которой соответствует абсолютная погрешность, равная 0,005, или относительная: $0,005 / 0,5 = 0,01$. Отсюда имеем оценку для n : $(2 / 0,01)^2 = 40\,000$. Выделив по секунде на бросок и регистрацию результата, мы обречем себя на полсутки подбрасывания монеты без единого перерыва. Это нижняя оценка; если же мы захотим увеличить абсолютную точность на порядок, нам потребуется в сто раз больше испытаний: либо задействовать сто экспериментаторов, либо три месяца непрерывно бросать монету.

На рисунке показаны результаты 40 000 испытаний для двух «монеток»: идеальной (с 50%-й вероятностью обоих исходов) и слегка неидеальной (в которой выпадение орла имеет вероятность 55%), проводимых с целью вычислить вероятность выпадения орла. Слово «монетка» взято в кавычки, потому что на самом деле использовался генератор случайных чисел, подчиняющийся распределению Бернулли. Видно, что только после 2000 испытаний «облака»

наблюдаемых значений среднего начинают отчетливо разделяться. Для простоты можно считать, что монетка — неплохой генератор случайного выбора из двух равновероятных вариантов (рис. 4.1).

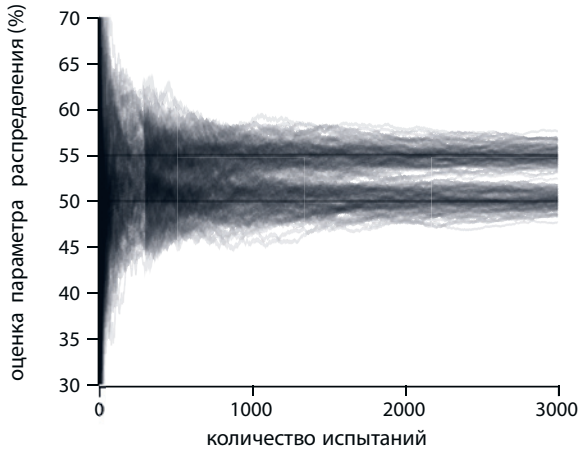


Рис. 4.1. Эксперименты с подбрасыванием идеальной и слегка неидеальной монетки с целью зафиксировать ее неидеальность

Правило 2σ для распределения Бернулли можно использовать в определении доверительного интервала при построении гистограмм. По сути, каждый столбик гистограммы представляет случайную величину с двумя значениями «попал» — «не попал», где вероятность попадания в выделенный интервал соответствует моделируемой функции вероятности. В качестве демонстрации сгенерируем множество выборок для трех распределений: равномерного, геометрического и нормального, — после чего сравним оценки разброса наблюдаемых данных с наблюдаемым разбросом. И здесь мы вновь видим отголоски центральной предельной теоремы, проявляющиеся в том, что распределение данных вокруг средних значений в гистограммах близко к нормальному. Однако вблизи нуля характер разброса изменяется, распределение точек становится близким к другому, часто встречающемуся *экспоненциальному* распределению. Этот пример хорошо показывает, почему я говорил, что в статистике мы имеем дело со случайными значениями параметров случайной величины.

Важно понимать, что правила 2σ и даже 3σ не избавляют нас от ошибок. Они не гарантируют истинности утверждения, это не доказательство. Статистика ограничивает степень недоверия к гипотезе, не более того (рис. 4.2).

Блестящий математик и автор прекрасного курса по теории вероятностей Джан-Карло Рота на своих лекциях в Массачусетском технологическом институте приводил такой пример. Представьте себе научный журнал, редакция которого приняла волевое решение: публиковать исключительно статьи с положительными результатами, которые удовлетворяют правилу 2σ или строже. При этом в редакционной колонке указано, что читатели могут быть уверены: с вероятностью 95% они не встретят на страницах этого журнала неверный результат! Увы, это утверждение легко опровергнуть теми же рассуждениями, что привели нас к вопиющей несправедливости при тестировании водителей на алкоголь. Пусть 1000 исследователей подвергнут опыту 1000 гипотез, из которых верна лишь какая-то часть, скажем 10%. Исходя из смысла проверки гипотез, можно ожидать, что $900 \times 0,05 = 45$ из неверных гипотез ошибочно не будут отвергнуты и войдут в журнал — наряду с $900 \times 0,95 = 95$ верными результатами. Итого из 140 результатов добрая треть окажется неверной!

Этот пример прекрасно демонстрирует наш отечественный закон подлости, который не вошел пока в хрестоматии мерфологии и сформулирован бывшим премьер-министром России Виктором Черномырдиным*:

Хотели как лучше, а получилось как всегда.

Легко получить общую оценку доли неверных результатов, которые войдут в выпуски журнала, при предположении, что доля верных гипотез равна $0 < \alpha < 1$, а вероятность принятия ошибочной гипотезы равна p :

$$x = \frac{(1 - \alpha)p}{\alpha(1 - p) + (1 - \alpha)p}.$$

* Существует также версия (не подтвержденная), что на самом деле автор этой фразы — последний председатель правительства СССР Валентин Павлов. *Прим. ред.*

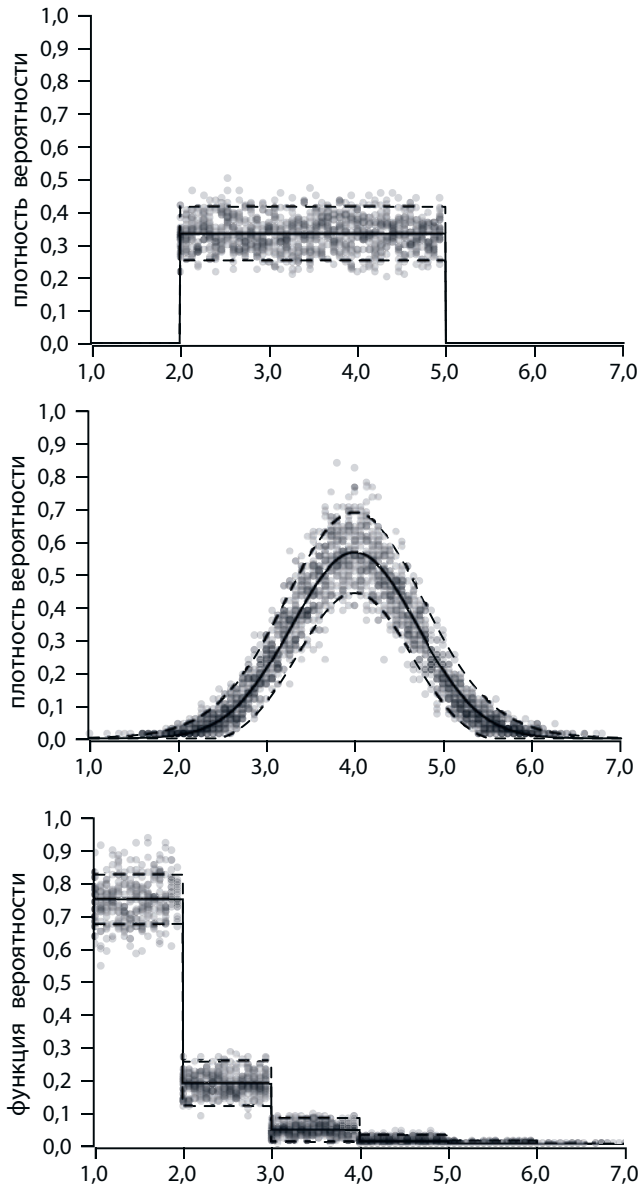


Рис. 4.2. Пример, показывающий соотношение оценки разброса, которая проведена по правилу 2σ , и наблюдаемого разброса для трех случайных величин. Здесь толстой линией показаны истинные распределения, а тонкими — оценка для наблюдаемых отклонений

Области, ограничивающие долю заведомо неверных результатов, которые смогут быть опубликованы в журнале, показаны на рис. 4.3.

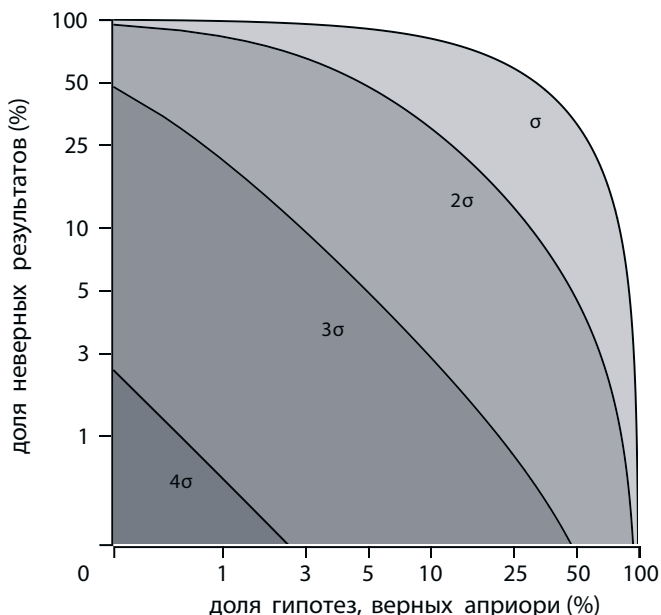


Рис. 4.3. Оценка доли публикаций, содержащих заведомо неверные результаты, при принятии различных критериев проверки гипотез. Видно, что принимать гипотезы по правилу 2σ рискованно, тогда как критерий 4σ уже может считаться весьма сильным

Конечно, мы не знаем этого α и не узнаем никогда, но оно заведомо меньше единицы, а значит, в любом случае утверждение из редакционной колонки нельзя принимать всерьез.

Можно ограничить себя жесткими рамками критерия 4σ , но он требует очень большого числа испытаний. Значит, надо увеличивать долю верных гипотез во множестве возможных предположений. На это и направлены стандартные подходы научного метода познания — логическая непротиворечивость гипотез, их согласованность с фактами и теориями, доказавшими свою применимость, опора на математические модели и критическое мышление.

Так правда ли, что дожди предпочитают выходные дни?

В начале главы мы говорили о том, что выходные и непогода совпадают чаще, чем хотелось бы. Попробуем завершить это исследование.

Каждый дождливый день можно рассматривать как наблюдение случайной величины — дня недели, подчиняющегося распределению Бернулли с вероятностью $1/7$. Примем в качестве нулевой гипотезы предположение, что все дни недели одинаковы с точки зрения погоды и дождь может пойти в любой из них равновероятно. Выходных у нас два, итого получаем ожидаемую вероятность совпадения непогожего дня и выходного равной $2/7$. Эта величина будет параметром распределения Бернулли. Как часто идет дождь? В разное время года по-разному, конечно, но в Петропавловске-Камчатском в среднем наблюдается девяносто дождливых или снежных дней в году. Так что доля дней с осадками составляет около $90/365 \approx 1/4$. Предположим на основании этого, что в течение некоторого периода (месяц, полгода, год) в среднем $1/4$ дней окажутся непогожими. Посчитаем, какое количество дождливых выходных мы должны зарегистрировать, чтобы быть уверенными в том, что существует некоторая закономерность. Результаты приведены в таблице.

Период наблюдений	Лето	Год	5 лет
Ожидаемое число наблюдений	23	90	456
Ожидаемое число положительных исходов	6	26	130
Значимое отклонение	4	9	19
Значимая доля непогожих в общем числе выходных дней	42%	33%	29%

О чем говорят эти цифры? Если вам кажется, что который год подряд «лета не было», злой рок преследует ваши выходные, насылая на них дождь, это можно проверить и подтвердить. Однако в течение лета уличить злой рок можно, лишь если больше двух пятых выходных окажутся дождливыми. Нулевая же гипотеза предполагает, что только четверть выходных должна совпасть с ненастной

погодой. За пять лет наблюдений уже можно надеяться подметить тонкие отклонения, выходящие за пределы 5%, и при необходимости приступить к их объяснению.

Я воспользовался школьным дневником погоды, который велся с 2014 по 2018 год, и выяснил, что за эти пять лет было 459 ненастных дней, из которых 141 пришелся на выходные. Это действительно больше ожидаемого числа на 11 дней, но значимые отклонения начинаются с 19 дней, так что это, как мы говорили в детстве, «не считается».

Вот как выглядят ряд данных и гистограмма, показывающая распределение непогоды по дням недели. Горизонтальными линиями на ней отмечен интервал, в котором может наблюдаться случайное отклонение от равномерного распределения при том же объеме данных (рис. 4.4).

Видно, что, начиная с пятницы, действительно наблюдается увеличение числа дней с плохой погодой. Но для поиска причины роста предпосылок недостаточно: такой же результат можно получать, перебирая случайные числа. Вывод: за пять лет наблюдения за погодой я накопил почти две тысячи записей, но ничего нового о распределении погоды по дням недели не узнал.

При взгляде на записи в дневнике явно бросается в глаза, что непогода приходит не отдельными днями, а двух-трехдневными периодами или даже недельными циклонами.

Это как-то влияет на результат? Можно попробовать принять это наблюдение во внимание и предположить, что дожди идут в среднем по два дня (на самом деле 1,7 дня); тогда вероятность перекрыть выходные увеличивается до 3/7. Тогда ожидаемое число совпадений для пяти лет должно составить 195 ± 21 , или от 174 до 216 раз. Наблюденная величина 141 не входит в этот диапазон, и, значит, гипотезу об эффекте сдвоенных дней непогоды можно смело отвергать. Узнали ли мы что-то новое? Да: казалось бы, очевидная особенность процесса не влечет никакого эффекта. Об этом стоит поразмыслить, и мы этим займемся чуть позже. Но главный вывод таков: какие-то более тонкие эффекты рассматривать нет резона, поскольку наблюдения и, главное, их количество согласованно говорят в пользу самого простого объяснения.

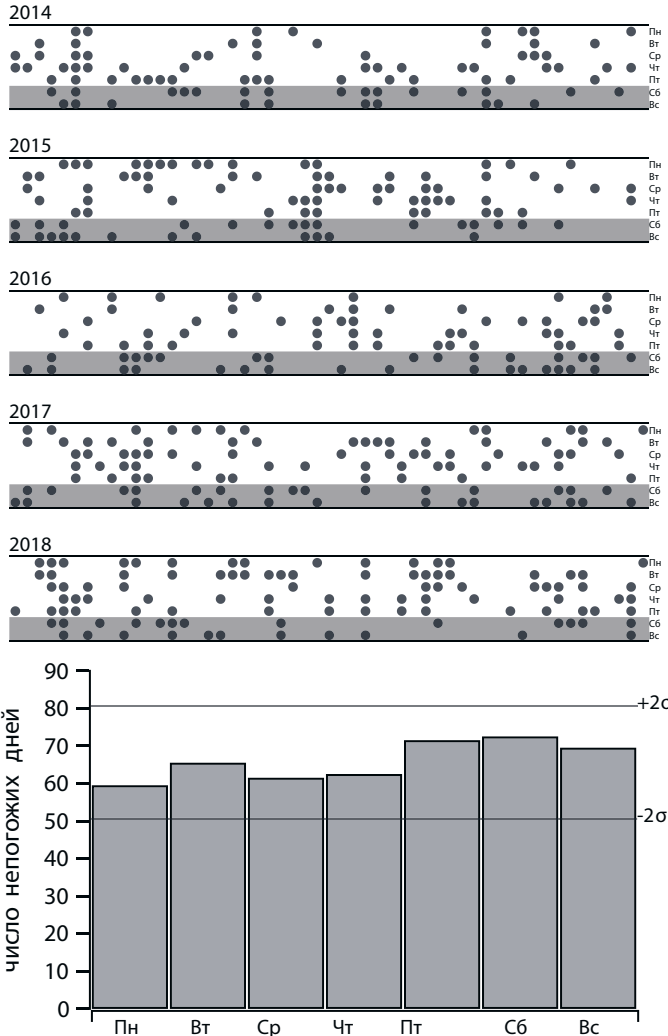


Рис. 4.4. Исходный ряд данных и распределение непогожих дней по дням недели, полученные за пять лет наблюдений

Но недовольство у нас вызывает не пятилетня и даже не годовая статистика: человеческая память не такая долгая. Обидно, когда дождливые дни выпадают на выходные три или четыре раза подряд! Как часто это может случаться? Особенно если вспомнить, что гадкая погода не приходит одна. Задачу можно сформулировать так: «Какова вероятность того, что n выходных подряд окажутся

дождливыми?» В главе 6 мы близко познакомимся с так называемыми случайными процессами как с моделями *случайных последовательностей событий во времени*. Один из них, особенно важный и вместе с тем особенно простой, называется *пуассоновским*. Его характерная особенность — независимость момента наступления следующего события от предыдущих, уже произошедших, а также то, что временные интервалы между событиями подчиняются *экспоненциальному распределению*. Такая последовательность характеризуется одним параметром, который называют *интенсивностью*: числом событий, в среднем случающихся за единичный интервал времени. Разумно предположить, что непогожие дни образуют пуассоновский поток с интенсивностью $1/4$. Это полностью соответствует нашему исходному положению, что в среднем четверть дней любого периода будет непогожей. Если рассматривать только выходные, процесс не должен изменить интенсивность, и из всех выходных непогожие дни должны составлять в среднем тоже четверть. Итак, выдвигаем нулевую гипотезу: ненастья формируют последовательность согласно пуассоновскому процессу с известным параметром, а значит, интервалы между пуассоновскими событиями описываются экспоненциальным распределением. Нас интересуют дискретные интервалы: 0, 1, 2, 3 дня и т. д., — поэтому мы можем воспользоваться дискретным аналогом экспоненциального распределения — *геометрическим распределением* с параметром $1/4$. На рисунке 4.5 показано, что у нас получилось. Очевидно: предположение о том, что мы наблюдаем пуассоновский процесс, нет резона отвергать.

Можно задаться таким вопросом: сколько лет нужно вести наблюдения, чтобы замеченную нами разницу в 11 дней можно было бы уверенно подтвердить или отвергнуть как случайное отклонение? Это легко посчитать: наблюдаемая вероятность $141/459 = 0,307$ отличается от ожидаемой $2/7 = 0,286$ на 0,02. Для фиксации различия в сотых требуется абсолютная погрешность, не превышающая 0,005, что составляет 1,75% от измеряемой величины. Отсюда получаем необходимый объем выборки $n \geq (4.5/7)/(0,01752 \cdot 2/7) \approx 32\,000$ дождливых дней. Это потребует около $4 \cdot 32000/365 \approx 360$ лет непрерывных метеорологических наблюдений, ведь только каждый четвертый день идет дождь или снег. Увы, данных за такой срок нет. Это даже

больше, чем время, которое Камчатка находится в составе России, поэтому шансов выяснить, как обстоят дела «на самом деле», у меня нет. Особенно если учесть, что за это время климат успел измениться разительно — из малого ледникового периода природа выходит в очередной оптимум.

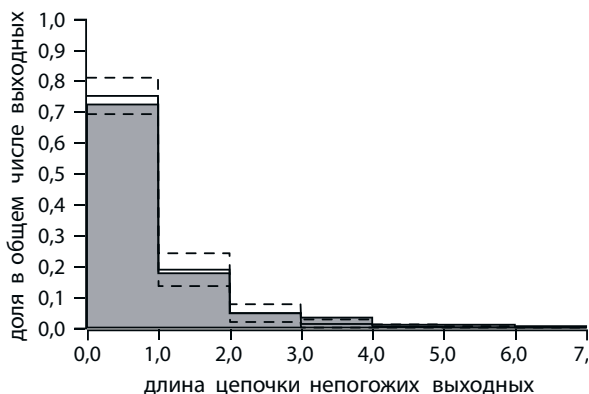


Рис. 4.5. Теоретическое и наблюдаемое распределение длины цепочек неудавшихся выходных. Тонкой линией показаны допустимые отклонения при имеющемся количестве наблюдений

Как же австралийским исследователям удалось зафиксировать отклонение температуры в доли градуса и почему имеет смысл всерьез рассматривать это исследование? Дело в том, что они использовали часовые данные температуры, которые не были «прорежены» каким-либо случайным процессом. Таким образом, за 30 лет метеонаблюдений удалось накопить более четверти миллиона отсчетов с нескольких датчиков, что позволяет уменьшить стандартное отклонение среднего в 500 раз по отношению к стандартному суточному отклонению температуры. Этого вполне достаточно, чтобы говорить о точности в десятые доли градуса. Кроме того, авторы использовали еще один красивый метод, подтверждающий наличие временного цикла: случайное перемешивание временного ряда. Такое перемешивание сохраняет статистические свойства, такие как интенсивность потока событий во времени, однако «стирает» временные закономерности, делая процесс истинно пуассоновским.

Сравнение множества синтетических рядов и экспериментального позволяет убедиться в том, что замеченные отклонения процесса от пуассоновского значимы. Таким же образом сейсмолог Александр Гусев показал, что землетрясения в каком-либо районе образуют своеобразный самоподобный поток со свойствами кластеризации*. Это означает, что землетрясения имеют обыкновение группироваться во времени, образуя весьма неприятные уплотнения потока. Позже выяснилось, что последовательность крупных вулканических извержений обладает тем же свойством.

Беспорядок внутри самих чисел

Конечно, погоду, как и землетрясения, нельзя описывать пуассоновским процессом. Это динамические процессы, в которых текущее состояние оказывается функцией предыдущих. Почему же наши наблюдения за погодой на выходных говорят в пользу простой стохастической модели? Мы отображаем закономерный процесс формирования осадков на множество дней недели, или, говоря на языке математики, на систему вычетов по модулю семь. Этот процесс способен порождать хаос из вполне упорядоченных рядов данных. Отсюда, например, происходит видимая случайность в последовательности цифр десятичной записи большинства вещественных чисел.

Мы уже говорили о рациональных числах, которые выражаются целочисленными дробями. Они имеют внутреннюю структуру, которая определяется двумя числами: числителем и знаменателем. Но при записи в десятичной форме можно наблюдать скачки от регулярности в представлении таких чисел, как $1/2 = 0,5$, или $1/3 = 0,3333... = 0,3$ до периодичного повторения уже вполне беспорядочных последовательностей в таких числах, как $1/17 = 0,0588235294117647$. Иррациональные числа не имеют конечной или периодической записи в десятичной форме, в последовательности цифр чаще всего царит хаос. Но это не значит, что

* Gusev A. A. Multiscale order grouping in sequences of Earth's earthquakes // *Izvestiya, Phys. Solid Earth*. 2005. Vol. 41. Pp. 798–812.

в таких числах нет порядка! Например, $\sqrt{2}$, одно из первых иррациональных чисел, встретившихся математикам, в десятичной записи порождает хаотический набор цифр. Однако, с другой стороны, это число можно представить в виде *бесконечной цепной дроби*:

$$\sqrt{2} = 1 + \frac{1}{2 + \frac{1}{2 + \frac{1}{2 + \dots}}}$$

Нетрудно показать, что эта цепочка действительно равна корню из двух, решив уравнение:

$$x - 1 = \frac{1}{2 + (x - 1)}$$

Цепные дроби с повторяющимися коэффициентами записывают коротко, подобно периодическим десятичным дробям, например: $\sqrt{2} = [1; 2]$, $\sqrt{3} = [1; 12]$. Знаменитое золотое сечение в этом смысле представляет собой проще всего устроенное иррациональное число: $\varphi = [1; 1]$. Все рациональные числа представляются в виде конечных цепных дробей; часть иррациональных — в виде бесконечных, но периодических, такие числа называют *алгебраическими*; те же, что не имеют конечной записи даже в такой форме, — *трансцендентными*. Самое, пожалуй, знаменитое из них — число π , оно порождает хаос как в десятичной записи, так и в виде цепной дроби: $\pi = [3; 7, 15, 1, 292, 1, 1, 1, 2, 1, 2, 1, 14, 2, 1, \dots]$. А вот число Эйлера e , будучи трансцендентным, в форме цепной дроби проявляет внутреннюю структуру, скрытую в десятичной записи: $e = [2; 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, 1, 1, 10, \dots]$.

Наверное, не один математик подозревал мир в коварстве, обнаруживая, что такое нужное, такое фундаментальное число π имеет столь неуловимо сложную хаотичную структуру. Конечно, его можно представить в виде более или менее изящных сумм, произведений, вложенных корней, но все эти ряды, в отличие, например, от цепных дробей, не универсальны и не характеризуют каких-либо особых классов чисел.

Я верю, что математикам будущего откроется какое-нибудь новое фундаментальное представление чисел — столь же универсальное, как цепные дроби, — которое позволит выявить строгий порядок, скрытый природой в числе π , и найти ему подобные.

Результаты этой главы по большей части отрицательные. И, как автор, желающий удивить читателя скрытыми закономерностями и неожиданными открытиями, я сомневался, стоит ли включать ее в книгу. Но наш разговор о погоде ушел в очень важную тему — о ценности и осмысленности естественнонаучного подхода.

Одна мудрая девочка, Соня Шаталова, глядя на мир сквозь призму аутизма, в десятилетнем возрасте дала очень лаконичное и точное определение: «*Наука — это система знаний, основанных на сомнении*». Реальный мир зыбок и норовит спрятаться за сложностью, видимой случайностью и ненадежностью измерений. Сомнение в естественных науках неизбежно. Математика представляется царством определенности, в котором, кажется, можно забыть о сомнении. И очень заманчиво спрятаться за его стенами; рассматривать вместо труднопознаваемого мира модели, которые можно исследовать досконально; считать и вычислять, благо формулы готовы переварить что угодно. Но все же математика — наука, и сомнение в ней отражает глубокую внутреннюю честность, не дающую покоя до тех пор, пока математическое построение не очистится от дополнительных предположений и лишних гипотез. В царстве математики говорят на сложном, но стройном языке, пригодном для рассуждений о реальном мире. Именно поэтому так важно хоть немного познакомиться с этим языком, чтобы не позволять цифрам выдавать себя за статистику, фактам — притворяться знанием, а невежеству и манипуляциям противопоставлять настоящую науку.

Глава 5

Закон арбузной корки и нормальность ненормальности

Глядя новости или читая комментарии к ним, мы порой недоумеваем: «Есть в этом мире нормальные люди?!» Вроде должны быть, ведь нас много и в среднем мы наверняка нормальны. Но при этом мудрецы говорят, что каждый из нас уникален. А подростки уверены, что они-то уж точно отличаются от серой массы «нормальных людей» и ни на кого не похожи.

Небольшое отступление о том, что такое «в среднем». Часто можно услышать шуточные фразы о «средней температуре по больнице» или «средней зарплате», не отражающей действительное распределение. В статистике встречаются несколько разных средних. Чаще всего применяются три вида — выборочное среднее (или просто среднее), выборочная медиана и мода.

Пусть у нас есть выборка $X = (x_1, \dots, x_n)$. Тогда выборочное среднее — обычное среднее арифметическое $(x_1 + \dots + x_n)/n$. Когда мы говорим о среднем росте или средней оценке в школе, обычно подразумеваем именно это.

Однако бывают случаи, когда выборочное среднее не отражает «нужную средность». Представьте, что вы считаете средний доход в городе. Если там живет Билл Гейтс, то вы получите

завышенный результат с точки зрения любой практической задачи. Для исправления ситуации можно использовать, например, медиану.

Возьмем ту же выборку и упорядочим числа по возрастанию: $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$. Такое представление называется вариационным рядом. Здесь $x_{(1)}$ — наименьшее число в выборке, $x_{(2)}$ — второе по величине и т. д. Выборочная медиана — среднее по номеру число в вариационном ряду. Если в нем нечетное число элементов ($n = 2k + 1$), то медиана — элемент $x_{(k+1)}$, а если четное ($n = 2k$), то медианой обычно считают полусумму двух средних элементов вариационного ряда $(x_{(k)} + x_{(k+1)})/2$. Иными словами, медиана — такое число, справа и слева от которого в вариационном ряду поровну элементов. Для оценки дохода (а также во всех иных случаях, когда в выборке могут быть значительные выбросы вверх и вниз) медиана подходит гораздо лучше: если в выборку добавить большое (или маленькое) число, то среднее арифметическое изменится сильно, а медиана гораздо слабее.

Наконец, мода — просто самое частое значение в нашей выборке. Приведем простой пример. Представим себе маленькую компанию, в которой работают пять человек. Директор получает 200 тысяч рублей, его заместитель — 100 тысяч, бухгалтер — 50 тысяч, а два рядовых работника — по 20 тысяч. Тогда выборочное среднее $(200\,000 + 100\,000 + 50\,000 + 20\,000 + 20\,000) / 5 = 78\,000$. Медиана — 50 000 (есть две зарплаты больше этого числа и две меньше). Мода — 20 000 (это значение встречается два раза — чаще других вариантов). Если компания будет зазывать новых работников и утверждать, что средняя зарплата в ней равна 78 000, то это будет формально верно, а на деле надувательство. Здесь нужно ориентироваться на моду: раз вас зовут, то, надо думать, рядовыми работниками, а не директорами.

В этой главе мы поговорим о средних значениях и их репрезентативности. До сих пор мы рассматривали *одномерные распределения* — распределения в одномерном пространстве исходов. Но жизнь многогранна и уж точно не одномерна! А при добавлении дополнительных размерностей порой происходят весьма неожиданные события.

Начнем с многомерного арбуза

Одна из особенностей многомерной геометрии — увеличение доли пограничных значений в ограниченном объеме. Вот что имеется в виду. Рассмотрим классическую задачу об арбузе в пространствах с различной размерностью и зададимся целью выяснить, сколько чудесной сахарной мякоти нам достанется от этого огромного, крепкого и аппетитного арбуза, если, надрезав его, мы выяснили, что толщина его корки не превышает 15% от его радиуса? Кажется, что это многовато, но посмотрите на рис. 5.1: пожалуй, арбуз с такими пропорциями мы сочтем вполне приемлемым. Рассмотрим сначала одномерный арбуз, в виде розового столбика. Его корка представляет собой два маленьких белых отрезочка по краям, ее суммарная длина будет мерой (обобщенным объемом) в одномерном мире и составит 15% от общей меры арбуза. У двумерного, блинообразного арбуза мера корки в виде площади белого кольца будет меньше, чем внутренняя часть, уже всего в три раза. В привычном нам трехмерном мире такая корка составит почти 40% общего объема. Чувствуете подвох?

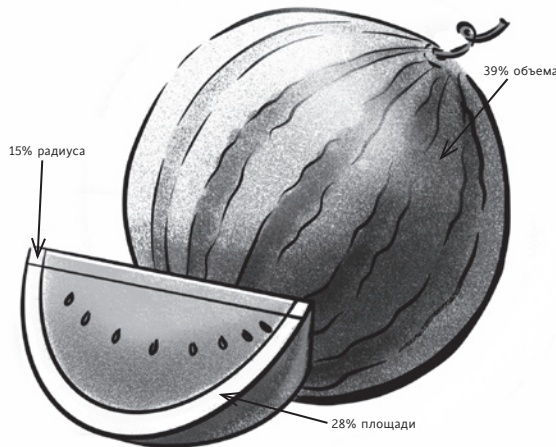


Рис. 5.1. Задача об арбузе

Такую возрастающую роль границ мы уже встречали, когда рассматривали туристический закон подлости. Но тогда мы

ограничились двумерным случаем, вполне естественным для топографических карт. Сейчас мы пойдем дальше.

Для шара, как, впрочем, и для тела произвольной формы, можно точно вычислить зависимость доли корки от общего объема тела. Ее легко получить и обобщить на произвольно многомерные пространства, вновь воспользовавшись анализом размерности и общим понятием меры. Для сплошного тела в пространстве размерности m его мера, или обобщенный объем, пропорциональна степенной функции от характерного размера тела d :

$$V \propto d^m.$$

Под знаком пропорциональности здесь скрывается константа, которая называется *формфактором*. Она зависит от формы тела и размерности пространства, но не зависит от размеров: для куба она равна 1, для шара того же размера выражается сложнее — через гамма-функцию: $\pi^{m/2}/\Gamma(m/2+1)$, которая для целых аргументов сводится к факториалу числа ($\Gamma(n+1) = n!$) и т. д. Ни конкретная форма, ни этот коэффициент для анализа нам не нужны. Под *сплошным* я понимаю тело, не относящееся к *фрактальным*. Такие объекты отличаются от сплошных именно тем, что их обобщенный объем пропорционален их размеру в некоторой дробной степени, отличной от размерности вмещающего пространства. С примерами фрактальных объектов — множеством Жюлиа и губкой Менгера — мы уже встречались раньше, когда рассматривали подмножества нулевой меры. Может показаться, что это экзотика, но природа находит фрактальные решения для очень многих задач: от роста кристаллов до разряда молнии, от корневой системы растений до устройства наших легких. Но, повторюсь, здесь мы будем рассматривать только сплошные тела.

С объемом как с мерой мы разобрались в главе 1, а что такое характерный размер? Мы можем сказать, что человек имеет характерный размер порядка метра, а муравей — миллиметра. В то же время характерный размер нашей Галактики — 100 тысяч световых лет. Все эти объекты имеют весьма сложную форму, но когда мы говорим о характерных размерах, она нас не интересует. Это понятие

можно строго определить как среднее геометрическое размеров тела в разных направлениях или как диаметр шара, имеющего такой же объем, как и рассматриваемое тело.

Объем корочки равен следующей разнице:

$$V_{\text{корки}} = V_{\text{общ}} - V_{\text{внутр}},$$

а отношение объема корки, составляющей долю δ от размеров тела, к общему объему выражается так:

$$\frac{V_{\text{корки}}}{V_{\text{общ}}} = 1 - \frac{V_{\text{внутр}}}{V_{\text{общ}}} = 1 - (1 - \delta)^m.$$

Как хорошо получилось — мы перешли от пропорциональности к точному равенству. Все благодаря отношениям, в которых сократились неизвестные нам формфактор и размеры тела. Таким образом, полученное соотношение объема корки и объема тела универсально и годится для арбузов сколь угодно сложной формы.

Вот как выглядит график роста доли пятнадцатипроцентной по радиусу корочки арбуза в его объеме при дальнейшем увеличении размерности пространства (рис. 5.2).

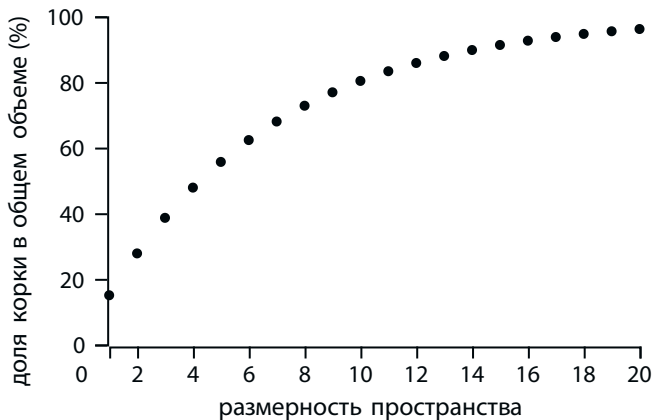


Рис. 5.2. В четырехмерном пространстве наш условно тонкокорый арбуз оставит нам уже лишь половину мякоти, а в одиннадцатимерном мы сможем полакомиться 15% арбуза, выбросив корочку, составляющую 15% его радиуса!

Итак, сейчас мы готовы сформулировать глубокомысленный закон **арбузной корки**:

**Покупая многомерный арбуз, ты приобретаешь
в основном его корку.**

Мне одному кажется, что я нормальный?

Обидно, конечно, но какое это имеет отношение к нормальности нашего мира и законам подлости? Увы, именно этот закон препятствует отысканию так называемой золотой середины, обесценивает результаты социологических опросов и повышает роль маловероятных неприятностей.

Дело в том, что пространство людей со всеми их параметрами существенно многомерно. В качестве различных размерностей можно рассматривать и очевидные рост, вес, возраст и достаток, а также уровни интеллектуального (IQ) и эмоционального (EQ) развития; наконец, наблюдаемые, хоть и плохо формализуемые черты лица либо характера — такие как уровень болтливости, упрямства или влюбчивости — тоже относятся к нашим параметрам. Мы без труда насчитаем с десяток-полтора величин, характеризующих человека. И для каждого из этих параметров существует некая статистически определяемая «норма» — самое ожидаемое и, более того, часто наблюдаемое значение. Сколько же в таком богатом пространстве параметров окажется людей, типичных во всех отношениях? Выражение, которое мы использовали для определения отношения объемов корки и арбуза, можно использовать и для вычисления вероятности попасть в число хоть в чем-то, но «ненормальных». Если мы сочтем все параметры независимыми (для некоторых пар параметров это может быть верно только приближенно), вероятность удовлетворить всем критериям типичности одновременно равна произведению вероятностей оказаться типичным по каждому критерию отдельно.

И вновь колмогоровское определение вероятности, которое мы ввели в самом начале, сильно упростит задачу, избавив нас

от пугающих формул, по которым нельзя ничего толком вычислить. Полученная нами формула арбуза работает для любых, сколь угодно сложных форм. В том числе не имеющих границы, подобно атмосфере Земли, уходящей далеко в космическое пространство, становясь все тоньше. Так что нам не нужно знать, каким именно распределениям подчиняются обсуждаемые качества людей, остается лишь предположить, что у них есть среднее значение (а это, как мы увидим, бывает не всегда). Если обозначить как P_{out} вероятность оказаться за пределами области, которую мы сочли бы нормой, то вероятность оказаться ненормальным в *чем-нибудь* при рассмотрении m критериев будет вычисляться по «арбузной» формуле (рис. 5.3):

$$P = 1 - (1 - P_{\text{out}})^m.$$

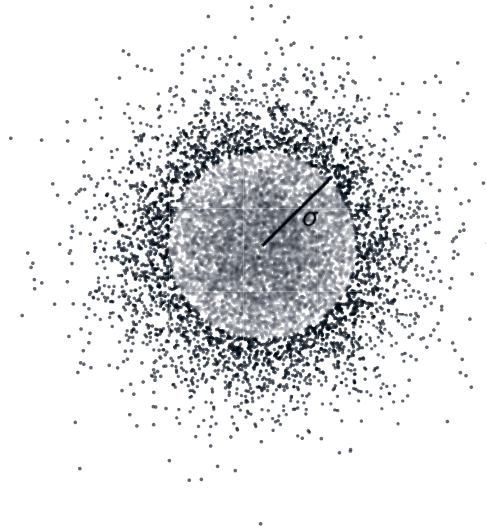


Рис. 5.3. Математическая модель арбуза

Вот она — сила правильно выбранной модели! Толщину корки арбуза мы измеряли линейкой, попадание случайной величины в какой-нибудь диапазон — вероятностью. Какой бы малой ни была вероятность P_{out} , при $m > \ln(1/2)/\ln(1 - P_{\text{out}})$, значение P превысит $1/2$.

Для внесения хоть какой-то конкретики можно предположить, что параметры, о которых мы говорим, имеют нормальное распределение. Это вполне разумно для наших целей, ведь мы не говорим о каком-то конкретном наборе характеристик, а, прямо скажем, фантазируем, стараясь сформулировать хоть что-то определенное в столь зыбкой теме. Выбор нормального распределения адекватно отражает степень нашего неведения, и загрузиться подробностями до тех пор, пока не видна самая общая картина, рановато. Итак, наш арбуз превратился в размытое туманное пятно, что не мешает нам вычислить долю его «корки». Для «хорошего» в каком-то смысле распределения за норму можно принять значения, не отклоняющиеся от среднего больше чем на величину стандартного отклонения. Для нормального распределения доля значений, выходящих за пределы нормы, имеет $P_{out} = 16\%$, примерно как в рассмотренном нами реальном арбузе. Применительно к нашему нечеткому арбузу здесь имеется в виду вероятность оказаться на удалении в одно стандартное отклонение от среднего, как показано на рис. 5.4. При более толерантном понимании нормы можно ограничиться двумя стандартными отклонениями, получив $P_{out} = 2,3\%$.

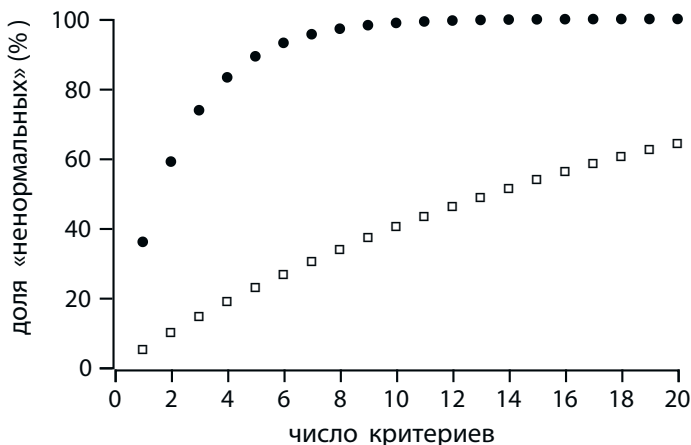


Рис. 5.4. Вероятности оказаться «ненормальным» для разного числа критериев сравнения и «строгости» определения нормы. Верхний и нижний графики различаются тем, что при определении «нормальности» используют радиус в одно и два стандартных отклонения соответственно

Что ж, выходит, это нормально — быть хоть в чем-то ненормальным. Оценивая людей по десятку параметров, будьте готовы к тому, что полностью заурядными окажутся лишь 2% общей популяции. Причем как только мы их разыщем, они тут же станут знаменитостями, утратив свою заурядность!

В погоне за Нормой

Нетипичность нормы и ментальные ошибки, к которым может привести попытка усреднения многопараметрических систем, подробно рассматриваются в книге Тодда Роуза «Долой среднее!»*. В частности, в ней приводится история времен начала Второй мировой войны. В попытке разобраться в причинах ошибок пилотов боевых самолетов командование ВВС США предприняло исследование, основной целью которого было уточнить средние характеристики летчиков. От этих параметров зависели конкретные инженерные решения по проектированию эргономики кабины. Считалось, что чем точнее будут известны эти характеристики, тем более эргономичной окажется разработанная на их основе техника. Каково же было удивление молодого антрополога Гилберта Дэниэлса, которому поручили эту работу, когда выяснилось, что из четырех тысяч обмеренных им пилотов не обнаружилось ни одного «среднего», для которого кабина самолета оказалась бы удобной по всем параметрам. Всего использовалось 10 физических характеристик, и Дэниэлс придерживался очень строгого критерия «нормальности»: выходящим за пределы нормы считалось отклонение от среднего, превышающее 30% от всей выборки. Мы теперь можем вычислить, что для десяти параметров вероятность попасть в нормальные значения по таким критериям составит 0,0006% — 1 человек на 170 тысяч! В конце концов Дэниэлс пришел к заключению, опубликованному уже после войны: в реальности среднего пилота не существует. Если вы проектируете кабину для него, то она не

* Издана на русском языке: Роуз Т. Долой среднее! Новый манифест индивидуальности. М. : Манн, Иванов и Фербер, 2018. *Прим. науч. ред.*

подойдет ни для кого. Чтобы повысить эффективность солдат, в том числе летчиков, рекомендуются радикальные изменения: окружение должно соответствовать индивидуальным параметрам, а не средним.

Кроме того, Тодд Роуз приводит историю из мирной жизни. Газета Plain Dealer объявила конкурс среди женщин и девушек. Им предлагалось прислать параметры своего тела, и победить должны были те представительницы прекрасного пола, которые окажутся ближе всего к параметрам «типичной женщины». Нормы, увековеченной в статуе из медицинского музея Кливленда (рис. 5.5). Норма родилась вследствие усреднения 15 000 женщин разного возраста и должна была олицетворять идеал, «определенный самой Природой». Всего рассматривалось девять параметров, и из 3864 конкурсанток ни одна не попала в средние значения. По пяти критериям «нормальными» оказались лишь 10% участниц, что дает нам возможность оценить использованную жюри «толщину корки» в 75%. С таким суровым подходом надеяться найти хотя бы один «идеал» в пространстве девяти измерений можно, лишь рассмотрев 260 тысяч красавиц. На все человечество таких «идеальных» барышень наберется от силы пара тысяч человек.

Далее Роуз отмечал: Дэниэлс и организаторы конкурса получили одинаковый результат, но сделали совершенно разные выводы. Большинство врачей и ученых того времени не сочли, что Норма представляет неправильный идеал. Наоборот: они решили, что большинство американских женщин нездоровы и не поддерживают нормальную форму. Одним из них был доктор Бруно Гебхард, директор медицинского музея Кливленда. Он сокрушался, что послевоенные женщины малопригодны к службе в армии, и упрекал их, ссылаясь на плохую физическую форму, в том, что они «плохие производители и плохие потребители». Дэниэлс говорил прямо противоположное: о том, что усреднение людей — ловушка, которая многих приводит к просчетам. Ведь почти невозможно найти среднего летчика не в силу каких-то индивидуальных черт его группы, а из-за большого разброса параметров в размерах тела у людей.

Тот самый закон подлости

Один из классических законов подлости, сформулированный в сердцах инженером Эдвардом Мёрфи, гласит:

Все, что может пойти не так, пойдет не так.

Сейчас мы можем взглянуть на него не только иронично.

Пусть для выполнения некоторой работы требуется совершить ряд действий, и для каждого из них существует маленькая, но отличная от нуля вероятность неудачи. Какова вероятность того, что все задуманное пройдет без сучка без задоринки? Мы имеем дело с пересечением множества событий, каждое из которых соответствует *успешному завершению* того или иного этапа работы. Как посчитать вероятность для пересечения двух событий, мы уже знаем: для этого нужно перемножить вероятность второго события при условии, что первое случилось, на вероятность первого события:

$$P(A_1 \cap A_2) = P(A_2 | A_1) \cdot P(A_1).$$

Операция пересечения *ассоциативна*: мы можем в произвольном порядке расставлять скобки для трех и более пересекающихся событий.

$$A_1 \cap A_2 \cap A_3 = (A_1 \cap A_2) \cap A_3 = A_1 \cap (A_2 \cap A_3).$$

Отсюда легко получить общую формулу для пересечения произвольного числа событий:

$$\begin{aligned} P(A_1 \cap A_2 \cap \dots \cap A_n) &= \\ &= P(A_1 | A_2 \cap \dots \cap A_n) \cdot P(A_2 | A_3 \cap \dots \cap A_n) \cdot \dots \cdot P(A_n). \end{aligned}$$

Если события независимы, то мы получаем произведение вероятностей наступления каждого из них:

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1) \cdot P(A_2) \cdot \dots \cdot P(A_n).$$

Но для нас важно, что вероятности, условные или нет, по определению должны быть меньше единицы, а значит, мы вправе использовать закон арбузной корки: чем больше число шагов, тем существеннее роль границ. В нашем случае границами становятся нештатные ситуации. Достаточно дюжины шагов, чтобы средняя вероятность такой ситуации или ошибки в 5% на одном шаге выросла до вероятности провала всего дела!

Эти наши рассуждения чрезвычайно просты, а закон Мёрфи — скорее эмоции, чем объективность, да и в целом кажется трюизмом. Но все же именно с этого наблюдения в сороковые-пятидесятые годы двадцатого века началась новая большая наука: *теория надежности*. Она добавила в рассмотрение время, взаимосвязь элементов систем, экономику, а также человеческий фактор, и нашла применение за пределами инженерных наук: в экономике, теории управления и, наконец, программировании. Мы еще вернемся к этой теме, когда будем изучать *проклятие режиссера*, заставляющее принтер барахлить именно в день сдачи проекта. Закон Мёрфи с учетом времени — поистине страшная сила!

В связи с рассуждениями о вероятности пересечения множества событий может возникнуть интересный и непростой вопрос. Если вероятность определена как мера, то она должна обладать свойством аддитивности. Иначе говоря, мера целого должна быть суммой мер его частей. Но мы рассмотрели вероятность успеха для некоего дела со множеством этапов и увидели другую картину: вероятность целого оказалась равна произведению вероятностей для его частей, а не сумме. Это соответствует свойству *мультипликативности*. Так аддитивна вероятность или мультипликативна? Тут следует различать *вероятностное пространство*, на котором вероятность играет роль аддитивной меры и в котором сложение целого из частей выполняется с помощью операции объединения событий, и *фазовое пространство* некоторой системы, содержащее все возможные ее состояния. Фазовое пространство измеримо, но вероятность мерой в нем не является. Чтобы произошло событие, соответствующее попаданию системы в заданное состояние, все ее составные части должны *одновременно* попасть в свои конкретные состояния — тогда возникнет пересечение соответствующих

событий. Таким образом, вероятности этих событий перемножаются. Однако превратить вероятность в «нормальную» аддитивную меру на фазовом пространстве можно и нужно. Мы совершим это превращение, когда будем говорить об *энтропии* систем и распределений случайных величин в главе 9.

Счастье — это найти друзей с тем же диагнозом, что и у тебя

А можно ли вообще ставить вопрос о соответствии какой-то норме, не пытаемся ли мы при этом оценивать и сравнивать? Вы спросите: что же в этом плохого? Мы все время кого-нибудь с кем-нибудь сравниваем, чаще всего себя с другими, но иногда позволяем себе оценить и кого-нибудь еще. Однако с точки зрения математики все не так просто. Чтобы сравнивать что-либо с чем-либо, нужно правильно определить *отношение порядка* или ввести *метрику*.

Определить отношение порядка — значит обозначить, что один элемент некоего множества в каком-то смысле предшествует другому. Этому мы научились еще в школе: 2 меньше 20, слон слабее кита, уговор дороже денег и т. п. Но вот вам ряд вопросов. Что идет раньше — понедельник или вторник? А воскресенье или понедельник? А какое воскресенье — то, что перед понедельником, или то, которое после субботы? А какое комплексное число больше: $2 + 3i$ или $3 + 2i$? Мы можем назвать по порядку цвета радуги и даже ассоциировать все промежуточные цвета с вещественным числом — частотой света. Но кроме этих цветов существует множество неспектральных. Они образуют хорошо знакомое типографам и дизайнерам цветовое пространство, в котором каждый цвет имеет три «координаты». Так можно ли все видимые глазом цвета выстроить по порядку?

Эти примеры показывают, что с отношением порядка бывают трудности. Например, для отношения «один день недели наступает после другого» не работает свойство *транзитивности* (из того, что воскресенье наступает позже четверга, а четверг — позже

понедельника, не следует, что воскресенье всегда наступает позже понедельника), так же как не транзитивно отношение «сильнее» в игре «камень-ножницы-бумага». Попытка ввести понятие больше/меньше на поле комплексных чисел не согласуется с арифметикой этих чисел, а цвета, которые можно параметризовать тремя «координатами» (тон, насыщенность, яркость), обладают обоими этими недостатками: и отсутствием транзитивности для тона — своеобразной «угловой» характеристики цвета, которая зациклена подобно дням недели; и существенной многомерностью. Даже на привычном нам множестве рациональных чисел отношение порядка хоть и определено, но не дает возможности указать наименьшее или наибольшее число на каком-либо открытом интервале.

Итак, мы видим, что отношение порядка вовсе не так просто, как мы привыкли думать, а главное — не универсально. Но мы все-таки можем сравнивать людей, книги, блюда, языки программирования и прочие объекты, имеющие множество параметров, пусть даже условно формализуемых? Можем, используя вместо сравнения другую концепцию — степень подобия объектов между собой, или *метрику*. Фильмы про Индиану Джонса ближе к «Пиратам Карибского моря», чем к комедиям Вуди Аллена или документалистике. Русский язык ближе к польскому, чем к немецкому, и совсем не похож на суахили. Числа $2+3i$ или $3+2i$ ближе друг к другу, чем к числу 100. Если мера обобщает размеры (длину, объем и т. д.), то метрика, введенная в математику Морисом Фреше в 1906 году, — это обобщение понятия «расстояние». Вот ее определение.

Пусть имеется произвольное множество X . Метрика — функция ρ , сопоставляющая любым двум элементам x и y множества вещественное число $\rho(x, y)$ и при этом удовлетворяющая таким условиям:

- 1) $\rho(x, y) = 0$ тогда и только тогда, когда $x = y$ (аксиома тождества: расстояние между точками равно нулю, если эти точки совпадают);
- 2) $\rho(x, y) = \rho(y, x)$ (аксиома симметрии: расстояние в обе стороны одинаково);

3) $\rho(x,z) \leq \rho(x,y) + \rho(y,z)$ (неравенство треугольника — аналог знакомого утверждения из курса геометрии: окружной путь не может быть короче прямого).

Множество X с введенной метрикой называется *метрическим пространством*. Из приведенных аксиом следует, что метрика — неотрицательная функция. Рассмотрим неравенство треугольника для случая $x = z$:

$$0 = \rho(x,x) \leq \rho(x,y) + \rho(y,x) = 2\rho(x,x), \text{ откуда } \rho(x,x) \geq 0.$$

Понятие метрики позволяет вводить аналог расстояния (или степени близости) в совсем неочевидных случаях, например на бесконечномерном пространстве функций, между строками текста или изображениями; наконец, между распределениями случайных величин. Введение метрики не решает всех проблем, но в отсутствие внятной и корректной метрики легко увязнуть в бесконечном, бурном и бессмысленном споре, который в околокомпьютерной среде известен как «холивар» (от англ. holy war — священная война). Увы, жаркие споры возникают чаще всего уже на этапе выбора метрик, поскольку они сами образуют некое множество, на котором тоже нужно определять отношение порядка «лучше/хуже». Впрочем, можно предложить вполне осмысленный способ рассуждений о сравнимости многомерных объектов, например людей.

В многомерном пространстве параметров каждый объект может быть представлен вектором — набором чисел, определяющих значения критериев, которые его характеризуют. Рассматривая ансамбль векторов (например, человеческое общество), мы увидим, что какие-то из них окажутся сонаправлены или по крайней мере близки по направлениям; вот их-то уже вполне можно сравнивать по длине. В то же время какие-то векторы *ортогональны* (в геометрическом смысле — перпендикулярны, в более широком — независимы), и соответствующие им люди попросту друг другу непонятны: они по ряду параметров в *сопряженных пространствах*, как пресловутые физики и лирики. Нет смысла рассуждать о том, что хороший поэт в чем-то лучше либо хуже талантливого инженера или одаренного природой

спортсмена. Единственное, о чём можно судить, — о длине вектора, то есть степени одаренности, расстоянии от среднего.

В связи с этим может возникнуть любопытный вопрос: а какая доля случайных векторов в пространстве заданной размерности будет сонаправленной, а какая ортогональной? Как много удастся найти единомышленников или хотя бы тех, с кем можно себя сравнить?

В двумерном мире каждому вектору соответствует одномерное пространство коллинеарных (сонаправленных) и одномерное пространство ортогональных векторов. Если мы рассмотрим «почти» сонаправленные и «почти» ортогональные векторы, то они образуют секторы одинаковой меры (неважно, площади или угла) при одинаковом выборе допустимого отклонения. Иначе говоря, похожих и непохожих объектов при рассмотрении двух критериев будет одинаковое количество (под количеством мы опять понимаем меру на множестве этих критериев, рис. 5.5).

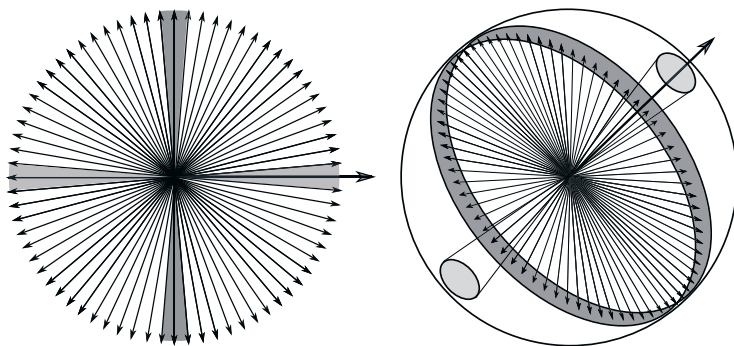


Рис. 5.5. Почти коллинеарные и почти ортогональные векторы в двумерном и трехмерном пространстве

В трехмерном мире картина поменяется. Сонаправленные векторы всё так же образуют одномерное пространство, а вот ортогональные уже заполняют плоскость, двумерное пространство. С точки зрения ортогональных векторов мера сонаправленных уже равна нулю, но все же позволим векторам немного отклониться от курса. Фиксируя их длину R и допуская небольшое отклонение от идеальных направлений на угол $\Delta\varphi$, можно количество почти сонаправленных векторов сопоставить с площадью круговых областей

вокруг полюсов $2\pi R^2 \Delta\varphi^2$, а число почти ортогональных — с площадью полосы вокруг экватора: $4\pi R^2 \Delta\varphi$. Их отношение $2/\Delta\varphi$ растет неограниченно при уменьшении отклонения $\Delta\varphi$.

В четырехмерном мире ортогональные векторы образуют уже трехмерное пространство, тогда как сонаправленные всё еще лежат в одномерном, и разница в их количестве растет уже пропорционально квадрату отклонения от идеала. Но на этом этапе лучше обратиться к теории вероятностей и выяснить, каковы шансы получить ортогональные или сонаправленные векторы, взяв наугад два вектора из пространства размерности m . Об этом нам расскажет распределение углов между случайными векторами (рис. 5.6). К счастью, рассуждая о площадях многомерных сфер, распределение можно вычислить аналитически и даже представить в конечной форме:

$$p(\varphi) = \frac{\Gamma\left(\frac{m}{2}\right)}{\sqrt{2}\Gamma\left(\frac{m-1}{2}\right)} \sin^{m-2}(\varphi), \text{ для } 0 \leq \varphi \leq \pi.$$

Здесь $\Gamma(x)$ — гамма-функция, обобщение факториала на вещественные (и даже комплексные) числа. Ее основное свойство: $\Gamma(x+1) = x\Gamma(x)$.

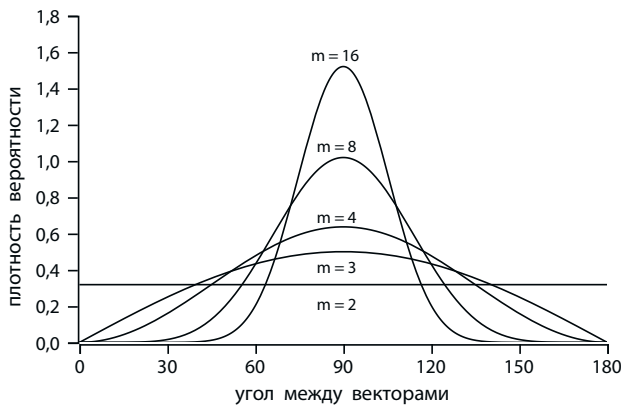


Рис. 5.6. Распределения углов случайных векторов в пространствах различных размерностей

Для двумерного пространства углы распределяются равномерно, для трехмерного — пропорционально синусоидальной функции. Свойства синуса приводят к тому, что плотность вероятности в нуле для $m > 2$ в точности равна нулю. Это согласуется с нашими рассуждениями о том, что сонаправленные векторы образуют множества нулевой меры. Для всех размерностей выше двух мода распределения приходится на 90° , и доля взаимно ортогональных векторов увеличивается по мере роста числа параметров. Самое же главное наблюдение — сонаправленных векторов (имеющих угол около 0° или 180°) практически не остается при достаточно высокой размерности пространства. Если считать более или менее похожими (сонаправленными, сравнимыми) векторы, имеющие угол менее 30° , то при сравнении по двум критериям похожей на какой-то выделенный вектор окажется треть всех случайных векторов, а при увеличении размерности пространства на единицу доля сравнимых векторов будет уменьшаться практически вдвое. Таким образом, мы приходим к векторной формулировке закона арбузной корки:

В пространствах высокой размерности почти все векторы ортогональны друг другу.

Или эквивалентно: на вкус и цвет товарищей нет.

Этот странный закольцованный мир

По мере повышения размерности распределение углов становится похожим на нормальное. Однако это не оно, несмотря на характерную колоколообразную форму. Нормальное распределение определено для всей вещественной числовой оси, в нашем же случае значение угла зациклено в пределах от 0 до 180° . Мы попали из поля вещественных чисел на кольцо вычетов — математическую структуру, подобную циферблату на часах, дням недели или остаткам от деления. Применяя привычные нам операции в этом

кольцевом мире, нужно быть аккуратным, даже выполняя простые расчеты. Скажем, чему равно среднее значение для двух углов: 30° и 350° ? Простое сложение даст ответ 190° , тогда как чертеж покажет, что правильным ответом будет 10° . А чему равно среднее значение равномерного распределения на всей окружности? Оно не определено, хотя площадь под кривой распределения конечна. Даже простое вычисление среднего для набора измеренных углов уже становится нетривиальной задачей, требующей перехода на плоскость (декартову или комплексную). Представьте себе, что вы исследуете зависимость числа обращений граждан в полицию от времени суток и получили гистограмму, показанную на рисунке слева (рис. 5.7).

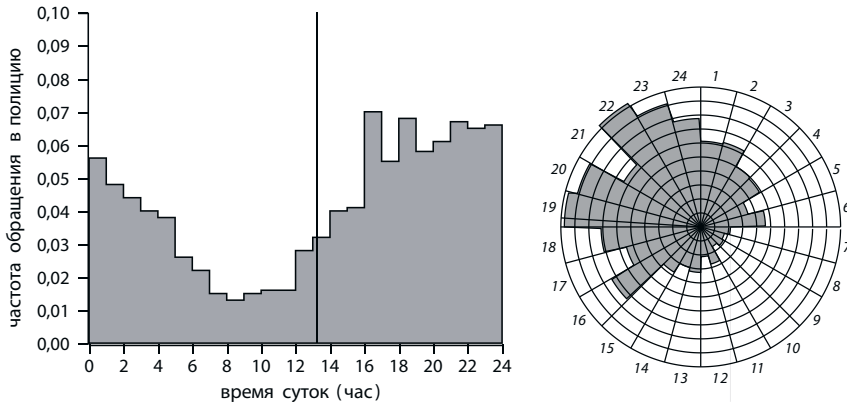


Рис. 5.7. Гистограмма, показывающая распределение числа событий по времени суток, не отражает цикличности времени и не дает возможности правильно найти среднее значение

Попытка вычислить математическое ожидание для самого неспокойного времени с помощью среднего арифметического даст невнятный результат. Он показан на рисунке вертикальной линией. Правильно будет изобразить нашу гистограмму в полярных координатах и там уже найти математическое ожидание, вычислив угловую координату положения *центра масс* получившейся фигуры. Ее можно визуализировать, построив из центра координат луч, проходящий через центр масс.

Привычные распределения вероятностей с хорошо известными свойствами на кольцах вычетов «зацикливаются» и становятся своеобразными. На рисунке 5.8 показано, как можно построить аналоги некоторых распределений на окружности. Числовая ось как бы наматывается на окружность, при этом каждый слой спирали суммируется, и в результате мы получаем циклический аналог распределения, имеющий единичную площадь.

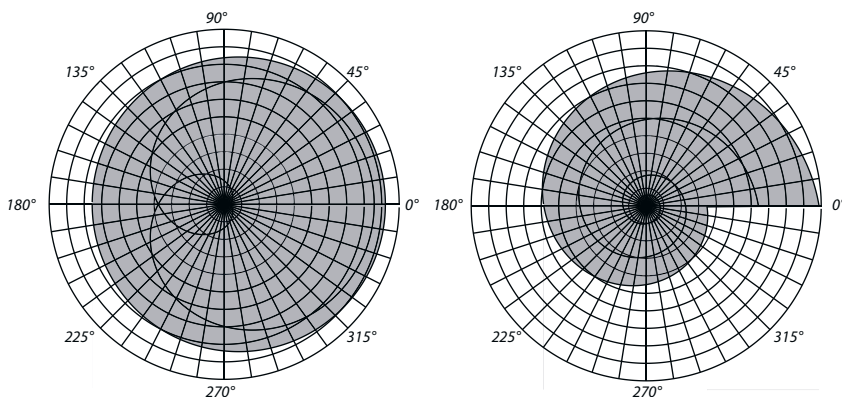


Рис. 5.8. Построение циклических экспоненциального (слева) и нормального (справа) распределений (показаны тонкой линией). Тут же приведены графики функций плотности для обыкновенных (линейных) распределений (показаны жирными линиями)

Например, циклическое экспоненциальное распределение (рис. 5.9) описывает *случайное положительное отклонение от заданного угла с заданным средним значением*. С его помощью можно описать *время суток*, в которое ожидается появление пуассоновского события. Циклическое нормальное распределение допустимо использовать для описания погрешностей в измерении углов. Хотя, если быть точным, они будут подчиняться другому распределению, но об этом чуть позже. Циклические распределения, хоть они и выглядят несколько однообразно, важны при анализе данных на земном шаре, если их дисперсии сравнимы с длиной экватора, а это характерно для широкого класса задач геофизики, климатологии и других наук о Земле.

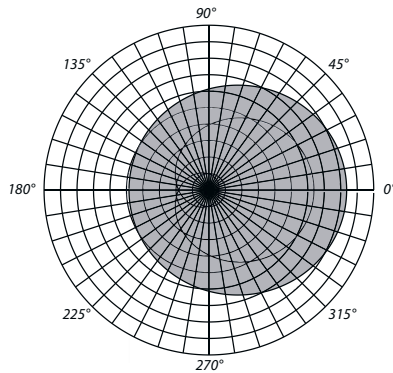


Рис. 5.9. Циклический аналог распределения Коши

Любопытно, что при зацикливании свойства распределения могут поменяться радикально. Например, относительная погрешность при измерении нулевой величины описывается распределением Коши. Оно примечательно тем, что ее функция плотности вероятности имеет бесконечную площадь под кривой, так что для этого распределения невозможно вычислить значения среднего и дисперсии: они, в отличие от моды и медианы, для распределения Коши просто не определены. Однако круговой аналог этого распределения ведет себя хорошо, интегрируется и имеет вычисляемые значения среднего и дисперсии. Это распределение встречается, например, в физике — при анализе явления дифракции.

Меняет свои свойства при зацикливании и нормальное (гауссовское) распределение. Его циклический аналог уже не будет устойчивым, а суммы случайных величин начнут сходиться не к нему. На окружности эту роль играет распределение фон Мизеса с такой функцией плотности вероятности:

$$P_{k,\mu}(x) = \frac{e^{k \cos(x-\mu)}}{2\pi I_0(k)}.$$

Среднее значение для этого распределения равно μ , а величина $1/k$ влияет на дисперсию. В приведенном выражении I_0 — модифицированная функция Бесселя, одна из целого семейства

специальных функций. Функции Бесселя обычно появляются, если в задаче есть осевая симметрия. Например, с их помощью описывается профиль круговых волн, разбегающихся по воде от упавшей капли.

Впрочем, когда дисперсия данных мала и x незначительно отклоняется от среднего значения μ , косинус можно разложить в степенной ряд, в котором главную роль играет квадратичный член. Таким образом, когда влияние цикличности становится незначительным, то и распределение фон Мизеса оказывается похожим на «обычное» гауссовское. Никуда от него не денешься — в отличие от мифического «среднего пилота» или усредненной «идеальной женщины», случайные величины, подчиняющиеся нормальному распределению, встречаются повсеместно!

Сравниваем и ищем с помощью вероятности

Наш опыт работы с вероятностным подходом учит тому, что вероятность можно вычислить, но дать ей однозначную интерпретацию непросто. Еще сложнее измерить эту характеристику явления или процесса. Хорошо, когда можно применить частотную интерпретацию: пронаблюдать достаточно долго за процессом или его моделью и получить оценку распределения управляющих параметров. Но вероятности возникают и по-другому.

В самом начале книги упоминался спам-фильтр, который каждому сообщению в электронной почте ставит в соответствие вероятность того, что это спам — назойливая реклама, рассылаемая безадресно. Это что за вероятность? К какому множеству сообщений она относится? Когда эксперт в интервью утверждает, что вероятность победы того или иного кандидата на государственный пост составляет, скажем, 75%, сколько раз он собирается проводить выборы, чтобы реально измерить это число и проверить свои выводы? А если это нельзя измерить, как проверять утверждение? Понятно, что в случае с выборами утверждение о вероятности чисто умозрительное и к математике не имеет отношения — число здесь

отражает в лучшем случае некую «уверенность по стобалльной шкале». Но бесстрастные автоматы, классифицирующие сообщения в почте, изображения с городских камер или предаварийные состояния сложной техники, выдают результат именно на языке вероятностей.

Эти вопросы заставляют рассуждать о вероятности уже не только как о мере, но и как о характеристике, позволяющей сравнивать трудно сравнимые вещи.

Спам-фильтр сообщает нам о степени благонадежности текста, честно вычисляя условную вероятность того, что сообщение рекламное, исходя и из частотности характерных слов, и, что очень важно, спама среди прочих сообщений (это позволяет избежать ошибок вроде тех, что обсуждались в главе 3: про тест на содержание алкоголя в крови и истинность научных публикаций). А в результате мы получаем некое число, по которому можем ранжировать сообщения, имея в виду степень «близости» или «похожести» текста на спам. Причем оно не показывает степень близости к какому-то одному «идеальному спаму», его и не существует вовсе; спам — некое очень сложное подпространство в пространстве возможных сообщений.

Можно посчитать, какая доля сообщений, принятая фильтром за спам, действительно им оказалась. Однако это измерение покажет некоторую суммарную характеристику эффективности фильтра и его настроек, например выбранного порога близости к спаму, но ничего не скажет о частотной интерпретации результата: «с вероятностью 87% данное сообщение — спам».

Вероятность в такой роли используется в современном подходе к поиску самолета или судна, потерпевшего крушение в океане. Эта методика называется *байесовским поиском*, поскольку в ее основе лежит понятие условной вероятности, рассчитываемой согласно теореме Байеса. В начале поисково-спасательных работ обследуемый участок территории разбивается на отдельные клетки (квадраты), потом с учетом направления движения судна или самолета строится априорное распределение вероятности того, что искомый объект находится в том или ином квадрате. Поисковые работы далее проводятся в двух основных направлениях: проверка наиболее

вероятных квадратов и отсекают маловероятных. Таким образом, вероятность превращается в метрику, позволяющую сравнивать квадраты между собой: не просто прочесывать местность, перебирая их по порядку, а сосредоточиваться на наиболее вероятных участках, экономя драгоценное время.

Во время поисковых работ квадраты, оказавшиеся пустыми, отсекаются. При этом они не просто вычеркиваются — остается некоторая ненулевая вероятность того, что искомый объект все же находился там, но не был обнаружен. Регулярно производится пересчет вероятностей для всех клеток с учетом этой новой информации, и карта меняется: на ней более отчетливо проступают области приоритетного поиска. Такие итерации могут продолжаться долго, сама длительность поиска тоже добавляет информации к вычисляемым вероятностям. В конце концов, если искомое будет найдено, оно, скорее всего, окажется в квадрате, для которого вычисленная таким образом вероятность заметно меньше единицы. И вовсе не обязательно заветный квадрат будет иметь наибольшую вероятность оказаться «тем самым».

Здесь можно вспомнить закон Бука из книги о законах Мёрфи: «Ключи всегда находишь в последнем кармане». Перебирая карманы в поиске ключей на пороге дома, вы не вычисляете вероятность их наличия в карманах, оставшихся не проверенными. Скорее всего, вы начинаете с самых привычных, потом переключаетесь на более оригинальные места (задние и внутренние карманы), а не найдя там, видимо, вновь станете перепроверять уже обследованные карманы. Интуитивно мы решаем задачу поиска ключей так же, как ведутся современные поисково-спасательные работы. Так что упоминавшийся в самом начале книги закон Бука, гласящий, что ключ всегда в последнем кармане, конечно, тривиален, но это связано с очень сильным упрощением процесса. Ключ запросто может оказаться в кармане, который вы уже проверили, но недостаточно аккуратно.

Сравнивайте разумно, не ищите в жизни норму и не бойтесь отклонений от нее. Сама математика подсказывает нам, что в сложном

мире людей корректно говорить можно лишь о степени подобия, но не о сравнении. Так что нет резона вести нескончаемые споры в поисках истины, стоит прислушаться и постараться услышать иное мнение, увидеть взгляд из другого, сопряженного пространства, обогащая тем самым свое восприятие мира.

Мудрецы правы: все мы уникальны и в своей уникальности абсолютно одинаковы.

Глава 6

Почему уж не везет так не везет?

Говорят, жизнь похожа на зебру: то белая полоса, то черная... А еще бывает, что к одной неприятности добавляется другая: и так все не просто в жизни, а тут еще кошка рожать принялась! То густо, то пусто! Одно к одному! Но самое печальное, что когда становится хорошо и в жизни наступает светлая полоса, то мысли закрадываются нехорошие: ох, не сглазить бы... ох, не придется ли за счастье расплачиваться... Знакомое ощущение? Об этом говорит один из законов мерфологии — **второй закон Чизхолма**:

Когда дела идут хорошо, что-то должно случиться в самом ближайшем будущем.

Но поскольку Френсис Чизхолм в своей оригинальной работе не дает детального анализа или доказательства этого закона, мы постараемся сами выяснить, кроется ли за этим какая-либо закономерность или нам так только кажется. А если это причуды математики, можно ли определить характерную длительность или частоту полосок на теле нашей зебры и от чего эти параметры зависят?

В жизни то и дело происходят *события*. Иногда они вовсе не связаны друг с другом, иногда образуют цепочки причинно-следственных взаимоотношений. Рассуждения об этих связях, цепочках и предопределенности жизненного пути могут увести нас очень

далеко, мы поговорим о них позже. А пока попробуем, как всегда, обойтись наименьшим количеством исходных данных для анализа нашего закона. Рассмотрим последовательность никак не связанных между собой событий и посмотрим, что удастся из нее добыть.

Синтезируем злодейку-судьбу

Наступление событий, которые никак не связаны между собой и происходят во времени случайно, описывается с помощью хорошо известного *пуассоновского потока*. Он соответствует многим случайным явлениям — от землетрясений до прихода покупателей в магазин.

Предположим, выполнены такие естественные условия.

1. Если есть два непересекающихся отрезка времени $[t_1, t_2]$ и $[t_3, t_4]$, то число событий в первом отрезке не зависит от числа событий во втором (отсутствие последействия).

2. Количество событий, произошедших на каком-либо отрезке времени, зависит только от длины отрезка, но не его положения (стационарность).

3. Вероятность, что два события происходят одновременно, пренебрежимо мала (ординарность).

Тогда можно показать, что число событий, попадающих на отрезок длины t , подчиняется *распределению Пуассона*. То есть вероятность P_m того, что на этом отрезке произойдет m событий, определяется так:

$$P_m = \frac{(\lambda t)^m e^{-\lambda t}}{m!}.$$

Число λ называется интенсивностью или плотностью потока и имеет смысл «среднего» числа наблюдений. Например, при измерении времени в днях значению параметра $\lambda = 1/7$ соответствует цепочка случайных событий, в среднем происходящих раз в неделю. Это вовсе не означает, что события будут происходить строго с *частотой* раз в неделю. Никакой определенной частоты у последовательности событий нет. Это среднее число событий: поскольку в году 52 недели, за год должно произойти около

52 событий (в среднем за много лет), но они будут разбросаны в году неравномерно. На рисунке 6.1 показаны 52 случайные равномерно распределенные даты в году, которые можно рассматривать как моменты появления пуассоновских событий.

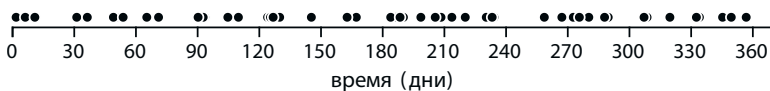


Рис. 6.1. Пример построения пуассоновского потока с интенсивностью 1/7 (время измеряется в днях). На отрезке в 365 дней случайным образом разбросали никак не связанные между собой 52 события

Как видите, о какой-либо периодичности в этих событиях речь не идет: когда пожелают, тогда и случатся. Но и в этом беспорядке статистика может нам показать определенные закономерности. Например, распределение длительности периодов между событиями, показанными на предыдущем рисунке, будет вовсе не равномерным (рис. 6.2).

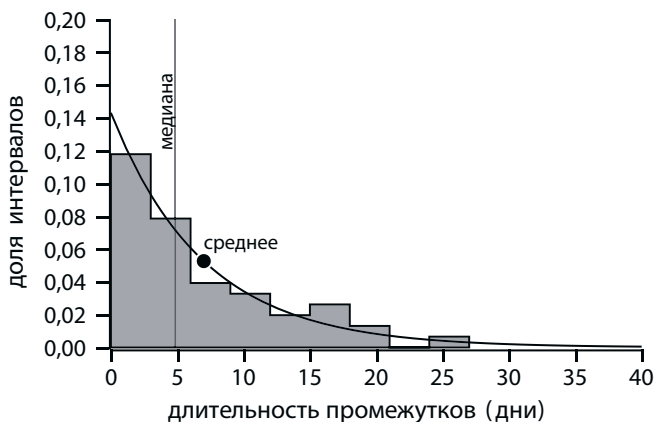


Рис. 6.2. Плотность распределения длительностей промежутков между 52 событиями, случайно разбросанными по отрезку в 365 дней

Промежутки времени между соседними пуассоновскими событиями имеют экспоненциальное распределение с плотностью $\lambda e^{-\lambda t}$

(на рисунке для нашего случая показана сплошной линией). У этого распределения максимум (мода) находится в нуле, а среднее значение равно $1/\lambda$, в нашем случае 7 дней. Более того, стандартное отклонение σ тоже равно 7 дням, поскольку дисперсия экспоненциального распределения $\sigma^2 = 1/\lambda^2$. Как видите, эти характеристики вовсе не гарантируют того, что между событиями будет проходить одна неделя. В среднем — да, но чаще всего меньше; к тому же могут наблюдаться и достаточно долгие промежутки без событий. Наконец, медиана показывает, что половина всех промежутков будет иметь длительность не более 5 дней. Интенсивность и частота — совсем не одно и то же; это очень важное замечание, к которому мы еще вернемся в этой главе.

Для справедливости положим, что хорошие и плохие события происходят равновероятно, но яркие и значимые (как хорошие, так и плохие) — существенно реже мелких и незначительных. Пусть это будет «обычная» жизнь, в которой эмоциональная окраска событий подчиняется нормальному (гауссовскому) распределению. Вот как может выглядеть год синтетической судьбы в виде череды случайных абсолютно независимых жизненных перипетий (рис. 6.3).

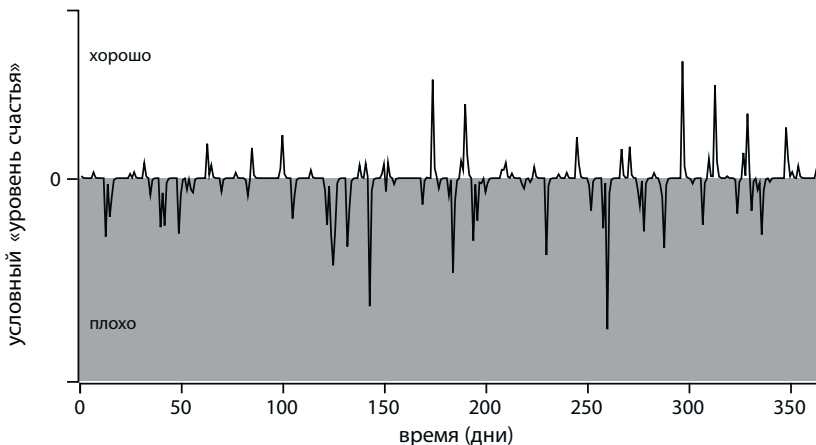


Рис. 6.3. Череда событий различной эмоциональной окраски, образующая пуассоновский поток с интенсивностью $2/7$ (2 события в 7 дней)

Знак пиков отражает эмоциональную окраску, а их высота соответствует важности события или глубине переживаний, с ним связанных. Пока никаких полос не наблюдается, есть некий шум. Каждое событие проходит бесследно, ничего не оставляя ни в памяти, ни в настроении. Так не бывает, поэтому наделим нашего модельного героя памятью — для начала идеальной. Каждое событие пусть навсегда врежется в его память и отразится на настроении, либо улучшая, либо ухудшая его. Вот какую картинку мы можем получить, понаблюдав за судьбой нашего героя на протяжении десяти лет (рис. 6.4). Текущий «уровень счастья» вычисляется суммированием вкладов всех предшествующих событий. Позитивные события эту сумму увеличивают, а негативные — уменьшают.

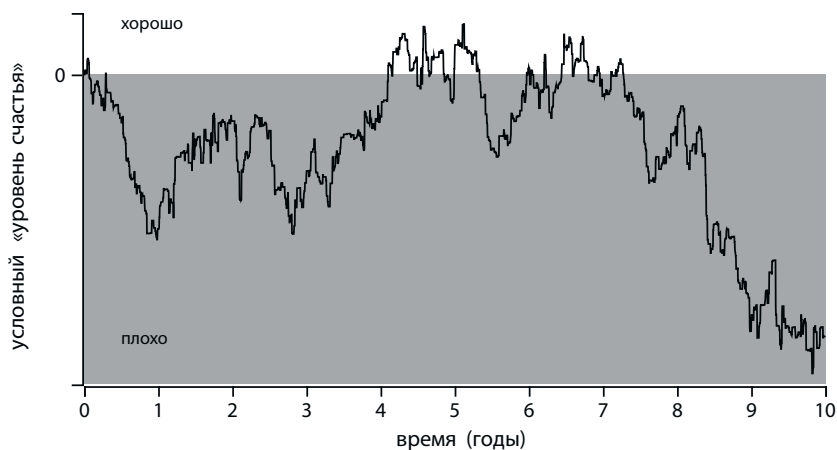


Рис. 6.4. События, сливаясь в памяти, образуют эмоциональную окраску «синтетической жизни»

Ну что же, мы уже видим какое-то чередование настроения, но картинка вышла не особо радостной. Наш герой после череды смен настроения впал в глубочайшую депрессию. Жаль. Попробуем сгенерировать еще несколько судеб (рис. 6.5). Все они проходят череду светлых и темных полос, но надолго увязают либо в беспросветной тоске, либо в запредельном счастье. Так бывает, конечно, но это явно ненормально.

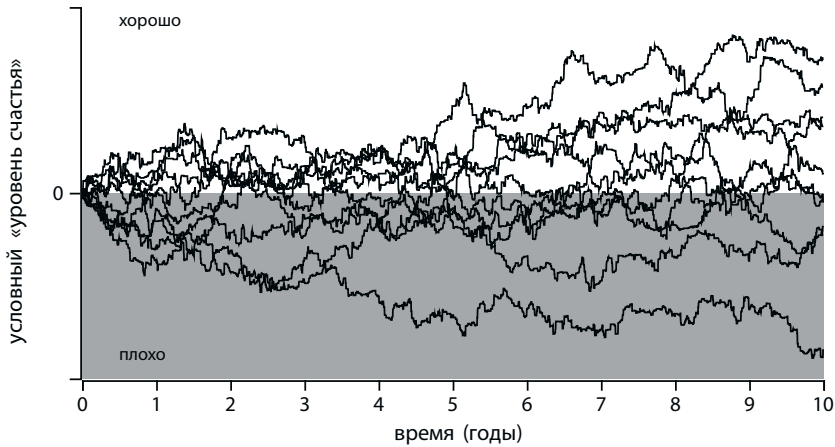


Рис. 6.5. Несколько примеров «синтетических судеб» людей с идеальной памятью

Ценность релаксации

Наши модельные судьбы мы описали очень примечательным процессом. Он называется *одномерным случайным блужданием* и имеет ряд необычных свойств, среди которых — *самоподобие*, то есть отсутствие какого-либо характерного временного масштаба. Получив в свое распоряжение неограниченное время, случайное блуждание способно увести неограниченно далеко. Более того, оно обязательно уведет вас на любое наперед заданное расстояние от начального значения! Таким образом, как бы хорошо ни шли ваши дела, но если они подчинены случайному блужданию, то обязательно скатятся до нуля и уйдут ниже — это просто вопрос времени! Правда, если речь о существенных отклонениях, то очень большого времени. Можно показать, что в рассмотренном нами процессе ожидаемая величина отклонения от начального состояния пропорциональна квадратному корню от времени. Это значит, что ожидаемое время, за которое система, отклонившаяся от нуля, вновь вернется в нулевое состояние, пропорционально квадрату начального отклонения.

Помните, как говорил кот Матроскин в известном мультфильме «Каникулы в Простоквашино»: «Я и так счастливый был, а теперь

в два раза счастливей стану. Потому что у меня две коровы есть!»
Таким образом, можно предположить, что рождение теленка (появление второй коровы) продлит счастье Матроскина в четыре раза.

Но все же идеальная эмоциональная память — это не очень хорошо. Наши герои не забывают ничего и тщательно хранят в памяти всё, даже самые давние события! На их настроение в старости влияет горе от поломанной игрушки в детстве или радость от поцелуя в юности. Причем все последующие поцелуи и игрушки имеют для них такую же важность. Надо этих бедолаг спасать. Эмоции со временем стихают, горе притупляется, радость, увы, тоже. Забывание во многом подобно остыванию, диффузии или замедлению движения в вязкой жидкости, поэтому разумно смоделировать его подобным образом. Перечисленные события относятся к *процессам релаксации*, о которых мы говорили в конце главы 2. Наделим же и наших героев способностью к релаксации!

Релаксирующая система возвращается к равновесному состоянию, причем тем быстрее, чем больше отклонение от равновесия. Это свойство можно смоделировать геометрической прогрессией или экспоненциальным законом. Введем в нашу модель новый параметр — скорость забывания μ . Его можно выразить через время (в отчетах нашей модели), за которое уровень эмоции уменьшится достаточно сильно. Например, для $\mu = 1/60$ эмоциональный след от события уменьшится на порядок через два месяца. И вот теперь жизнь стала по-хорошему «полосатой» (рис. 6.6)!

Меня «степень забывчивости», мы можем получить более или менее эмоционально уравновешенных подопытных. Кажется, мы нашли источник зеброобразности! Это, во-первых, случайные блуждания, склонные к расползанию во все стороны; во-вторых, целительная забывчивость, возвращающая настроение в норму. Результатом становится волнообразное меандрирование* настроения.

* Меандр в математике — замкнутая кривая без самопересечений, которая при этом пересекает прямую несколько раз. *Прим. ред.*

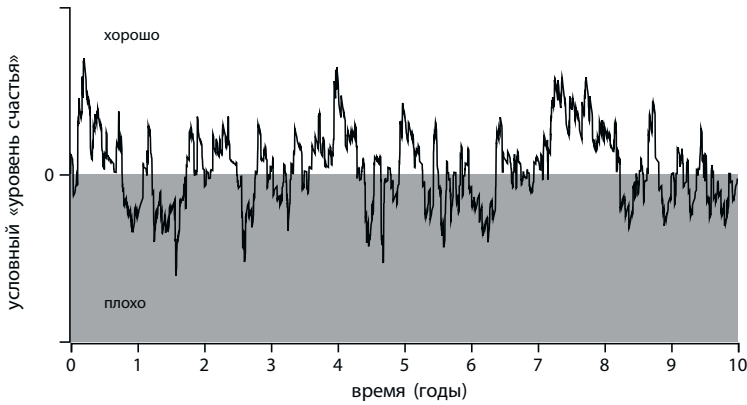


Рис. 6.6. Ограничение памяти приводит к тому, что череда событий и их следов в памяти, сливаясь, образует череду эмоционально окрашенных полос

Изучим свойства полученных нами «синтетических» житейских полос. Построим гистограмму, показывающую распределение их длительностей для длинющей жизни (или множества обычных) с параметрами $\lambda=1/7$, $\mu=1/60$ (рис. 6.7).

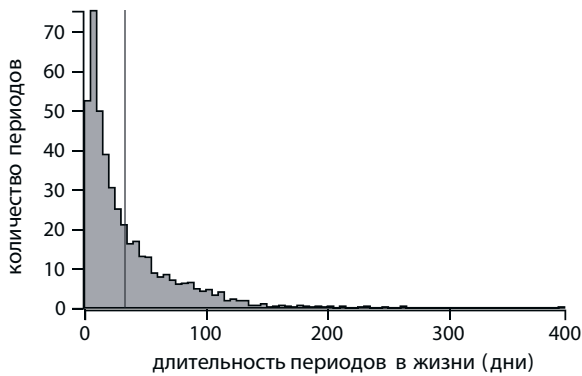


Рис. 6.7. Распределение длительностей периодов счастья и горя для большого числа синтетических судеб. Вертикальной линией отмечено среднее значение, равное 33

Первое, что бросается в глаза, — максимум распределения (мода) находится вблизи нуля. Значит, чаще всего времена счастья и несчастья очень коротки, однако встречаются и периоды

длительностью более года. В среднем же их продолжительность составляет 33 дня со стандартным отклонением в 36 дней. Это распределение близко к экспоненциальному (на самом деле оно неплохо описывается более общим *гамма-распределением* с такими параметрами, которые приближают его к экспоненциальному). В свою очередь, экспоненциальное распределение длительностей полос в жизни означает, что смены настроений можно рассматривать как пуассоновский поток — цепочку независимых случайных событий, не имеющих определенной частоты, но случающихся с некоторой известной интенсивностью. Например, в рассмотренном нами примере темные и светлые полосы сменяются с интенсивностью раз в 33 дня, но гораздо чаще в жизни наблюдаются короткие периоды: половина их не дольше десяти дней.

В случае отсутствия «памяти» (для $\mu = 0$) распределение перестает быть экспоненциально убывающим и описывается *распределением Юла*, которое можно приблизить степенным распределением (*распределением Парето*) для длительности меандров T (рис. 6.8).

$$p(T) = \frac{\Gamma(T)}{\Gamma(T + 2/3)\Gamma(3 + 1/3)} \approx \frac{1}{3} T^{-\frac{2}{3}}.$$

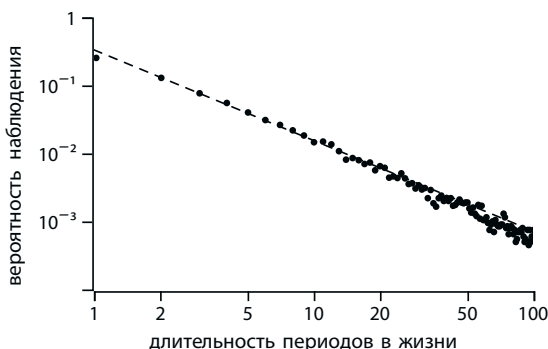


Рис. 6.8. Распределение длительностей меандров для случайного блуждания имеет характер степенного. Двойной логарифмический масштаб графика позволяет распознать степенную зависимость

Статистики говорят, что у таких распределений *тяжелый хвост*, делающий вполне вероятными очень большие отклонения от среднего значения. Мы наблюдали их в виде долгих «погружений» в то или иное настроение. У полученного распределения есть одно непривычное и странное свойство: для него не определены ни среднее значение (математическое ожидание), ни стандартное отклонение. В предыдущей главе мы уже упоминали, что такое бывает, например, у распределения Коши. Дело в том, что все соответствующие интегралы для распределения Юла расходятся. В связи с этим можно слышать, что и среднее значение в таком случае бесконечно, но это не так. Посмотрите, что произойдет при попытке вычислить математическое ожидание длительности меандров случайного блуждания (рис. 6.9).

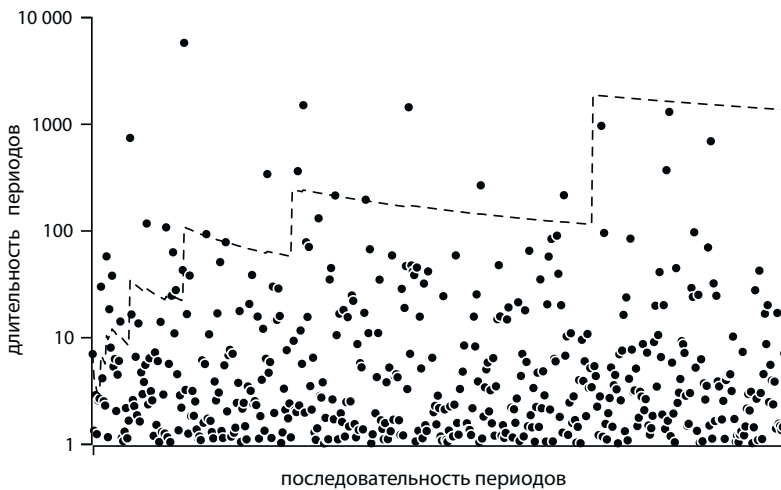


Рис. 6.9. Попытка вычислить среднее значение для последовательности длительностей периодов между сменами настроения при отсутствии памяти. Появляющиеся экстремальные значения из тяжелого хвоста распределения приводят к тому, что значение среднего не сходится к какому-либо пределу

Огромные скачки из тяжелого хвоста то и дело сбивают значение среднего, и последовательность усреднений не сходится ни к какому пределу. Значение среднего вовсе не бесконечно, но

интеграл не сходится и о каком-то конкретном значении говорить нельзя. Именно в невозможности вычислить среднее для длительности меандров отражается свойство самоподобия случайного блуждания, или отсутствие собственного масштаба времени.

Мы моделировали приспособляемость к житейским неурядицам с помощью релаксации — затухания эмоциональных всплесков. Можно истолковать этот процесс иначе — как приспособляемость человека к жизненным обстоятельствам. При обработке зашумленных сигналов или последовательностей часто для сглаживания и выделения полезного сигнала используют *метод скользящего среднего*, рассматривая в каждый момент не сам сигнал, а усредненное его значение за некоторый промежуток времени. Так удается избавиться от шума и получить представление о долгосрочных тенденциях сигнала. Применяя такое усреднение к житейским неурядицам, мы можем моделировать приспособляемость человека. Люди влюбляются и находят повод для радости даже во время войн, а жизнь богатых бездельников не безоблачна. Смещается локальное представление о норме (привычном состоянии дел), от которой настроение отклоняется в ту или иную сторону. Рассматривая разницу между последовательностью эмоций и сглаженной линией фона, мы получим такую же картину полос, какую дала предыдущая модель, с теми же статистическими характеристиками. Это неудивительно, ведь концептуально они практически не различаются, описывая систему с релаксацией (рис. 6.10).

Какие выводы можно сделать из нашего несерьезного исследования? Череда светлых и темных полос в жизни не иллюзия, они существуют на самом деле. Но в них нет особенных закономерностей. Чаще всего они коротки, но бывают и затяжными. Все зависит от легкости характера и способности отпускать прошлое. Более того, если события будут происходить редко, то жизнь станет серой чередой исчезающих в прошлом воспоминаний. Так что в наших интересах запоминать прожитое и в наших силах сделать так, чтобы жизнь не становилась случайным блужданием. Мы можем добиться того, чтобы хороших событий становилось больше и происходили они почаще, пусть даже они окажутся и незначительными.

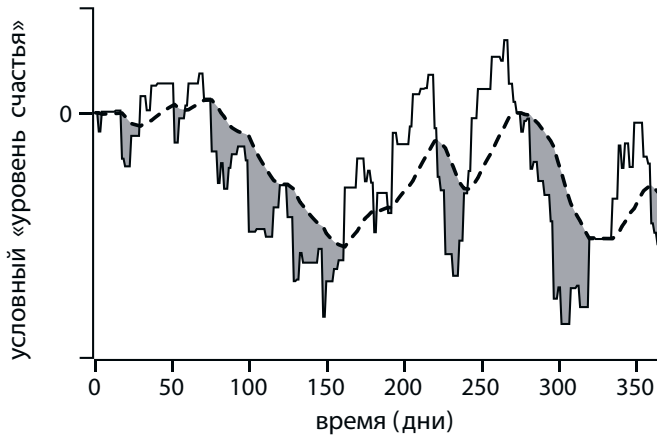


Рис. 6.10. Меандрирование и смену настроений можно получить, моделируя скользящим средним приспособляемость человека к обстоятельствам

Лыжная прогулка, искренняя улыбка прохожего, билет на концерт, чашка горячего шоколада в холодный день — все это поможет создать положительный тренд и продлит светлую полосу в жизни. Правда, неизбежные грустные события обязательно сменяют настроение. Но не надо винить в этом свое счастье. Это не расплата за него и не сглаз. Это свойство релаксирующих систем — склонность к колебаниям при стохастическом внешнем воздействии.

О марковских цепях и пессимистах с оптимистами

В рассмотренных моделях мы получали пуассоновский поток смены настроений, генерируя события с помощью его же. В этом можно усмотреть подтасовку: пуассоновский случайный процесс оказался изначально «вшит» в модель. Насколько при этом универсален результат? Можно ли получить его как-нибудь совсем иначе?

Житейский опыт — штука плохо формализуемая, его можно подогнать под различные математические инструменты, внося

не только упрощающие допущения, но и спекуляции. В науке такой подход недопустим, но в путешествии по методам теории случайных процессов мы можем позволить себе поиграть с ними, чтобы познакомиться получше.

Выше для объяснения полос в жизни мы учитывали память, то есть вклад предыдущих состояний в текущее. Но можно получить характерное «полосатое» поведение и полностью исключив влияние прошлого.

Для этого полезны объекты, называемые цепями Маркова.

Последовательность дискретных случайных величин x_1, x_2, \dots называется цепью Маркова, если распределение величины x_{n+1} зависит только от распределения величины x_n , но не от предыдущих величин x_1, \dots, x_n . Иными словами, будущее зависит от настоящего, но не от прошлого. Область значений наших величин x_n называется пространством состояний цепи. Переходы между состояниями определяются числами p_{ij} — вероятностями перейти из состояния с номером i в состояние с номером j . Мы ограничимся случаем, когда эти вероятности не зависят от номера n (тогда цепь Маркова называется однородной). Числа p_{ij} образуют так называемую матрицу переходов, о которой мы поговорим позже.

Такие цепи удобно представлять в виде *взвешенных графов**. Вершинами графа оказываются состояния цепи, а ребрами — возможные переходы между ними. Например, однородная марковская цепь, описывающая динамику настроения, может быть представлена в следующем виде. Пусть для простоты у человека есть всего два состояния (радостное и печальное) и он каждый день может оказаться либо в одном, либо в другом. При этом вероятность остаться на следующий день в прежнем состоянии равна 0,75, а вероятность поменять его — 0,25 (рис. 6.11).

Почему мы выбрали такие вероятности? Наблюдая за динамикой настроения и мировосприятия, можно заметить, что человеку свойственно «залипать» в определенном состоянии духа.

* Граф — это еще одна универсальная математическая структура, пожалуй, наиболее общая из всех. Это абстракция структуры как таковой. Теория графов достойна отдельного большого разговора, поэтому я предупреждаю читателя, с ней не знакомого: во-первых, вы рано или поздно с ней обязательно познакомитесь, а во-вторых, получите удовольствие!

Если дела идут в целом хорошо, то и дурная новость может быть воспринята с оптимизмом. И напротив, меланхолическое настроение, однажды поглотив человека, способно испортить даже радостное известие. С математической точки зрения это значит, что вероятность остаться в текущем настроении выше вероятности его изменить.

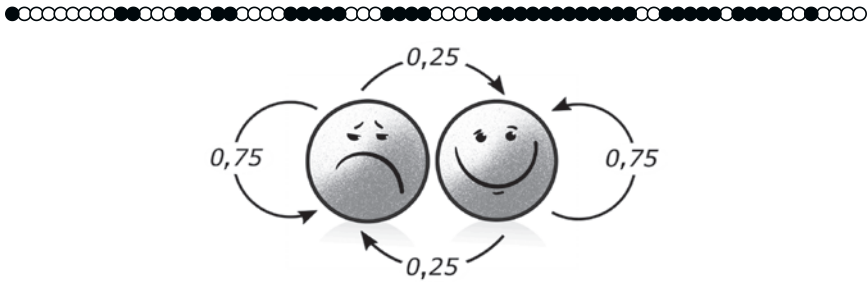


Рис. 6.11. Цепь Маркова с двумя состояниями («радостное» и «печальное»). Стрелки обозначают переходы и их вероятности. В нашем симметричном случае вероятность остаться в существующем настроении превышает вероятность его смены, но не зависит от самого настроения. Переходы случаются раз в день

Наша цепь способна генерировать последовательности состояний, и, конечно, в ней появятся полосы житейской зебры. Самое интересное — выяснить, какому распределению будут подчиняться длительности этих полос. Для нашей более чем простой модели можно получить точный ответ — это *геометрическое распределение*, описывающее вероятность наблюдать заданное количество испытаний до первого «успеха».

Геометрическое распределение — дискретный аналог экспоненциального в том смысле, что ему подчиняются округленные значения экспоненциально распределенной случайной величины. Существует связь между параметром геометрического распределения и интенсивностью соответствующего экспоненциального. Так мы опять получаем пуассоновский поток смен настроения, и для описанной нами марковской цепи его интенсивность равна $\lambda = -\ln(0,75) \approx 2/7$ (рис. 6.12).

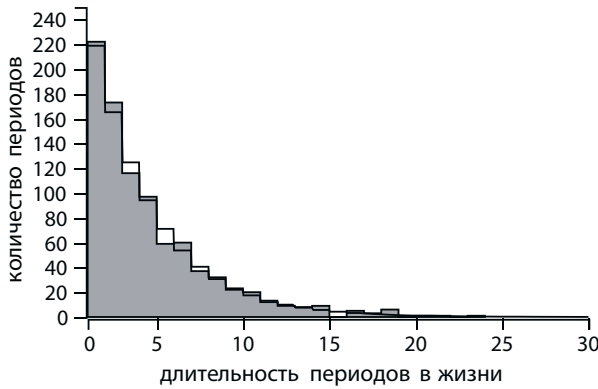


Рис. 6.12. Гистограмма для длительностей периодов одинакового настроения в последовательности ежедневных смен состояний, сгенерированной симметричной цепью Маркова, и функция вероятности геометрического распределения с параметром, равным вероятности перехода между состояниями. Последовательность имеет длительность в 10 лет

Если мы нарушим симметрию цепи, то сможем описать «оптимиста» либо «пессимиста», охотнее «залипающего» в том или ином настроении. Распределение длительностей полос отклонится от геометрического, но при этом большая часть полос будет короткой и какой-либо выделенной периодичности мы не отметим (рис. 6.13).

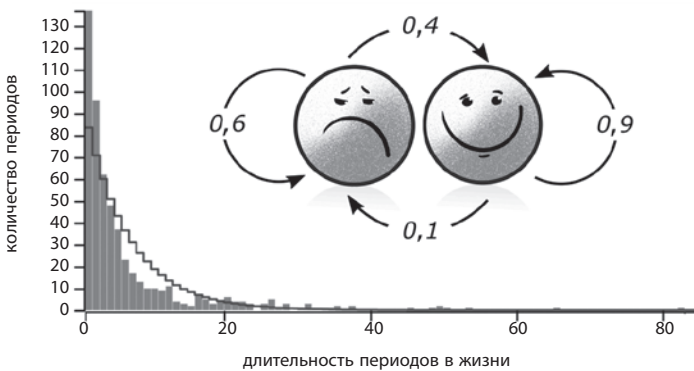


Рис. 6.13. Гистограмма для длительностей периодов постоянного настроения в последовательности, сгенерированной асимметричной цепью Маркова. Ступенчатая линия показывает геометрическое распределение из предыдущего примера

Цепи Маркова — мощный инструмент анализа случайных процессов, в которых кроется некий алгоритм или сценарий. Они дают нам своеобразный взгляд на процессы, привычно относимые к циклическим. Например, известная максима «история человечества ходит по кругу» часто трактуется так: в истории существуют некие циклы или даже периодичности. Доводится слышать, например, о том, что начало века сулит потрясения и войны. Рискуя уйти не в свою тему, возьму на себя смелость предположить, что на самом деле имеет смысл говорить не о буквальных циклах, а о более или менее устойчивых сценариях — закономерных цепочках, которые можно описать цепью Маркова. Среди таких цепей есть класс циклических, которые в самом деле способны создавать повторяющиеся последовательности. Однако настоящей детерминистической периодичности в их поведении нет. Случайно возникая в разные исторические периоды и в разных контекстах, такие циклы похожи друг на друга и могут создать ощущение исторического «дежавю». Изучать и описывать их полезно, но ожидать строгого календарного плана, пожалуй, не стоит.

«Лила» и игра с бесконечностью

Характерную цикличность в случайном на первый взгляд процессе я наблюдал, принимая участие в игре «Лила» (рис. 6.14). Это разновидность игры «Лестницы и змеи», у которой, как говорят, древние индийские корни. Участники перемещают свои фишки (амулеты) согласно выпадающим числам на кубике, следуя переходам — «лестницам» или «стрелам», ведущим вперед, и «змеям», возвращающим игрока назад. Основной смысл заключается в философских и эзотерических толкованиях траектории, которую проходит игрок. В нашей компании были опытные люди, они делились впечатлениями от прошлых игр и восхищались «явно неслучайными» совпадениями траекторий игры и реальной жизни, точному их повторению от партии к партии — как у одного и того же, так и у разных участников.

В игре 72 состояния, и правила бросания кубика нетривиальны: они делают более вероятными близкие переходы, но допускают

и далекие скачки; кроме того, «лестницы» и «змеи» добавляют путаницы. Действительно, в игре много элементов случайности, но она все равно остается марковской, поскольку ближайшее будущее игрока определяется только текущим его состоянием. А значит, сам процесс можно анализировать на предмет наличия в нем повторяющихся последовательностей или наиболее вероятных состояний.

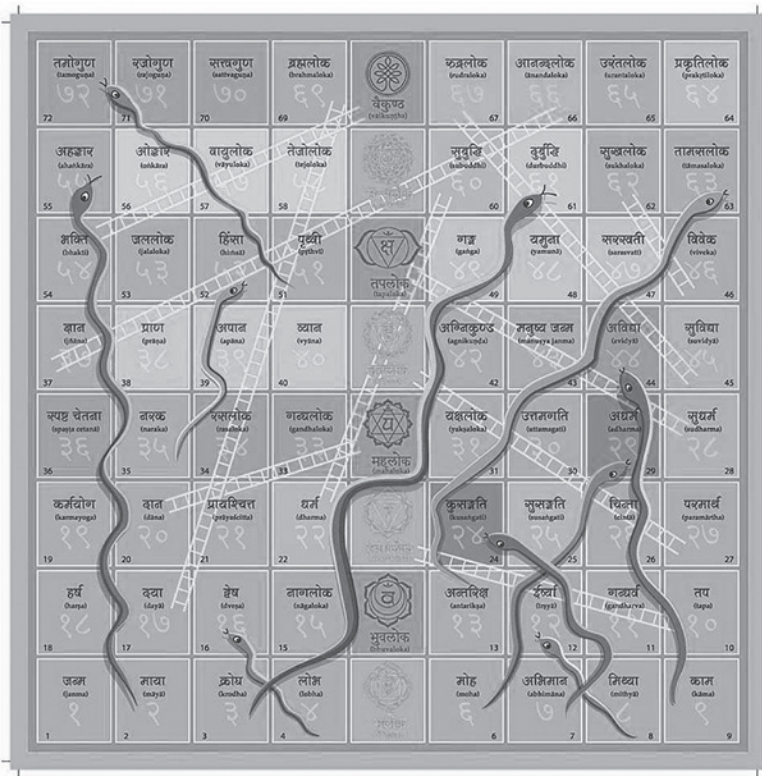


Рис. 6.14. Доска для игры «Лила»

Несложно написать программу, которая могла бы играть в «Лилу», не задумываясь о сокровенном смысле состояний и переходов, и которую можно было бы использовать в анализе методом Монте-Карло. Приведу для тех, кому, как и мне, любопытно поэкспериментировать, алгоритм для одного шага.

Переходы по лестницам и змеям могут быть описаны ассоциативным массивом

```
Jumps = { 10:23, 16:4, 61:3, 20:32, 22:60, 24:7,
          27:41, 28:50, 29:6, 37:66, 45:67, 46:62,
          52:35, 54:68, 55:2, 61:3, 63:13, 72:51, 68:1 }
```

Вход: текущее состояние (номер клетки) s

```
если (jumps содержит состояние s), вернуть jumps[s]
m := случайное целое число от 1 до 6
если (m = 6), m := m + случайное число от 1 до 6
если (s > 60), m := min(m, 72-s)
вернуть s + m
```

Вот что можно сказать после сотни тысяч партий. Средняя продолжительность игры (то есть достижения 68-й клетки) составляет 41,5 шага, при этом в половине партий игра закончится после 31 шагов. Это довольно много: учитывая, что шаги совершаются по очереди четырьмя-пятью участниками, игра может длиться несколько дней. Клетки посещаются неравновероятно, и разброс вероятностей достаточно велик.

Но любому математику интереснее не получить ответ из эксперимента, а вывести из свойств исследуемой системы. Мы рассмотрим *матрицу переходов* M для игры, она показана на рис. 6.15.

Эта квадратная матрица имеет столько строк, сколько существует состояний (клеток) игры. Насыщенность цвета каждой клетки показывает вероятность перехода с позиции, указанной по вертикали, на позицию по горизонтали. Стрелки приводят пример, соответствующий вероятности перехода с 40-й клетки на 50-ю. Широкая полоса вокруг диагонали соответствует переходам с помощью кубика, прочие отмеченные точки — прыжкам, диктуемым «стрелами» и «змеями». Игра имеет одно *поглощающее состояние*: достигнув ячейки 68, игрок заканчивает партию. Но пока мы это правило заменим другим: пусть игрок, попав в клетку 68, вновь начинает с первой позиции. Этот переход показан незакрашенным кружком на матрице. Позже я объясню, для чего нам потребовалось таким способом закольцевать игру.

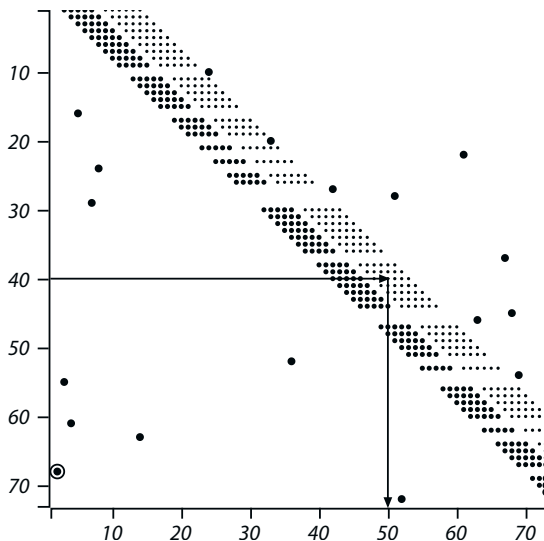


Рис. 6.15. Графическое представление матрицы переходов для «Лилы». Ненулевые элементы показаны кружками, размеры отражают их величину

Точные параметры можно получить не прибегая к методу Монте-Карло, а используя только матрицу переходов. Квадратные матрицы образуют *алгебру*: их можно по определенным правилам складывать и вычитать, умножать на число, перемножать и «делить» (умножать на обратную матрицу). Как и для чисел, многократное умножение матрицы на себя можно рассматривать как возведение в целочисленную степень. В случае с матрицей переходов для цепи Маркова возведение в степень n дает нам распределение вероятностей для всех переходов из клетки в клетку через n шагов. Так мы получаем своего рода «машину времени», способную мгновенно переместить нас в будущее. Вот как выглядят матрицы переходов игры «Лила» после 2, 3, 10 и, как это ни странно звучит, бесконечного числа умножений (рис. 6.16).

Необычно видеть что-то конечное и нетривиальное, возведенное в бесконечную степень. Привычные для нас вещественные числа (положительные) при возведении в большие степени либо увеличиваются до бесконечности, либо стремятся к нулю, и только числа 0 и 1 не изменяются.

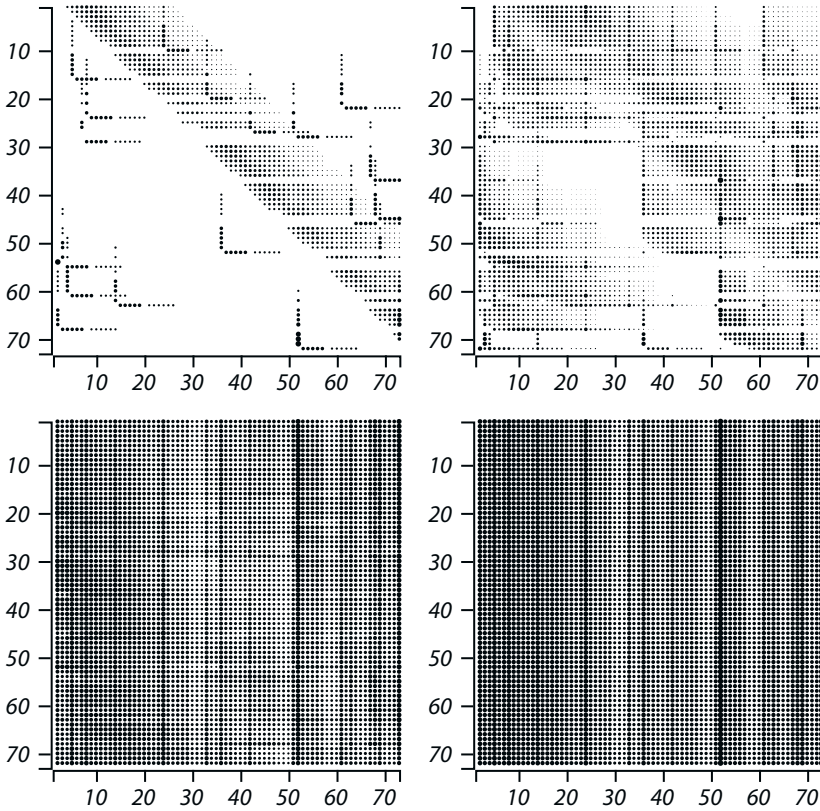


Рис. 6.16. Матрицы переходов, возведенные в степени 2, 4, 10 и ∞

Матрицы существенно раздвигают горизонты математического сознания, порождая необычные, порой причудливые, но полезные алгебраические системы*. Матрица переходов относится к классу *стохастических*, их характеризует то, что сумма элементов любой их строки равна единице. Это связано с тем, что каждая строка соответствует какому-то состоянию системы, а ее элементы — вероятностям перехода из этого состояния в другие. Рассматриваются все возможные варианты переходов, поэтому сумма всех вероятностей равна единице. Возведение стохастической матрицы в целочисленную степень оставляет ее стохастической. В пределе же

* С помощью матриц изящно описываются такие полезные понятия, как комплексные числа, вращения, кватернионы, конечные группы и т. д.

мы получили матрицу, которая не изменяется при умножении на саму себя:

$$M^\infty \cdot M^\infty = M^\infty.$$

Такие матрицы называют *идемпотентными*. Может показаться, что это какой-то экзотический случай, но идемпотентны все преобразования проекции, а значит, и представляющие их матрицы. Вообразите преобразование, для каждого трехмерного объекта возвращающее его тень на некой фиксированной стене. В процессе часть информации о форме объекта неизбежно теряется, а другая остается неизменной. На этом основаны занимательные задачи, в которых нужно определить, тело какой формы может отбросить указанные тени. А что случится с тенью, если мы еще раз спроецируем ее на эту же стену? Ровным счетом ничего, она не изменится. Можно точно показать, что любое преобразование проекции идемпотентно, но пример с тенью уже позволяет понять, что это означает и что это свойство не такая уж редкость. Многократное перемножение матрицы перехода для нашей марковской цепи привело нас к такому случаю. Эта предельная матрица отражает все мыслимые партии сразу. Впечатляет, но игра, определяемая такой матрицей, становится уже неинтересной.

Предельная матрица получилась «полосатой»: все ее столбцы одинаковы, и полоски говорят нам, что вероятность перехода определяется только конечной клеткой и не зависит от начала пути: прошлое в марковском процессе теряется безвозвратно (как форма тела в его тени). Любая строка этой предельной матрицы дает точное распределение «популярности» клеток. Полученный набор вероятностей для состояний игры образует особый вектор π , который называется *стационарным состоянием* цепи (рис. 6.17). Это и есть своеобразная «тень» игры, которая не меняется под действием матрицы перехода*: $M \cdot \pi = \pi$. Величины, обратные найденным нами вероятностям, характеризуют ожидаемое время достижения для

* Мы получили стационарное состояние в результате многократного умножения матрицы перехода. Это не универсальное свойство стохастических матриц. Если в игре есть безусловные циклы, то многократное перемножение может не дать какой-то одной предельной матрицы, хотя инвариант в этом случае отыскать возможно.

каждой клетки. Например, для клетки 68, конечной в игре, инвариантный вектор дает вероятность достижения 2,4%. Обратная величина равна 41,5, что совпадает со средней продолжительностью игры, полученной в эксперименте.

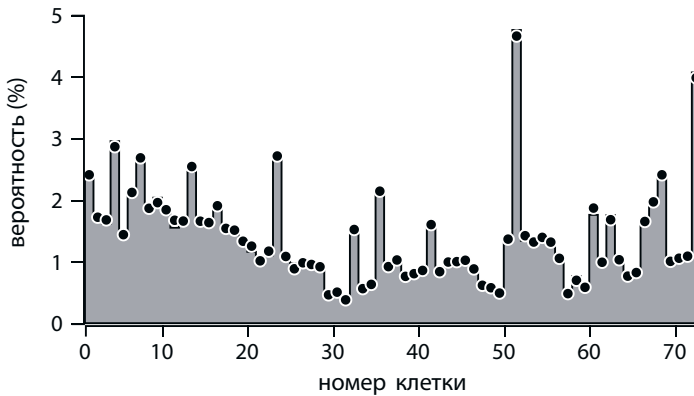


Рис. 6.17. Стационарное состояние игры отражает распределение вероятности посещения клеток. Точками показаны точные значения вероятностей, а столбиками — полученные после ста тысяч шагов игры

Если бы мы оставили состояние 68 поглощающим, как предписывают правила игры, в бесконечном будущем можно было бы ожидать, что все партии сойдутся к нему. Инвариантом в этом случае был бы вектор, в котором от нуля отлична лишь 68-я позиция. Но и такая матрица перехода может быть полезна. Она дает нам возможность проанализировать время окончания игры. Матрица M^n соответствует n шагам в игре, а значит, элемент $(M^n)_{ij}$ покажет вероятность достижения состояния j из состояния i за n шагов. Таким образом, мы можем построить точное распределение времени окончания игры, нарисовав график зависимости $p(n) = (M^n)_{1,68}$, как показано на рис. 6.18.

Так можно не играя вычислить, что изменится при каких-либо поправках к правилам: например, смене поглощающего состояния, добавлении или удалении переходов, усложнении выбрасывания кубика и т. п. Матричные вычисления, в том числе точные, можно выполнять очень быстро, почти мгновенно, в отличие от

имитационного моделирования, так что допустимо поручить машине оптимизацию правил игры с целью сделать ее интереснее, создавать маловероятные «ценные клетки» и контролировать при этом длительность партии.

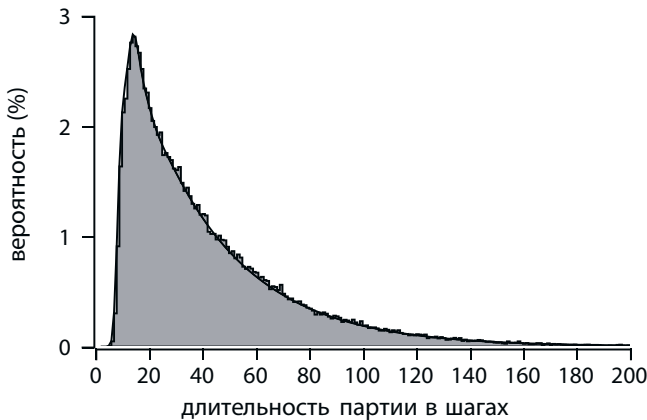


Рис. 6.18. Распределение длительности партии в игру «Лила», полученное в ходе ста тысяч экспериментов и теоретически

Кстати, в вычислениях для этой главы я использовал один красивый прием, имеющий отношение к нашей второй сквозной теме: алгебраическим структурам. С давних пор известен способ умножения целых чисел, который зовется то египетским, то способом русского крестьянина и представляет интерес не только своим практическим смыслом, но и глубокой математической основой и следующей из нее универсальностью. Вы без труда найдете его описание во многих книгах по популярной математике. Метод основан на двух очень простых равенствах, вполне очевидных даже для школьника:

$$(2n)a = 2(na) = na + na,$$

$$(n + 1)a = na + a.$$

Первое равенство позволяет уменьшить множитель n за счет удвоения произведения, а второе — перейти к первому, если уменьшаемый множитель нечетный. Сами по себе эти равенства обладают

свойствами ассоциативности и дистрибутивности* умножения, то есть носят фундаментальный характер, но поскольку единица — нейтральный элемент для умножения, они образуют весьма эффективную рекурсивную схему вычисления произведения. Эффективность связана с тем, что умножение — или многократное сложение — заменяется операцией удвоения, которая увеличивает результат существенно быстрее. Например, при перемножении чисел в пределах миллиона потребуются не более 20 шагов этого алгоритма.

Но вот что делает этот метод по-настоящему замечательным: число a можно заменить любым другим объектом, для которого определена ассоциативная операция сложения с нейтральным элементом. Такие объекты образуют структуру, называемую *полугруппой с единицей*, или *моноидом*. Дело в том, что умножение элемента моноида на целое число эквивалентно многократному сложению этого объекта с самим собой. А это значит, что, имея любой моноид, мы можем применить к нему метод русского крестьянина! Числа образуют моноид не только с операцией сложения, но и с операцией умножения, и тогда метод можно использовать для быстрого возведения в степень. Моноид с операцией умножения формируют и матрицы, а также представляемые ими линейные преобразования. Это позволяет очень быстро вычислить результат возведения матрицы в очень большую степень без потери точности. Чем я и воспользовался.

В завершение разговора об игре «Лила» перейдем к часто повторяющимся мотивам. Их тоже можно изучать не играя, а анализируя матрицу переходов. Вероятности для любой цепочки вычисляются как произведения вероятностей переходов, умноженных на вероятность попадания в начальную позицию:

$$P(3 \rightarrow 5 \rightarrow 13 \rightarrow 15) = \pi_3 M_{3,5} M_{5,13} M_{13,15}.$$

Так можно перебрать все цепочки длины 3, 4, 5 и т. д. и найти наиболее вероятные. Но такой поиск занял бы слишком много времени. Возможно отыскивать такие цепочки более целенаправленно.

* Более того, таким образом определяется операция умножения чисел на самом базовом уровне, так что это аксиома умножения, а не следствие из определения.

Для любой начальной клетки можно, пользуясь матрицей переходов, создать дерево возможных шагов, оставляя по мере построения несколько наиболее вероятных ветвей. Такой процесс называется *поиском оптимального пути в ширину с отсечением*. Действуя таким способом, можно отыскивать самые часто наблюдаемые цепочки и выяснить, как распределяются цепочки по вероятности их наблюдения (рис. 6.19).

Вероятность для цепочки	Число цепочек
> 25%	3
> 10%	10
> 5%	64

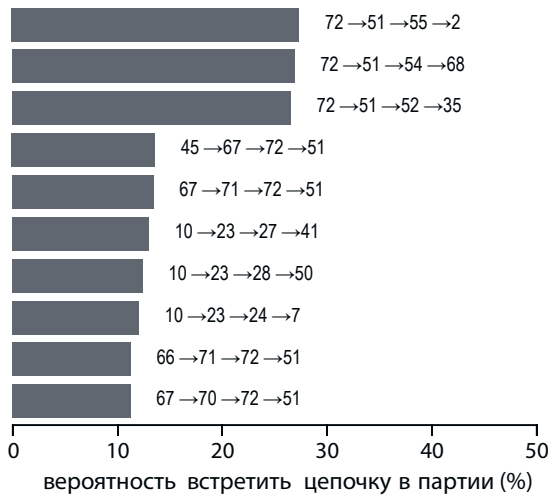


Рис. 6.19. Наиболее часто наблюдаемые цепочки в игре «Лила»

Пример с игрой «Лила» напрямую не касается вопроса о полосах в реальной жизни, но заставляет задуматься. Должно быть, для всемогущего божества, способного видеть сколь угодно далекое будущее, играющего во все игры сразу, мир предстает достаточно скучной вырожденной идемпотентной матрицей. Впрочем, оставим наше мифическое божество разбираться с этой проблемой самостоятельно. Я привел этот пример здесь потому, что мне хотелось показать, как математика позволяет проанализировать структуру довольно сложной и стохастической игры. Предпринимались

попытки анализа известной игры «Монополия», но здесь становится существенной роль эксперимента, поскольку процесс накопления игроками денег добавляет в процесс память — и он перестает быть марковским.

Несмотря на простоту и некоторую ограниченность, трудно переоценить важность концепции цепей Маркова. Если взяться перечислять области, в которых они используются, получится внушительный перечень не на одну страницу. В нем окажутся и симуляции реальности более сложной, чем игры; генерация текстов, музыки, речи, тестовых заданий для систем автоматического управления; поиск страниц в сети интернет; физика, химия, биология, генетика, экономика, социология, безопасность дорожного движения... даже в спорте используются цепи Маркова!*

Почему автобуса все нет?!

Говоря о пуассоновском процессе, мы различали частоту и интенсивность потока событий. Это важно понимать, слушая новости или читая результаты научных исследований. Например, на сегодняшний день сейсмологи, увы, не могут предсказать конкретное землетрясение: его время, место и силу. Зато наработаны методики долгосрочного сейсмического прогноза для какого-то региона, но их результаты формулируются на языке теории вероятностей. Что с ними делать — не всегда очевидно.

Например, для Авачинского залива, на берегах которого расположен Петропавловск-Камчатский, в 2018 году был дан такой прогноз: «Суммарная вероятность землетрясений с магнитудой более 7,7, которые могут иметь силу 7–9 баллов в г. Петропавловске-Камчатском, может достигать на следующее пятилетие 52,3%». Что это значит? Завтра тряхнет? А когда? А где? Увы, на такие прямые вопросы мы ответить пока не в силах. Интерпретируя это сообщение, не стоит мыслить о вероятности как о мере частоты событий. Конечно, если повторить пятилетний период сто раз, то

* Сама идея цепи появилась при работе Андрея Андреевича Маркова над темой, как кажется, весьма далекой от математики: анализом сочетаний гласных и согласных звуков в тексте романа А. С. Пушкина «Евгений Онегин».

можно заключить, что в ближайшие 500 лет произойдет примерно 52 землетрясения. Но этот вывод будет верным только при условии неизменности потока, а уже через месяц прогноз изменится. Интенсивность похожа в этом смысле на мгновенную скорость движения: чтобы измерить, что вы двигаетесь со скоростью 60 км/ч, не обязательно ехать целый час именно с таким показателем на спидометре. И, главное, данный учеными прогноз не говорит о том, что между землетрясениями проходит десять лет, как можно предположить, разделив 500 лет на 52 события. Таким образом, если на протяжении десяти лет не было сильного землетрясения, это не значит, что оно произойдет не сегодня-завтра. Оно будет, конечно. Но сколько именно придется ждать — неизвестно.

Посмотрите, как меняется уровень сейсмической активности Камчатского региона для разных масштабов времени (рис. 6.20, изображение взято с сайта Монитора сейсмической активности Камчатского филиала Единой геофизической службы РАН).

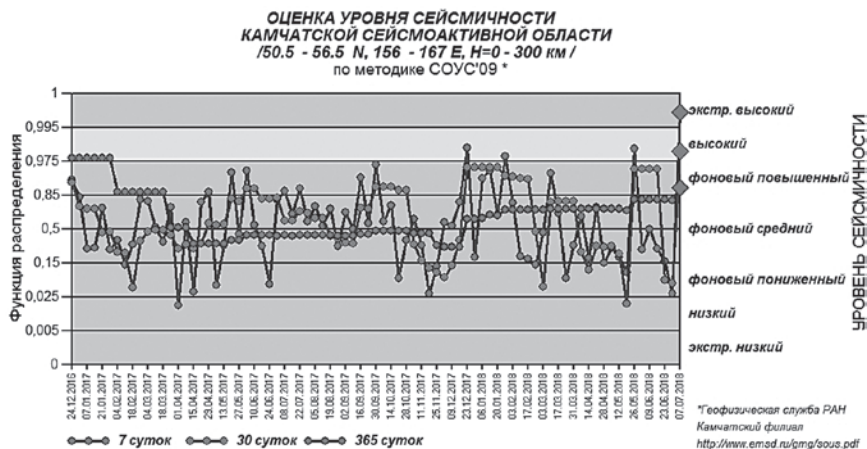


Рис. 6.20. На смену пониженному уровню активности приходит повышенный, активность «дышит», но не периодически, а подобно все тому же случайному блужданию с релаксацией

Но землетрясения — всё же неприятные явления, и пусть бы их не случалось подольше. Бывают события, которых ждешь с большим нетерпением, например прибытие автобуса. Приходя на остановку,

мы, конечно, желаем мгновенно сесть на нужный маршрут, но чаще всего это не удается. Тогда, если в этом месте действует четкое расписание, мы смотрим на него, потом на часы, а затем погружаемся в книжку или телефон. Но где-нибудь в середине маршрута часто вместо расписания указывается интервал движения транспорта, например 15 минут. Это значит, что мы уже далеко от станции, с которой автобусы выходят точно по расписанию, и накапливается некоторая ошибка, делающая прибытие транспорта случайным. И вот тут надо иметь в виду, что в среднем придется ждать именно четверть часа, независимо от того, когда вы приходите на остановку. Вот если бы автобусы приходили с *периодичностью* 15 минут, среднее время ожидания составило бы половину периода — 7,5 минуты. Но с интенсивностью так не выйдет! При отсутствии дополнительных условий движение транспорта моделируют пуассоновским потоком, а это значит, что время ожидания автобуса будет подчиняться экспоненциальному закону с той же интенсивностью. Но математическое ожидание для экспоненциально распределенной величины с интенсивностью λ равно $1/\lambda$, откуда и следует наш вывод. И что совсем обидно — количество времени, уже проведенного вами на остановке, никак не влияет на вероятность того, что автобус вот-вот подойдет. Это свойство экспоненциального распределения — *отсутствие памяти*, связанное с независимостью пуассоновских событий.

Впрочем, если быть точным, то дела с ожиданием автобуса обстоят еще хуже. Измеряемый наблюдателем случайный отрезок времени между машинами статистически больше $1/\lambda$, и вероятность длительного интервала выше, чем среднего. Такой парадокс мы уже встречали — это *парадокс наблюдателя* или *инспектора*.

Подведем итог. Приходя на остановку, нужно четко принять решение: ждать или идти пешком. Размышлять на тему: подождать еще или уже пойти — только обрекать себя на встречу с законом подлости. Ведь если вы, прождав 17 минут, плюнете и пойдете пешком, вас, весьма вероятно, обгонит долгожданный автобус, а то и два.

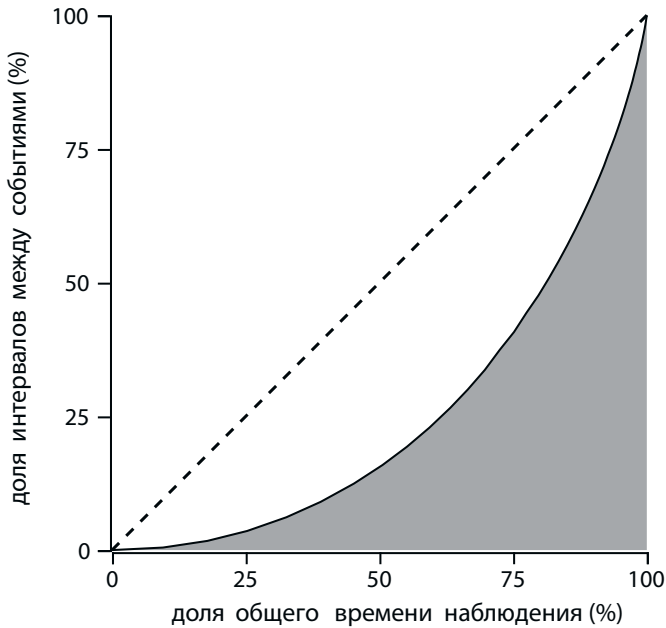


Рис. 6.21. Кривая Лоренца для экспоненциального распределения не зависит от его параметра (интенсивности)

Несправедливость, к которой приводит парадокс инспектора, демонстрирует кривая Лоренца (рис. 6.21). Интересно, что она в случае экспоненциального распределения одинакова для любых интенсивностей. Таким образом, для всех пуассоновских процессов верно утверждение: половина общего времени наблюдения приходится на 20% случаев, когда это очередное событие задерживается. К этому выводу можно прийти, увидев, что на кривой Лоренца 50% общего времени приходится на 80% интервалов, в оставшиеся 20% попали длинные интервалы, поглощающие половину времени ожидания. Коэффициент Джини для экспоненциального распределения равен в точности $1/2$.

Глава 7

Прелести чужой очереди

Я размышляю о законах подлости, стоя в аэропорту в очереди на регистрацию пассажиров и оформление багажа. Хвост длинный, люди разные и заметные со всеми своими сумками, детьми или клетками. Сзади слышу ворчание: «Как обычно, наша очередь тормозит. Вон, гляди, тот усатый в кепке наравне с нами стоял, а теперь вон где... Вот ведь закон подлости!» Этот закон зовется **наблюдением Этторе**:

Соседняя очередь всегда движется быстрее.

Что же это — психологический эффект или причуды математики?

Еще раз про пуассоновский процесс

Мы уже достаточно знаем о случайных процессах, чтобы немного проанализировать очередь, в которой стоим. За неимением других данных, разумно предположить, что выход из нее происходит пуассоновски: пассажиры подходят к стойке регистрации и проводят там какое-то время, не зависящее от времени обработки данных других пассажиров. Перемещение наблюдателя, стоящего в очереди,

будет иметь вид монотонно изменяющейся ступенчатой линии, с одинаковыми шагами через случайные промежутки времени, подчиненные экспоненциальному распределению. Пара реализаций примеров пуассоновских процессов с одинаковой интенсивностью приведена на рис. 7.1. Обычно пуассоновский процесс накапливает события, и его изображение выглядит как «лесенка», растущая со временем. Но, стоя в очереди, мы заинтересованы в ее скорейшем уменьшении, так что шаги нашего процесса ведут вниз.

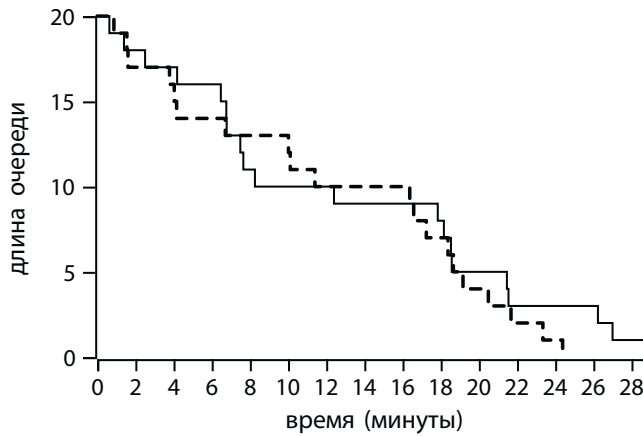


Рис. 7.1. Перемещения двух очередей как пуассоновских процессов с равной интенсивностью. То одна, то другая «вырывается вперед» на какое-то время

Разница двух одинаковых пуассоновских процессов — а именно ее наблюдает человек, скупающий в хвосте и исследующий соседнюю очередь, — представляет собой своеобразное случайное блуждание. В описанном нами случае величина отставания одной очереди от другой подчиняется *распределению Скеллама*. Для двух одинаковых очередей, пропускающих μ человек в единицу времени, вероятность отставания одной из них на k шагов равна:

$$P(k) = e^{-2\mu} I_{|k|}(2\mu),$$

где $I_k(x)$ — встречавшаяся нам в предыдущей главе модифицированная функция Бесселя. Она возникла здесь не из-за круговой

симметрии, а как результат сложения двух случайных величин, подчиняющихся распределению Пуассона.

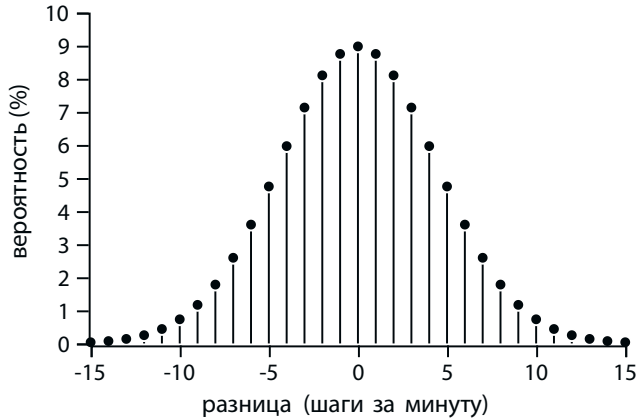


Рис. 7.2. Вероятность накопления разницы между двумя одинаковыми очередями со средней скоростью 5 шагов в минуту

Распределение Скеллама имеет симметричный колоколообразный вид (рис. 7.2), практически не отличимый от биномиального распределения. А раз так, мы уже готовы сделать некоторые качественные выводы, основываясь на опыте, полученном в предыдущей главе.

Во-первых, расстояние между одновременно вставшими в одинаковые очереди людьми будет то увеличиваться, то уменьшаться, при этом станут образовываться характерные меандры с постоянно меняющейся длительностью. Во-вторых, из-за самоподобия случайного блуждания длительность меандров — как для коротких очередей, так и для длинных — окажется соизмеримой со временем стояния в очереди, и, значит, они будут заметны. А меандры — уже повод для недовольства. В-третьих, заранее неизвестно, какая очередь пройдет быстрее, ведь случайное блуждание равновероятно уходит как вверх, так и вниз. И наконец, четвертое заключение: очереди движутся независимо, то и дело опережая и нагоняя друг друга, но в среднем одинаково, и ожидаемая разница между ними стремится к нулю, однако разброс вокруг среднего со временем растет пропорционально квадратному корню из времени.

Выходит, нет никаких подлых штучек злодейки-судьбы, а есть только честное случайное блуждание. Правда, если нам не повезло и мы оказались во временно отстающей очереди, то мы в ней проведем больше времени и, согласно закону велосипедиста, у нас будет больше возможностей посетовать на судьбу! А теперь, внимание, хорошие новости: в любой выбранный интервал времени тех, кому повезет попасть в быструю очередь, будет больше, чем невезунчиков, ведь быстрая очередь может пропустить больше людей! Но, увы, это ничуть не утешит того, кто надолго застрял в хвосте.

Теория для заскучавших в коридоре

Тем и хороша математика, что она способна сделать увлекательным даже стояние в очереди. Например, можно прикинуть, сколько еще ждать своей очереди, но для этого, как ни странно, надо посмотреть не вперед, а назад, на растущий хвост. Если подождать какое-то время, скажем 10 минут, и посчитать, сколько человек выстроилось за вами, то, разделив количество людей перед вами на полученное число, вы вычислите среднее время ожидания в десятках минут. Например, пусть за десять минут хвост вырос на пять человек; если в момент подсчета перед вами семь человек, то ожидаемое время ожидания составит $10 \times 7/5 = 14$ минут. Понятно, что эта оценка будет весьма грубой, но любопытно, что она действительно соответствует среднему времени ожидания. Об этом говорит *теорема Литтла* — один из самых ранних и самых общих результатов *теории очередей*, известной в России как *теория массового обслуживания*.

Теория очередей появилась в самом начале XX века, с первых работ датского математика Агнера Эрланга (1878–1929), который занимался зарождающейся областью телекоммуникаций. За сотню лет результаты исследований Эрланга прочно вошли в нашу жизнь — настолько, что возникает ощущение, будто мы вошли в мир телекоммуникаций. Несколько позже большой вклад в развитие этой науки внес советский математик Александр Яковлевич

Хинчин (1894–1959), который вместе с Андреем Николаевичем Колмогоровым (1903–1987) заложил основы современной теории вероятностей. Результаты теории массового обслуживания важны для проектирования магазинов и залов ожидания, оптимального управления операционной системой компьютера и операционным залом банка, для грамотной разработки бюрократической машины, управления дорожной сетью и в оценке рисков страховой компании. В очередях могут стоять люди (покупатели, клиенты, пассажиры), автотранспорт и грузы, задачи и документы; а обрабатывать их — кассиры, операторы, регистраторы, серверы и бюрократы. Чтобы не путаться и не утопать в деталях, будем называть стоящих в очереди *клиентами*, а того, кто их обслуживает, — *оператором*.

Представьте себе очередь, в которую люди встают согласно некоторому распределению временных интервалов $p_{in}(t)$ со средним значением $1/\lambda$. Время, которое оператор тратит на работу с клиентами, подчинено распределению $p_{out}(t)$ со средним значением $1/\mu$. На рисунке 7.3 показана очередь, в которой ожидают два клиента под номерами 1 и 2, один с номером 0 обслуживается, а клиент номер 3 готов в нее встать. Ее можно описать как марковский процесс, в котором состояние определяется длиной очереди: состояние 0 — в очереди никого, состояние 1 — один клиент, состояние 2 — два клиента и т. д. В идеальном мире ничто не запрещает очереди стать сколь угодно длинной; значит, мы получаем цепь с *бесконечным числом состояний*, и для анализа очереди придется иметь дело с матрицей переходов, содержащей бесконечное число строк и столбцов. В предыдущей главе мы уже имели дело с марковскими процессами, и для анализа стационарного состояния цепи нам понадобилось возводить матрицу переходов в бесконечную степень. Так что же, надо вычислить *бесконечную матрицу, возведенную в бесконечную степень*? Математиков эта задача не испугала, и уже в 1930-е были придуманы методы для таких вычислений. Результатом анализа будут свойства *стационарного состояния* очереди. Оно не меняется со временем, но все параметры очереди, такие как длина или время ожидания в ней, — случайные величины. Они могут постоянно меняться, но при этом всегда остаются в рамках

каких-то распределений, от времени не зависящих. И чего только не придумаешь, скучая в зале ожидания!



Рис. 7.3. Модель очереди

Свойства очереди сильно зависят от соотношения λ и μ . Если $\lambda > \mu$, хвост будет расти неограниченно, как пробка на дороге, в которую въезжает больше автомобилей, чем может выехать. Она попросту перекрывает поток клиентов, накапливая их в себе. Для $\lambda < \mu$ очередь *устойчива*. Она может расти или уменьшаться по мере того, как клиенты добавляются и выходят из нее, но клиенты в ней не накапливаются неограниченно: сколько их вошло в зону ожидания, столько же выйдет. Иными словами, устойчивая очередь может затормозить тех, кто в ней стоит, но неспособна изменить интенсивность потока людей, проходящих сквозь нее. И если на входе мы имеем в среднем λ человек в единицу времени, то и на выходе должны получить такой же поток, независимо от скорости работы оператора. Случай $\lambda \approx \mu$ рассматривается отдельно. Такая метастабильная очередь ведет себя неустойчиво и моделируется процессом случайного блуждания — с той только разницей, что длина очереди не может быть отрицательной. У блуждающей таким образом системы есть непроницаемая стенка снизу, которая, однако, не мешает практически неограниченному росту длины очереди. И хотя рано или поздно она сократится и даже исчезнет, отклонения времени ожидания и времени работы оператора от среднего будут столь велики, что счесть такое обслуживание удовлетворительным никак не получится. Далее мы будем рассматривать только устойчивые очереди. От характера распределений $p_m(t)$

и $p_{out}(t)$ зависят динамика очереди и ее характеристики, такие как распределение для ее длины, времени ожидания клиентом и времени занятости оператора. Для очередей создана система обозначений, называемая нотацией Кендалла. Например, простая очередь, в которую люди входят равномерно и так же уходят, как, например, в аэропорту при посадке на рейс, обозначается D/D/1 (буква D здесь обозначает детерминированный процесс, соответствующий вырожденному распределению, а единица — одного оператора). Въезд и выезд автомашин на территорию аэропорта через три автоматических шлагбаума можно описать очередью M/D/3. Буквой M обозначается пуассоновский (марковский) процесс, то есть случайный процесс без памяти. В очередь на регистрацию билетов и оформление багажа новые люди приходят по-пуассоновски, и багаж у всех разный, так что клиенты будут выходить из очереди тоже по-пуассоновски. Для пяти стоек такая очередь обозначается M/M/5. Собственные обозначения существуют и для других видов распределений. Если же мы вообще ничего не знаем о распределении появления клиентов или методах их обслуживания, то обозначаем такой произвольный процесс буквой G (от слова General — общий).

В этой главе мы будем исследовать неприятности и неожиданности, наблюдаемые в очередях, на примере очереди с $\lambda = 30$ чел./ч и $\mu = 34$ чел./ч. В среднем новые клиенты будут поступать в нее с интервалом в 2 минуты, а обрабатываться оператором примерно за 1 минуту 45 секунд. Это похоже на очередь у стойки регистрации в аэропорту. На рисунке 7.4 показан пример того, как могут «жить» M/D/1- и M/M/1-очереди с такими параметрами.

В стационарном состоянии длина M/M/1-очереди n описывается *геометрическим распределением*:

$$p(n) = \left(1 - \frac{\lambda}{\mu}\right) \left(\frac{\lambda}{\mu}\right)^n.$$

Мы встречали его в предыдущей главе, рассматривая простейшую несимметричную марковскую цепь. Зная это распределение, можно вычислить ожидаемую длину $L = \frac{\lambda}{\mu - \lambda}$.

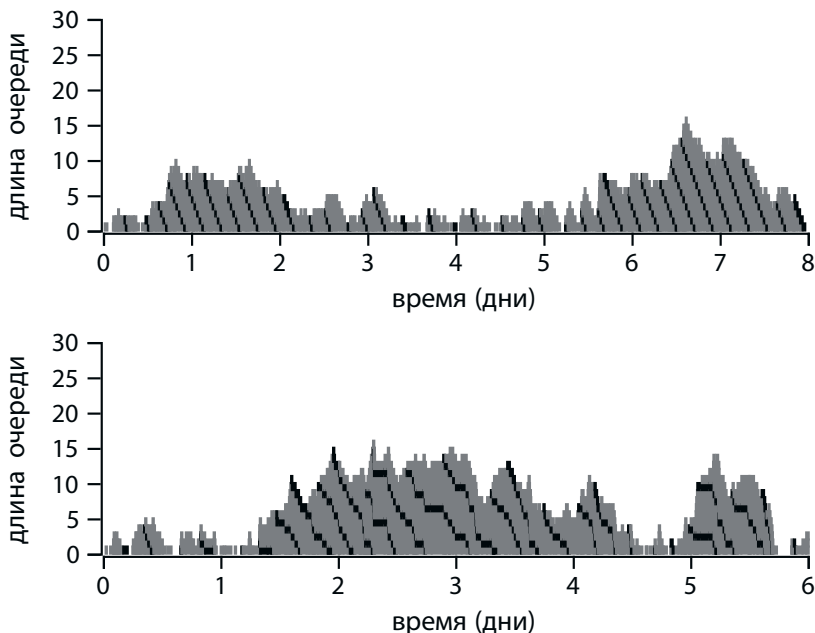


Рис. 7.4. Динамика M/D/1 и M/M/1 очередей. Более темным цветом выделены траектории каждого седьмого клиента в очереди. Длина очереди склонна к своеобразным колебаниям: она «дышит», то удлиняясь, то сокращаясь, оставаясь при этом в стационарном состоянии

Для нашего примера средняя длина очереди составит 7,5 человек. *Время обслуживания клиента* (сумма времени ожидания своей очереди и собственно времени работы с оператором) в M/M/1-очереди описывается экспоненциальным распределением с параметром $\mu - \lambda$. Это приводит к значению среднего времени ожидания $W = \frac{1}{\mu - \lambda}$.

Среднее время работы с каждым клиентом не превышает 2 минут, однако среднее время ожидания для нашего примера равно 15 минутам. Как видно, для стационарной M/M/1-очереди выполняется равенство:

$$\lambda W = L.$$

Это и есть формула Литтла, которой мы воспользовались, стоя в очереди и от нечего делать занявшись подсчетами. Будучи очень простой, формула на удивление сильна: она выполняется для очень

широкого класса очередей и в самых разных задачах. То, что в формулу Литтла входит только λ , а не μ , отражает основное свойство стабильной (устойчивой) очереди: она может задерживать клиентов, но не меняет их поток, который определяется значением λ . И даже если скорость работы оператора μ будет очень велика, среднее время ожидания все равно определяется входным потоком и уже скопившимся числом клиентов. А поскольку для устойчивых очередей $\lambda < \mu$, мы получаем еще один закон подлости:

**Даже с идеальным кассиром время стояния в очереди
в кассу определяется бестолковыми покупателями.**

Важная характеристика очереди — *время занятости оператора*, или длительность непрерывных периодов времени, в которые он обслуживает клиентов. Обозначим это время B . Периоды занятости перемежаются *периодами простоя*, когда по какой-то причине клиентов в очереди не оказывается. Клиенты приходят, ждут и уходят, а оператор остается работать, поэтому разумно предположить, что $B > W$. В действительности ожидаемое, среднее время занятости для $M/M/1$ -очередей равно среднему времени ожидания, то есть $B = W$. Уже не вполне интуитивно понятный результат, но и это еще не всё: при той же интенсивности труда среднее время обслуживания клиента может стать существенно больше среднего времени работы оператора! Вот это уже кажется парадоксом. Получается, оператор в среднем умудряется работать меньше, чем в среднем обслуживается клиент!

Как мы уже говорили, средние значения надо использовать осторожно. Объяснить этот парадокс и понять, что происходит в очереди, можно, привлекая дисперсию распределения времени обслуживания одного клиента $p_{out}(t)$. Еще в 1930-е австрийскому математику Феликсу Поллачеку удалось в общем виде вычислить отношение W/B для произвольной $M/G/1$ -очередей:

$$\frac{W}{B} = 1 + \frac{\lambda}{2\mu} (\mu^2 \sigma^2 - 1).$$

Здесь σ — дисперсия распределения $p_{out}(t)$. В случае $M/M/1$ -очередей $\sigma = 1/\mu$, и это отношение равно 1. Но может случиться, что при

том же значении среднего распределение $p_{out}(t)$ будет иметь бóльшую дисперсию, и тогда W окажется больше B . На рисунке 7.5 показан пример, в котором $p_{in}(t)$ распределено экспоненциально с $\lambda = 30$ чел./ч, а $p_{out}(t)$ описывается гамма-распределением, соответствующим интенсивности $\mu = 34$ чел./ч с дисперсией $\sigma = 2/\mu$.

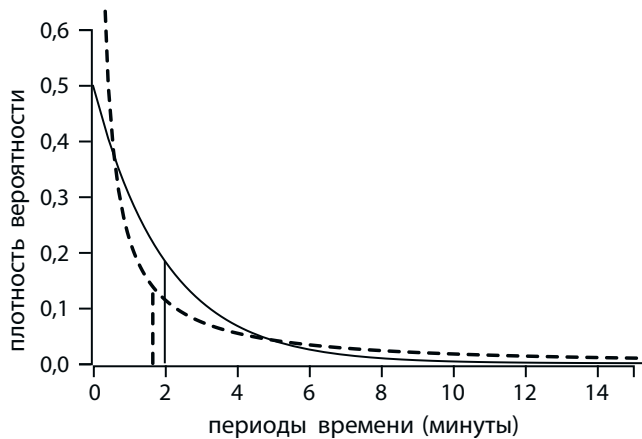


Рис. 7.5. Распределения для периодов между появлением новых клиентов (сплошная линия — экспоненциальное распределение) и времени обслуживания одного клиента (пунктирная линия — гамма-распределение)

Очередь остается стабильной, поскольку $\lambda < \mu$ и клиенты в среднем обслуживаются быстрее, чем приходят новые. Оператор работает хорошо: большинство клиентов обслуживаются очень быстро; но обратите внимание на долю «трудных» клиентов, которые формируют достаточно толстый хвост распределения. Их мало, но каждый отнимает много времени, и все в очереди вынуждены их ждать. Для примера, приведенного на рисунке, среднее время ожидания оказалось равно 35 минутам, хотя среднее время занятости оператора прежде (15 минут). Получается, что, не переставая работать, оператор в среднем флонирует, пока мы страдаем в очереди от безделья!

Динамика такой очереди отличается от динамики M/M/1. Для нее характерен несимметричный пилообразный рисунок с плавной восходящей линией и резким сбросом. Пока оператор занят

«трудным» клиентом, постепенно вырастает длинный хвост, а потом, освободившись, оператор очень быстро с ним справляется (рис. 7.6).

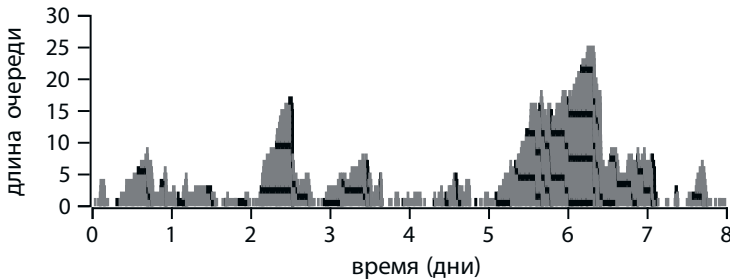


Рис. 7.6. Динамика M/G/1-очереди, где время ожидания клиентов вдвое превосходит время занятости оператора. Горизонтальные темные полосы показывают периоды долгого ожидания очередного «трудного» клиента

Совсем немного о случайных функциях

Здесь мы ненадолго остановимся и обсудим, что же все-таки такое случайный процесс.

Все очереди движутся по-разному. Ступеньки пуассоновского процесса не повторяют друг друга, и мы располагаем только какими-то статистическими свойствами случайных процессов. Но это уже явно не просто случайное число, а кое-что посложнее. С чем же мы имеем дело? Случайный процесс порождает некую последовательность. Его повторение приведет к новой последовательности, скорее всего с другим числом точек. А можно ли обобщить все эти случайные последовательности? Главным свойством случайных величин мы считаем их непостоянство: от раза к разу, от эксперимента к эксперименту каждая из них меняет свое значение, оставаясь при этом одним объектом. Мы смогли однозначно характеризовать его распределением случайной величины — функцией, сопоставляющей каждое значение случайной величины (или диапазон значений) и его вероятность.

Говоря о стохастических последовательностях, мы имеем дело уже не со случайной величиной, а со *случайной функцией*. Например, для пуассоновского процесса это функция от времени, возвращающая случайную величину — число отсчетов, наблюдаемых за указанное время. Можно ли такую случайную функцию характеризовать так же однозначно и точно, как случайная величина определяется своим распределением?

Построим на одном графике большое число пуассоновских «лесенок» одинаковой интенсивности, а потом для каждого момента времени создадим срез всех этих данных и усредним их, получив одну точку. Вот что мы увидим (рис. 7.7).

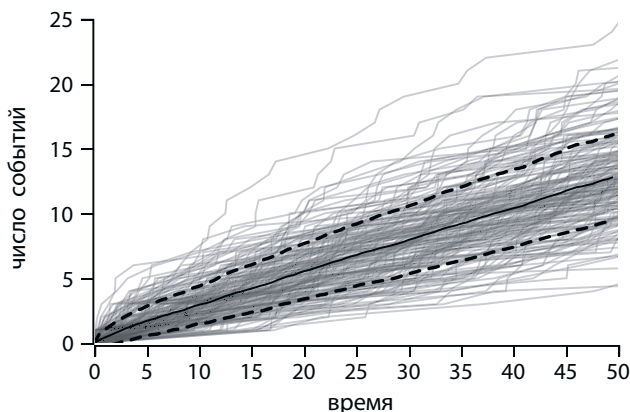


Рис. 7.7. Черная сплошная линия — результат усреднения множества реализаций пуассоновского процесса с интенсивностью $1/4$

Облаком всевозможных последовательностей оказалась окружена прямая линия, имеющая наклон, равный интенсивности потока. Это график *математического ожидания случайной функции*. В отличие от настоящего пуассоновского процесса, то есть подсчета числа событий, значения этой функции — уже не целые числа. Как и среднее значение случайной величины, она характеризует случайную функцию, но вовсе не полностью. Например, можно рассмотреть аналог дисперсии, показав, насколько велик ожидаемый разброс значений от среднего. Стандартное

отклонение показано на рисунке пунктиром. Но и две функции — среднее и дисперсия — не дадут полной характеристики. Одна и та же случайная функция способна породить бесчисленное множество последовательностей одинаковой интенсивности. Вновь перенесемся в аэропорт и представим себе две одинаковые очереди, идущие параллельно, например к стойке регистрации. Их движение описывается идентичными случайными функциями, средние графики неразличимы, однако наблюдаемая разница в шагах между двумя параллельными одинаковыми очередями подчиняется не тривиальному распределению Скеллама.

Может быть, если для каждого среза времени мы выясним распределение случайной величины $F(t)$ (скажем, найдя его плотность вероятности $p_F(t)$), то получим исчерпывающую информацию о случайной функции F ? Наконец, можно ли синтезировать случайный процесс, генерируя случайные числа согласно распределениям $p_F(t)$?

Ответ на все эти вопросы: нет. Случайные функции устроены сложнее, чем случайные числа. Рассуждая о марковских цепях, мы говорили, что они порождают случайные процессы, *не имеющие памяти*. При этом мы имели в виду, что на будущее в этих процессах влияет не прошлое, а только настоящий момент. Это свойство — отсутствие памяти — характерно для экспоненциального распределения и связанного с ним пуассоновского процесса. Характеристика памяти процесса — величина, называемая *автокорреляционной функцией*, которая определяется как среднее от произведения двух значений функции, вычисленных в разделенные известным промежутком τ моменты времени:

$$K(\tau) = M[F(\tau) \cdot F(t - \tau)].$$

Здесь символ $M[F(\tau)]$ обозначает математическое ожидание (среднее значение) функции $f(t)$. Величина временного лага τ показывает, насколько далеко мы заглядываем в прошлое. Для важного класса случайных функций, которые называются *эргодическими*, усреднение может производиться не по множеству реализаций случайного процесса, как для множества пуассоновских процессов,

а по одному достаточно длинному ряду наблюдений за единственной реализацией. В физике, экономике или климатологии эргодичность случайных последовательностей очень важна, поскольку мы располагаем одним-единственным миром и можем наблюдать за ним долго, но неспособны исследовать множество его различных реализаций.

Автокорреляция позволяет различать истинно стохастические процессы, детерминированные процессы с наложенным на них шумом и процессы, порождаемые динамическим хаосом. С ее помощью можно отделять в экспериментальных данных основные временные закономерности, присущие процессу, порождающему эти данные, от случайного, не связанного с ними шума. Это один из основных инструментов *анализа временных рядов*. С его помощью сейсмологи расшифровывают запись землетрясения, выделяя из, казалось бы, совершенно беспорядочного сигнала первичные волны, пришедшие непосредственно от землетрясения, волны, отраженные от границ внутренних слоев Земли, вплоть до самого ядра, и обменные волны, рождающиеся на этих границах. Так сильные землетрясения на несколько часов делают нашу планету «прозрачной», как бы подсвечивая ее изнутри лучами сейсмических волн.

Корреляция в переводе с латыни — «отношение»; получается, что автокорреляция — «отношение к самому себе» или «связь с самим собой в прошлом». Согласитесь, это красивый образ не только для случайной функции, но и для человека.

Мне только спросить!

Но вернемся к очередям и проблемам, с ними связанным. Есть в нашей жизни досадное явление — «обочечники». Это ушлые водители, объезжающие пробку по обочине и потом встречающие в поток. Есть настырные посетители поликлиник и касс, норовящие просочиться к заветному окошку или двери с формулой «Мне только спросить...». В любую отлаженную бюрократическую систему то и дело врываются неотложные дела, не терпящие промедления. Понятно, что порой без таких случаев не обойтись:

в больницах бывают неотложные пациенты, в операционной системе компьютера есть задачи с очень высоким приоритетом; наконец, на дороге мы обязаны пропускать спецтранспорт, едущий по экстренному случаю. Но как внеочередники влияют на всю очередь? Подобные случаи моделируются *очередями с приоритетом* (рис. 7.8), и для них тоже есть развитая теория, поскольку в жизни они встречаются чуть ли не чаще простых очередей.



Рис. 7.8. Очередь с приоритетом

Пусть в нашей $M/M/1$ -очереди с вероятностью ε могут появляться особые клиенты, назовем их VIP (very impatient person — очень нетерпеливые персоны), которые встают не в конец очереди, а вклиниваются в ее начало, заставляя ждать всех стоящих позади. При этом они всё же дают оператору завершить работу с текущим клиентом, не прерывая его. Если внеочередников наберется несколько, они могут образовать свою VIP-очередь. Вспомним, что пуассоновский поток можно представить как случайное «разбрасывание» по временному интервалу какого-то известного количества событий. Поскольку все клиенты приходят независимо, то, согласно нашему условию, мы получим поток нетерпеливых клиентов $\varepsilon\lambda$ и поток обычных клиентов $(1-\varepsilon)\lambda$, при этом общий поток останется неизменным. Среднее время ожидания для VIP будет равно $W_{VIP} = \frac{1}{\mu - \varepsilon\lambda}$, как в простой $M/M/1$ -очереди, поскольку они в своей VIP-очереди «не замечают» присутствия обычных клиентов. Для того, кто ждет на общих основаниях, время ожидания вырастет и составит уже:

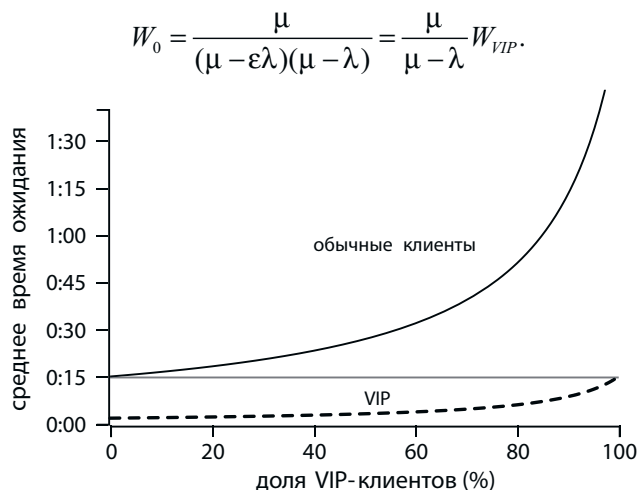


Рис. 7.9. Соотношение средних времен ожидания для очереди с нетерпеливыми VIP-клиентами

Как показывает рисунок 7.9, пока VIP-ов немного, очереди они мешают не сильно. Но если доля внеочередников оказывается близкой к единице, то никакого преимущества они уже не имеют, зато немногочисленным скромным очередникам приходится ждать существенно дольше. При ε , стремящемся к единице, среднее время ожидания рядовых очередников стремится к $\mu/(\mu - \lambda)^2$ (больше двух часов в нашем случае!); и вообще, если μ лишь немного превышает λ , очередь остается устойчивой, однако время ожидания в ней вырастает катастрофически!

Но вот что любопытно. Можно найти среднее время ожидания для всей группы клиентов как взвешенную сумму $\varepsilon W_{VIP} + (1 - \varepsilon)W_0$, и она окажется равной $1/(\mu - \lambda)^2$, то есть такой же, как для обыкновенной M/M/1-очереди без всяких VIP-ов. Выходит, системе в целом внеочередники не мешают. На время занятости оператора они тоже не влияют, распределение времен ожидания остается экспоненциальным. Мы уже говорили в предыдущей главе, что для экспоненциального распределения кривая Лоренца и, соответственно, коэффициент Джини не зависят от параметра распределения, а значит, все M/M/1-очереди имеют одинаковую степень несправедливости — 0,5. Отсюда следует, что наш обобщенный

критерий несправедливости для всех ожидающих в очереди также останется равным 0,5.

Стационарный бардак

А теперь немного изменим политику очередности. Пусть внеочередники будут сверхнаглыми, и если так случится, что один такой клиент придет вслед за другим, то вместо формирования нормальной очереди второй вклинится перед первым. Эта задача уже отличается от классического подхода к очередям с приоритетом. Давайте сразу рассмотрим предельный случай, когда доля наглых клиентов равна единице. Тогда наша очередь (рис. 7.10) превращается в то, что программисты называют *стеком*, — последовательность элементов, подчиняющуюся правилу «*первым вошел, последним вышел*» (FILO — first in, last out) — в противовес очереди, для которой выполняется правило «*первым вошел, первым вышел*» (FIFO — first in, first out).

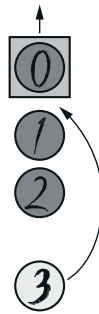


Рис. 7.10. Очередь как стек

Такая «очередь наоборот» выглядит неестественно, но если вместо людей мы рассмотрим пачку документов, то можем увидеть знакомую картину на рабочем столе, когда входящие документы не сортируются по времени, складываются в стопку по мере поступления, а потом обрабатываются начиная сверху. Удивительно, но в стационарном состоянии все средние значения основных параметров — и длины очереди, и времени ожидания, и времени

занятости оператора — будут точно такими же, как и в FIFO-очереди. Что же поменяется? Посмотрим на пример работы такой очереди, он показан на рис. 7.11. Мы видим, что вместо целенаправленного движения к оператору клиенты могут то приближаться к нему, то отдаляться. Время ожидания для самого последнего клиента существенно удлинится, однако, пока он ждет, через оператора проходит много вновь поступающих клиентов, которые обрабатываются почти мгновенно. В среднем же мы получаем примерно такое же время ожидания, как для «нормальной» очереди. Но мы уже много раз убеждались в том, что среднее значение не может характеризовать случайную величину в полной мере.

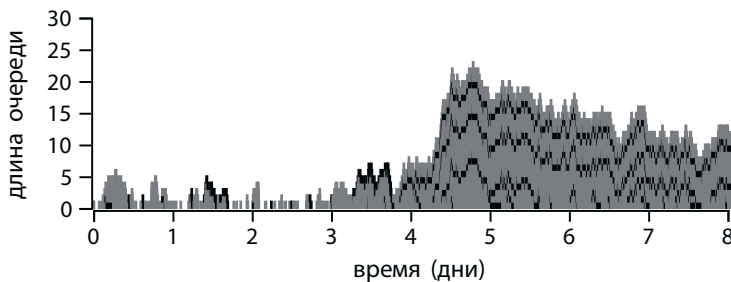


Рис. 7.11. Динамика FIFO-очереди или стопки документов, которые при поступлении кладутся наверх и обрабатываются начиная сверху. Как и прежде, темные точки соответствуют каждому седьмому клиенту

Глядя на динамику FIFO-очереди, легко понять, что время ожидания клиента должно быть близким к времени занятости оператора. Действительно, время занятости определяется как период от момента прихода первого клиента до момента выхода последнего, но в стеке первый клиент и оказывается последним. Нужно еще учесть, что очередной клиент не прерывает оператора, поэтому ко времени его ожидания добавится время обслуживания клиента, с которым уже работает сотрудник. Если оно распределено экспоненциально, то, как уже обсуждалось в связи со временем ожидания автобуса, добавочное время будет распределено точно так же. В итоге время ожидания окажется распределено как сумма времени занятости оператора и периода работы с одним клиентом. На рисунке 7.12

показаны распределения времени ожидания для М/М/1-очереди: обыкновенной, с политикой FIFO, придерживающейся правила FILO. В обоих случаях $\lambda = 30$ и $\mu = 34$ человека в час.

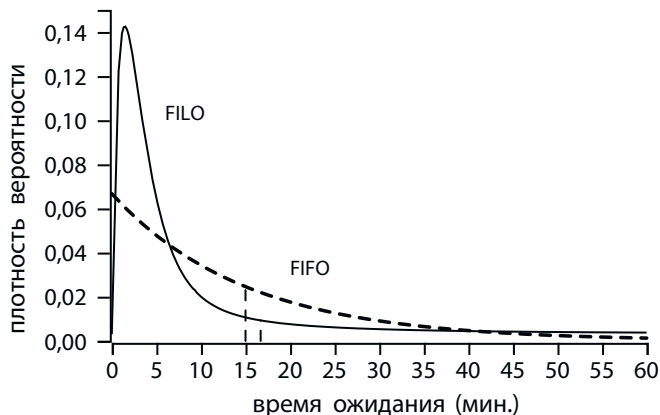


Рис. 7.12. Распределения времени ожидания для М/М/1-очереди с различной политикой

Распределения сильно различаются, но средние значения у них практически одинаковые. Хотя распределение времени ожидания для FILO-очереди кажется сконцентрированным около моды (близкой к $1/\mu$, времени работы с одним клиентом), у него длинный тяжелый хвост, который сильно увеличивает дисперсию и повышает среднее значение. Медиана этого распределения равна 3 минутам. Это значит, что в половине случаев клиент будет ждать немного, но если уж застрянет, так застрянет: 5% клиентов потратят больше часа, а самые невезучие 2% вместо 2 минут вынуждены будут ждать своей очереди больше 2 часов! Для FIFO-очереди с такими же параметрами вероятность застрять на 2 часа составляет не более 0,04%.

При этом оператор, то есть бюрократ, обрабатывающий бумаги, не заметит разницы между очередью и стеком: распределение его времени занятости не изменится. Руководитель бюрократа тоже увидит, что из кабинета подчиненного бумаги выходят с нормальной интенсивностью в силу устойчивости очереди. И большинство документов даже окажутся обработаны очень оперативно. Но то и дело какая-то их часть внезапно «проваливается» на дно стопки

и задерживается там очень надолго. Такие дела приходится «двигать вручную» тем, кто в них заинтересован.

Подобная картина наблюдается и в шкафу, куда мы складываем вещи с мыслью разобрать потом. Но мы задвигаем то, что уже лежит там, поглубже и добавляем новые вещи. Так что даже если мы и станем их постепенно разбирать, до «ископаемых» у самой стенки руки дойдут очень и очень нескоро.

Для демонстрации несправедливости распределения времени между различными делами в ведомстве бюрократа (или вещами в шкафу) изобразим кривые Лоренца для FIFO- и FILO-очереди, описываемых формулой $M/M/1$. Если для всех FIFO-очереди кривая Лоренца одинакова, несправедливость FILO-очереди зависит от соотношения λ и μ . Чем ближе их отношение к единице, тем ближе к ней и индекс Джини. На рисунке 7.13 показаны кривые Лоренца для этих двух типов очередей с теми же параметрами, что и для предыдущего рисунка.

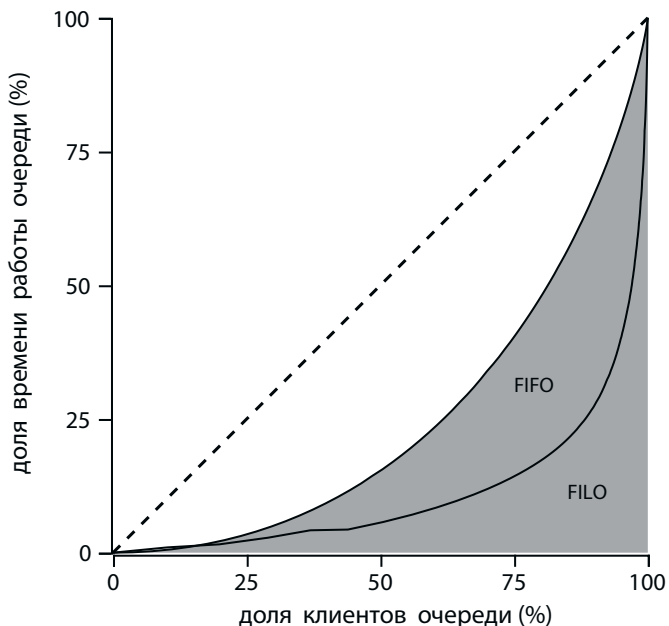


Рис. 7.13. Кривые Лоренца для времени ожидания в двух типах очередей. Коэффициент Джини для FIFO-очереди равен 0,5, а для FILO-очереди — 0,78

Закон сохранения клиентов, или равенство входного и выходного потоков, может сыграть еще одну злую шутку. Представьте себе контору, через которую проходит за рабочий день, скажем, 15 человек; при этом на каждого клиента в среднем уходит полчаса. В конторе два клерка, один трудится с интенсивностью 16 человек в день, а второй — 14. Вместе они могли бы обслужить человек тридцать, но, кажется, столько и не нужно. Люди в такой конторе почти не ждут своей очереди, среднее время ожидания составляет 18 минут, средняя длина очереди всего 1,2 человека*. Очень часто бывает так, что, пока с клиентом работает кто-то из клерков, второй в это время отдыхает. На рисунке 7.14 показан пример динамики очереди в конторе. При этом надо иметь в виду, что этот незначительный поток распределен между двумя операторами: каждый из них наблюдает поток, еще в два раза менее интенсивный. Про такую работу говорят: «не бей лежачего».

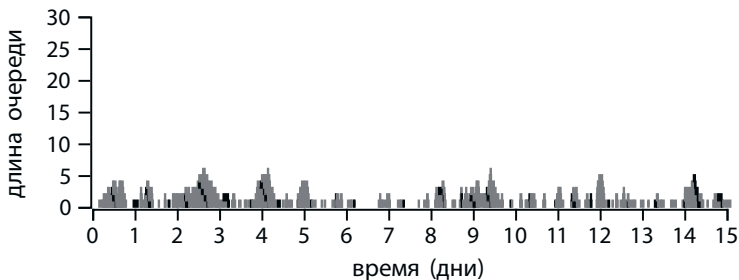


Рис. 7.14. Две недели в конторе с двумя клерками

Руководство, проведя все эти замеры и наблюдения, решает, что клерки живут уж больно вольготно, трудясь только половину рабочего времени, и в стремлении оптимизировать работу учреждения увольняет нерасторопного клерка. И вот все стало совсем иначе! Система приблизилась к опасному состоянию, когда $\mu \approx \lambda$. При таких условиях очередь становится метастабильной: она может какое-то время вести себя «хорошо», а потом внезапно наступит коллапс (рис. 7.15).

* Параметры стационарной M/M/2-очереди можно рассчитать по общим формулам, которые я здесь не привожу из-за их громоздкости.

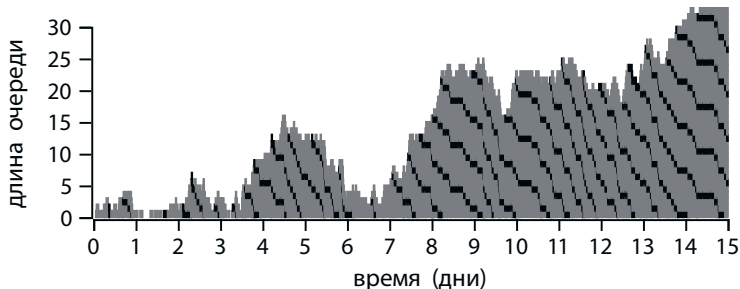


Рис. 7.15. В течение недели один клерк вполне справлялся с объемом работ, но потом все превратилось в кошмар

При $\lambda = 15$ и $\mu = 16$ средняя длина очереди будет как раз равна 15 клиентам, а среднее время занятости оператора составит 1 день. Руководство может быть довольно своей оптимизацией. Но мы-то знаем, что средние показатели не показывают толком почти ничего. Посмотрите, с какой вероятностью время занятости клерка превысит указанное количество дней (рис. 7.16).

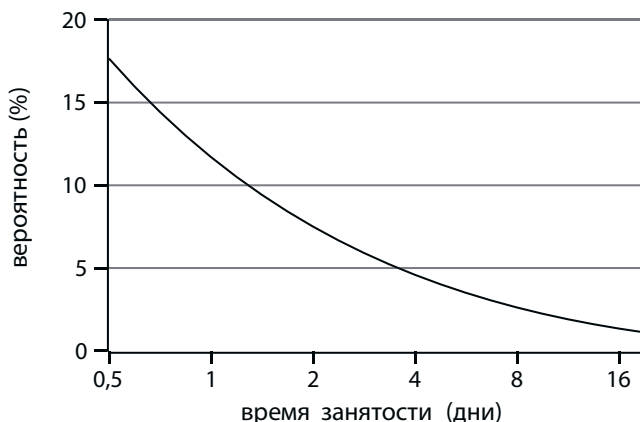


Рис. 7.16. Вероятность для одного клерка не уложиться с текущими делами в указанный период времени

Что еще хуже, среднее время ожидания одного клиента вырастает тоже до одного дня! Вместо 18 минут он застрянет со своим делом на час с вероятностью 88%, вероятность проторчать в конторе подлня составит 60%, а ухлопать на это весь день — 37%.

Таким образом, вроде бы разумное решение может иметь неожиданно неприятные последствия.

Лучшее — враг хорошего

Наконец, говоря об очередях и неприятностях, с ними связанных, нельзя не упомянуть о совершенно возмутительном парадоксе Браеса. Этот эффект приводит к тому, что в коммуникационной сети, содержащей очереди, добавление новых простых связей, даже не стохастических, может привести к уменьшению пропускной способности всей сети.

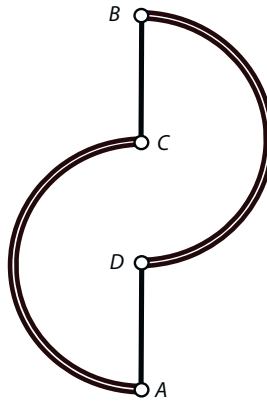


Рис. 7.17. Модель дорожной сети

Простейшей моделью, в которой наблюдается этот эффект, может быть дорожная сеть, где два населенных пункта А и В соединены двумя дорогами так, как показано на рис. 7.17. При этом выделяются четыре участка дорог, два из которых, АС и DB, — достаточно широки и свободны, так что среднее время пути по ним занимает известное постоянное время t_0 . Два других плеча — AD и CB — короче, но имеют склонность к образованию заторов. Дорожный поток и пробка во многом похожи на очередь, и для них тоже работает теорема Литтла, позволяющая связать время пути по загруженному (или узкому) участку с числом машин на дороге. Таким образом, для загруженных участков время пути можно считать

пропорциональным числу участников дорожного движения: $t = \lambda N$. И последнее важное условие: пассажиропоток между городами таков, что $t_0 > \lambda N/2$.

Тут мы впервые вынуждены сделать предположение, которое выходит за рамки темы книги. Оно касается того, как люди принимают решения. Это тоже можно описывать математически с помощью методов теории игр — области знания, которая получила широкое развитие в последние десятилетия. Здесь я не хочу вдаваться в подробности самой теории, а сразу воспользуюсь одним из ее результатов — понятием *равновесия Нэша*. Поведение людей во многих случаях можно считать оптимизирующим: они пытаются уменьшить потери и увеличить свои преимущества. Но во взаимодействии с такими же оптимизирующими группа игроков может нащупать некое равновесное состояние — не лучшее, но хотя бы удовлетворительное. Применительно к нашим дорогам приход к равновесию Нэша выразится в том, что водители будут стремиться распределиться по обоим плечам дорог ACB и ADB поровну. Так что если обычно из города А в город В ездит N автомобилистов, время в пути можно выразить как $\lambda N/2 + t_0$.

Теперь, стремясь оптимизировать движение в этой сети, мы построим связку CD, причем постараемся сделать ее как можно шире и лучше, чтобы время на ее преодоление было существенно меньше, чем t_0 или $\lambda N/2$ (рис. 7.18). Воспользовавшись ею, автомобилист сможет попасть из пункта А в пункт В за время порядка $2t_0$ (двигаясь по пути ACDB) либо $2 \times \lambda N/2 = \lambda N$ (в случае пути ADCB). Но, правда, только при условии, что он окажется на дороге один. Проблема в том, что, как только люди прознают о новой дороге, естественно, какая-то часть водителей постарается пользоваться только ею. И вот к чему это приведет. В равновесии Нэша часть публики αN предпочтет путь ADCB — как более короткий, так что мы должны получить следующие характерные времена: ACB, ADB — $\lambda \alpha N + t_0$, ADCB — $2\lambda \alpha N$, ACDB — $2t_0$. Подвох в том, что все эти времена превышают прежний средний результат $\lambda N/2 + t_0$ для любого $\alpha > 1/2$.

Рассмотрим конкретный пример. Пусть $t_0 = 30$ мин., $\lambda = 1/100$ мин./чел., $\alpha = 2/3$, $N = 5000$. Это означает, что из пункта

А выехало 5000 человек. В отсутствие связки CD среднее время пути от А до В составит 55 минут. Наличие короткого пути приведет к таким вариантам среднего времени: АСВ, АDB — 63 минуты, ADCB — 67 минут, ACDB — 60 минут. Иначе говоря, ни по одному из этих путей не удастся добраться из города А в город В быстрее, чем до строительства новой скоростной дороги. Если водители каким-то усилием воли распределятся по обеим дорогам поровну, то все вернется к первоначальному состоянию. Но тогда, выходит, не было смысла строить новую связку CD!

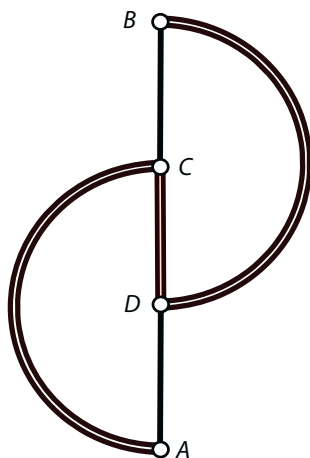


Рис. 7.18. Модель дорог с дополнительной связкой

Парадокс Браеса долгое время казался мне не очень интуитивным и ярким: слишком многое нужно принять во внимание, чтобы понять, что же в нем парадоксального. Мое мнение о нем изменилось, когда я увидел физическую модель этого явления... на пружинках. Удлинение пружины пропорционально приложенной к ней силе; это мы изучали в школе, называя законом Гука. Увеличение времени пути пропорционально загруженности трассы; об этом нам говорит теорема Литтла. Можно рассмотреть две схемы соединения пружин, которые будут эквивалентны двум схемам соединения дорогами населенных пунктов, как показано на рис. 7.19. Физическая модель делает разницу между этими двумя схемами очевидной. В первом случае мы имеем *параллельное* соединение

участков с линейной зависимостью (пружин или затрудненных участков дороги), а во втором — *последовательное*. Для двух одинаковых пружин жесткостью k эффективная жесткость первой (параллельной) схемы будет равна $2k$, а для второй (последовательной) — $k/2$. Таким образом, при одной и той же нагрузке растяжение второй системы больше, чем первой, — при условии, что длина нерастяжимых нитей окажется не меньше длины растянутых пружин. Это условие в точности соответствует требованию $t_0 > \lambda N/2$.

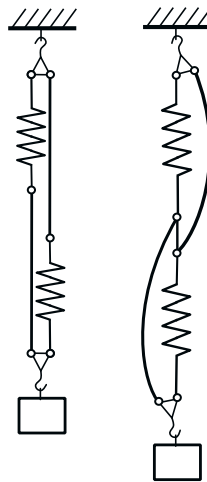


Рис. 7.19. Модель парадокса Браеса на пружинах

Этот парадокс оставался бы на страницах учебников по теории игр, если бы не проявлялся в реальной жизни, и не только в дорожном строительстве. Такое парадоксальное уменьшение пропускной способности сети при добавлении новых соединений встретилось и в механике, и в электрических сетях*, и в полупроводниковых структурах на микроуровне**. А исследования случайных графов, которые важны для анализа социальных сетей и сети интернет,

* Cohen J. E., Horowitz P. Paradoxical behavior of mechanical and electrical networks // Nature. 1991. Vol. 352. Pp. 699–701.

** Pala M., Sellier H. et al. A new transport phenomenon in nanostructures: a mesoscopic analog of the Braess paradox encountered in road networks // Nanoscale Research Letters. 2012. Vol. 7. P. 472.

показали, что эффект Браеса почти наверняка проявляется в них начиная с определенного уровня сложности*. Применив аналогию с пружинами, нетрудно представить себе сложную сеть, в которой есть как упругие, так и нерастяжимые связи. При перерезании каких-то коротких нитей часть упругих связей начнет работать параллельно, распределив между собой нагрузку, и вся сеть станет более жесткой, менее зависимой от нагрузки.

В этой главе мы разбирались с не самыми приятными сторонами нашей жизни — очередями и бюрократией. И хотя они часто вызывают у нас раздражение, всё же эти явления призваны помогать в организации по-настоящему сложных процессов, они поддаются исчислению и избавляют нас от гораздо более неприятного и даже опасного неуправляемого хаоса.

* Valiant G., Roughgarden T. Braess's Paradox in large random graph // Random Structures & Algorithms. 2010. Vol. 37. Pp. 495–515.

Глава 8

Проклятие режиссера и проклятые принтеры

*Настоящее — самое подходящее
время что-то отложить.*

Четвертый закон Хечта

*Любое стоящее дело стоило сделать
вчера.*

Дилемма Гроссмана

Наше время принято считать нелегким, чересчур суетливым и полным стрессов. Так все и заявляют: «В наше нелегкое время...». Уверен, что так говорили, говорят и будут говорить всегда. И основной претензией к любому времени постоянно будет то, что его катастрофически не хватает! Мчатся поезда и самолеты, компьютеры подыскивают и доставляют нам прямо в постель мегабайты информации, всё, что нам нужно, — от светских новостей до рабочих сводок. Поисковые системы мгновенно отвечают как на самые глубокие, так и на самые дурацкие вопросы, и нам все еще не хватает времени. В основном на себя: на прогулку ради прогулки, на то, чтобы послушать музыку — не на бегу в наушниках, не в машине, а дома в кресле с единственной целью: послушать музыку!

Некогда! Но уверен я также и в том, что это вовсе не болезнь века, в отличие, например, от гиподинамии, которая, несмотря на суетливость, беготню и стресс, преследует современного человека. Эта наша спешка и связанная с ней нервозность математически обусловлены и потому вечны, как ворчание стариков на «нынешнее бестолковое поколение».

В этой главе мы поговорим о том, почему нам не хватает времени на задуманное. Почему жизнь так коротка. Почему даже у добросовестного студента к концу учебного года остается лишь одна ночь на выполнение доброй половины всех заданий и почему, в конце концов, именно в эту ночь сломается принтер или его девушка задумает выяснить отношения.

Стратегия балбеса

Для анализа суеты нам опять потребуются случайные процессы. Один из самых простых из них, требующих минимума дополнительных предположений, — пуассоновский поток. Напомню, что его можно реализовать, случайно распределяя известное количество независимых событий по ограниченному временному интервалу. Хорошими примерами могут быть удары капель дождя по крыше, поток частных автомобилей на дороге, сильные землетрясения и т. п.

Но что мы получим, если события перестанут быть независимыми и начнут образовывать упорядоченную цепочку? Скажем, пусть в цепочке событий $\{A, B, C\}$ B может случиться только после A , но перед C . При этом моменты, в которые эти события произойдут, останутся случайными. Посмотрим, как смогут разместиться такие упорядоченные цепочки на ограниченном временном интервале.

Первое событие мы расположим в произвольной точке, второе — тоже случайно, но обязательно после первого, третье — после второго и т. д. Для каждого следующего этапа будет оставаться все меньше времени, так что к правой части интервала (перед дедлайном) должно наблюдаться заметное увеличение

интенсивности процесса. Рано или поздно время для выполнения задач закончится, и цепочка завершится. Назовем построенный нами процесс *стохастической цепочкой с дедлайном*, а выбранную безалаберную стратегию выполнения работы — *стратегией балбеса*. На рисунке 8.1 показан пример построенной таким образом цепочки из пяти этапов работы, на которую было отпущено 20 дней.

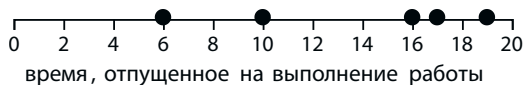


Рис. 8.1. Пример стохастической цепочки с дедлайном.

В данном случае пять дел выполнить удалось, можно успеть шестое, а на семь времени уже не хватит

Понятно, что, выполняя задачи в соответствии с мерфологической аксиомой Дехэя: «Простую работу можно отложить, потому что всегда будет время ее сделать потом», — непросто уложиться в сроки. Но можно ли как-то проанализировать это явление? Сформулируем задачу, взяв в качестве испытуемого, скажем, театрального режиссера. Пусть в распоряжении режиссера и его труппы имеется n дней для постановки некоего действия. Подготовка разбивается на k последовательных репетиционных этапов, каждый из которых требует день на выполнение. Какова вероятность не уложиться в срок, если действовать согласно описанному нами процессу выполнения работ? Подготовка мероприятия требует вовлечения разных людей и различных производственных процессов, возможны накладки, болезни или попросту хандра — все предпосылки к реализации нашей стохастической цепочки с дедлайном.

Для начала я обратился к имитационному моделированию, чтобы выяснить, как распределяется длина цепочек, которые удастся выполнить в ограниченный промежуток времени заданной длины, пользуясь стратегией балбеса.

Вычисления состояли в генерации стохастических цепочек и подсчете их длин для различных ограничений по времени по следующему алгоритму.

Вход: число дней n

Повторять, пока не набрано нужное число цепочек

$x := n$

$k := 0$

Повторять, пока $x > 0$

выбрать случайное целое число $x \sim$

$\text{Uniform}([0, x])$

увеличить счетчик k

конец

добавить k в гистограмму

конец

Вот такая гистограмма получается, например, для $n = 10$ (рис. 8.2).

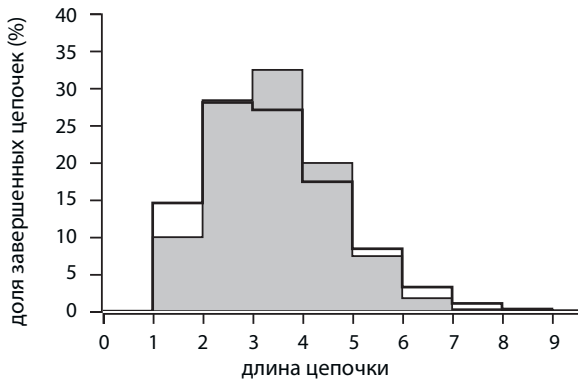


Рис. 8.2. Гистограмма функции вероятности для длины цепочек, которые удается выполнить в отведенный срок. Синей линией показано распределение Пуассона с интенсивностью, соответствующей наблюдаемой средней длине цепочек

Подсчитывая события в настоящем пуассоновском потоке с интенсивностью λ , мы приходим к упоминавшемуся уже распределению Пуассона:

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

которое, напомним, описывает вероятность получить ровно k событий в единичном интервале времени. Распределение внешне

похоже на пуассоновское, но оказалось, что это все же не оно. Разберемся, откуда взялись именно такие доли.

Отвлечемся от дел и сроков и формально опишем исследуемый процесс. Рассмотрим ряд из n пронумерованных ячеек. Процесс состоит в последовательном случайном размещении точек по ним. Первая может оказаться в любой ячейке с равной вероятностью; пусть это будет ячейка с номером i_1 . Следующая точка может оказаться в любой ячейке с номером $i_2 > i_1$. Для всех последующих точек $i_k > i_{k-1}$. Процесс завершится, когда $i_k = n$. Нас интересует вероятность того, что для заданного $n > k$ удастся разместить менее k точек.

Мы будем рассуждать, рассматривая размещение точек «с конца», в обратном порядке. Любая цепочка завершается размещением последней точки в последней ячейке. Шансов не разместить какую-то одну точку нет, поскольку по условиям для первой точки все ячейки свободны. Короткие цепочки из двух точек устроены так: в последней ячейке располагается вторая, последняя точка (с вероятностью $1/n$), а на расположение первой точки ограничений нет, так что вероятность для $k=2$ равна $1/n$. Дальше можно действовать индуктивно. Для произвольного k последняя точка обязательно должна оказаться в последней ячейке; это может случиться с вероятностью $1/n$. Потом мы можем поместить предпоследнюю точку в любую из свободных ячеек, скажем с номером m , сведя при этом задачу к случаю $(k-1)$ точек и $(n-m)$ ячеек. Выбор m ограничен сверху числом $(k-2)$, поскольку две точки — последняя и предпоследняя — уже на местах. У нас уже есть способ получить точное решение искомой задачи, но для этого нужно знать решения всех входящих в нее подзадач:

$$p_n(1) = 0,$$

$$p_n(2) = \frac{1}{n},$$

$$p_n(k) = \frac{1}{n} [p_{n-1}(k-1) + p_{n-2}(k-1) + \dots + p_{k-2}(k-1)].$$

Такое определение функции называется *рекуррентным*. Чтобы им можно было воспользоваться, необходимо знать решение некоторых *базовых подзадач*; в нашем случае это выражения для $k=0$ и 1 .

Полученное рекуррентное соотношение позволяет вычислить точное распределение, но его трудно анализировать. Нужно привести его в *конечную форму* — формулу, содержащую фиксированное конечное число арифметических действий над хорошо известными функциями. Мне удалось получить такую форму, оказавшуюся весьма компактной:

$$p_n(k) = \frac{1}{n!} S(n, k).$$

Здесь символ $S(n, k)$ обозначает так называемые *числа Стирлинга первого рода*. Они возникают в комбинаторике при подсчете циклических перестановок и в задачах о распределении рекордов*. По правде говоря, числа Стирлинга тоже вычисляются рекуррентным соотношением:

$$\begin{aligned} S(0, 0) &= 1, \\ S(0, k) &= S(n, 0) = 0, \\ S(n, k) &= (n-1)S(n-1, k) + S(n-1, k-1), \end{aligned}$$

но они используются уже с середины XVIII века, и достаточно широко, чтобы можно было счесть их «хорошо известными». А главное, известны свойства этих чисел, позволяющие анализировать полученное решение. Благодаря этому удалось вывести точные выражения для математического ожидания длины цепочек и ее дисперсии; собственно, ради вычисления этих значений я и исследовал получившееся распределение:

$$M[k] = H_n, \quad D[k] = H_n - H_{n,2}.$$

Эти величины выражаются через очень интересные *гармонические числа*: $H_n = 1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n}$, или в конечной форме $H_n = \frac{1}{n!} S(n+1, 2)$,

* Рекордной случайной величиной (или просто рекордом) в последовательности случайных величин называется величина, которая превосходит все предыдущие. Вероятность того, что среди n непрерывных случайных величин будет зарегистрировано k рекордов, описывается точно таким же выражением. Подробнее об этом можно прочесть в работе: Balakrishnan N., Nevzorov V. B. Stirling numbers and records // Advances in Combinatorial Methods and Applications to Probability and Statistics. Ed. N. Balakrishnan. Boston: Birkhauser, 1997. Pp. 189–200. Автор благодарит профессора Санкт-Петербургского государственного университета Валерия Борисовича Невзорова за любезно предоставленную информацию.

и $H_{n,2} = 1 + \frac{1}{2^2} + \frac{1}{3^2} + \dots + \frac{1}{n^2}$. Эти числа играют важную роль в такой неожиданно сложной области математики, как *теория чисел*.

Казалось бы, что может быть проще, чем изучение чисел, тем более целых? Арифметику проходят в школе; со свойствами чисел, такими как делимость, мы знакомимся на личном опыте, пытаюсь честно разделить пять рублей на троих. Но именно эта область математики ставит перед исследователем чрезвычайно сложные проблемы. Одна великая теорема Ферма чего стоит! От гармонических чисел дорожка ведет к *дзета-функции Римана*, а от нее — к великой загадке распределения простых чисел. Нам не потребуются результаты теории чисел явным образом, но свойства гармонических чисел мы используем. Средняя длина цепочек с ростом n растет очень медленно, хоть и неограниченно: имея бесконечное время, можно в среднем успеть сделать бесконечное число дел. Не сильно ошибившись, можно сказать, что она растет логарифмически. В свою очередь, дисперсия не сильно отличается от среднего, а добавочный коэффициент $H_{n,2}$ стремится к константе $\pi^2/6$. Немного позже нам пригодится это наблюдение.

На наш вопрос: «Какова вероятность не уложиться в n дней, имея перед собой k последовательных этапов выполнения задачи?» — поможет ответить функция распределения, то есть кумулятивная кривая для распределения Стирлинга. Построим такие кривые для $n = 7, 30, 365$ и $25\,000$, соответствующие неделе, месяцу, году и (конечно, условно) всей жизни (рис. 8.3).

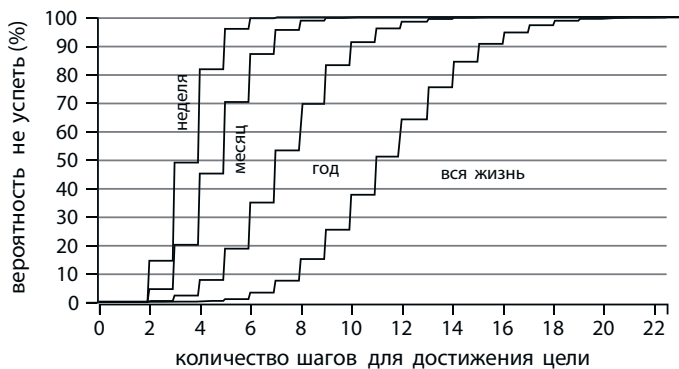


Рис. 8.3. Вероятность не успеть выполнить цепочки различной длины в тот или иной срок

Эти графики показывают, что вероятность не уложиться в месяц с заданием, состоящим из пяти шагов, превышает 80%. Неорганизованному балбесу на неделю лучше не планировать более трех дел, а десяток он не выполнит с вероятностью, превышающей 50%, и за всю жизнь! Мы убеждаемся в том, что при увеличении сроков на несколько порядков число дел, выполняемых как попало, увеличивается незначительно. Жизнь так коротка!

О методе пристального всматривания

Немного отвлекусь от основной темы и расскажу о том, как именно мне удалось перейти от рекуррентного соотношения к конечной форме распределения Стирлинга. Эта история может быть поучительной, особенно в свете нашей основной темы — законов подлости.

Повторюсь, что я не взаврадашний математик, а физик и вулканолог, использующий математику как инструмент. Но я этот свой инструмент очень люблю. Он красивый, изящный и мощный. Владение им делает меня счастливым и даже немного гордым от причастности к великим людям, создававшим его на протяжении столетий. Но при всем при том математика — инструмент, требующий особого к себе отношения. Она подобна породистой лошади или дорогому автомобилю, а то и легкомоторному самолету. Без умения, особого подхода и, если хотите, уважения к себе они испортятся и гордость от владения ими сменится горечью утраты. Конечно, я утрирую, но что-то в этом есть. Я имею в виду, что с математикой можно играть, а не только использовать в серьезной работе. Но в обоих случаях нужно как можно дольше оставаться настоящим математиком и ценить драгоценную точность и полноту результатов.

Я в принципе мог бы и остановиться, получив экспериментальную гистограмму, отражающую распределение числа последовательных дел, которые можно завершить в ограниченный срок. Это же скорее развлекательная книга, а не учебник и не научная статья. Но, поверьте, я просто не смог этого сделать: отсутствие точного решения

не давало мне покоя. Я готов был вообще выбросить этот эпизод из книги — и не потому, что не верил в точность результата, а потому что не считал это каким-то результатом. Я исписал множество листов, пытаясь вывести точную формулу, но ничего не выходило! Повторю, я не настоящий математик, у которого есть последовательное базовое математическое образование. Мне не доставало не инструментария или методик — я легко отыскивал их в учебниках и статьях. Но они заводили меня в дебри и тупики. Мне не хватало *интуиции* математика — той самой штуки, которая либо возникает от многих лет непрерывной работы, постоянного поиска внутренних связей и закономерностей, либо дается от рождения, примерами чего могут быть такие потрясающие люди, как Сриниваса Рамануджан Айенгор или Карл Фридрих Гаусс. Но большинство великих, замечательных и просто видных математиков были вооружены не врожденным талантом, а любовью к этой науке, предельной честностью перед собой и, главное, невероятным трудолюбием, благодаря которым их математическая интуиция превращалась в самую настоящую магию! И я убежден, что она доступна всем, но требует непрерывных упражнений: как говорили в моем родном Новосибирском государственном университете, «приседания мозгами». А силу для этих упражнений может дать только любовь. Ни чувство долга, ни страх провалить сессию, ни осознание полезности математики как инструмента не станут достаточной мотивацией для такой удивительно кропотливой, незаметной и чаще всего непрактичной работы.

Задача о проклятии режиссера вряд ли спасет чьи-то жизни или принесет мне славу и много денег, но без точного результата я чувствовал себя не вправе говорить о ней, поэтому я вновь и вновь выписывал столбцы известных мне точных значений функции вероятности (для $k = 1, 2$ и n), дополняя эмпирическими цифрами, приведенными к рациональному виду (мне быстро стало ясно, что нормировкой искомой функции будет $n!$), пытаюсь то угадать закономерность, то получить ее, подходя так или эдак. В конце концов решение пришло ко мне так же, как решения больших и чудовищно сложных задач приходят к настоящим математикам. Итогом моего пристального всматривания и вживания в ряды чисел стала искра интуиции. Блуждая уже практически бесцельно по страницам справочника комбинаторики,

я наткнулся на числа Стирлинга, о существовании которых до этого и не подозревал. Они происходят из совсем другой задачи и поначалу вызвали просто любопытство. Хорошо, что в справочнике приводились некоторые примеры рядов этих чисел. Мой взгляд выхватил знакомые цифры, и после недолгих проверок мне уже было ясно: мое распределение выражается через числа Стирлинга настолько просто и лаконично, что это стало настоящей наградой! Решение нашлось и, более того, оказалось удивительно простым и красивым! Но, конечно, и этого было мало. Совпадения чисел недостаточно для утверждения о том, что решение найдено. Однако, зная, что искать, я уже без труда смог строго свести рекуррентное соотношение для моего распределения к соотношению, определяющему числа Стирлинга, после чего задачу можно было считать решенной.

Мне очевидно, что это достаточно скромный результат, а специалисту по комбинаторике он, скорее всего, покажется простым упражнением. Но я могу им гордиться. После долгих упорных усилий и из моей волшебной палочки вылетели наконец искры и перышко взлетело на пару сантиметров над столом! Это значит, что я действительно делал все верно и когда искал решение, и, главное, когда не допускал возможности опубликовать простую эмпирику, претендуя на объяснение пусть даже шуточного эффекта. Я пишу эти строки не для того, чтобы похвастаться, а чтобы вдохновить тех, кто чувствует в себе настоящую любовь к математике, на долгий, кропотливый, но счастливый труд.

К законам подлости эти мои рассуждения имеют вот какое отношение. Метод пристального всматривания в расчете на интуицию работает только тогда, когда к волшебной палочке прилагается аналитический аппарат, который позволит строго проверить результат «озарения». В известной книге «Физики шутят» приводился анекдот о том, как строятся рассуждения представителей различных специальностей.

— Взгляни на этого математика, — сказал логик. — Он замечает, что первые девяносто девять чисел меньше сотни, и отсюда с помощью того, что он называет индукцией, заключает, что любые числа меньше сотни.

— Физик верит, — сказал математик, — что 60 делится на все числа. Он замечает, что 60 делится на 1, 2, 3, 4, 5 и 6. Он проверяет несколько других чисел, например 10, 20 и 30, взятых, как он говорит, наугад. Поскольку 60 делится на них, он считает экспериментальные данные достаточными.

— Да, но взгляни на инженера, — возразил физик. — Он подозревает, что все нечетные числа простые. Во всяком случае, 1 можно рассматривать как простое число, доказывает он. Затем идут 3, 5 и 7, все, несомненно, простые. Затем идет 9 — досадный случай; по-видимому, 9 не является простым числом. Но 11 и 13, конечно, простые. Возвратимся к 9, — говорит он, — я заключаю, что 9 должно быть ошибкой эксперимента*.

Это забавно, конечно, но вот вам такой числовой ряд:

$$1, 2, 4, 8, 16, \dots$$

Продолжите его. «Это же, очевидно, степени двойки! — воскликнете вы. — Следующим числом будет 32, а за ним 64 и т. д.». Но что, если я скажу вам, что следующим должно быть 31? И это не степени двойки, а значения вот такого выражения:

$$f_n = \frac{1}{24} (n^4 - 6n^3 + 23n^2 - 18n + 24) = \sum_{k=0}^4 C_n^k.$$

При $n = 0, 1, 2, 3, \dots$ здесь под знаком суммы стоит биномиальный коэффициент. Первые тринадцать членов этого ряда выглядят так:

$$1, 2, 4, 8, 16, 31, 57, 99, 163, 256, 386, 562, 794, \dots$$

Приведенное мною выражение дает число областей, на которые разбивается круг, если расположить на его окружности n различных точек и соединить их каждую с каждой**. И эта простая и аб-

* Пойа Д. Математика и правдоподобные рассуждения. М.: Издательство иностранной литературы, 1957.

** «Магическое число» 4 возникает в сумме биномиальных коэффициентов из того обстоятельства, что любым четверем из этих n точек на окружности соответствует одна точка внутри круга, в которой должны пересечься соединяющие их отрезки. Далее вывод строится на знаменитой формуле Эйлера, связывающей число узлов и ребер некоторого планарного графа с числом областей, на которые он разбивает конечную область (например, сферу).

солютно понятная задача имеет столь коварную «подсказку»! Ведь на проверку даже первых пяти чисел уже должно уйти достаточно много времени, чтобы заключить, что число областей выражается степенью двойки. Ну а если упорство возобладает, то подсчет областей при $n = 6$ неизбежно вызовет недоумение и поиск ошибки в подсчете, ведь 31 так близко к 32 (попробуйте сами нарисовать и сосчитать эти области). Забавно то, что десятый член ряда опять равен степени двойки. Понять, откуда эти степени взялись и почему ряд начинается столь многообещающе, поможет хорошо известный арифметический треугольник, или треугольник Паскаля. Его элементы — биномиальные коэффициенты, а сумма всех чисел каждого ряда в точности равна степени двойки (это обстоятельство используется для нормировки функции вероятности биномиального распределения). Поскольку число областей, на которые разбивается круг, выражается суммой пяти первых биномиальных коэффициентов (на рис. 8.4 они выделены черным цветом), первые пять таких сумм содержат в себе полные ряды в треугольнике, однако начиная с шестого ряда суммирование идет не по всем коэффициентам. Отсюда и взялось «коварное» число 31. В десятом же ряду первые пять коэффициентов составляют ровно половину ряда, общая сумма которого равна степени двойки (2^9), и, значит, половина тоже будет степенью двойки. Если где-то еще они и встретятся, то это уже будет случайным совпадением.

				1									
				1	1								
				1	2	1							
				1	3	3	1						
				1	4	6	4	1					
				1	5	10	10	5	1				
				1	6	15	20	15	6	1			
				1	7	21	35	35	21	7	1		
				1	8	28	56	70	56	28	8	1	
				1	9	36	84	126	126	84	36	9	1

Рис. 8.4. Треугольник Паскаля

Ричард Ги из Университета Калгари в 1988 году опубликовал статью, озаглавленную «Сильный закон малых чисел»*, в которой приводит и этот пример (с полным доказательством), и теорему, достойную иных законов подлости:

Просто посмотреть недостаточно.

В ней есть еще более трех десятков примеров последовательностей и «фактов», которые выглядят многообещающими, но никак не могут быть законами.

Мне очень понравился такой пример: при использовании знаменитого метода Евклида для доказательства бесконечности ряда простых чисел последние получаются не всегда. Здесь речь о том, что, предположив конечность ряда простых чисел, мы можем вычислить произведение всех членов этого ряда, увеличить его на единицу и получить число, превышающее все имеющиеся, но не делящееся ни на одно из них. Можно подумать, что произведение нескольких первых простых чисел, увеличенное на единицу, всегда порождает простое число, и убедиться в этом на нескольких примерах.

$$2+1=3$$

$$(2\times 3)+1=7$$

$$(2\times 3\times 5)+1=31$$

$$(2\times 3\times 5\times 7)+1=211$$

$$(2\times 3\times 5\times 7\times 11)+1=2311$$

$$(2\times 3\times 5\times 7\times 11\times 13)+1=30031=59\times 509.$$

Последний, да и последующие примеры дают осечку! Получается, доказательство Евклида неверно? Нет, оно совершенно справедливо, поскольку ничего не говорит о простоте результата, но утверждает существование числа, не делящегося ни на одно из полного (по нашему предположению) множества простых чисел. Число 30 031 и вправду не делится ни на одно из перемножаемых чисел.

* Guy R. K. The Strong Law of Small Numbers // Amer. Math. Monthly. 1988. Vol. 95.

Позже, в 1990 году, тот же Ричард Ги выпустил в свет еще одну статью «Второй сильный закон малых чисел»*, в которой приводит еще полсотни примеров последовательностей, ломающих интуицию математика!

Воспетая мной математическая интуиция без строгого доказательства может сыграть злую шутку. Более того, и в строгое, но очень сложное доказательство может вкрасться незаметная коварная ошибка, чему есть множество примеров. Обязательно прочтите чудесную книгу «Великая теорема Ферма» Саймона Сингха, чтобы почувствовать, с какими поистине циклопическими законами подлости приходится иметь дело в большой математике. Но удивительное дело: именно эти примеры и рассказы вдохновляют меня на добросовестный поиск математической истины там, где вполне хватило бы наблюдения или приблизительного результата.

Быстрее, еще быстрее!

Давайте теперь исследуем само явление цейтнота и его выматывающие свойства. Для этого обратимся к методу Монте-Карло и построим несколько тысяч стохастических цепочек, после чего усредним их, получив некую гладкую функцию. Она показана сплошной линией на рис. 8.5 и представляет собой математическое ожидание случайной функции, описывающей наш нестационарный стохастический процесс. Назовем эту случайную функцию *темпом выполнения работы*.

В предыдущей главе мы говорили о таких функциях, рассматривая очень простой случай стационарных процессов с неизменной интенсивностью. Сейчас же мы видим иную картину. Наша функция имеет переменную дисперсию, уменьшающуюся ближе к дедлайну. Это говорит о том, что последовательности, порождаемые случайной функцией, при приближении к правому краю сливаются и становятся неотличимы друг от друга.

Обратите внимание на то, что оси графика приведены к общему числу дел и всему отпущенному времени. Это, с одной стороны,

* Guy R. K. The Second Strong Law of Small Numbers // Mathematics Magazine. 1990. Vol. 63.

позволяет нам сравнивать как разные сроки, так и различные по длине цепочки, а с другой — мы опять получили что-то подобное кривой Лоренца: некое формализованное отражение несправедливости.

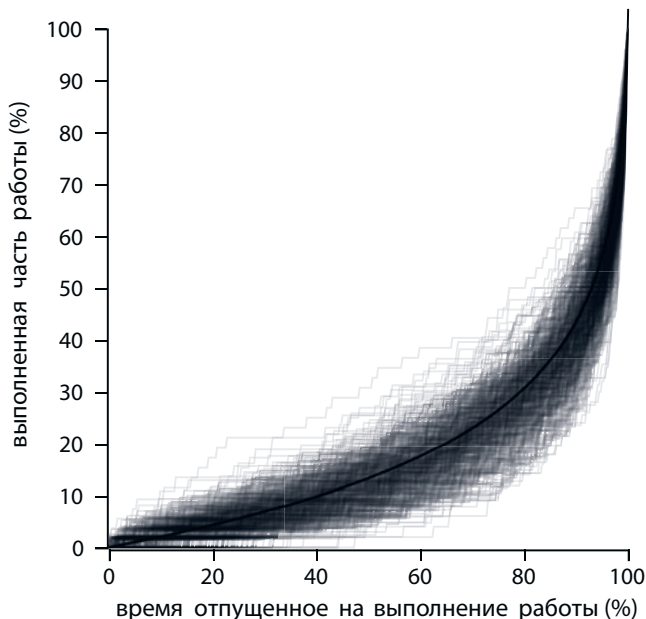


Рис. 8.5. Множество стохастических цепочек с дедлайном и ожидаемый темп выполнения работы

Наблюдаемый темп, увы, очень неравномерен: в первую половину срока будет сделано едва ли 10% работы, а добрую половину всех дел придется выполнять, имея в распоряжении менее 10% времени. Но главная особенность: темп, вернее его наклон, стремительно увеличивается при приближении к дедлайну! Мы получили модель предновогоднего ража или паники в преддверии годового отчета, а также нащупали закон подлости, знакомый всякому, кому приходилось организовывать концерт, костюмированный вечер или иное мероприятие:

Сколько бы времени ни было отпущено на подготовку мероприятия, большая часть дел останется на последнюю ночь!

Прекрасные живые примеры таких процессов описаны, например, в рассказах Карела Чапека «Как делают газету» и «Как ставится пьеса». Неужели причина этого проклятия кроется только в нашей неорганизованности и безалаберности? Это, конечно, основные причины, но мы не настолько в них виноваты, чтобы нельзя было попробовать оправдаться каким-нибудь математическим законом. Стратегия балбеса, конечно, выглядит глупо, но взрывной рост темпа — это не шутки! Можно ли вообще с ним справиться?

Имея в распоряжении функцию вероятности для распределения Стирлинга, ожидаемый темп выполнения работы можно вычислить точно. Формула не слишком изящна, однако примечательно, что в нее входит число дней n и не входит число запланированных дел:

$$T_n(x) = -\frac{1}{H_n + 1} \log_2 \left[1 - x(1 - 2^{-H_n - 1}) \right].$$

Логарифм — функция медленная, если только его не прижать к стенке. В последние дни перед дедлайном темп растет катастрофически — с такой же скоростью, с которой логарифм проваливается в бездну при приближении к нулю. Однако от числа выделенных дней он все же зависит. Можно посмотреть, как выглядит ожидаемый темп для недели, месяца и года (рис. 8.6).

Обратите внимание на то, что жесткое ограничение по времени благотворно. Имея в запасе всего неделю, мы, скорее всего, станем выполнять работу равномернее (к середине срока будет готова треть), а если впереди целый год, то можно и расслабиться, а потом об этом пожалеть. У идеального исполнителя-перфекциониста, который выполняет работу равномерно, темп соответствует диагонали (пунктирная линия на рисунке). Это похоже на кривую равенства на диаграмме Лоренца, знаменующую справедливость. Подобно тому как мы вычисляли коэффициент Джини для диаграммы Лоренца, мы можем, основываясь на площади между кривой темпа выполнения работ и идеальной кривой, определить некий коэффициент подлости, который покажет, насколько мы далеки от идеала. Он зависит от длины выделенного срока и потихоньку увеличивается с ростом n .

В приведенных нами примерах для недели, месяца и года коэффициент подлости равен соответственно 0,37, 0,49 и 0,63. Этот индекс увеличивается с ростом n очень медленно, но если устремить число дней к бесконечности, он будет стремиться к единице. Итак, мы приходим к парадоксальному, но по-своему красивому результату: имея в распоряжении бесконечное время, балбес может запланировать бесконечное число дел, однако ожидаемый темп выполнения будет почти всюду равен нулю. Это значит, что *почти наверняка* он не выполнит ничего из запланированного, отложив все дела на бесконечное будущее! Вспоминаются привычные сетования: «Целое лето (каникулы, жизнь) пролетело, а я так ничего и не успел!» Что ж, даже этому есть математическое объяснение.

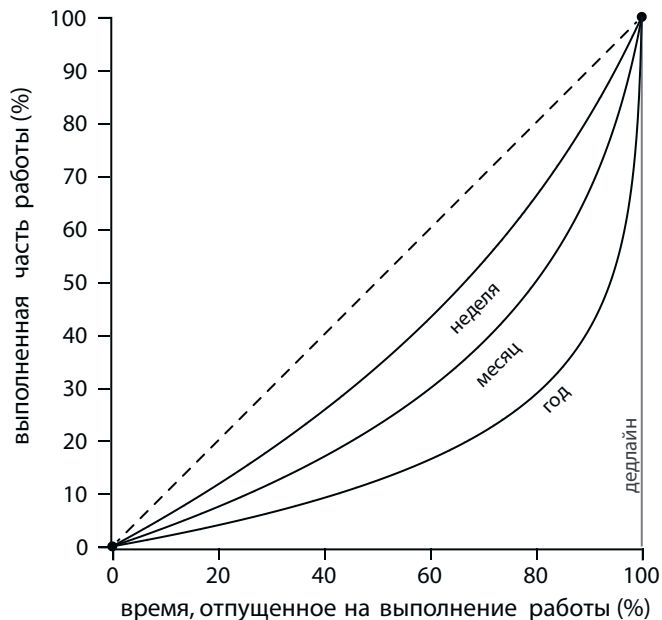


Рис. 8.6. Наиболее вероятный темп выполнения работы в ограниченный срок

Даосы в Древнем Китае крепко размышляли о вечной жизни, причем очень грамотно: наряду с упражнениями тела, необходимыми для решения такой задачи, они занимались упражнениями ума, чтобы приспособить его к вечному существованию, породив

оригинальную и интересную философию. Как видно, вечная жизнь требует большой дисциплины, иначе даже вечность — весьма вероятно — можно потратить впустую.

Мостим дорогу благими намерениями

Но как же бороться с нарастающей волной забот и цейтнотом? Например, взять себя в руки. Человек с синдромом отличника может стремиться выполнить следующее дело как можно раньше, насколько это возможно, конечно. Правдоподобной моделью будет выбор момента для выполнения следующего дела, следуя экспоненциальному распределению с интенсивностью, обратно пропорциональной оставшемуся времени. Это не исключит некоторой неопределенности, присущей нашей жизни, но выразит благие стремления делать всё как можно скорее. Назовем эту стратегию *стратегией благих намерений*. Вот какими будут распределения вероятностей выполнения заданий в срок для ее приверженца, который в половине случаев сделает очередное дело в первую четверть оставшегося времени (рис. 8.7).

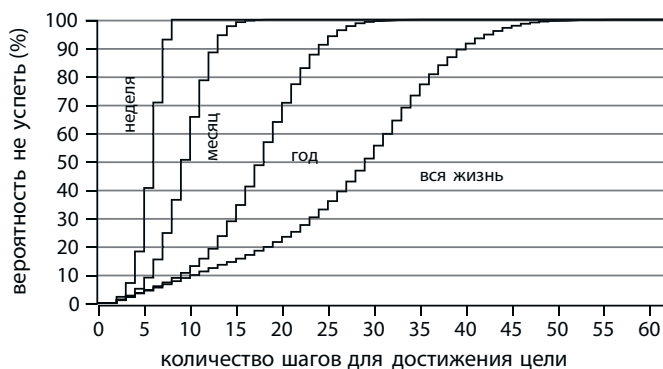


Рис. 8.7. Распределение вероятности не успеть в срок для стратегии благих намерений

Что же, существенно лучше! В течение недели можно с неплохой вероятностью успеть сделать пять дел и оставить себе два выходных

дня. Но все же для больших периодов увеличение возможностей не революционное. Проблема в том, что ожидаемое число успешно завершаемых дел все равно остается пропорциональным логарифму отпущенного времени, а логарифм растет крайне медленно! Так что, планируя многое, нужно иметь в виду, что интенсивность процесса будет неизбежно возрастать, а времени в преддверии дедлайна, скорее всего, станет не хватать. В любом случае необходимо помнить, что жизнь коротка. Чтобы успеть реализовать задуманное, нужно действовать прямо сейчас!

Полюбуемся на математическое ожидание темпа благонамеренного отличника (рис. 8.8).

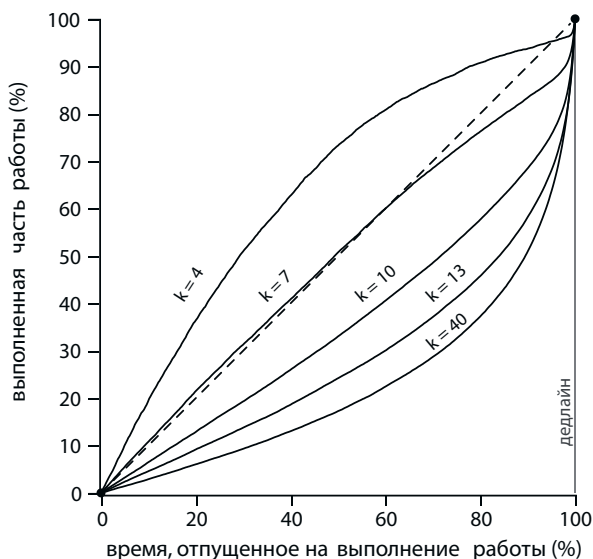


Рис. 8.8. Ожидаемый темп выполнения работы методичным человеком, старающимся приступить к следующему этапу работы как можно скорее. Число k показывает количество запланированных задач

Нашему аккуратисту удалось более равномерно распределить работу и выполнить существенно больше дел, но его все равно ожидает цейтнот. Короткие цепочки такой человек будет реализовывать с существенным перевыполнением плана, а цепочку из семи дел — практически идеально. Однако по мере увеличения числа

дел ожидаемый темп быстро стремится к теоретическому, полученному с помощью стратегии балбеса! Увеличилась общая производительность, но запарка перед самым дедлайном никуда не делась. Так что нагрузкой можно доконать и заправского зануду!

Впрочем, существует еще один широко известный способ внести дисциплину в выполнение работ: *вместо одного дедлайна сделать много*. Разобьем срок выполнения работы на две равные части и будем придерживаться этого нового дедлайна, считая его, скажем, промежуточным отчетом. Для каждой из этих частей мы можем построить кривую ожидаемого темпа выполнения работ, как показано на рис. 8.9.

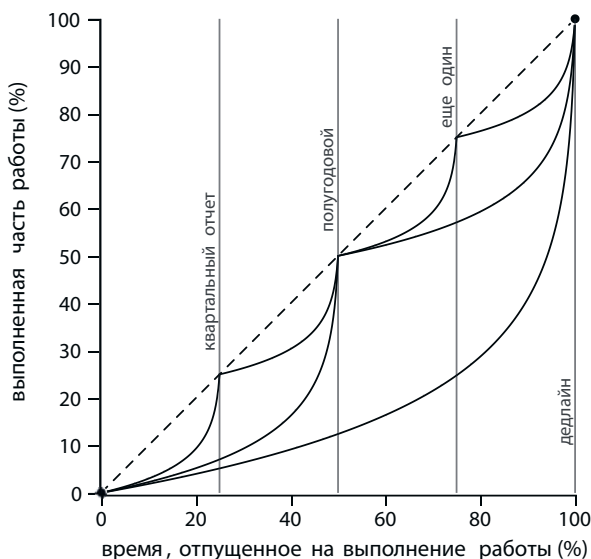


Рис. 8.9. Разбиение времени выполнения на несколько промежуточных отчетных периодов позволяет сделать работу более равномерно, но добавляет стресс при приближении каждого нового отчета

Несмотря на нервотрепку с промежуточным отчетом, мы достигли своей цели: площадь между общей кривой темпа выполнения и диагональю сократилась, и коэффициент подлости уменьшился с 0,65 до 0,3. Кроме того, сокращение срока (вместе с сокращением числа дел, разумеется) приближает ожидаемый темп выполнения

работы к идеальному, поэтому коэффициент подлости уменьшился более чем в два раза. Добавление еще двух, скажем, квартальных отчетов уменьшит его уже до 0,13, но тем самым мы вгоним наших исполнителей сразу в четыре стрессовых периода, и они все равно станут громко страдать, жалуясь на судьбу и руководство! Что же, мы можем показать работникам наши выкладки и доказать, что, введя ежеквартальную отчетность, в пять раз понизили коэффициент подлости их жизни — если это, конечно, станет им утешением. Более того, при стремлении количества промежуточных дедлайнов к числу дней, отпущенных на работу, темп приблизится к идеальному, но очень занудному.

Ну вот! Еще и принтер сломался!

Добавим еще пару слов о стратегии балбеса. Числа Стирлинга при увеличении n имеют асимптотическое разложение, которое сводит распределение длин цепочек с дедлайном к смещенному распределению Пуассона (рис. 8.10).

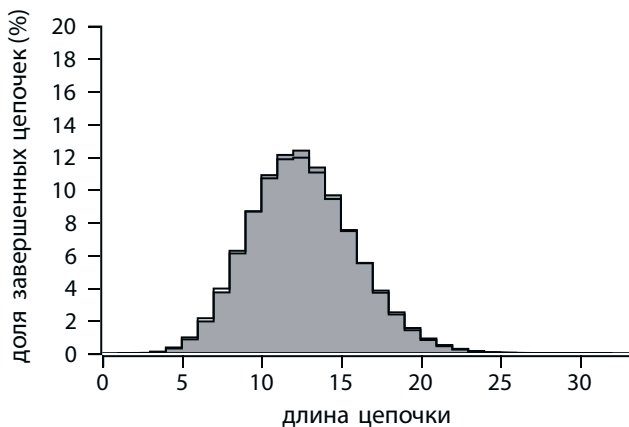


Рис. 8.10. Распределение Стирлинга (гистограмма) и Пуассона (ступеньки) для $n = 100\,000$ становятся очень близки друг к другу

Таким образом, наш стохастический процесс с дедлайном можно рассматривать либо как пуассоновский на сгущающейся временной

сетке, либо как неоднородный пуассоновский, интенсивность которого монотонно и стремительно растет. И хотя, строго говоря, наш процесс не пуассоновский, поскольку события в нем не независимы, нужные нам статистические свойства у них схожи. Об этом говорит и подмеченная ранее близость среднего значения и дисперсии распределения Стирлинга, характерная именно для пуассоновского распределения.

Этот вывод позволяет задать вопрос. Что, если добавить к построенному нами процессу выполнения цепочки дел какие-либо не зависящие от нас редкие неприятности: пургу, жуткую пробку, насморк, поломку принтера или всенародный праздник?

Для пуассоновского процесса определен *процесс случайного прореживания*, заключающийся в удалении событий из потока с какой-то известной вероятностью. Случайное прореживание с вероятностью $(1 - p)$ оставляет процесс пуассоновским, но его интенсивность уменьшается, умножаясь на p . События, соответствующие совпадению неприятности и какого-либо этапа выполнения работы, сами образуют пуассоновский процесс — с существенно меньшей интенсивностью, но в нашем случае также монотонно и стремительно растущей. Так стремительно, что, какой бы малой ни была вероятность неприятности, для достаточно большого числа дел (или срока, отведенного на работу) ближе к дедлайну она может увеличиться до вполне наблюдаемой. И принтер забарахлит именно накануне сдачи курсовика! Разумеется, это работает для достаточно длинных цепочек.

Не удивляйтесь, если автобус сломается именно тогда, когда вы уже опаздываете. Он не желает вам зла. Просто если вы девушка, то последовательность дел: выбрать платье, съесть конфетку, умыться, надеть выбранное платье, накраситься, надеть цепочку, переложить вещи из сумочки в клатч, почистить туфли и прочее... подходит к самому главному и волнительному дедлайну — к свиданию! И темп, с которым вы летите навстречу судьбе, уже настолько сумасшедший, что начинают происходить самые маловероятные чудеса. В конце концов, а что же такое чудо, как не реализация невероятного!

Глава 9

Термодинамика классового неравенства

*Среди экономистов реальный мир зачастую
считается частным случаем.*

Наблюдение Хонгрена

Современная экономика — большая, серьезная, но своеобразная наука. Несомненно, она жизненно необходима как дисциплина, изучающая реальное и важное явление нашего мира — *экономическую действительность*. Она стремится к доказуемости и формализации, в ней много математики, подчас сложной и интересной. Однако, открыв серьезный экономический учебник, вы, скорее всего, обнаружите какие-то сравнительно несложные выкладки, готовые рецепты и тонны неформальных рассуждений в таком духе: «...но на самом деле все может быть не так и вообще как угодно, если на то будет воля ключевых игроков или правительства». В конце концов порой складывается ощущение, что в этой дисциплине интуиция, знание психологии и умение воспринимать общий контекст важнее, чем точный расчет и скрупулезное рассмотрение деталей (речь об экономике, а не о бухгалтерии). Наконец, в наше время почти половина липовых диссертаций пишется именно по экономике, а значит, не так уж и сложно научно обоснованно рассуждать на подобные темы. Попробуем и мы свои силы на этом поприще, благо

нигде так остро не воспринимается несправедливость этого мира, как в вопросе распределения богатства. К тому же чем бы ни занимался человек, какой бы профессией ни владел, он вовлечен в экономику и ее игры. От ее законов, как и от законов физики или математики, никуда не спрятаться.

Из всей массы задач, решаемых математической экономикой, мы рассмотрим лишь одну — как выходит так, что даже при равных условиях для всех участников рынка и справедливом обмене средствами бедных становится больше, чем богатых, и почему даже идеальное математическое общество склонно к финансовому неравенству. Ну и, конечно, узнаем кое-что новое и полезное о распределениях случайных величин.

Как говорить об экономике?

На протяжении всей книги мы задаем себе одни и те же вопросы. Как рассуждать о том или ином предмете, чтобы наши слова имели смысл? Какую математическую структуру стоит использовать для моделирования интересующего нас объекта?

Я физик по образованию и по профессии. Моя профессиональная деформация выражается в своеобразном взгляде на мир как на множество разнообразных физических систем и процессов. С точки зрения физика, реальный рынок — *существенно нестационарная открытая система со множеством степеней свободы*, в которой важную роль играют стохастические (случайные) процессы. В этом смысле он похож на предмет изучения таких разделов физики, как *термодинамика* и *статистическая физика*, в которых, ввиду невозможности рассмотреть всё неисчислимое количество деталей и поведение всех составляющих частей системы, переходят к обобщающим и измеримым ее свойствам, таким как *энергия*, *температура* или *давление*. Неудивительно, что попытки термодинамического описания экономических систем и создания такой смежной дисциплины, как *эконофизика*, предпринимаются уже более ста лет. Но вот беда: пока ученые рассматривают детали, обобщают полученные знания и ведут споры о фундаментальных законах, основной объект изучения — экономическая действительность — успевает

поменяться до неузнаваемости. Ее поведение как будто стремится сохранить, а то и увеличить свои неопределенность и непредсказуемость.

Хорошим примером служит двухвековая история использования технического анализа при игре на фондовой бирже. Когда появляется новый мощный инструмент, позволяющий нащупать скрытые закономерности и предсказать курс ценной бумаги или акции, он начинает приносить прибыль тем, кто его использует. Но вскоре рынок «чувствует» новых игроков и подстраивается под их стратегию, тогда точность предсказаний нового замечательного метода падает. Спустя какое-то время он попадает в длинный список устаревших и не слишком надежных инструментов. Ни современные гибкие самообучающиеся нейросетевые алгоритмы, ни сверхскоростные роботы-трейдеры, совершающие миллионы операций в минуту, не поменяли за минувшие два десятилетия основное свойство биржевой игры — ее непредсказуемость. И до сих пор основными достоинствами профессионала в этой отрасли остаются воля, выдержка характера, несклонность к азарту... ну или владение биржей. Всё как в казино, где игры основаны на чистой случайности! С одной стороны, это, конечно, обидно, а с другой — дает повод постоянно совершенствовать методы и подходы. Когда-то и теория вероятностей, и математическая статистика родились из попыток анализа азартных и экономических игр. Только потом они нашли применение практически во всех естественных науках.

Итак, вслед за физиками мы будем моделировать экономическую действительность *макросистемой* — ансамблем взаимодействующих частиц, обменивающихся ценностями. В дальнейших рассуждениях под ценностями мы будем иметь в виду деньги, но даже эта привычная повседневно используемая категория на удивление сложна и неоднозначна. Смысл и ценность денег зависят от множества факторов: называя вне контекста некую сумму, мы ничего не говорим о ее реальной ценности. Это отличает денежные величины от большинства физических, описывающих наш мир, и мешает проводить строгие рассуждения в экономике. Но цель нашего разговора — математические основы законов подлости, повседневных,

понятных и простых. Именно поэтому в дальнейшем мы будем говорить о неких «рублях», имея в виду формальный билетик или монетку и подразумевая, что чем больше этих «рублей» у кого-то, тем он «богаче». Прочие же рассуждения о покупательской способности, нематериальных или неликвидных ценностях, наконец, о «не в деньгах счастье» мы оставим за рамками разговора.

Подходите, всем хватит!

Начнем мы с того, что станем раздавать деньги некой конечной группе людей и сравним между собой справедливость различных способов это сделать. И наконец-то мы станем применять кривую Лоренца и индекс Джини в экономическом контексте — именно так, как это было задумано их создателями!

Первая, самая очевидная стратегия: «взять всё, да и поделить», выделить каждому члену группы по равной доле общей суммы, скажем по 100 рублей. Такое распределение называется *вырожденным*, оно имеет индекс Джини, равный нулю, и соответствует кривой равенства на диаграмме Лоренца (рис. 9.1).

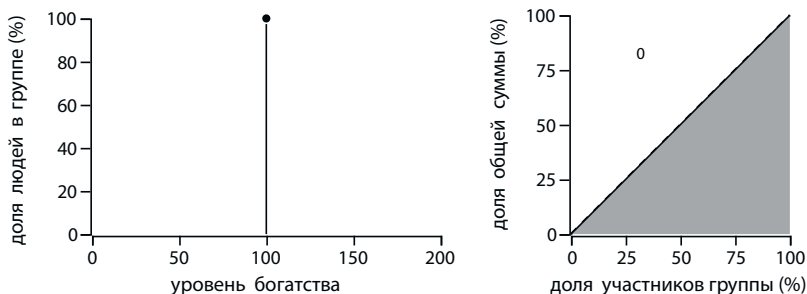


Рис. 9.1. Абсолютно справедливое вырожденное распределение денег: у всех поровну. Кривая Лоренца совпадает с кривой равенства, а число 0 показывает индекс Джини

Прекрасный вариант! Назовем его «стратегией Шарикова» в честь героя повести Михаила Булгакова «Собачье сердце», который именно таким способом предлагал решить все экономические вопросы молодой советской республики.

Вторая стратегия, несколько более реалистичная, заключается в многократной раздаче всем по одному рублю в случайном порядке. Кому как повезет. Можем назвать эту стратегию пуассоновской, поскольку именно так распределяются по временной шкале независимые случайные события в процессе Пуассона. Для группы из n человек вероятность каждого из участников получить рубль составляет $1/n$. После раздачи таким образом M рублей каждый должен получить сумму, равную количеству таких «положительных» исходов. Функция вероятности для подобной суммы хорошо известна — это *биномиальное распределение*, похожее на колокол, который симметрично разбегается вокруг среднего значения $m = M/n$. Обычно студенты знакомятся с этим распределением на примере вычисления вероятности получить указанную сумму при бросании игральных костей. В нашем случае мы бросаем честную кость с n гранями M раз. Для больших значений биномиальное распределение становится практически неотличимым от нормального (рис. 9.2).

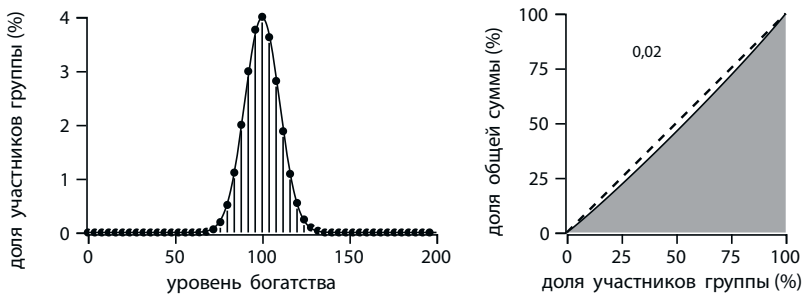


Рис. 9.2. Результат раздачи денег по принципу «на кого бог пошлет» — биномиальное распределение. Чем больше денег мы раздаем, тем больше кривая Лоренца приближается к кривой равенства. Здесь $M = 10\,000$, $n = 100$

Это распределение с точки зрения справедливости выглядит очень неплохо; более того, оно становится тем справедливее, чем больше денег мы раздаем публике! Просто замечательно! Жаль, что общество устроено не так и денежный дождь не сыплется на всех нас поровну.

Для полноты картины рассмотрим еще одно простое искусственное распределение денег — такое, чтобы в группе были как бедные, так и богатые, и чтобы вероятность иметь тот или иной достаток была одинакова для всех уровней достатка (рис. 9.3). Иными словами, чтобы распределение оказалось *равномерным*. При этом мы вынуждены ввести ограничение на максимальный уровень достатка для участника группы. Думаю, затей мы социологический опрос, многие респонденты с улицы согласились бы, что это звучит справедливо.

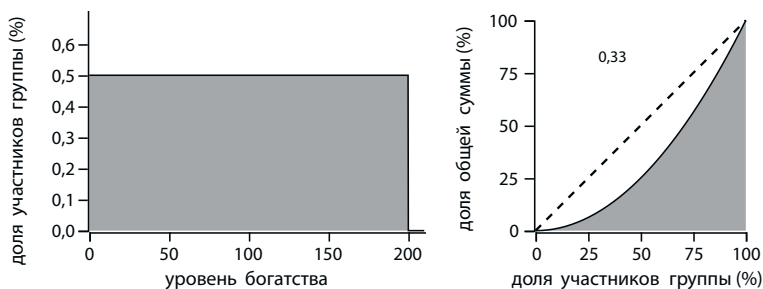


Рис. 9.3. Равномерное распределение не означает, что деньги распределяются всем равномерно. При таком распределении число богатых, бедных и середнячков одинаково, но деньги в основном принадлежат богатым: половина всех средств сосредоточена лишь у четверти группы

Однако кривая Лоренца показывает, что такое распределение уже далеко от справедливости. Для равномерного распределения она представляет собой квадратичную параболу. Если левая граница распределения равна 0, как в нашем случае, то из-за нормировки параболы становится независимой от положения правой границы. Иными словами, для всех равномерных распределений с нулевой левой границей она будет одинаковой, и индекс Джини для всех таких распределений равен в точности $1/3$. Такое значение индекса (но не такое же распределение!) было, например, у экономики Австралии в 2000-е — это вполне неплохой показатель, но далекий от совершенства.

Рассмотренные нами способы распределить средства по группе людей очень просты и вполне естественны. Но может

возникнуть вопрос: а смогут ли они как-нибудь реализоваться в жизни? Насколько сами эти распределения вероятны? Ведь рынок есть рынок: если дать людям волю обмениваться деньгами, менять их на услуги, копить их и проматывать в одну ночь, смогут ли эти идеальные распределения сохранить устойчивость? Не превратятся ли они в какие-нибудь другие? Что нужно сделать с рынком, чтобы он сам, без принудительной раздачи средств, приблизился, например, к биномиальному или нормальному распределению, очень привлекательному с точки зрения справедливости?

Мы уже встречались с такой постановкой вопроса, говоря о центральной предельной теореме — одной из основ математической статистики. Согласно этой теореме, распределение для суммы одинаково распределенных случайных величин стремится к нормальному независимо от распределения этих величин. Таким образом, можно сделать вывод, что нормальное распределение и будет наиболее вероятным и устойчивым. Мы уже говорили, что оно соответствует минимальной информации о случайной величине, а раздавая деньги всем без дополнительных условий, мы и получили распределение, неотличимое от нормального. Так что, возможно, и в реальных обществах должно наблюдаться такое распределение богатства? Почему же индекс Джини для большинства государств, считающихся весьма успешными, почти никогда не бывает ниже 0,25, а для всего мира он близок к 0,4? Откуда берется столь существенное неравенство? Кто мешает наступлению устойчивого золотого века? Неужели это заговор богачей или непреодолимая жадность человека?!

Мы привыкли судить о роде человеческом плохо, упрекать его в стяжательстве и прочих грехах, но сейчас я хочу выступить в роли адвоката и показать, что греховность людей тут ни при чем. Все дело в математических законах, которым подчиняются не только слабый смертный, но и бесстрастная физика. Если бы не мысль и не воля человека разумного, придумавшего и внедрившего ряд рыночных механизмов, получить экономическую систему с индексом Джини меньше 0,5 было бы крайне непросто. Именно ради поиска фундаментальных законов экономики и создавалась эконофизика.

Чтобы немного разобраться в них, нам предстоит погрузить нашу группу испытуемых в модель рынка.

Новая экономическая политика

Вновь рассмотрим группу из n человек и раздадим всем участникам эксперимента по равной денежной сумме — по t рублей каждому, получив самое справедливое в мире шариковское распределение средств в обществе. После раздачи в нашей системе будет находиться $M = nt$ денежных единиц. Теперь предоставим им свободу богатеть и беднеть по воле собственной судьбы, согласно следующей примитивной модели рынка. Попросим кого-нибудь, выбранного случайно, отдать один рубль любому человеку из группы, также выбранному случайно. Можно счесть это приобретением некой услуги по фиксированной цене $\Delta t = 1$. Распределение богатства ожидаемо изменится: у кого-то денег станет меньше, у кого-то больше. Станем повторять эту процедуру снова и снова и посмотрим, как будет изменяться распределение богатства в группе.

Пусть вас не смущает нереалистичная примитивность описанной нами модели. Ее достоинство в том, что она требует минимальной априорной информации и соответствует некоторой базовой системе. Если мы обнаружим какие-то закономерности на этой модели, то они проявятся и в более сложных моделях.

Разумно перед проведением эксперимента поразмыслить, что же мы ожидаем увидеть. Получение денег участниками происходит равновероятно, как в случае пуассоновской стратегии раздачи, но в то же время игроки и теряют деньги, причем по такому же пуассоновскому принципу и с той же интенсивностью. Если вместо одного шага мы будем рассматривать сразу сотню, то вместо фиксированного количества денег участники группы будут обмениваться какими-то случайными суммами. Из опыта с пуассоновской раздачей денег следует заключить, что как положительные, так и отрицательные приращения будут распределены практически нормально и расположены симметрично относительно нуля. Каждый игрок в итоге будет получать разность этих приращений, которая для

двух нормально распределенных случайных величин будет тоже нормально распределена*, в данном случае вокруг нуля, поскольку потери и выигрыши симметричны (рис. 9.4).

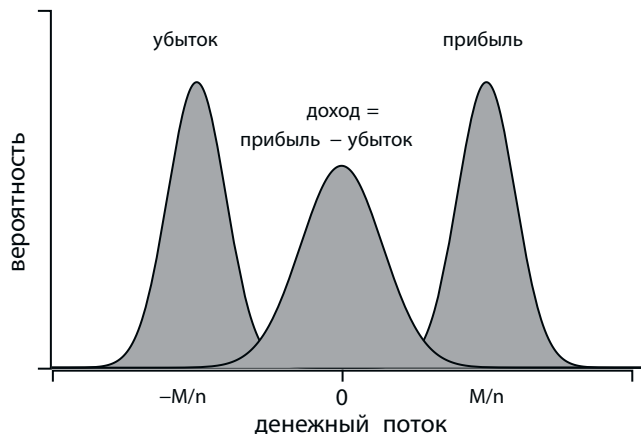


Рис. 9.4. После множества обменов каждый игрок получит и потеряет суммы, которые подчиняются распределению, близкому к нормальному. Суммарный доход также будет нормально распределен вокруг нуля

Таким образом, мы получаем классическое случайное блуждание с нормально распределенными приращениями. Нам уже знаком этот процесс, окрашивающий жизнь в темные и светлые полосы. Поведение множества случайно блуждающих частиц подобно диффузии: их плотность будет расплываться гауссовым колоколом вокруг неизменного среднего значения, увеличивая дисперсию пропорционально квадратному корню из числа обменов (времени). Вроде бы все просто. Если нет каких-то механизмов, сдерживающих эту диффузию, колокол расплывется по всей числовой оси. Таким же образом диффузия выравнивает неоднородности концентрации веществ в некотором замкнутом объеме или теплообмен распределяет температуру в изначально неравномерно нагретом стержне.

* То, что сумма или разность нормально распределенных случайных величин тоже будет подчиняться нормальному распределению, называется *устойчивостью* этого распределения. О смысле и ценности этого понятия мы поговорим чуть позже.

Но есть нюанс. Если по каким-то причинам у кого-либо из группы не осталось средств, он не сможет приобретать услуги, отдавая деньги, но по-прежнему может получать их. Возможное значение благосостояния ограничено слева нулем, а это значит, что диффузия богатства не сможет распространяться во все стороны бесконечно и наблюдаемая функция вероятности рано или поздно перестанет быть симметричной.

Есть еще один нюанс. Количество денег в нашей замкнутой системе ограничено и неизменно; это значит, что случайные блуждания не независимы. Какой-нибудь везучий игрок сможет получить очень большие суммы и уйти от ансамбля очень далеко, но только если общая масса настолько же обеднеет. Участников эксперимента стягивает невидимой сетью закон сохранения денежной массы в системе. К чему же будет стремиться распределение богатства в таких условиях? Похоже, ответ не столь очевиден, как может показаться на первый взгляд. Обратимся к имитационному моделированию и посмотрим, что у нас получится.

Для любопытных читателей, которые захотят сами провести этот эксперимент, приведу алгоритм процесса перераспределения денег для фиксированного Δt , равного для всех участников группы (рис. 9.5).

Исходные данные: xs — массив из n элементов, инициализированный значениями m .

Повторять для каждого i от 0 до n

если $xs[i] > 0$

$j \leftarrow$ случайное целое от 0 до n

$xs[i] \leftarrow xs[i] - 1$

$xs[j] \leftarrow xs[j] + 1$

Сначала действительно наблюдается явление, подобное диффузии, однако по мере достижения левой границы распределение искажается и начинает стремиться к характерной несимметричной форме. Если эту книгу читает физик, то он сможет уверенно предположить, что это может быть за распределение: он назовет его *распределением Гиббса*. Внимательный читатель вспомнит, что

мы уже встречались с ним, когда изучали фрустрацию во время ожидания автобуса. Тогда мы рассматривали распределение интервалов между пуассоновскими событиями, которое описывалось *экспоненциальным распределением*. Оба этих проницательных господина будут правы, называя разными именами одно и то же замечательное распределение.

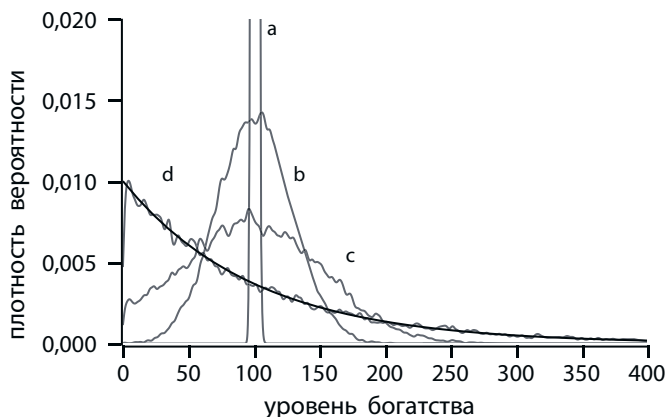


Рис. 9.5. Результат имитационного моделирования процесса перераспределения для фиксированного Δt для $m = 100$ рублей и $n = 5000$ человек. *a* — 10 шагов, *b* — 5000 шагов, *c* — $5 \cdot 10^{10}$ шагов, *d* — 10^8 шагов алгоритма

Люди — молекулы

Распределение Гиббса — из области статистической физики. Здесь описываются свойства систем, называемых красивым словом «ансамбль», которые состоят из великого множества взаимодействующих элементов — чаще всего физических частиц. Под частицами понимаются такие объекты (или их модели), внутренняя структура которых несущественна: на первый план выходит взаимодействие между ними. В ансамбле можно выделять произвольные подсистемы (например, отдельные частицы или их группы) и ставить им в соответствие некие функции состояния (это могут быть обобщенные координаты, скорости, концентрации, химические

потенциалы и многое другое). С помощью методов статистической физики удастся объяснить и вычислить параметры самых разнообразных явлений: химических и каталитических процессов, турбулентности, ферромагнетизма, поведения жидких кристаллов, сверхтекучести и сверхпроводимости и т. д.

Нелишним тут будет повторить слова великого физика и блестящего лектора Ричарда Фейнмана.

Если бы в результате какой-то мировой катастрофы все накопленные научные знания оказались уничтоженными и к грядущим поколениям живых существ перешла бы только одна фраза, то какое утверждение, составленное из наименьшего числа слов, принесло бы наибольшую информацию?

Я считаю, что это атомная гипотеза: все тела состоят из атомов — маленьких телец, которые находятся в непрерывном движении, притягиваются на небольших расстояниях, но отталкиваются, если одно из них плотнее прижать к другому. В одной этой фразе содержится невероятное количество информации о мире, стоит лишь приложить к ней немного воображения и чуть соображения*.

Исходя из этой гипотезы, статистическая физика дает фундаментальное объяснение практически всему, что мы наблюдаем и измеряем в масштабах кристалла, человеческого тела или звезды.

В рамках этой науки распределение Гиббса отвечает на вопрос, какова вероятность встретить некое состояние подсистемы, если даны: а) энергия состояния; б) макроскопические (условно говоря, глобальные) свойства системы, например температура; в) известно, что система находится в термодинамическом равновесии. В последней фразе достаточно много терминов, не характерных для нашей книги: *энергия, температура, равновесие...* Но как в самом начале мы положились на интуитивное понимание вероятности, а потом дополнили его строгими определениями, так и сейчас я предполагаю, что читатель знаком с этими понятиями хотя бы из школьного

* Цит. по: Фейнман Р., Лейтон Р., Сэндс М. Фейнмановские лекции по физике. М., 1977. Вып. 1, 2. С. 23–24.

курса физики. Чуть позже мы разберемся с тем, какое отношение все это имеет к нашим экономическим моделям.

Распределение Гиббса может быть схематично выражено следующей формулой:

$$p(x) = Ce^{-\frac{E(x)}{kT}},$$

где x — некое состояние подсистемы, $E(x)$ — энергия этого состояния, T — абсолютная температура системы (или ее аналог), а C и k — величины, необходимые для нормировки и соответствия размерностей. Очень важное условие равновесия означает, что из рассмотрения исчезает время и что вся система окажется в наиболее вероятном своем состоянии для заданных условий.

Строгий вывод выражения для распределения Гиббса нам здесь не нужен, вместо него я покажу красивейшее, чисто математическое рассуждение, приводящее к его экспоненциальной форме.

Поскольку рассматриваются части системы, которые в сумме дают всю систему, то и в качестве их характеристики стоит выбрать какую-нибудь аддитивную величину, играющую роль меры. Напомню, что значение аддитивной величины для ансамбля равно арифметической сумме значений этой величины для его частей. В качестве такой величины в механике можно использовать *энергию*. С другой стороны, мы вычисляем вероятность того, что будем наблюдать некоторое состояние системы. Если ее можно разбить на части, то вероятность наблюдать их все одновременно будет равна произведению вероятностей для состояния каждой из частей. Таким образом, нам нужна функция, превращающая аддитивную величину в мультипликативную:

$$f(x+y) = f(x)f(y).$$

Если отбросить тривиальное решение $f(x) \equiv 0$, то таким свойством обладает только *показательная функция* $f(x) = a^x$, которая сумму аргументов превращает в произведение значений: $a^{x+y} = a^x a^y$. Ну а из всех показательных функций наиболее удобна экспонента, поскольку она очень хорошо ведет себя при интегрировании и дифференцировании.

Насколько универсально распределение Гиббса? Напомню, что это распределение количества частиц по энергиям. Такое распределение можно получить, рассматривая тепловое движение молекул газа, а потом только из него можно вывести (не пронаблюдать в эксперименте, а получить математически) *уравнение состояния идеального газа*, знакомое со школы под названием уравнения Менделеева — Клапейрона. В твердом теле, например кристалле, к энергии движения частиц добавляется сила упругости (притягивания и отталкивания), но базовым распределением по полной энергии все равно останется распределение Гиббса. Если мы сосредоточимся на энергии частиц в поле силы тяжести, то вновь получим экспоненциальное распределение. На этот раз оно будет носить имя Людвиг Больцмана, автора точного выражения для энтропии. Распределение Больцмана покажет нам, как изменяется плотность газа с высотой. Экспоненциальное распределение — как распределение с максимальной энтропией — база, с которой начинается исследование сложной физической системы.

Если быть совсем точным и вспомнить, что деньги в нашем эксперименте — величина дискретная, то мы наблюдаем *геометрическое распределение* — дискретный аналог экспоненциального. Эти два распределения подобны и сливаются при уменьшении вероятности выигрыша. В нашем эксперименте шансы получить рубль равны $1/5000$; это настолько малая величина, что геометрическое и экспоненциальное распределения можно считать неотличимыми друг от друга.

Измеряем температуру у рынка

В нашей модели рынка мы имеем аддитивную величину — количество денег у каждого игрока; это аналог энергии. При описанном нами обмене эта величина у всей системы, как и энергия в замкнутой физической системе, сохраняется. А какой смысл здесь у температуры? Это можно выяснить, посмотрев на выражение для плотности вероятности экспоненциального распределения:

$$p(x) = \lambda e^{-\lambda x},$$

и вспомнив, что среднее значение для него равно $1/\lambda$. Поскольку число игроков в ходе торгов неизменно, сохраняется и среднее количество денег у них, равное первоначально раздаваемой каждому сумме m . Отсюда естественным образом следует, что $\lambda = 1/m$ и, значит, в роли температуры в нашей экономической модели выступает среднее количество денег у игроков m . На рисунке 9.6 показаны примеры равновесных состояний рынков, соответствующих низкой и высокой температуре при одинаковом количестве участников.

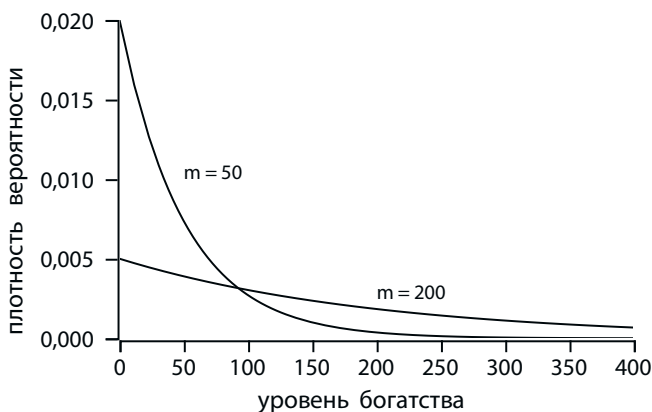


Рис. 9.6. Распределения достатка, соответствующие «горячему» ($m = 200$) и «холодному» ($m = 50$) рынкам

На «разогретом» рынке с большой ликвидностью мы сможем наблюдать и больший разброс в уровне благосостояния, чем на «холодном», ведь у экспоненциального распределения дисперсия равна $1/\lambda^2$. Как говорил Остап Бендер в «Золотом теленке» Ильи Ильфа и Евгения Петрова: «Раз в стране бродят какие-то денежные знаки, то должны же быть люди, у которых их много».

А что случится, если мы приведем «холодный» и «горячий» рынки в соприкосновение, позволив членам этих двух групп производить обмен между последними? Путь в одной группе n_1 участников владеют суммой M_1 , а в другой — n_2 участников располагают общей денежной массой M_2 . Средние значения $m_1 = M_1/n_1$

и $m_2 = M_2/n_2$ характеризуют абсолютную температуру рынков. Через какое-то время суммарная система придет к равновесию, и мы получим одну группу с числом участников $n = n_1 + n_2$ и с денежной массой $M = M_1 + M_2$. Отсюда можно найти температуру комплексной системы, она будет равна

$$m = \frac{M}{n} = \frac{n_1 m_1 + n_2 m_2}{n_1 + n_2}.$$

Если вы помните, именно так вычисляется температура, получающаяся, например, при смешивании двух объемов воды, нагретых по-разному. Так что аналогия среднего достатка и температуры вполне пригодна для использования.

Завершим мы рассказ о температуре рынка еще одним примером, в котором эта концепция совпадает по смыслу с физической величиной. Представьте себе, что наша система становится *открытой* и может выпускать членов группы, набравших определенную денежную сумму. Иными словами, разрешим богачам, как говорится, «линять» из системы, прихватив с собой «золотой парашют». Что мы должны наблюдать? По мере исчезновения самых богатых количество денег в группе станет убывать. Если бы из нее могли выбывать любые участники, то средний достаток практически не менялся бы из-за одинакового уменьшения как количества участников, так и общей денежной массы. Но, поскольку по нашим правилам выбывают именно богатые, будет убывать и средний уровень благосостояния, а это приведет к тому, что температура нашего рынка станет падать.

Описанный процесс очень похож на остывание жидкости при испарении: помните, как охлаждает руку тонкий слой спирта, наносимый врачом перед уколom? Молекулы, толкая друг друга случайным образом, могут какой-то из них придать такой импульс, что она окажется в состоянии преодолеть общее притяжение и покинуть систему, унеся при этом и энергию, подаренную ей соседями. В «холодной» рыночной системе возрастает доля бедных по сравнению с «горячей», так что остающимся в группе участникам этот процесс не сулит ничего хорошего.

Постигаем Дао энтропии

Осталось разобраться с равновесностью итогового состояния рынка. Термодинамическое равновесие можно описать разными способами. Во-первых, равновесным должно быть *стационарное* состояние, в котором система может находиться неограниченно долго, не изменяя своих макроскопических параметров и не образуя внутри себя упорядоченных потоков вещества и энергии. Во-вторых, такое состояние должно быть *устойчивым*: если вывести систему из него, она будет стремиться к нему вернуться. В-третьих, оно соответствует наиболее вероятному состоянию системы из всех возможных. Оно чаще наблюдается, и система со временем будет стремиться попасть в устойчивое равновесие из любого другого состояния.

Наш эксперимент демонстрирует все эти критерии равновесности: придя к экспоненциальному распределению, система в нем и остается. К тому же в эксперименте легко убедиться, что из любого произвольного распределения мы по истечении какого-то времени снова придем к экспоненциальному. Но это еще не доказательство, а только намек, что мы, скорее всего, имеем дело с равновесием. Нужен формальный измеримый критерий, который однозначно укажет нам, что система равновесна, без необходимости ждать бесконечно долго или перебирать все возможные первоначальные распределения. Это был бы полезный критерий, который допустимо применять и к реальному рынку — без необходимости проводить рискованные эксперименты на живых людях.

Размышления о равновесии привели физиков к одному фундаментальному понятию, о котором слышали, наверное, все, но объяснить и тем более с толком использовать способны немногие, — *энтропии*. Она постепенно вышла за пределы термодинамики и так понравилась ученым всех направлений, философам и даже широкой публике, что это сугубо термодинамическое понятие получило нынче ореол загадочности, непостижимости и бог знает еще чего. Простое и специальное, в сущности, понятие приобрело в сознании широких масс репутацию необъяснимо

управляющей миром концепции. Это связано с тем, что термодинамика описывает на очень высоком уровне абстракции системы самой разной природы: от физических, химических и биологических до социальных, экономических и даже чисто гуманитарных. После школьного курса, правда, остается ощущение, что термодинамика — это про скучный идеальный газ, какие-то поршни и невозможный цикл Карно. Такое однобокое представление связано с тем, что термодинамика, будучи одним из самых абстрактных и универсальных разделов естествознания, элегантно решает прикладные задачи, которые могут быть поняты школьниками и при этом оказаться полезными в промышленности. Этого не скажешь, например, о теории категорий или топологии — тоже весьма абстрактных, универсальных и, несомненно, полезных дисциплинах, но в повседневных задачах почти не встречающихся.

Итак, на сцену выходит энтропия. Создателю термодинамики Рудольфу Клаузиусу (1822–1888), а позже физикам Джозайе Гиббсу (1839–1903) и Людвигу Больцману (1844–1906) потребовалась количественная характеристика равновесности, которая говорила бы о вероятности наблюдать указанное состояние системы или ее частей. Причем эта величина, которая отражает вероятность, мультипликативную для ансамбля, должна быть аддитивной функцией состояния, чтобы можно было вычислить ее для системы, складывая установленные значения ее частей. Когда мы искали подходящую функцию для распределения Гиббса, мы исходили из того, что она должна превращать аддитивный аргумент в мультипликативное значение. При поиске выражения для энтропии мы нуждаемся в функции, мультипликативной по аргументу и аддитивной по значению:

$$f(ab) = f(a) + f(b).$$

Это функциональное уравнение решает логарифмическая функция, обратная показательной. Энтропия состояния сложной системы может быть выражена как ожидаемое значение для логарифма вероятности наблюдения состояния всех ее частей, или,

по Больцману, как логарифм числа способов, которыми можно реализовать это состояние системы. При этом более вероятному состоянию соответствует большее значение энтропии, а *равновесному* — максимальное из возможных.

Число способов, которыми можно реализовать то или иное состояние, зависит от числа ограничений или условий, при которых это состояние может реализоваться. Чем их меньше, тем более вероятно состояние и тем выше значение его энтропии. Эти ограничения и условия имеют смысл *информации о состоянии*. Отсюда возникла идея о том, что энтропия отражает *степень* нашего *незнания* о системе: чем меньше нам о состоянии известно, тем больше его энтропия. Позже Клод Элвуд Шеннон (1916–2001) обобщил это понятие для любых систем, содержащих в себе информацию, в том числе распределений случайных величин. Вот что у него получилось. Для случайной величины X , определяемой функцией вероятности $p(x)$, энтропия определяется следующим образом:

$$H \equiv -M[\ln p(x)] = -\sum p(x) \ln p(x),$$

где суммирование производится по всем значениям x , в которых $p(x) > 0$. Таким образом, мы имеем возможность вычислить энтропию состояния любой сложной системы, располагая ее статистическим описанием.

Каждому распределению случайной величины — неважно, задаваемому аналитически или полученному экспериментально в виде гистограммы — можно поставить в соответствие положительное число — его энтропию. Это, в свою очередь, задает метрику на пространстве распределений, давая нам возможность сравнивать их между собой, определяя более или менее равновесные и вероятные распределения для заданных условий. Более того, для некоторого класса распределений можно выделить одно с максимальной энтропией — и только одно. Классы определяются ограничениями, или мерой нашего знания о статистических свойствах системы. Приведем самые важные примеры распределений, имеющих наибольшую энтропию.

Что нам известно о случайной величине x	Распределение с максимальной энтропией	Выражение для энтропии
$x \in [a, b]$	равномерное распределение на отрезке $[a, b]$	$\ln(b - a)$
$x \in \{0, 1\}$	распределение Бернулли с параметром p , $q = 1 - p$	$-p \ln p - q \ln q$
$x \in [0, \infty)$ + среднее μ	экспоненциальное (геометрическое) распределение	$1 - \ln\left(\frac{1}{\mu}\right)$
$x \in [x_0, \infty)$ + среднее геометрическое $x_m e^{1/k}$	распределение Парето (степенное) с параметром k	$\ln\left(\frac{k}{x_m}\right) - \frac{1}{k} - 1$
$x \in [x_0, \infty)$ + среднее + среднее геометрическое	гамма-распределение	выражается через специальные функции
$x \in (x_0, \infty)$ + среднее геометрическое + дисперсия для среднего геометрического	лог-нормальное распределение с параметрами μ и σ	$\log_2\left(\sigma e^{\mu + \frac{1}{2}} \sqrt{2\pi}\right)$
$x \in (-\infty, \infty)$ + среднее + дисперсия	нормальное (гауссово) распределение	$\ln(\sigma \sqrt{2\pi} e)$

Знакомые всё лица! Это очень часто используемые распределения, которые статистики применяют к широчайшему классу задач. Их универсальность обусловлена именно тем, что они, имея максимальную энтропию, наиболее вероятны и наблюдаются чаще других. К ним, как к равновесным, стремятся многие распределения реальных случайных величин.

Наиболее свободно от ограничений *нормальное распределение*: оно требует минимума информации о случайной величине. Меньше уже не получится: если мы укажем лишь среднее значение, то при попытках увеличить энтропию распределение «размажется» по всей числовой оси. Зато если мы знаем лишь среднее, но при этом ограничим случайную величину положительными значениями, то равновесное распределение будет однозначным — *экспоненциальным*. Именно этот случай мы и наблюдали в нашем эксперименте с рынком. Нам заранее было известно лишь то, сколько денег мы выдали каждому игроку, и то, что их количество в системе неизменно. Эта информация фиксирует среднее значение. А поскольку количество

денег у нас — величина положительная, то, вероятнее всего, в равновесии мы получим именно экспоненциальное распределение.

В численном эксперименте можно вычислять энтропию нашей системы по мере приближения модели рынка к равновесию. Пример такого графика приведен на рис. 9.7. Обратите внимание на то, что ось X логарифмическая. Благодаря этому мы сможем одинаково внятно увидеть как начальные этапы развития модели, так и ее поведение для очень большого числа обменов, и в то же время логарифмическая шкала позволяет четко выделить отдельные этапы эволюции модельной системы. Буквы здесь соответствуют распределениям, показанным на рис. 9.5.

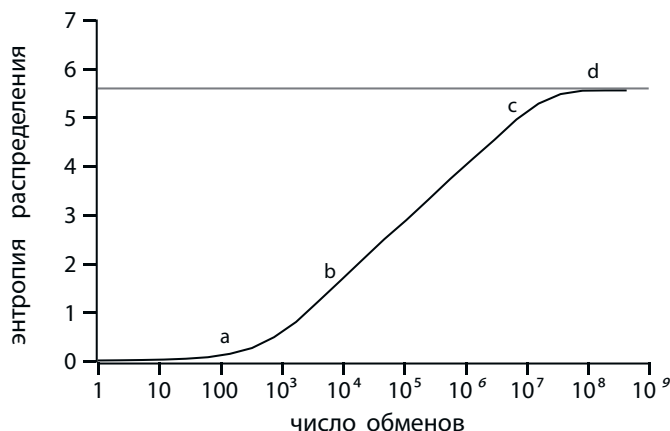


Рис. 9.7. Рост энтропии, наблюдающийся по мере приближения рынка к равновесному состоянию. Горизонтальной линией на графике показано теоретическое значение энтропии для экспоненциального распределения

Начальное состояние (вырожденное, при котором все участники группы располагают равными суммами) имеет нулевую энтропию; о том, что это значит, мы скажем чуть позже. Первые десятки обменов до состояния (a) лишь немного ее увеличивают, распределение все равно остается близким к вырожденному. Но далее оно становится очень похожим на нормальное, начинается диффузионный процесс, сопровождающийся линейным ростом энтропии на нашем графике. Если вы заглянете в таблицу выше, то увидите, что

энтропия нормального распределения пропорциональна логарифму от стандартного отклонения. Именно эту пропорциональность и показывает нам график энтропии в выбранном нами логарифмическом масштабе. Теперь мы можем интерпретировать появление здесь нормального распределения как наиболее вероятного для случайной величины, о которой мы знаем лишь ее среднее (оно остается неизменным) и дисперсию (она растет, как в процессе случайного блуждания). Наконец, в состоянии (с) система начинает «чувствовать» дно и симметричность распределения нарушается, после чего оно постепенно достигает равновесного.

Не знаю, как читателю, а мне показалось обидным, что изначально справедливое распределение после серии абсолютно симметричных и беспристрастных обменов само по себе приходит к несправедливости. Мы уже говорили, что коэффициент Джини для экспоненциального распределения в точности равен $1/2$ и при таком распределении половина всех денег принадлежит богатейшим 20% группы. С другой стороны, может порадовать то обстоятельство, что эта несправедливость возникает не вследствие греховной человеческой натуры, а из-за натуры больших ансамблей взаимодействующих частиц.

Наша модель предельно проста. Существует множество ее модификаций: передаваемая сумма Δt может быть не фиксированной, а случайной величиной, ограниченной состоянием участника; при этом можно не давать деньги какому-то одному игроку, а распределять случайным образом. Пока мы не вводим новых параметров, все эти модификации не меняют форму равновесного распределения богатства — оно остается экспоненциальным. Многие исследователи отмечали эту особенность моделей рынка. В устойчивости решения можно убедиться с помощью имитационного моделирования, но приводить картинки для различных способов обмена неинтересно — все они будут одинаковыми. Любопытна модель, построенная Адрианом Драгулеску и Виктором Яковенко из Мэрилендского университета*. В ней игроков объеди-

* Dragulescu A., Yakovenko V. M. Statistical mechanics of money // Eur. Phys. J. 2000. Vol. B 17. Pp. 723–729.

няют в некие «компании», а далее имитируется взаимодействие компаний с игроками-работниками и игроками-покупателями. Но и в этом, уже достаточно сложном случае равновесным оказывается экспоненциальное распределение, безразличное к выбираемым параметрам модели.

Загадочная и могущественная энтропия — это, конечно, солидно и, возможно, даже убедительно. Но почему же при симметричном обмене бедных становится больше, чем богатых? Почему мода равновесного распределения равна нулю? Надо, как говорят физики, разобраться в *кинетике процесса*, в судьбе отдельных частиц.

Мы не ошиблись, предположив, что модель случайного блуждания описывает изменение состояния отдельного участника торгов: он с равной вероятностью совершает шаги как вверх, так и вниз. Мы уже говорили о том, что случайно блуждающая частица обязательно окажется в любом наперед указанном месте. При этом ожидаемое расстояние, на которое частица удалится от какой-либо начальной точки, оказывается пропорционально квадратному корню от числа шагов. Все это приводит к тому, что если частица начинает свой путь вблизи нуля, то она с высокой вероятностью его достигнет, а поскольку ноль в нашей задаче — непроницаемая граница, она будет вынуждена вновь и вновь начинать свой путь около нулевой точки, с большой вероятностью быстро к ней возвращаясь. По мере удаления частицы от нуля вероятность к нему вернуться уменьшается и у богатых становится больше шансов сбросить свое состояние.

Но тогда что же мешает частице удалиться сколь угодно далеко, а конкретному игроку стать сколь угодно богатым? Вообще-то ничего, кроме конечности денег в системе: экспоненциальное распределение отлично от нуля на всей положительной полуоси. Но чтобы достичь невероятного богатства по правилам нашей игры, нужно, чтобы какой-то ее участник случайно получил систематическое преимущество перед остальными. Выбор, кому отдать деньги в нашей модели, падает на всех одинаково, а это значит, что доставаться они будут не только богатым, но и бедным. Есть в этом мире справедливость, хоть и торжествующая совсем недолго, для того, кто растерял все свое богатство.

Игры с энтропией

Если понятие энтропии помогло предсказать и объяснить экспоненциальное распределение в простейшей модели рынка, то, быть может, оно окажется полезным и в более сложных моделях? Мы станем добавлять ограничения в модель рынка, делать предположение о форме распределения исходя из принципа максимума энтропии, а потом проверять результат с помощью имитационного моделирования.

Для начала искусственно ограничим сверху уровень богатства отдельного игрока, запретив ему получать деньги, если у него уже есть некая фиксированная сумма x_{\max} . В случае, если $t = x_{\max}/2$, мы приходим к варианту, описанному в первом ряду таблицы распределений с максимальной энтропией. Действительно, ограничивая случайную величину конечным отрезком и не указывая больше ничего, мы не можем предположить никакого другого ожидаемого значения среднего, кроме середины этого отрезка (рис. 9.8). Следовательно, равновесным распределением при таком варианте должно быть *равномерное*. Проверим, так ли это, воспользовавшись следующим алгоритмом.

Исходные данные: xs — массив из n элементов, инициализированный значениями m , $xMax$ — максимальная разрешенная сумма.

Повторять

```

i <- случайное целое от 0 до n
если xs[i] > 0
    j <- случайное целое от 0 до n
    если xs[j] < xMax
        xs[i] <- xs[i] - 1
        xs[j] <- xs[j] + 1

```

Надо заметить, что мы получили довольно любопытный результат. Каждый из участников группы все еще испытывает случайное блуждание, но никто не «прилипает» к границам и в группе происходит равномерное перемешивание. Напомню, что коэффициент

Джини для равномерного распределения равен $1/3$, что уже существенно лучше, чем $1/2$ для экспоненциального распределения, так что ограничения могут пойти на пользу.

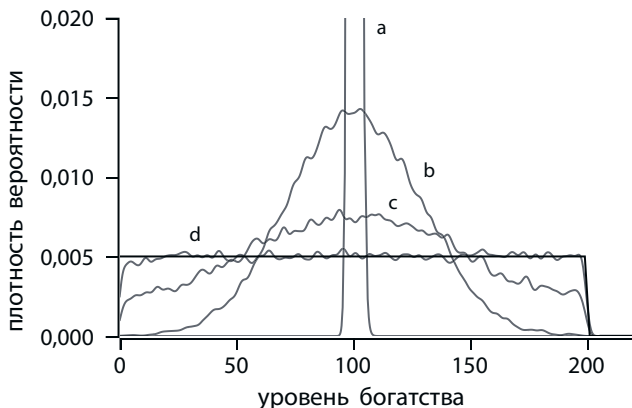


Рис. 9.8. Вот что происходит при ограничении сверху возможного уровня богатства игроков, причем таким образом, что верхняя граница ровно вдвое превышает среднее значение

А что случится при нарушении симметрии, то есть при сдвиге правой границы вправо или влево от значения $2m$? Распределение достатка в таком случае перестанет быть равномерным и приобретет некоторый перекосяк в сторону смещения среднего относительно середины разрешенного диапазона уровня богатства. Принцип максимума энтропии позволяет получить точные выражения для этих распределений — это всё те же распределения Гиббса (экспоненциальные), но отличные от нуля лишь на заданном отрезке и соответствующим образом нормированные (рис. 9.9). Правда, в конечной форме (в виде алгебраического выражения) показатели экспонент уже не выражаются, но их всегда можно получить численно с необходимой точностью.

Очень необычный вид распределения получается при смещении среднего относительно середины отрезка вправо: богатых игроков в равновесии становится больше, чем бедных. Показатель, характеризующий температуру в этом распределении, имеет отрицательный знак! В обычной жизни под

отрицательной мы понимаем температуру ниже точки замерзания воды — 0°C — и ничего странного в ней не находим. Однако в термодинамике речь идет об *абсолютной температуре* (по Больцману) как о характеристике внутренней энергии системы. Таким образом, для частиц, не взаимодействующих между собой (как в идеальном газе), говорить об отрицательной температуре нет смысла: модуль количества движения не может быть меньше нуля. Но в других физических системах такая ситуация уже возможна. В статистической физике отрицательной считается температура, характеризующая равновесные состояния термодинамической системы, где вероятность обнаружить систему в микросостоянии с более высокой энергией выше, чем в микросостоянии с более низкой. Это становится возможным лишь при ограниченном объеме фазового пространства; именно такой случай мы и наблюдаем. Примерами систем с отрицательной абсолютной температурой могут быть лазер в возбужденном состоянии, частицы газа в сложных внешних силовых полях, например в стоячей световой волне, и другие непростые квантовые системы.

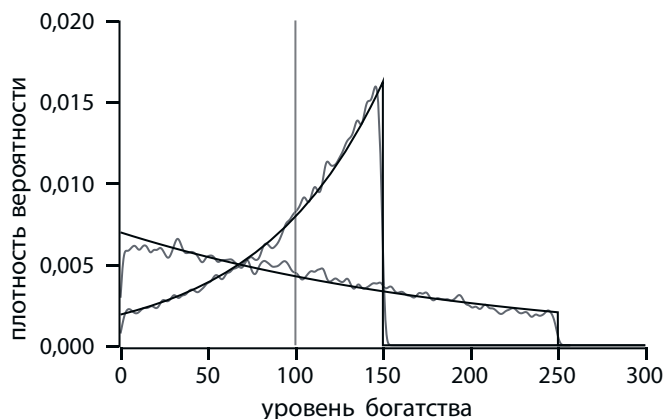


Рис. 9.9. Варианты равновесных распределений для обмена с ограничением сверху. Вертикальной линией показано значение среднего (начального) богатства участников эксперимента. Коэффициенты Джини для полученных нами двух случаев равны 0,2 (для правого смещения среднего) и 0,43 (для левого смещения среднего)

Внимательный читатель может возмутиться: мы же говорили, что в нашем случае роль температуры играет среднее количество денег у членов группы, какой же смысл может быть в отрицательном среднем количестве денег? Введение верхнего предела оставило распределение экспоненциальным, но поменяло форму показателя в экспоненте. Теперь он хоть и зависит от среднего значения t , но не равен ему. Если нам будет угодно, мы и дальше можем называть величину, обратную показателю, аналогом температуры, но делать это следует с большой осторожностью. Показатель в экспоненте получается пропорциональным значению $1/(t - x_{\max}/2)$, и эта величина уже может менять знак. Более того, он меняется при переходе знаменателя через ноль! Получается, что равномерному распределению $(t - x_{\max}/2)$ соответствует бесконечная температура? Это не совсем так. На ноль, как мы уже упоминали, делить нельзя, так что о какой-либо температуре — в смысле показателя экспоненты — для равномерного распределения говорить тоже нельзя, ведь распределение вовсе перестает быть экспоненциальным. Выбранная нами математическая модель меняется, и в ней нет аналога термодинамической температуры.

Я хочу здесь еще ненадолго остановиться на вопросе применимости математических и физических аналогий. Часто привычные и, как нам кажется, простые понятия имеют очень глубокие и фундаментальные основания. Так, знакомое всем нам с детства понятие температуры физикам удалось глубоко понять и осознать только с развитием методов теории вероятностей и математической статистики. После этого стало возможным осмысленно рассуждать о термодинамике лазеров, биологических и социальных систем, звезд и даже черных дыр. В данной книге мы постигаем природу несправедливости с помощью этих же методов. Но не нужно буквально понимать наши достаточно вольные рассуждения о температуре рынка, ее знаке и возможности бесконечных значений. Мы много раз говорили об удивительной способности математики обнаруживать одинаковые модели и структуры для самых разнообразных явлений. Построенная нами статистическая модель рынка и модель ансамбля физических частиц, имея много общего, все же не одно и то же. Именно поэтому то, что в физике называется

и является температурой, имеет аналог в экономфизике, но собственно температурой в этой дисциплине не считается — как и величина, обратная интенсивности, в экспоненциальном распределении пауз между машинами на автостраде.

Экономика должна быть экономной

Пока наша модель обмена никак не учитывает достатка игроков, она остается нереалистичной. В действительности богатые тратят больше, а бедные меньше; более того, разумные люди стараются сохранить какую-то часть своего состояния. В качестве следующего усложнения модели потребуем, чтобы игроки в процессе перераспределения отдавали некую известную долю своего состояния $\Delta m = \lfloor \alpha m \rfloor$, где $0 < \alpha < 1$. При этом дробные денежные единицы округляются до ближайшего целого вниз (это значит, что если αm окажется меньше единицы, то $\Delta m = 0$). Иными словами, добавим нашим участникам желание быть экономными.

В систему вводятся новый параметр и новое ограничение; следовательно, равновесное состояние должно как-то отклониться от экспоненциального. Опирируя долями от уровня благосостояния, мы переходим к мультипликативным характеристикам, таким, например, как *доходность вложения*, *возврат инвестиций* и т. д. Во всех учебниках по экономике указывается, что если вы желаете вычислить среднюю доходность вложения, скажем, за много лет, то следует определять *среднее геометрическое* для доходностей каждого года. В нашем случае среднее геометрическое однозначно, хоть и нетривиально, определяется значением α . Таким образом, добавляя новый параметр, мы фиксируем среднее геометрическое распределения дохода игроков, или среднюю доходность модели рынка. Значит, согласно таблице распределений с максимальной энтропией, мы можем ожидать, что равновесное распределение богатства должно неплохо описываться *гамма-распределением*. В этом мы можем убедиться, проведя имитационное моделирование (рис. 9.10).

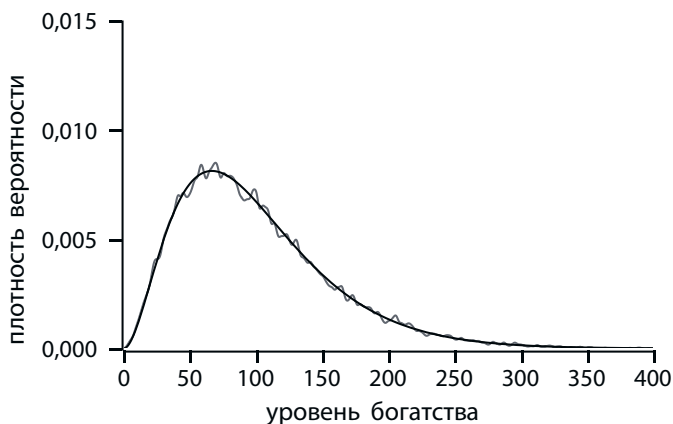


Рис. 9.10. Если расходы при обмене пропорциональны достатку, равновесное распределение стремится к характерному несимметричному колоколообразному гамма-распределению. В данной модели $\alpha = 1/3$

Для имитационного моделирования был реализован такой алгоритм пропорционального обмена.

Исходные данные: xs — массив из n элементов, инициализированный значениями m , α — доля капитала, которая тратится при обмене.

Повторять

```

i <- случайное целое от 0 до n
если xs[i] > 0
    dx <- floor(xs[i]*alpha)
xs[i] <- xs[i] - dx
j <- случайное целое от 0 до n
xs[j] <- xs[j] + dx

```

Эта книга — хоть и популярная, но все же математическая. Это значит, что все результаты, попавшие на ее страницы, имеют доказательства или строгий вывод, пусть зачастую и остающиеся за пределами изложения в силу их громоздкости. И хотя для дальнейшего изложения этот результат не нужен, я приведу точное и довольно изящное выражение для распределения,

которое мне удалось получить для модели пропорционального обмена.

Гамма-распределение $Gamma(k, \theta)$ — двухпараметрическое распределение, которое часто используется как обобщение экспоненциального и сводится к нему при $k = 1$. Оно имеет ряд замечательных свойств, делающих его полезным. Об одном из них мы уже говорили — это распределение с максимальной энтропией в своем классе. Другое важное свойство — его *бесконечная делимость* и связанная с этим *устойчивость*. Случайная величина называется бесконечно делимой, если для любого числа $n \geq 1$ ее можно представить в виде суммы n независимых одинаково распределенных случайных величин. А если эти слагаемые имеют то же распределение, что и исходная случайная величина, последняя называется устойчивой. Яркий пример устойчивого распределения — нормальное. И именно это его свойство вместе с тем, что оно является распределением с максимальной энтропией в самом широком классе распределений, делает его героем центральной предельной теоремы.

Но вернемся к гамма-распределению. Для него верно, что:

если $x \sim Gamma(k_1, \theta)$, $y \sim Gamma(k_2, \theta)$, то $x + y \sim Gamma(k_1 + k_2, \theta)$.

Наконец, гамма-распределение масштабируемо:

если $x \sim Gamma(k, \theta)$, то $ax \sim Gamma(k, a\theta)$.

Все эти свойства позволили получить распределение благосостояния для нашей модели со средним значением m и коэффициентом α в таком виде:

$$Gamma\left(\frac{1}{\alpha} - 1, \frac{1}{m} \left(\frac{1}{\alpha} - 1\right)\right).$$

В модели обмена фиксированной суммой вероятность потратить все деньги была достаточно велика. В модели пропорционального обмена она оказывается равна нулю. Это связано с тем, что бедные тратят в среднем меньше, чем получают от богатых,

ведь и те и другие обмениваются долями своего капитала. Но этот социальный лифт действует только при $\alpha < 1/2$. Если тратить больше половины того, что имеешь, вероятность оказаться в бедняках становится не просто отличной от нуля, а весьма ощутимой. Для различных значений можно получить различающиеся по форме распределения с широким диапазоном несправедливости (рис. 9.11).

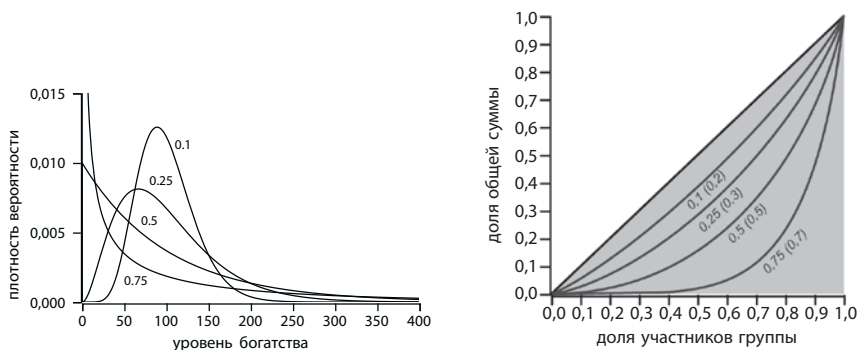


Рис. 9.11. Различные варианты равновесных распределений при расходах, пропорциональных достатку. Графики помечены значениями α , на правом графике в скобках приведены еще и значения индекса Джини

Получается, что чем бóльшую часть своего капитала игроки вынуждены тратить (например, на повседневные нужды или еду), тем больше становится доля бедных и тем менее справедливым оказывается общество. Любопытно, что при $\alpha = 1/2$ равновесное распределение становится экспоненциальным, как в модели при равном обмене. Напомню, что экспоненциальное распределение — частный случай гамма-распределения с параметром $k = 1$, так что это превращение само по себе не удивительно. Но тут есть одна любопытная тонкость: энтропия этого частного случая превышает энтропию распределений с любыми другими значениями α . Посмотрите, как изменяется энтропия по мере развития ситуации при $\alpha = 0,75$ (рис. 9.12).

Поначалу значение энтропии монотонно увеличивается; потом, практически достигнув теоретического максимума,

соответствующего экспоненциальному распределению, рост энтропии останавливается, и она начинает уменьшаться. Нет ли в этом противоречия с определением равновесного состояния как состояния с максимумом энтропии? Противоречия нет, поскольку равновесное состояние должно быть, во-первых, *стационарным*, не создающим направленных потоков энергии, а во-вторых, *устойчивым* или, говоря языком теории динамических систем, *притягивающим к себе систему*. В конце концов, среди всех стационарных состояний равновесным будет состояние с максимальной энтропией. А в нашем случае при $\alpha = 0,75$ экспоненциальное распределение соответствует нестационарному состоянию.

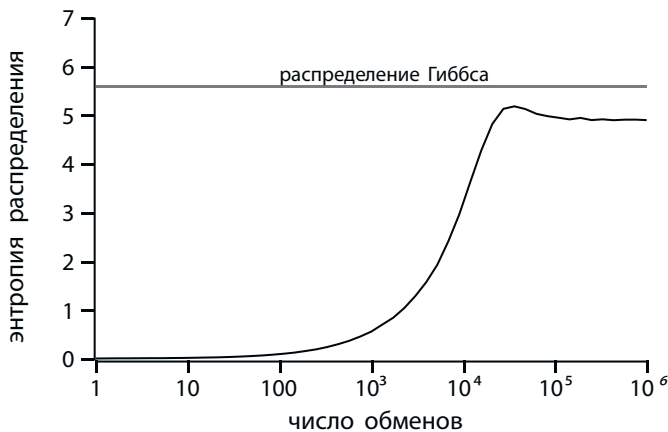


Рис. 9.12. В процессе перехода к равновесию система «проскакивает» состояние с максимальной энтропией

Исследователи из Бостонского университета Слава Исполатов и Павел Крапивский* усложнили модель пропорционального обмена так, чтобы обмен происходил с учетом благосостояния не только тратящего, но и получающего. Миллионер редко покупает что-то у зеленщика, и зеленщик нечасто имеет большой доход; с другой стороны, производитель автомобилей экстра-класса будет взаимодействовать лишь с богатыми клиентами, но и сам не

* Ispolatov S., Krapivsky P. L., Redner S. Wealth Distributions in Models of Capital Exchange // Eur. Phys. J. B. 1998. Vol. 2. P. 267.

останется внакладе. Алгоритм такого обмена остается достаточно простым.

Исходные данные: x_s — массив из n элементов, инициализированный значениями m , α — доля капитала, которая тратится при обмене, β — доля капитала, приобретаемого при обмене.

Повторять

```
i <- случайное целое от 0 до n
если  $x_s[i] > 0$ 
     $dx <- \text{floor}(x_s[i]*\alpha)$ 
 $x_s[i] <- x_s[i] - dx$ 
    повторять, пока  $dx > 0$ 
        j <- случайное целое от 0 до n
 $d = \min(dx, \text{floor}(x_s[j]*\beta))$ 
         $x_s[j] <- x_s[j] + d$ 
         $dx <- dx - d$ 
```

И вот в моделях, в которых богатые начинают платить преимущественно богатым, а бедные — бедным, общество «разваливается» окончательно. Если денежные потоки оказываются зависимы от капитала, система теряет устойчивость и приводит к постоянному обнищанию группы и все большему нарастанию классового неравенства. В ней существует только одно стационарное состояние: когда все игроки не имеют (и, следовательно, не получают) ровным счетом ничего, а все богатство достается кому-нибудь одному. Коэффициент Джини в таком состоянии практически равен единице, и оно очень далеко от нормального равновесного — его энтропия почти равна нулю. Спасти положение можно различными способами. Например, ввести ограничение снизу, запрещающее игрокам терять абсолютно все сбережения, и в этом случае равновесное распределение становится снова экспоненциальным либо гамма-распределением. Или организовать подобие налогообложения, обеспечивающее стабильный поток средств от богатых ко всем, в том числе бедным. Модель «дикого рынка» вполне применима к рынку ценных бумаг без каких-либо ограничений, но

на реальных биржах с этим борются, вводя ограничения на объем сделок, совершаемых за день, и на максимальные уровни роста или падения цены на тот или иной актив.

Все эти печальные выводы говорят не в пользу свободного рынка. То ли дело модель, предложенная Шариковым! А какова же энтропия у вырожденного распределения? Согласно стандартной формуле, она в точности равна нулю. Это самое неравновесное, самое маловероятное распределение, и в любой модели обмена оно нестационарно, так что получить подобное общество можно только искусственно. Дикий рынок, конечно, не подарок: он неустойчив и тяготеет к вопиющему неравенству. Требуется множество взаимосогласованных ограничений и тонко настроенных связей для построения устойчивого рынка и более или менее справедливого общества. Человечество исследует этот вопрос еще не очень долго и в основном на ощупь, методом проб и ошибок, но одно ясно: несправедливость в экономическом пространстве — не следствие поганой человеческой природы, а объективное свойство системы, в которую входим все мы. Более того, попытки создать абсолютную справедливость по-шариковски всегда проходили с боем и кровью, а результаты, в силу ее неравновесности, существовали недолго.

Вряд ли молекулы и атомы рассуждают о несправедливости своего мира, да и физики с инженерами за двести лет смирились с тем, что, какую бы идеальную тепловую машину они ни построили, хаос не позволит полностью преобразовать тепло в механическую работу. Когда понятно, то не так обидно. Надеюсь, эта глава поможет читателю понять и принять свойства нашего сложного и несправедливого мира. Принять не смирившись, а оттолкнувшись от них как от условия задачи, и постараться найти такие решения, которые помогли бы уменьшить эту несправедливость. На то нам и дан разум!

Заключение

Учитателя, который только знакомится с математикой, может возникнуть странное ощущение, что наша книжка ни о чем. В школе задачи имеют ответ. И он один, даже если состоит из системы решений, как, например, для квадратного уравнения. Причем этот ответ можно сравнить с правильным вариантом, который приведен в конце задачника или учебника.

Разбираясь с законами подлости, мы рассматривали разные задачи, но ни одна из них не получила ответа такого рода. И это не связано ни со случайностью как основным предметом нашего разговора, ни с тем, что мы плохо старались. Мы увидели, что о случайных величинах и функциях вполне можно рассуждать так, чтобы получать точное знание. Дело в том, что математика чаще всего интересуется не решением задачи, а свойства этого решения. Ему мало отыскать корень уравнения. Надо понять, единственное ли это решение, а если нет, то какую систему образует их множество и при каких условиях. Так из решения алгебраических уравнений — с некоторыми из них мы знакомимся еще в школе — родилась теория Галуа, которая расширила взгляд не только на сами уравнения, но и практически на все области математического знания! Именно поэтому мы не ограничивались результатами имитационного моделирования, хотя они давали вариант решения. Но настоящее решение — нечто иное. Настоящий математический анализ модели заменяет бесконечное число экспериментов, очерчивает границы ее применимости и подсказывает направление для ее расширения.

А главное, в работе математика или физика не будет ответа в конце задачника. Сам-то ответ есть, природе он известен, но он редко имеет бинарный характер вроде «ложно/истинно» или «да/нет». Скорее вы обнаружите какие-то отношения, функциональные зависимости, сеть связей между категориями, в которых формулируется задача. И эта «подкладка» ценнее любого частного ответа, пусть и имеющего большое практическое значение. Если начинающий свой путь ученый это почувствует, ему проще будет понять, как совместить практическую пользу математики с искусством, абстрактные построения с «производственной необходимостью», математический стиль мышления — с решением повседневных вопросов и проблем.

Не ждите даже от «царицы наук» волшебного ответа на все вопросы. Но не упускайте случая лишний раз услышать ее мнение и насладиться ее красотой!

Рекомендуемая литература

Арнольд В. И. Что такое математика? М. : МЦНМО, 2012.

Арнольд В. И. Экспериментальная математика. М. : МЦНМО, 2018.

Колмогоров А. Н. Основные понятия теории вероятностей. Изд. стереотип. URSS, 2019.

Курант Р., Роббинс Г. Что такое математика? (Элементарный очерк идей и методов) / Пер. с англ. под ред. А. Н. Колмогорова. 3-е изд., испр. и доп. М. : МЦНМО, 2001.

Мазур Дж. Игра случая. Математика и мифология совпадения. М. : Альпина Диджитал, 2017.

Млодинов Л. (Не)совершенная случайность. Как случай управляет нашей жизнью. М. : Гаятри, 2008.

Сингх С. Великая теорема Ферма. М. : МЦНМО, 2000.

Строгац С. Удовольствие от x . М. : Манн, Иванов и Фербер, 2017.

Френкель Э. Любовь и математика. СПб. : Питер, 2015.

Элленберг Дж. Как не ошибаться. Сила математического мышления. М. : Манн, Иванов и Фербер, 2017.

Об авторе

Сергей Самойленко, к.ф.-м.н., PhD. Вулканолог, педагог, популяризатор науки. Родился и вырос на Камчатке, учился в Новосибирской физико-математической школе, закончил Новосибирский государственный университет, физик по специальности. Проведя несколько лет за рубежом, решил связать свою жизнь с изучением вулканов и преподаванием точных наук (математики, физики, программирования) на родной земле. Был научным сотрудником и заместителем директора Института вулканологии Дальневосточного отделения Российской академии наук, заведовал кафедрой геологии, географии и геофизики Камчатского государственного университета имени Витуса Беринга. В 2017 году совместно с женой основал музей вулканов «Вулканариум», где сейчас активно ведет просветительскую и популяризаторскую работу для жителей Камчатки и гостей, приезжающих со всего мира.

Серия «Наука для всех»
Научно-популярное издание

Самойленко Сергей Борисович

Вероятности и неприятности
Математика повседневной жизни

Руководитель редакции *Светлана Мотылькова*
Шеф-редактор *Ренат Шагабутдинов*
Ответственный редактор *Юлия Потемкина*
Литературный редактор *Ольга Свитова*
Арт-директор *Алексей Богомолов*
Дизайн обложки *Алексей Галкин*
Верстка *Екатерина Матусовская*
Корректоры *Лев Зелексон, Евлалия Мазаник*

ООО «Манн, Иванов и Фербер»
123104, Россия, г. Москва, Б. Козихинский пер., д. 7, стр. 2
www.mann-ivanov-ferber.ru
www.vk.com/mifbooks



Кажется, что вас преследует злой рок: бутерброд падает маслом вниз, дожди разрушают давние планы на выходные, а дорожные пробки только и ждут вашего появления, чтобы материализоваться.

Но под канвой ежедневных событий кроется Математика. С научной точки зрения большинством мелких неурядиц можно буквально любоваться, что автор и делает. Теория вероятностей и математическая статистика, теория мер, марковские цепи, стохастические процессы, динамический хаос встроены в нашу жизнь.

Эта книга помогает развить математическое мышление. Если вы школьник, студент или преподаватель математики, она для вас. Если просто хотите правильно интерпретировать цифры и факты — тоже.

Сергей Самойленко — кандидат физико-математических наук, преподаватель Камчатского государственного университета, научный сотрудник Института вулканологии, педагог дополнительного образования. Профессионально занимается популяризацией науки: создал культурно-просветительский центр «Вулканариум» в Петропавловске-Камчатском, выступает с научно-популярными лекциями для широкой аудитории.

