

Лекции
по Теории Вероятностей

Г.А. Малиновская, А.И. Саичев

Оглавление

1	Основания теории вероятностей	4
1.1	Случайные явления	4
1.1.1	Испытание и случайное событие	4
1.1.2	Статистическая устойчивость	6
1.1.3	Относительная частота и вероятность	8
1.1.4	Бросание монеты и статистическое моделирование	9
1.1.5	Математический взгляд на вероятность	11
1.2	Схема шансов	12
1.2.1	Игра в кости	13
1.2.2	Формула шансов	14
1.2.3	Бросание двух костей	15
1.2.4	Исчисление вероятностей средствами комбинаторики	17
1.3	Вероятность и ее аргументы	24
1.3.1	События и множества	24
1.3.2	Алгебра событий	26
1.3.3	Свойства вероятностей	28
1.4	Геометрическое исчисление вероятностей	30
1.4.1	Задача о встрече	31
1.4.2	Задача Бюффона	31
1.4.3	Диаграммы Венна	34
1.4.4	Портретная галерея событий	36
1.4.5	Теорема сложения вероятностей	37
1.5	Вероятностное пространство	39
1.5.1	Минимальные алгебры событий	39
1.5.2	Алгебра Бореля	40
1.6	Условные вероятности	43
1.6.1	Скачки на ипподроме	44
1.6.2	Свойства условных вероятностей	46
1.6.3	Формула умножения вероятностей	48

1.6.4	Формула полной вероятности	48
1.6.5	Формула Байеса	50
1.7	Статистическая независимость	52
1.7.1	Независимость и несовместность	54
1.7.2	Парадокс Бернштейна	54
1.7.3	Вероятность объединения независимых событий	55
1.7.4	Мнемоническое правило	57
2	Случайные величины	60
2.1	Дискретные случайные величины	60
2.1.1	Распределения дискретных величин	60
2.1.2	Интегральная функция распределения	63
2.1.3	Плотность вероятностей	65
2.1.4	Связь дискретности и непрерывности	67
2.2	Статистические средние	69
2.2.1	Математическое ожидание	69
2.2.2	Среднее по ансамблю	72
2.2.3	Средние функций случайных величин	74
2.2.4	Моменты и дисперсия	76
2.2.5	Характеристическая функция	80
2.2.6	Свойства характеристической функции	83
2.3	Испытания Бернулли	86
2.3.1	Независимые испытания	86
2.3.2	Биномиальное распределение	87
2.3.3	Закон редких событий	92
2.3.4	Закон Пуассона в физике	96
2.4	Законы большого числа испытаний	97
2.4.1	Локальная теорема Муавра-Лапласа	97
2.4.2	Интегральная формула Муавра-Лапласа	102
2.4.3	Закон Больших Чисел	104
2.5	Примеры непрерывных величин	105
2.5.1	Распределение Гаусса	106
2.5.2	Экспоненциальное и родственные распределения	109
2.6	Функции случайных величин	115
2.6.1	Строго монотонные функции	115
2.6.2	Нарушение строгой монотонности	123
2.6.3	Немонотонные функции	124

3	Совокупности случайных величин	128
3.1	Совместные вероятностные свойства	128
3.1.1	Интегральная функция распределения	128
3.1.2	Совместная плотность вероятностей	130
3.1.3	Статистические средние	131
3.1.4	Общие свойства совместных функций распределения	132
3.1.5	Условные функции распределения	133
3.1.6	Независимые случайные величины	137
3.2	Функции совокупностей	139
3.2.1	Интегральная функция распределения	139
3.2.2	χ^2 -распределение	142
3.2.3	Распределение суммы	145
3.2.4	Преобразование распределений с помощью дельта-функций	147
3.2.5	Распределение Стьюдента	149
3.3	Отображения совокупностей	152
3.3.1	Взаимно-однозначное отображение	152
3.4	Корреляции случайных величин	157
3.4.1	Корреляция и ковариация	157
3.4.2	Ковариационная матрица	160
3.5	Еще раз о Законе Больших Чисел	162
3.5.1	Неравенство Чебышева	162
3.5.2	Закон Больших Чисел в форме Чебышева	164
3.5.3	Построение гистограмм	165
3.6	Характеристическая функция совокупности случайных ве- личин	169
3.6.1	Многомерные характеристические функции	169
3.6.2	Кумулянты	171
3.6.3	Центральная предельная теорема	175
3.7	Гауссовы совокупности	179
3.7.1	Определение гауссовой совокупности	179
3.7.2	Необходимые сведения из теории матриц	180
3.7.3	Плотность вероятностей гауссовой совокупности	183
3.7.4	Эллипсоиды рассеяния	185
3.7.5	Двумерное гауссово распределение	189
3.7.6	Условные гауссовы распределения	193
3.7.7	Характеристическая функция гауссовой совокупности	197

Глава 1

Основания теории вероятностей

В этой главе читатель ознакомится с ключевыми понятиями теории вероятностей, необходимыми для решения проблем, возникающих в самых разнообразных инженерных, экономических, и многих других областях человеческой деятельности.

1.1 Случайные явления

Данный курс посвящен изучению теории вероятностей, случайных процессов, математической статистики и их приложениям в науке и инженерии. Но прежде чем перейти к непосредственному обсуждению вероятностных и статистических закономерностей, напомним причины, по которым теория вероятностей и статистика играют такую важную роль в современной науке и технике.

1.1.1 Испытание и случайное событие

Что же изучает теория вероятностей? До ее возникновения основным объектом исследования физиков, химиков и инженеров были явления, в которых условия опыта однозначно определяли его исход. Например считалось неоспоримым, что если на камень действует заданная сила, а его координаты и скорость в начальный момент известны, то дальнейший полет камня однозначно описывается уравнением Ньютона. Этот детерминистский подход, однако, не всегда адекватно отражал реальную ситуацию. Возьмем к примеру полет пули. Опытным стрелкам хорошо известно, что ее траектория от выстрела к выстрелу разная. Дело в том, что начальная скорость пули, в силу многих причин, для разных выстрелов различна. Кроме того невозможно предугадать порывы ветра в

разных участках траектории пули. В итоге, даже одинаково прицеливаясь, стрелок попадает в разные точки мишени.

Пока вариации начальной скорости и окружающая среда слабо влияют на итоги стрельбы (дают отклонения пули, много меньшие размера мишени), детерминированная модель полета пули вполне оправдана. В противном случае от нее приходится отказываться и искать иные, более реалистичные, способы предсказания результатов стрельбы.

Подобная неопределенность исхода опыта при соблюдении основных условий его проведения наблюдается для очень широкого круга явлений. Так, подбрасывая монету или игральный кубик, мы не можем заранее угадать исход. Различны значения тока, протекающего в одних и тех же условиях через электрическую цепь. Все это происходит потому, что влияние множества разнообразных неконтролируемых факторов, каждый из которых в отдельности может быть и мало влияет на результат опыта, в сумме приводит к непредсказуемому исходу. Подобные явления, исход которых не удается предсказать заранее, называют *случайными*. Систематическим изучением подобных явлений и занимаются теория вероятностей и математическая статистика.

Читателю уже ясно, что истоком случайных явлений служат неконтролируемые факторы, влияющие на поведение изучаемого явления, и в конечном итоге приводящие к непредсказуемому заранее – случайному – исходу. С другой стороны имеется набор условий, которые удается довольно точно контролировать, такие как направление ружья и начальная скорость пули. Наблюдения, проводимые при некотором комплексе контролируемых условий, который удается многократно воссоздать, называют в теории вероятностей *статистическим испытанием (опытом, экспериментом)*. Всякий исход испытания при неизменных контролируемых условиях называют *случайным событием* или просто *событием*.

К хорошо известным в повседневной жизни случайным событиям относятся: выпадение герба на монете, трех очков при бросании кубика, и иные самые разнообразные события, список которых читатель легко может пополнить сам.

В дальнейшем условимся обозначать случайные события большими латинскими буквами A , B , C , D и так далее. Оговоримся однако, что некоторые из них зарезервированы за ключевыми понятиями теории вероятностей. Прежде всего это относится к букве P , которой принято обозначать *вероятность* случайного события. Происхождение обозначения очевидно: это первая буква слова Probability – Вероятность.

Заметим еще, что случайные события, особенно в инженерных приложениях, как правило выражают *случайными величинами*. Сюда относятся уже упомянутые случайная величина тока, сила порывов ветра, ме-

ханические напряжения в конструкциях зданий и многие многие другие. Более того, при количественном описании статистических закономерностей самых разнообразных случайных событий, стараются сопоставить им те или иные случайные величины. Последние принято обозначать большими латинскими буквами, привычно ассоциирующимися с математическими переменными. Например, буквами X , Y , Z , а также N и M для целочисленных случайных величин.

1.1.2 Статистическая устойчивость

Прежде чем определить ключевое понятие вероятности случайного события, сделаем несколько важных замечаний. Инженерам, применяющим в своей практике теорию вероятности и статистику, полезно сознавать, что не все случайные явления, с которыми приходится повседневно сталкиваться, поддаются вероятностному и статистическому описанию. Чтобы пояснить это важное утверждение, напомним основной принцип физической теории. Как известно, ее объектом могут служить лишь *устойчивые* явления, которые можно экспериментально обнаружить любое число раз, достаточно тщательно соблюдая условия эксперимента. Случайными же явлениями называют те, что на первый взгляд не удовлетворяют этому фундаментальному условию воспроизводимости в эксперименте. В самом деле, при достаточно аккуратных измерениях скорости течения газа в трубке или воды в реке, измеренные значения ее скорости, из-за непредсказуемых турбулентных пульсаций, будут от замера к замеру разными.

Другим физическим примером такого рода служит явление радиоактивного распада. Как известно, атом радиоактивного вещества, например радия, рано или поздно распадается на ядра более легких элементов. Но наблюдая за отдельным атомом в течение заданного промежутка времени длительностью τ , невозможно заранее с достоверностью предсказать, распадется или нет атом по прошествии данного времени.

Тем не менее, множество встречающихся в инженерной практике случайных явлений обладает замечательным свойством *статистической устойчивости*. Поясним его на элементарном примере.

Пример: Сопоставим со случайным явлением A некоторую случайную величину I – так называемый *индикатор* случайного явления. Применительно к атому радиоактивного элемента будем считать ее равной единице, если атом распался за время наблюдения τ , и нулем в противном случае. Из сказанного выше ясно, что невозможно заранее предсказать

в каждом отдельном наблюдении, какое значение примет величина I . Однако если сделать большое число измерений (в нашем случае – наблюдения за распадом $N \gg 1$ атомов за данный промежуток времени τ), измерить среднее арифметическое данных всех наблюдений

$$\bar{I}_N = \frac{1}{N} \sum_{k=1}^N I_k, \quad (1)$$

а затем вновь провести серию из N наблюдений и найти новое среднее \bar{I}'_N , то обнаружится, что различие между \bar{I}_N и \bar{I}'_N оказывается довольно малым. Более того, при увеличении числа N наблюдений в разных сериях, различие между \bar{I}_N и \bar{I}'_N становится все меньше и меньше, и стремится к нулю при безграничном увеличении числа наблюдений.

Именно в силу статистической устойчивости случайного явления радиоактивного распада, физики оперируют вполне определенным понятием – *периода полураспада* – относящимся к радиоактивному образцу, состоящему из громадного числа атомов радиоактивного вещества. ★

Указанное свойство исчезновения различия между средними, полученными в разных сериях из многих испытаний, и называют статистической устойчивостью. Это свойство присуще многим, встречающимся в инженерной практике, случайным явлениям и процессам.

Свойство статистической устойчивости случайных явлений и величин позволяет построить для них теорию, предсказывающую результаты многократно воспроизводимых, при неизменных условиях, испытаний. Только объектом предсказания теории здесь являются не индивидуальные значения сопутствующих тем или иным случайным явлениям случайных величин, а их различные средние, определенные на совокупности большого числа испытаний. В итоге, как и обычная физическая теория, теория вероятностей оперирует со вполне определенными, детерминированными, экспериментально измеримыми величинами: средними, вероятностными распределениями, корреляционными функциями и т.д.

Замечание: Заметим однако, что вокруг нас совершается множество самых разнообразных случайных явлений, не поддающихся вероятностному анализу. К ним относятся единичные случайные явления, такие как встреча близкого друга на улице незнакомого города, или вовремя пришедшая в голову удачная мысль. К явлениям подобного рода относятся большинство исторических событий. В самом деле, теория вероятностей неуместна, когда речь идет о Наполеоне: историческая наука еще не настолько всеильна, чтобы многократно воспроизвести условия

его рождения в соответствующей исторической обстановке. Теория вероятностей мало что может сказать даже по поводу несчастного случая – падения кирпича на голову Пети, гулявшего около строящегося дома. Просто потому, что с тех пор Петя обходит стройки далеко стороной. Подобные и другие единичные события, вроде падения громадного астероида на землю или явления Христа народу, к которым неприменимо понятие статистической устойчивости, лучше называть не случайными, а чудесными. При описании их теория вероятностей бессильна. ■

1.1.3 Относительная частота и вероятность

Как уже понял читатель, теория вероятностей занимается изучением статистически устойчивых случайных событий. Сама же статистическая устойчивость того или иного события принципиально недоказуема. Поэтому математики принимают ее за аксиому. Физики же и инженеры пытаются убедиться в наличии этого свойства у конкретного случайного события, скажем события A , чисто эмпирически. При этом они опираются на понятие *относительной частоты* появления случайного события. Введем его, переписав соотношение (1) в более прозрачном виде. Применительно к произвольному событию A (не обязательно радиоактивному распаду) правая часть равенства (1) равна отношению M/N , где M – случайная величина, равная числу испытаний из общего числа N , при которых реализовалось событие A . Для указанного отношения вводят специальное обозначение P^* и называют его *относительной частотой* появления события A в серии из N испытаний. В более аккуратных обозначениях

$$P^*(A, N) = \frac{M}{N}, \quad (2)$$

где $P^*(A, N)$ – относительная частота появления случайного события A в N испытаниях. Если обсуждаемое случайное событие статистически устойчиво, относительная частота становится, с увеличением числа испытаний N , все менее случайной, а при безграничном увеличении N стремится к некоторой детерминированной величине $P(A)$. На математическом языке этот факт записывают в виде

$$\lim_{N \rightarrow \infty} \frac{M}{N} = \lim_{N \rightarrow \infty} P^*(A, N) = P(A). \quad (3)$$

При этом детерминированную величину $P(A)$ как раз и называют *вероятностью* случайного события A .

Вычисление вероятности некоторого случайного события как предела (3) называют *статистическим определением* вероятности. Будучи

вполне приемлемым для экспериментаторов, оно вызывает у математиков множество вопросов. Прежде всего, в каком смысле понимать стремление к пределу? Совершенно очевидно, что здесь неприемлемы принятые в математическом анализе критерии сходимости. Кроме того, поскольку каждое испытание занимает определенное время, точное значение вероятности случайного события в принципе не может быть найдено, так как вычисление предела (3) требует бесконечного времени.

Математики частично снимают проблемы, связанные со статистическим определением вероятности (3), вводя специальные виды сходимости. В дальнейшем мы обсудим вероятностные подходы к понятию сходимости случайных последовательностей. А пока заметим только, что многочисленные испытания демонстрируют удивительную устойчивость самых разнообразных случайных явлений.

К примеру, относительная частота рождения мальчиков в любой наперед выбранной стране очень близка к 0.515, и это значение слабо зависит от местности или этнического состава населения. Относительная частота распада атома ^{226}Ra за 100 лет составляет 0.04184. В последнем случае число испытаний еще более велико, так как совпадает с числом атомов радия в типичном образце ($N \simeq 10^{23} \div 10^{24}$).

1.1.4 Бросание монеты и статистическое моделирование

Чтобы лучше свыкнуться с ключевым понятием статистической устойчивости случайных событий, детально обсудим классический эксперимент с подбрасыванием симметричной монеты. Его возможные исходы состоят в падении монеты на стол орлом или решкой вверх. Будем интересоваться частотой случайного события A - выпадения орлов. Иными словами примем, что индикатор I принимает значение 1, если выпал орел, и 0 в противном случае.

К бросанию монеты прибегали многие, желавшие удостовериться в наличии статистической устойчивости этого, очевидно случайного, события. Один из основоположников теории вероятностей, француз Бюффон еще в XVIII веке произвел 4040 бросаний монеты. Герб выпал 2048 раз, что дало

$$P^*(A, N = 4040) \approx 0.5069$$

К известным людям, занимавшимся в научных целях игрой в орла или решку, относится знаменитый специалист по статистике Пирсон, не поленившийся подбросить монету $N = 12000$ раз, и получивший выпадение орлов в $M = 6019$ бросаниях. Соответственно, относительная

частота выпадения орлов получилась равной $P^* \approx 0.5016$. Затем Пирсон подбросил монету уже $N = 24000$ раз. В итоге орлы выпали в $M = 12012$ случаях, что дает $P^* \approx 0.5005$.

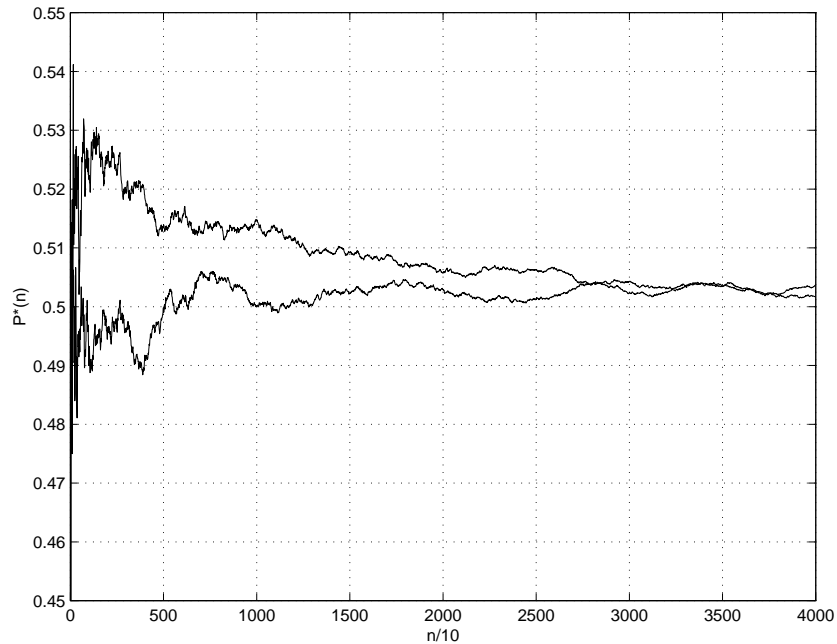


Рис. 1.1

Типичный график зависимости относительной частоты единиц, в моделируемой генератором случайных чисел случайной последовательности, от числа элементов последовательности. График построен для двух случайных последовательностей, каждая из которых содержит $N = 40000$ элементов. Хороша видна тенденция обоих графиков приблизиться к вероятности выпадения орла при бросании монеты $P = 0.5$.

Видно, что полученные в разных сериях испытаний и в разные века, относительные частоты выпадения орла мало отличаются друг от друга и близки к величине

$$P(A) = \frac{1}{2}, \quad (4)$$

которую и принимают за вероятность выпадения орла при бросании симметричной монеты.

В наши дни нетрудно проверить равенство (4) еще аккуратнее, чем Бюффон и Пирсон, опираясь на метод *статистического моделирования* и пользуясь мощностью наших компьютеров. При статистическом моделировании роль падающей монеты играет *генератор случайных чисел*.

Он создает последовательность нулей и единиц, которую вполне можно считать совершенно случайной. Такой генератор “вшит” в любую маломальски серьезную математическую программу. Например в известной программе MATLAB вызов случайного нуля или единицы осуществляется командой:

```
>> randint
```

На Рис. 1 изображен смоделированный компьютером график относительной частоты $P^*(N)$ выпадения орла в зависимости от числа испытаний, для двух серий испытаний, каждая “длиной” $N = 40000$. В отличие от Бюффона и Пирсона, которым требовались многие часы для бросаний монеты, процедура вычислений и построения графика на компьютере заняла всего несколько минут.

1.1.5 Математический взгляд на вероятность

Мы уже определили понятие вероятности случайного события $P(A)$ как предела (3) относительной частоты. В дальнейшем выяснится однако, что в большинстве встречающихся на практике ситуаций имеются более экономичные и эффективные способы вычисления вероятностей тех или иных случайных событий. Но чтобы освоить их, требуется ясное понимание математической структуры вероятности.

С позиций математиков вероятность $P(A)$ это функция. Только в отличие от стандартных функций, например $y = \sin x$, отображающих точки числовой оси x на числовую же ось y , аргументами вероятности служат всевозможные случайные события $\{A, B, C, \dots\}$. При этом значениями вероятностной функции $P(A)$, как и обычных функций, являются числа. В математическом анализе, при изучении любой функции, начинают с нахождения ее областей определения и значений, а затем приступают к исследованию свойств самой функции. Исследование области определения функции $P(A)$ и ее свойств, по сути центральный вопрос теории вероятностей, требующий вдумчивого анализа, и мы посвятим ему много времени и места ниже. Область же значений функции $P(A)$ уже сейчас легко указать, опираясь на статистическое определение вероятности (3). Действительно, поскольку число испытаний M , при которых реализовалось случайное событие A , удовлетворяет очевидным неравенствам

$$0 \leq M \leq N,$$

то и вероятность любого события, согласно (3), подчиняется неравенствам

$$0 \leq P(A) \leq 1. \quad (5)$$

Другими словами функция вероятности отображает всевозможные случайные события на единичный отрезок числовой оси $[0, 1]$. При этом если вероятность какого-то события равна единице, то такое событие иногда называют *достоверным*. Если же вероятность события равна нулю, говорят о *невозможном* событии.

Замечание: Полезно иметь в виду, что не всякое достоверное в указанном смысле событие реализуется в любом испытании, а невозможное событие порой происходит. В самом деле, событие, встречавшееся конечное число раз в бесконечной череде испытаний, согласно использованной терминологии, невозможно. По данной терминологии невозможно любое событие обыденной жизни. Убедитесь в этом сами, попробовав попасть с закрытыми глазами тонкой иглой в заранее помеченную на стене точку. Наверняка промажете, хотя попадете в другую, тем самым реализовав иное “невозможное” событие. В дальнейшем мы строго определим невозможное и достоверное события. Обсужденные же выше события будем называть скромнее: событиями с нулевой и единичной вероятностью. ■

1.2 Схема шансов

Как уже упоминалось, статистическое определение (1.3) образует слишком зыбкий фундамент для построения здания теории вероятностей: Опыт показывает, что значения относительных частот появления одного и того же события даже при большом числе испытаний, пусть незначительно, отличаются друг от друга. В итоге определить точное значение вероятности оказывается невозможным. Кроме того, чтобы вычислить вероятность каждого случайного события, метод статистических испытаний предписывает проведение громадного числа испытаний, что неприемлемо, а порой просто невозможно, для большинства практически значимых случайных событий. Поэтому развитие теории вероятностей пошло по другому пути. Он сводится к анализу связей между вероятностями реальных случайных событий, и более простых событий, вероятности которых, как правило, известны заранее.

Исторически, первым шагом на этом пути стал анализ азартных игр, и прежде всего игры в кости. Кто-то справедливо заметил, что азартные игры дали науке гораздо больше, чем игра в шахматы. Исследование игры в кости позволит нам ввести ключевые понятия теории и установить основные свойства вероятностей.

1.2.1 Игра в кости

Приступим к изучению области определения вероятностной функции $P(A)$ и ее свойств. Значительно продвинуться в этом направлении поможет детальный анализ игры в кости. В простейшем варианте она сводится к бросанию игроками игральной кости, вроде тех, что изображены на Рис. 2. Выигрывает тот, у кого выпало большее число очков.

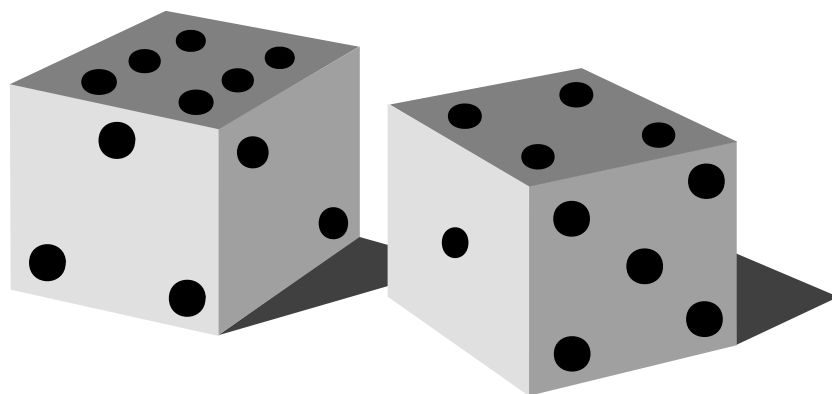


Рис. 1.2

Общепризнанный символ теории вероятностей: игральные кости, сильно способствовавшие становлению теории вероятностей как важной области науки.

Будем пока бросать одну игральную кость. Пусть нас интересует вероятность выпадения, при ее бросании, наибольшего числа очков $k = 6$, но нет времени оценить вероятность статистическим методом, многократно подбрасывая кость. Разрешить проблему помогает мысленный анализ возможных исходов испытаний (бросаний кости). Ясно, что имеется множество Ω , состоящее из шести событий – всевозможных вариантов падения кости

$$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}, \quad (1)$$

где событие ω_k состоит в том, что при бросании кости выпало k очков. Это так называемые *элементарные события*, которые только и могут произойти с игральной костью. Поскольку кость никогда не встает на ребро или вершину, перечисленные события называют *полной группой* элементарных событий, отвечающих испытанию (бросанию кости).

Если кость симметрична, то указанные события (1) могут произойти в любом испытании с равной вероятностью, а значит события (1) *равновозможны*. Кроме того, все они взаимоисключают друг друга: Если в

испытании произошло одно из событий (1), то заведомо не реализуются остальные. По этим причинам события (1) называют *несовместными*. Таким образом множество событий (1) образует *полную группу равно-возможных несовместных элементарных событий*.

Если при анализе возможных исходов какого-либо испытания удастся выделить подобную полную группу равновозможных несовместных элементарных событий, то каждое из них называют кратко: *шансом*. Пусть число шансов, отвечающих некоторому испытанию, равно n . Интуитивно ясно, что в длинной серии N испытаний каждый из шансов реализуется примерно одинаковое число $M \simeq N/n$ раз. А если так, то вероятность любого шанса естественно вычислять с помощью простой формулы

$$P(\omega) = \frac{1}{n}. \quad (2)$$

Пример: При бросании, сделанной из однородного материала, тонкой монеты, неспособной становиться на ребро, имеется всего $n = 2$ шанса – выпадение орла или решки:

$$\Omega = \{\omega_h, \omega_t\}.$$

Следовательно вероятность выпадения орла равна $1/2$. ★

Пример: Еще одним классическим примером шанса служит извлечение бубнового валета из тщательно перемешанной колоды, содержащей 52 карты. В данном случае всего шансов $n = 52$, а значит вероятность прикупить бубнового валета оказывается равной $1/52$. ★

Наконец, интересовавшее нас вначале выпадение шестерки (как и любого другого числа очков) при бросании игральной кости равно $1/6$.

1.2.2 Формула шансов

До сих пор мы обсуждали лишь элементарные события. На практике чаще интересуются вероятностями более сложных, составных, событий. Так игрок, поставивший при бросании кости на чет (выпадение четного числа очков), легко оценит вероятность благоприятного исхода. Для этого ему надо поделить число шансов с четным числом очков на полное их число. Количество указанных шансов в данном случае $m = 3$ (выпадение 2, 4, и 6). А значит искомая вероятность равна

$$P(\text{even}) = \frac{3}{6} = \frac{1}{2}.$$

В общем случае вычисление вероятности некоторого случайного события A методом шансов сводится к следующей процедуре: Вначале пытаются выделить полную группу шансов, отвечающих данному событию. Если это удастся сделать, указывают все те из них, реализация которых в испытании равносильна событию A . Такие шансы называют *благоприятствующими* событию A . Пусть число благоприятствующих шансов равно m . Поделив их на полное число шансов n , приходим к формуле вычисления вероятности случайного события по схеме шансов:

$$P(A) = \frac{m}{n}. \quad (3)$$

Замечание: В частном случае одного, благоприятствующего событию A , шанса формула (3) вырождается в (2). При этом событие A не обязательно идентично шансу. Например, это может быть выигрыш, при выпадении шестерки, ста долларов ($A = \$100$). ■

Замечание: Обратим внимание, что схема шансов вычисляет вероятность выпадения орла (1.4) куда быстрее, чем даже статистическое моделирование с помощью компьютера. Уже этот простой пример убедительно демонстрирует потенциальную эффективность исчисления вероятностей с помощью схемы шансов, по сравнению с методом статистических испытаний. ■

Замечание: Заметим однако, что схема шансов почти всегда связана с теми или иными идеализациями и упрощениями. К примеру, вычисляя с помощью формулы (2) упомянутую вероятность выпадения орла, мы полагаем, что обе стороны монеты абсолютно равноценны. Очевидно, это не совсем так. Более того, встречаются монеты со смещенным центром тяжести, падающие орлом и решкой с разной частотой. Можно полагать также, что упорная тренировка или специальные приемы бросания монеты, могут сделать ее разные стороны неравноправными. В таких ситуациях схема шансов дает осечку. Все это надо иметь ввиду, пользуясь идеализированной схемой шансов. ■

1.2.3 Бросание двух костей

Читатель уже убедился в потенциальной эффективности схемы шансов при исчислении вероятностей разных случайных событий. Поэтому надеемся, что он с интересом отнесется к дополнительным примерам, иллюстрирующим ее возможные применения, и позволяющим набить руку для использования схемы шансов в самых разнообразных ситуациях.

В средние века, когда игра в кости была чрезвычайно распространена, бросали обычно не одну а как минимум две игральные кости, и учи-

события	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	A_{11}	A_{12}
шансы	1+1	1+2 2+1	1+3 2+2 3+1	1+4 2+3 3+2 4+1	1+5 2+4 3+3 4+2 5+1	1+6 2+5 3+4 4+3 5+2 6+1	2+6 3+5 4+4 5+3 6+2	3+6 4+5 5+4 6+3	4+6 5+5 6+4	5+6 6+5	6+6
m	1	2	3	4	5	6	5	4	3	2	1
P	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$

Таблица 1.1: Подсчет вероятностей выпавших очков при бросании двух игральных костей.

тывали суммарное число очков, выпавших на обеих костях. В данном случае в испытаниях реализовывались 11 событий $\{A_2, \dots, A_{12}\}$, состоящие в том, что выпало соответственно 2, 3 или 12 очков. Поскольку при игре в кости делались значительные ставки, игрокам было небезынтересно знать, какова вероятность выпадения того или иного числа очков. Например числа 10, как в ситуации, изображенной на Рис. 2.

Проблема заключается в том, что перечисленные события неравноправные и составные: Каждое из них реализуется при нескольких разных комбинациях более простых равноправных событий -выпадениях, при бросании двух костей, соответственно, k и l числа очков. Поэтому задачу исчисления вероятностей событий $\{A_2, \dots, A_{12}\}$ удастся решить, лишь явно указав группу шансов – несовместных равновозможных элементарных событий, и подсчитав, сколько из них благоприятствует тому или иному суммарному числу очков.

Очевидно шансами в данном случае являются события ω_{kl} , состоящие в том, что на первой кости выпало k очков, а на второй l очков ($k = 1, \dots, 6; l = 1, \dots, 6$). Всего таких шансов $n = 6 \times 6 = 36$.

Число шансов m , благоприятствующих тому или иному суммарному числу очков, здесь удобно подсчитать вручную, простым перебором всех возможных ситуаций. С этой целью полезно составить таблицу, в которой под каждым событием выписаны благоприятствующие ему шансы.

Из Таблицы 1 видно, что наиболее вероятно выпадение суммарного числа очков 7. Может быть это одна из веских причин, по которым число 7 считают счастливым? Заметим еще, что вероятность выпавшего на Рис. 2 “соседнего” числа очков $P(A_8) = 5/36 \approx 0.139$ заметно меньше

вероятности выпадения “счастливого” числа $P(A_7) = 1/6 \approx 0.167$

Замечание: Иногда при построении Таблицы 1 спрашивают – почему большинство комбинаций двух чисел появляется в таблице дважды (например $1 + 3$ и $3 + 1$), а некоторые (к примеру $2 + 2$), лишь один раз? Вопрос не так наивен, как может показаться. Дело в том, что при подсчете шансов мы (явно или неявно) полагали, что кости различимы и разделяли их на первую и вторую. В физике встречаются принципиально неразличимые элементарные частицы, для которых шансы находиться в том или ином состоянии приходится считать по другому. ■

1.2.4 Исчисление вероятностей средствами комбинаторики

Использованный при решении предыдущей задачи непосредственный перебор всех благоприятствующих шансов интересующих событий, порой чересчур громоздок или вообще нереален, если число объектов, участвующих в испытании (например тех же игральные кости), велико или произвольно. К счастью, во многих подобных ситуациях общий ответ удастся выразить с помощью комбинаторных формул. Поэтому остановимся на них подробно.

Нагляднее всего суть формул комбинаторики демонстрируют задачи об извлечении шаров из урны. Возьмем простейшую задачу такого рода: Пусть в урну помещены N шаров, а вынимают один из них. Спрашивается – сколькими способами это можно сделать? Как ни удивительно на первый взгляд, но возможны разные ответы. Так если шары *различимы*, к примеру пронумерованы, то имеется N способов извлечь шар: под номером 1, 2 и так далее. Если же все шары абсолютно похожи, например все белые и без номеров, то исход испытания всегда один – извлечение шара белого цвета. Таким образом, чтобы дать определенный ответ, требуется дополнительная информация о различимости шаров.

Будем полагать, что все шары различимы (пронумерованы) и выясним, сколькими способами можно извлечь R из них. Поскольку роль шаров могут играть предметы совершенно произвольной природы: яблоки, члены спортивной команды или даже разные взгляды на жизнь, используем общую терминологию. Будем говорить об *упорядоченной выборке* R элементов из *генеральной совокупности* элементов

$$\{b_1, b_2, \dots, b_N\}$$

объема N . Вне зависимости от природы объектов генеральной совокупности, итог упорядоченной выборки можно представить в виде последо-

вательности чисел – номеров последовательно выбранных объектов.

Упорядоченные выборки с возвращением

Обсудим две процедуры извлечения объектов из генеральной совокупности. Первую из них называют *упорядоченной выборкой с возвращением*, когда записывают номер вынутого из урны шара, а затем кладут его обратно в урну. В этом случае один и тот же номер может встретиться в выборке несколько раз – столько, сколько раз вынимается тот же самый шар.

Сосчитаем число всевозможных различных упорядоченных выборок с возвращением R шаров из N . Для этого заметим, что первый шар можно вынуть N различными способами. Поскольку его возвращают в урну, то и второй шар можно вынуть N способами и так далее. В итоге, полное число различных упорядоченных выборок равно

$$n = \underbrace{N \times N \times \dots \times N}_R = N^R. \quad (4)$$

Если шары вынимают наугад, то число n (4) – это сумма шансов, сопутствующих испытанию – извлечению упорядоченной последовательности R различных шаров из N , а вероятность реализации любого из шансов равна

$$P = \frac{1}{N^R}.$$

Так при бросании монеты ее стороны выполняют роль “шаров”, “извлекаемых” при бросании. А значит $N = 2$. Следовательно, вероятность выпадения всех орлов при R кратном бросании монеты равна $P = 2^{-R}$. Такова же вероятность того, что первый раз выпадет орел, а в остальных бросаниях решка.

Упорядоченные выборки без возвращения

Вторая естественная процедура выборки – упорядоченная выборка R шаров из N *без возвращения*, когда после очередного извлечения шара из урны в ней остается одним шаром меньше. А значит 1-й элемент выборки можно выбрать N различными способами, второй $N - 1$, а последний $(N - R + 1)$ способами. Таким образом полное число *размещений*: всевозможных различных упорядоченных выборок без возвращения R шаров из N равно

$$n = N(N - 1) \dots (N - R + 1) = \frac{N!}{(N - R)!}. \quad (5)$$

В частности при $R = N$ получаем количество *перестановок*

$$n = N! \tag{6}$$

Столькими различными способами, например, можно выстроить в шеренгу группу из N человек.

Пример: Уже приведенная пара комбинаторных формул (4) и (5) позволяет довольно глубоко исследовать интересный вопрос о несовпадающих днях рождения в группе из R студентов. Попробуем оценить вероятность что дни рождения всех R студентов разные.

Как и в реальных прикладных проблемах, чтобы получить ответ, надо сделать разумные допущения, чтобы выработать достаточно адекватную, и вместе с тем сравнительно простую, математическую модель. Такими допущениями примем, что количество дней в году равно $N = 365$, а люди с равной возможностью рождаются в любой день года.

Сформулируем поставленную математическую задачу на языке упорядоченных выборок. Для этого мысленно представим всевозможные дни рождения как шары в урне, пронумерованные по дням в году. Всего таких “шаров” 365, а значит имеется генеральная совокупность из $N = 365$ элементов.

Набор дней рождения R студентов, переписанный из стопки анкет, разложенных в алфавитном порядке, составляет упорядоченную выборку с возвращением, поскольку дни рождения разных школьников могут совпадать. Следовательно, полное число n разных комбинаций дней рождения – шансов – дано формулой (4). Число же благоприятствующих обсуждаемому событию шансов равно полному числу упорядоченных выборок без возвращения

$$m = N(N - 1) \dots (N - R + 1),$$

поскольку в соответствующих последовательностях дней рождения каждая дата встречается не более одного раза. Поделив число благоприятствующих шансов на полное их число, найдем искомую вероятность

$$P(R) = \left(1 - \frac{1}{365}\right) \left(1 - \frac{2}{365}\right) \dots \left(1 - \frac{R-1}{365}\right). \tag{7}$$

График этой функции целочисленного аргумента изображен ступеньками на Рис. 3. ★

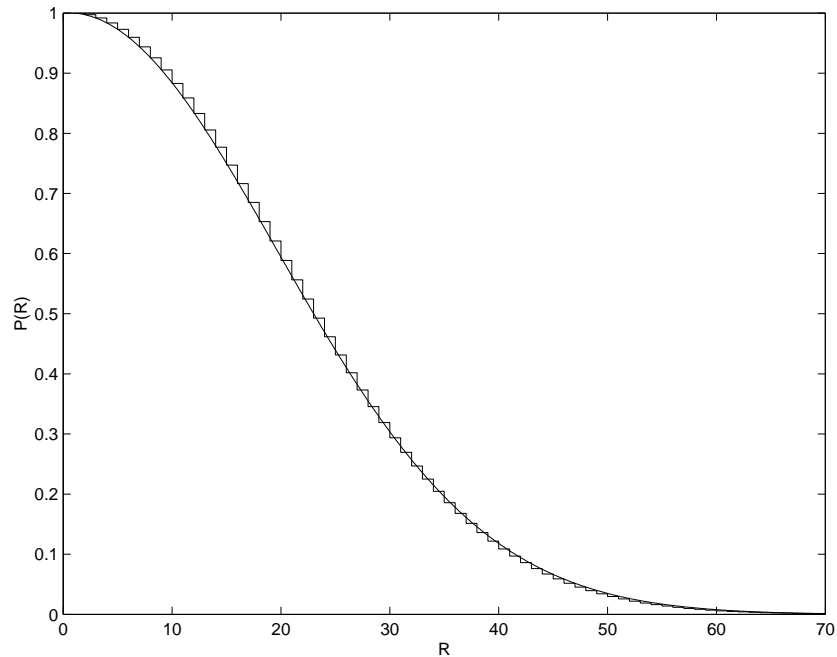


Рис. 1.3

График зависимости вероятности $P(R)$ (7) от числа школьников R в группе (ступенчатая линия). Гладкой линией изображен график, построенный по приближенной формуле (9).

Формула (7) вероятности несовпадения дней рождения “непрозрачна” и тем более громоздка, чем больше R . Поэтому часто пользуются более удобной аппроксимирующей формулой. Ее удастся получить, не внося большой погрешности, пока

$$R \ll N. \quad (8)$$

Тогда справедливы приближенные равенства

$$\left(1 - \frac{k}{365}\right) \approx \exp\left(-\frac{k}{365}\right) \quad (k = 1, 2, \dots, R).$$

Заменим с их помощью сомножители в (7) на экспоненты. В итоге, после суммирования арифметической прогрессии в показателе экспоненты, придем к аппроксимирующей формуле

$$P(R) \approx \exp\left(-\frac{R(R-1)}{730}\right). \quad (9)$$

Ее график, показанный на Рис. 3 гладкой линией, близок к ступенчатому графику функции (7) при любых R , где исследуемая вероятность

еще заметно отлична от нуля. Поэтому формула (9) хорошо работает не только при условии (8), а по сути для любых R , вплоть до $R = N$. В частности, для $R = 30$, формула (9) дает: $P(30) \simeq 0.304$. Следовательно, примерно в семи из десяти студенческих групп по 30 человек, найдутся студенты с одинаковыми днями рождения.

Замечание: В дальнейшем мы еще не раз воспользуемся плодотворным приемом замены близких к единице сомножителей экспонентами для вывода важных аналитических соотношений теории вероятностей. ■

Замечание: Выведенная выше формула (7) вероятности несовпадения дней рождения служит прекрасной иллюстрацией того, как математические (и вероятностные в том числе) выводы порой приходят в вопиющее противоречие со “здравым смыслом”. В самом деле, последний подсказывает, что вероятность несовпадений дней рождений в группе из 150 человек должна быть близка к $1/2$. Подсчет же по формуле (7) дает:

$$P(R = 150) \simeq 2.45 \times 10^{-16}.$$

Иначе говоря, за всю историю человечества вряд ли встречалась группа из 150 человек (первобытного племени, роты солдат или сотрудников небольшой фирмы), в которой все дни рождения были бы разные. Конечно если группа не подбиралась специально по этому признаку. ■

Сочетания

Вернемся к выборкам без возвращения. До сих пор, при подсчете их полного числа по формуле (5), принимался во внимание порядок, в каком элементы генеральной совокупности появлялись в выборке. К примеру, полагались разными выборки

$$(4, 7) \quad (7, 4) \tag{10}$$

четвертого и седьмого шаров, извлеченных из урны с семью шарами. Во многих ситуациях порядок извлечения элементов генеральной совокупности безразличен. Так, публикуя таблицу выигрышных лотерейных билетов, их номера указывают в порядке возрастания, а не в том, в котором их извлекли при розыгрыше лотереи. То же происходит при игре в карты, когда игрок раскладывает сданные ему карты по мастям и старшинству, а не в той последовательности, в какой они были сданы. Аналогично, считая голоса, поданные за того или иного кандидата в президенты, совершенно не учитывают, в каком порядке они были поданы.

Во всех подобных случаях разными естественно считать выборки, отличающиеся элементами, а не порядком их следования. Такие выборки без учета порядка следования элементов называют *сочетаниями*.

Чтобы найти полное число сочетаний R элементов из N , составляющих генеральную совокупность, поделим число упорядоченных выборок без возвращения (5) на количество всевозможных перестановок одних и тех же R элементов. Последнее, согласно (6), равно $R!$. Поделив (5) на $R!$, найдем, что число сочетаний R элементов из N равняется так называемому *биномиальному коэффициенту*:

$$C_N^R = \binom{N}{R} = \frac{N!}{R!(N-R)!}. \quad (11)$$

Гипергеометрическое распределение

Как и другие формулы комбинаторики, формулу (11) напрямую используют в теории вероятностей. Для иллюстрации решим следующую задачу:

Пример: Из колоды в 36 карт наугад вынимают три карты. Какова вероятность, что среди них окажется один и только один туз.

Число шансов в данном случае равно количеству всевозможных сочетаний 3-х карт из 36:

$$n = \binom{36}{3} = \frac{36 \times 35 \times 34}{1 \times 2 \times 3} = 7140.$$

Сосчитаем число шансов, благоприятствующих обсуждаемому событию. Для этого заметим, что один туз можно выбрать $\binom{4}{1} = 4$ способами. Оставшиеся две карты могут быть выбраны $\binom{32}{2} = 496$ способами. Таким образом, число благоприятствующих шансов равно

$$m = \binom{4}{1} \binom{32}{2} = 4 \times 496 = 1984,$$

а искомая вероятность такова

$$P = \frac{m}{n} = \frac{1984}{7140} \approx 0.278.$$

Иными словами, чуть чаще чем каждый четвертый раз, игрок, прикупающий три карты, получит туза. ★

Обратим внимание, что при решении задачи мы по сути разбили карты на два сорта: тузов и не тузов. Подобная ситуация, когда генеральная

совокупность состоит из элементов двух сортов, типична для приложений. Например, при производстве некоторых изделий, их подразделяют на качественные и бракованные, а при контроле качества интересуются числом бракованных в партии продукции, представленной на проверку. Возникающие при этом вероятностные проблемы решают с помощью *гипергеометрического распределения*.

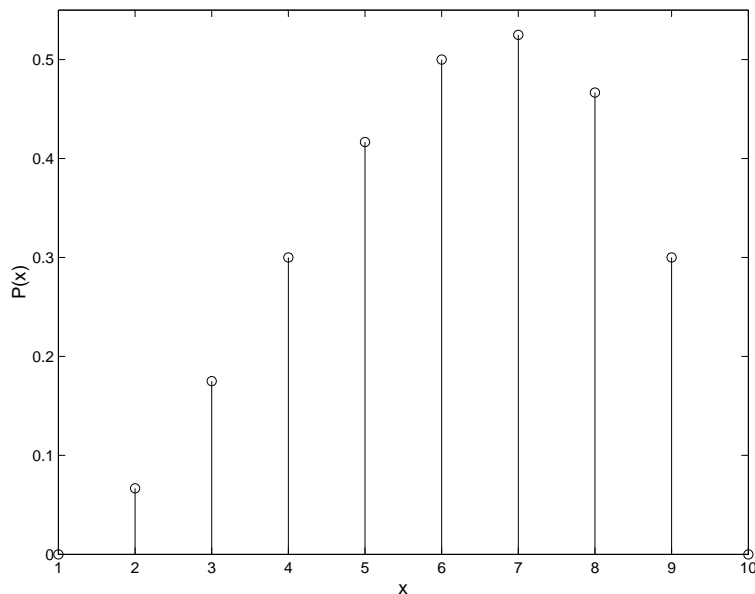


Рис. 1.4

График зависимости вероятности извлечения двух красных шаров из урны с семью зелеными и тремя красными шарами, в зависимости от числа x вынутых шаров.

Выведем его, пользуясь терминологией шаров, извлеченных из урны. Пусть в урне находится g зеленых шаров и r красных. Наугад вынимают a шаров. Требуется найти вероятность, что среди них ровно b красных ($a - g \leq b \leq \min\{a, r\}$). Рассуждения, аналогичные тем, которые привели нас к решению предыдущей задачи, дают

$$P_{r,g}(a, b) = \frac{\binom{r}{b} \binom{g}{a-b}}{\binom{r+g}{a}}. \quad (12)$$

Пример: Пусть в урне 7 зеленых шаров и 3 красных. Пользуясь гипергеометрическим распределением, выясним, сколько надо вынуть шаров,

чтобы вероятность извлечь ровно два красных шара была наибольшей. Обозначим число вынимаемых шаров за x ($x = a$). Тогда, согласно (12), зависимость вероятности извлечь два красных шара от числа извлеченных шаров описывается функцией

$$P(x) = 3 \binom{7}{x-2} \binom{10}{x}^{-1} \quad (13)$$

целочисленного аргумента x . График ее построен на Рис. 4. Для значений $x = 1$ и $x = 10$, при которых числитель в (13) не определен, использованы очевидные равенства $P(1) = P(10) = 0$. Из графика видно, что наибольшая вероятность извлечь два красных шара $P(7) = 0.525$ приходится на $x = 7$. При бóльших значениях аргумента вероятность уменьшается просто потому, что становится более вероятным извлечь не два, а все три красных шара. ★

1.3 Вероятность и ее аргументы

Мы уже знаем, что возможные значения вероятностной функции $P(A)$ образуют единичный интервал $P \in [0, 1]$. Приступим к обсуждению аргумента $P(A)$. Им может быть любое случайное событие A . Схема шансов предоставляет прекрасную возможность наглядно проиллюстрировать основные свойства аргумента функции $P(A)$ - всевозможных случайных событий, а также самой функции $P(A)$. Начнем с анализа событий.

1.3.1 События и множества

Напомним, схема шансов исходит из того, что имеется полная группа n несовместных равновозможных элементарных событий – шансов

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}, \quad (1)$$

с помощью которых выражается любое другое событие, какое только может произойти в испытании. При этом множество Ω (1) называют *пространством элементарных событий*, а любое его подмножество просто событием.

Чтобы описать случайные события математически, ознакомимся с основными понятиями теории множеств. Согласно ей каждое событие есть некоторое подмножество множества Ω элементарных событий. Так событие

$$A = \{\omega_3, \omega_5\}$$

есть множество, состоящее из благоприятствующих событию A элементарных событий ω_3 и ω_5 . В теории множеств этот факт записывают в виде

$$A = \omega_3 \cup \omega_5. \quad (2)$$

Здесь знак \cup выражает операцию *объединения* множеств. Иногда операцию объединения символизируют более привычным нам знаком $+$, записывая соотношение (2) в виде

$$A = \omega_3 + \omega_5,$$

а событие A называют *суммой* событий ω_3 и ω_5 .

Операция объединения применима не только к элементарным событиям, а к любым другим множествам. Так множество

$$C = A \cup B \quad (3)$$

состоит, по определению, из всех элементов, входящих хотя бы в одно из множеств A и B . К примеру, пусть даны события

$$A = \omega_3 \cup \omega_5 \quad \text{и} \quad B = \omega_2 \cup \omega_5 \cup \omega_7. \quad (4)$$

Тогда событие C (3) равно

$$C = \omega_2 \cup \omega_3 \cup \omega_5 \cup \omega_7.$$

В теории вероятностей, заимствуя операции теории множеств, пользуются несколько иной терминологией: говорят, что случайное событие $C = A \cup B$ произойдет в некотором испытании, если в нем реализуется хотя бы одно из событий A или B (или оба вместе).

Другая, важная для теории вероятностей, операция теории множеств, операция *пересечения* (*произведения*) событий. Ее обозначают символом \cap пересечения (или знаком умножения). По определению, событие

$$D = A \cap B \quad (\text{или} \quad D = A \times B)$$

имеет место тогда и только тогда, когда в испытании реализуются сразу оба события A и B . Так результатом пересечения событий (4) будет шанс $D = \omega_5$.

Замечание: Объединение множеств \cup обычно ассоциируют с союзом “или”, а пересечение \cap выражают союзом “и”. ■

Если события A и B несовместны, то им благоприятствуют группы шансов, не имеющие общих элементов. Тогда говорят, что пересечение A и B образует *пустое множество* и записывают этот факт в виде

$$A \cap B = \emptyset. \quad (5)$$

Здесь \emptyset символ *пустого множества*. В теории вероятностей именно его называют *невозможным событием*, не наблюдающимся ни в одном испытании. Напротив, событие Ω , объединяющее все шансы, называют *достоверным событием*, поскольку в любом испытании всегда реализуется хотя бы один шанс.

Нам понадобится еще операция *вычитания* множеств. Назовем событие E *разностью* событий A и B , и запишем данный факт в форме

$$E = A/B \quad \text{или более привычно:} \quad E = A - B, \quad (6)$$

если событие E состоит в том, что происходит событие A , но не происходит B . Очевидно, в случае (4) $E = \omega_3$.

Операция вычитания дает возможность определить еще одно важное понятие теории вероятностей, понятие *противоположного события*. А именно, событие

$$\bar{A} = \Omega - A. \quad (7)$$

называют *противоположным* событию A . Смысл названия станет ясен, если сообразить, что событие \bar{A} реализуется тогда и только тогда, когда не происходит событие A . Очевидно, справедливы следующие тождества теории множеств:

$$A \cup \bar{A} \equiv \Omega, \quad A \cap \bar{A} \equiv \emptyset, \quad A - \bar{A} \equiv A.$$

Подобные и более сложные алгебраические соотношения между различными событиями играют важную роль в теории вероятностей.

1.3.2 Алгебра событий

Обозначим множество всевозможных подмножеств пространства элементарных событий Ω (1), содержащее само пространство Ω и пустое множество \emptyset , буквой \mathcal{F} . Таким образом,

$$\Omega \subset \mathcal{F}, \quad \emptyset \subset \mathcal{F}. \quad (8a)$$

Кроме того очевидно, что

$$A \subset \mathcal{F} \quad \Longleftrightarrow \quad \bar{A} \subset \mathcal{F}. \quad (8b)$$

Очевидно также, что любые объединения и пересечения входящих в \mathcal{F} множеств также принадлежат множеству \mathcal{F} :

$$A \cup B \subset \mathcal{F}, \quad A \cap B \subset \mathcal{F}. \quad (8c)$$

Множество множеств \mathcal{F} , для всех элементов которого справедливы соотношения (8а-с), называют *алгеброй множеств* или *алгеброй событий*.

Замечание: Когда пространство Ω имеет n элементов, легко подсчитать число элементов алгебры событий \mathcal{F} . Сделаем это, сопоставив каждому элементу алгебры \mathcal{F} последовательность n нулей и единиц. Например

$$A = \underbrace{\{0, 0, 1, 0, 1, 1, \dots, 0, 1, 0\}}_n, \quad (9)$$

где единица в k -й позиции означает, что элементарное событие ω_k благоприятствует событию A . Если же нет, то на этой позиции стоит нуль. Во взаимно-однозначном соответствии между всевозможными последовательностями нулей и единиц длины n и элементами множества \mathcal{F} легко убедиться, заменив единицы соответствующими элементарными событиями ω_k . Причем последовательность из одних только нулей отвечает невозможному событию, а из единиц – достоверному событию:

$$\emptyset = \underbrace{\{0, 0, \dots, 0\}}_n, \quad \Omega = \underbrace{\{1, 1, \dots, 1\}}_n.$$

Следовательно, число элементов \mathcal{F} равно числу всевозможных последовательностей n нулей и единиц. Всех таких последовательностей 2^n . Действительно, есть два варианта выбора первого числа (0 или 1), два варианта выбора второго и так далее. Следовательно, всего вариантов

$$\underbrace{2 \times 2 \times \dots \times 2}_n = 2^n.$$

Это и есть число элементов алгебры событий \mathcal{F} . В частности, бросанию игральной кости отвечает 6 элементарных событий, а значит алгебра \mathcal{F} состоит из $2^6 = 64$ событий. При бросании одной монеты ($n = 2$), может реализоваться $2^2 = 4$ события:

$$\mathcal{F} = \{\emptyset, \text{орел}, \text{решка}, \Omega\}, \quad (10)$$

а при бросании 2-х монет $2^4 = 16$ событий. Полезно для упражнения выписать и прокомментировать все элементы этой алгебры событий. ■

Замечание: Если кому-то невозможное событие \emptyset , содержащееся в алгебре (10), все еще кажется слишком абстрактным, пусть пока считает, что это падение монеты на ребро (или рассыпание монеты в пыль). ■

1.3.3 Свойства вероятностей

До сих пор, при исчислении вероятностей разных событий, мы прибегали к схеме шансов. Ее возможности не очень велики. Во многих реальных ситуациях невозможно не только сосчитать, но даже указать все шансы того или иного события. Примером служит стрельба по мишени. Пуля пробивает в мишени отверстие, и существует бесчисленное множество возможных расположений отверстия на мишени. Тем не менее, ценность схемы шансов не ограничивается узким классом вероятностных задач, которые удается решить с ее помощью. Дело в том, что свойства вероятностной функции $P(A)$, которые нетрудно установить в рамках схемы шансов, и совершенно невозможно доказать в более общих ситуациях, тем не менее остаются справедливыми всегда, и их принимают за аксиомы теории вероятностей. Иными словами, схема шансов выполняет в теории вероятностей важную эвристическую роль.

Перечислим несколько таких свойств вероятностной функции, предельно очевидные в рамках схемы шансов. Это прежде всего уже упомянутое неравенство

$$0 \leq P(A) \leq 1, \quad (11)$$

вытекающее из формулы (2.3) и из того факта, что число m благоприятствующих любому событию шансов удовлетворяет неравенству

$$0 \leq m \leq n,$$

где n общее число шансов. Во-вторых, для достоверного события

$$P(\Omega) = 1. \quad (12)$$

Действительно, достоверному событию благоприятствуют все шансы: $m = n$. А значит, по формуле (2.3), справедливо равенство (12). В общем случае оно выражает одну из аксиом теории вероятностей: *Вероятность достоверного события равна единице.*

Аналогично, вероятность невозможного события равна нулю:

$$P(\emptyset) = 0, \quad (13)$$

поскольку ему благоприятствует нуль шансов ($m = 0$).

Пусть события A и B несовместны, тогда вероятность объединения этих событий равна

$$P(A \cup B) = P(A) + P(B) \quad (A \cap B = \emptyset) \quad (14)$$

сумме вероятностей. Это замечательное свойство вероятностей несовместных событий элементарно доказать в рамках схемы шансов: Пусть

событию A благоприятствует a шансов, а событию B – b шансов. Из несовместности событий A и B следует, что эти группы шансов не содержат общих элементов. А значит число шансов, благоприятствующих объединенному событию $C = A \cup B$, равно сумме шансов каждого из объединяемых событий ($c = a + b$). Таким образом, согласно схеме шансов, вероятность объединенного события C равна

$$P(C) = \frac{a + b}{n} = \frac{a}{n} + \frac{b}{n} = P(A) + P(B).$$

В общем случае равенство (14) недоказуемо и его принимают за аксиому. Это так называемая *аксиома конечной аддитивности* теории вероятностей. Функцию $P(A)$, обладающую свойством (14), называют *аддитивной функцией множеств*.

Пусть нам известна вероятность некоторого события A . Найдем вероятность противоположного ему события \bar{A} . Для этого достаточно заметить, что если событию A благоприятствует m шансов, то событию $\bar{A} = \Omega - A$ благоприятствуют все оставшиеся $n - m$ шансов. Следовательно

$$P(\bar{A}) = \frac{n - m}{n} = 1 - \frac{m}{n}$$

или, на вероятностном языке,

$$P(\bar{A}) = 1 - P(A). \quad (15)$$

Последнее равенство, справедливое для произвольного события A и противоположного ему события \bar{A} , мы уже можем доказать в самом общем случае, опираясь на аксиомы (12) и (14). Действительно, события A и \bar{A} несовместны, следовательно, согласно (14),

$$P(A \cup \bar{A}) = P(A) + P(\bar{A}).$$

С другой стороны, событие $A \cup \bar{A} = \Omega$ есть достоверное событие, вероятность которого равна единице. Таким образом

$$P(A) + P(\bar{A}) = 1 \quad \Longleftrightarrow \quad P(A) = 1 - P(\bar{A}).$$

Точно также, в общем случае удастся доказать равенство (13). В самом деле, поскольку события \emptyset и Ω несовместны, из аксиомы конечной аддитивности имеем

$$P(\emptyset \cup \Omega) = P(\emptyset) + P(\Omega).$$

Далее $\emptyset \cup \Omega = \Omega$. Следовательно, последнее равенство принимает вид:

$$P(\Omega) = P(\emptyset) + P(\Omega) \quad \Longrightarrow \quad P(\emptyset) = 0.$$

1.4 Геометрическое исчисление вероятностей

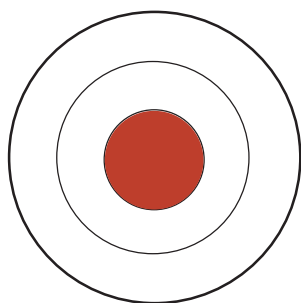


Рис. 1.5

Мишень для стрельбы.

Как уже неоднократно упоминалось, схема шансов буксует при решении многих вероятностных задач. Убедимся в этом на примере посредственного стрелка: Он никогда не промахивается, но с равной возможностью попадает в любую точку мишени. Пусть на мишени радиуса R нарисованы концентрические окружности, радиусами $R/3$ и $2R/3$, делящие мишень на три части (см. Рис. 5). Надо найти вероятность того, что стрелок попадет в центральную часть мишени.

Полную группу шансов Ω в данном случае образуют все точки мишени. Шансы же, благоприятствующие событию A , составляют точки центрального круга. Число и тех и других бессчетное, поэтому здесь нельзя напрямую применить формулу (3). Положение спасает тот факт, что надо знать не сами количества точек, а лишь их отношение. Последнее можно найти, измеряя количество точек некоторой области A более подходящей в данном случае мерой – площадью области. В итоге удастся определить вероятность интересующего события по формуле

$$P(A) = \frac{S(A)}{S(\Omega)}. \quad (1)$$

В исследуемой задаче площади мишени и центрального круга равны

$$S(\Omega) = \pi R^2 \quad \text{и} \quad S(A) = \frac{\pi}{9} R^2.$$

Следовательно, искомая вероятность такова

$$P(A) = \frac{1}{9}.$$

На формулы типа (1) опирается *геометрический метод исчисления вероятностей*. Чтобы применить его, стараются сопоставить пространству элементарных событий Ω некую область, точки которой изображают равновозможные элементарные исходы испытания, а интересующему событию A подобласть пространства Ω ($A \subset \Omega$). При этом вероятность события вычисляют как отношение *мер* областей события и пространства элементарных событий (их длин, площадей, объемов и так далее).

Сфера применения геометрического исчисления вероятностей довольно значительна. Последнее связано с тем, что геометрическими образами удается выразить самые разнообразные события, на первый взгляд не имеющие касательства к площадям и объемам. Покажем это на примере двух известных вероятностных задач.

1.4.1 Задача о встрече

Два друга условились встретиться у фонтана в течение часа после полудня. Каждый из них с равной возможностью может явиться в назначенное место в любой момент условленного часа, после чего ждет 20 минут и уходит. Спрашивается, какова вероятность встречи друзей.

Геометрически выразим возможность встречи, обозначив момент прихода одного из друзей за x , а другого за y . Элементарные события: пары моментов (x, y) приходов друзей, будем трактовать как координаты точек на плоскости. В итоге пространство элементарных событий примет вид, изображенного на Рис. 6, квадрата с длиной сторон 60 (минут).

Друзья встретятся, если моменты их прихода различаются не более чем на 20 минут: $|x - y| \leq 20$ и кроме того принадлежат упомянутому квадрату. На Рис. 6 соответствующая область квадрата закрашена.

Чтобы найти вероятность встречи, осталось найти площади квадрата $S(\Omega) = 60 \times 60 = 3600$ и закрашенной области. Последнюю удобно сосчитать как разность площадей большого квадрата и малого, составленного из неокрашенных кусков большого квадрата:

$$S(A) = 60^2 - 40^2 = 2000.$$

Следовательно, согласно формуле (1),

$$P(A) = \frac{20}{36} = \frac{5}{9}.$$

1.4.2 Задача Бюффона

Исследуем, с помощью геометрического исчисления вероятностей, еще одну вероятностную задачу, предложенную Бюффоном (Buffon) в его знаменитой работе, вышедшей в свет в 1777 году.

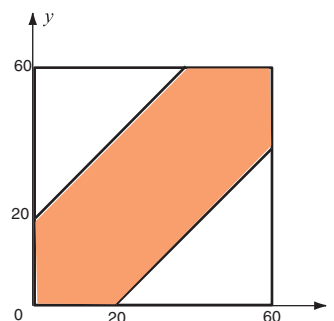


Рис. 1.6
Иллюстрация к задаче о встрече.

Формулировка задачи Бюффона такова: Тонкую иглу длины $2l$ бросают на плоскость, разграфленную параллельными прямыми, отстоящими на расстоянии $2a$ ($l < a$, чтобы игла могла пересечь не более одной прямой). Найти вероятность того, что игла пересечет какую-либо прямую.

Решим задачу геометрическим методом, для чего введем два параметра, характеризующие положение иглы на плоскости. За X обозначим расстояние от центра иглы до ближайшей прямой, а за Φ - угол между прямыми и иглой, отсчитанный так, как это показано на Рис. 7. Очевидно, координаты X и Φ полностью определяют взаимное расположение иглы и ближайшей прямой.

Изобразим пространство элементарных событий прямоугольником

$$0 \leq X \leq a, \quad 0 \leq \Phi \leq \pi$$

на вспомогательной плоскости с декартовыми координатами (x, φ) (см. Рис. 8). С другой стороны, как видно из Рис. 7, для пересечения иглы с ближайшей прямой необходимо и достаточно, чтобы выполнялось неравенство

$$X \leq l \sin \Phi,$$

то есть чтобы точка с координатами (X, Φ) попала в заштрихованную на Рис. 8 область A , расположенную под графиком синусоиды $x = l \sin \varphi$. Следовательно, согласно формуле (29), вероятность, что игла пересечет прямую, равна площади фигуры A , деленной на площадь прямоугольника $S(\Omega) = a\pi$:

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

Ответ, к которому мы пришли, решая задачу Бюффона, содержит в зародыше идею важного численного метода решения математических, физических, инженерных и экономических задач, завоевавшего, почти через двести лет после Бюффона, всеобщее признание. Укажем эту идею, для чего напомним, что согласно свойству статистической устойчивости вероятность события A , что игла пересечет прямую,

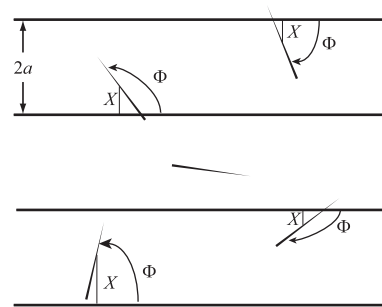


Рис. 1.7 Иглы, брошенные на разноманованный пол.

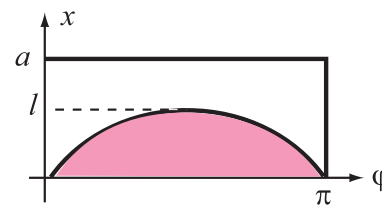


Рис. 1.8 Пространство элементарных событий и область события в задаче Бюффона.

можно оценить по формуле

$$P(A) \approx \frac{M}{N},$$

где N число бросаний иглы на разлинованную плоскость, а M количество бросаний, при которых игла пересекает прямую. Заменяв в (2) P на M/N , после несложных выкладок приходим к приближенной формуле

$$\pi \approx \frac{2l}{a} \times \frac{N}{M}. \quad (3)$$

Она выражает тот удивительный факт, что число π , равное отношению длины окружности к ее диаметру, и казалось бы не имеющее ничего общего со статистическими испытаниями, можно приближенно вычислить, подсчитывая число пересечений прямых иглой, при многократном бросании последней на разлинованную плоскость.

Подобное, с помощью многократно воспроизведенных испытаний тех или иных случайных событий, численное решение разнообразных научных и прикладных проблем, получило название *метода Монте-Карло*, в честь французского казино, где игроки испытывают случай, выигрывая большие деньги или теряя громадные состояния. С появлением компьютеров, способных генерировать длинные последовательности случайных чисел за сравнительно короткое время, метод Монте-Карло стал одним из эффективных методов решения обширного класса сложных проблем.

Замечание: Еще до появления компьютеров, исследователи многократно бросали иглу, чтобы таким способом оценить значение числа π , а заодно и проверить правильность формулы (2). Результаты некоторых подобных испытаний сведены в Таблице 2. Интересно, что таблица позволяет оценить, кроме прочего, и добросовестность самих испытателей. Им, как и всем нам, легко было заранее узнать истинное значение числа π со сколь угодно точностью (будем здесь ориентироваться на $\pi \simeq 3.14159$). Поэтому, в стремлении добиться наиболее впечатляющего результата, они могли поддаваться соблазну остановиться в подходящий момент, когда число (3) достаточно близко к истинному значению π . Может быть поэтому Фокс, с его гораздо меньшим, чем у Вольфа, числом бросаний, фактически повторил результат последнего (хотя конечно это могло быть и счастливой случайностью). Особенно впечатляет в этом

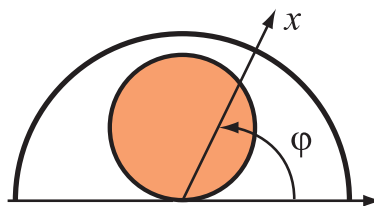


Рис. 1.9

К решению задачи Бюффона в полярных координатах.

Экспериментатор	l/a	N	M	Оценка
Вольф, 1850	0.8	5,000	2,532	3.1596
Смит, 1855	0.6	3,204	1,218.5	3.1553
Де Морган, 1860	1	600	382.5	3.137
Фокс, 1884	0.75	1,030	489.5	3.1595
Рейна, 1925	0.5419	2,520	859	3.1795
Гриджеман, 1960	0.7857	2	1	3.143

Таблица 1.2: Итоги испытаний по бросанию иглы на разливную плоскость.

смысле достижение Гриджемана, переставшего испытывать судьбу уже после двух бросков. Нам же следует осознавать, что выбор числа испытаний из условия близости результата к некоторому предполагаемому значению, делает сомнительными сами статистические испытания. Так, обладая “удачей” или большим упорством, можно показать, что $\pi \approx 2$ или, к примеру, $\pi \approx 4$. ■

Замечание: Кое-кто сомневается в справедливости формулы (30) для вероятности пересечения иглы и прямых, исходя из наглядных геометрических соображений. Почему, замечают они, пространство элементарных событий изображено на плоскости с декартовыми координатами (x, φ) ? Кажется естественнее считать упомянутые координаты полярными. Тогда пространство элементарных событий будет полукругом радиуса a , а область исследуемого события образует, изображенный на Рис. 9, круг диаметром l . Соответственно, вероятность пересечения иглы и прямых, найденная по формуле (29), будет равна

$$P' = \frac{1}{2} \left(\frac{l}{a} \right)^2 .$$

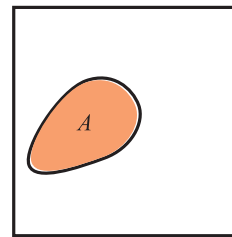
Советуем читателю поразмыслить над тем, отчего неверны такие “естественные” геометрические построения. Осмысление возникшего парадокса поможет освоению последующего материала. ■

1.4.3 Диаграммы Венна

Безотносительно к вычислению вероятностей различных событий, геометрическая трактовка последних помогает зримо представить многие

свойства как самих случайных событий, так и их вероятностей. Создадим геометрический образ события и его вероятности, взяв квадрат единичной стороны и вообразив, что кто-то беспорядочно колет его иглой, с равной возможностью попадая в любую точку квадрата (и только в него). Пусть случайное событие состоит в том, что острие иглы вонзается внутрь некоторой области A , закрашенной на Рис. 10. Из равной возможности попадания иглой в любую точку квадрата следует, что вероятность реализации указанного события в испытании (одиночном уколе квадрата) равна площади данной области

$$P(A) = S_A. \quad (4)$$



Подобные ассоциации между событиями и областями на плоскости, а также вероятностями и площадями областей, облегчают восприятие многих нетривиальных фактов теории вероятностей. Картинки, иллюстрирующие отношения между событиями и их вероятностями, называют *диаграммами Венна*. В дальнейшем мы будем не раз подкреплять ими выводы тех или иных вероятностных законов.

Рис. 1.10
Диаграмма Венна.

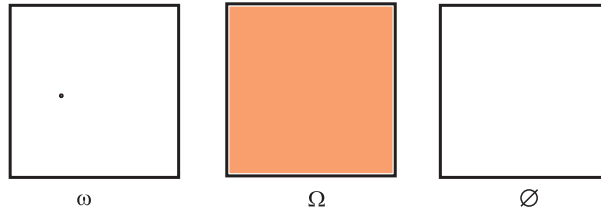


Рис. 1.11

Портреты элементарного события (слева), невозможного события (справа) и достоверного события (в центре).

Замечание: Уже только готовясь применять геометрические аналогии при анализе случайных событий, удается разрешить парадокс, возникший при решении задачи Бюффона: Из равновозможности разных значений угла Φ и расстояния X наугад брошенной иглы вытекает, что вероятность их попадания в область

$$\Phi \in [\varphi, \varphi + \Delta\varphi), \quad X \in [x, x + \Delta x) \quad (5)$$

пропорциональна произведению $\Delta\varphi\Delta x$ и не зависит от значений x и φ (внутри пространства элементарных событий). Эта независимость вероятности попадания в область (5) адекватно передается графиком Рис. 8

с декартовыми координатами (x, φ) , где областям (5) соответствуют равновеликие прямоугольники. В полярной же системе координат (x, φ) размеры областей (5) существенно зависят от значений x и φ . Понятен источник разрешенного нами и других подобных парадоксов. Дело в том, что при бесконечном числе шансов вероятность каждого шанса равна нулю. Поэтому, формируя геометрические образы событий, надо апеллировать к группам шансов, имеющим ненулевую вероятность. ■

1.4.4 Портретная галерея событий

Пользуясь геометрическими образами, легко создать своего рода портретную галерею случайных событий. Так на Рис. 11 изображены портреты элементарного, достоверного и невозможного событий. На Рис. 12 наглядно представлены несовместные и противоположные события, а на Рис. 13 даны портреты объединения, пересечения и разности событий A и B .

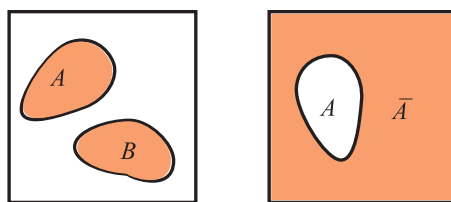


Рис. 1.12

Несовместные события (слева). Событие A и противоположное ему событие (справа).

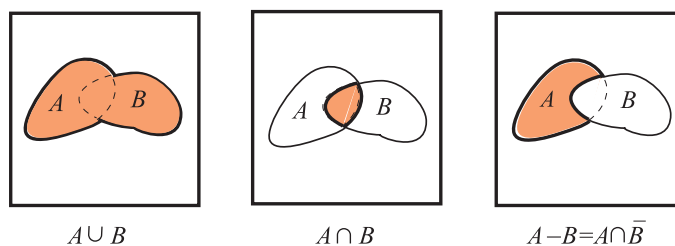


Рис. 1.13

Закрашенные области изображают, соответственно, объединение, пересечение и разность событий. Из диаграммы справа видно еще, что разность $A - B$ равна пересечению событий A и \bar{B} .

Портреты событий позволяют, в частности, зримо убедиться в справедливости многих общих свойств операций теории множеств, например коммутативности объединения и пересечения множеств

$$A \cup B \equiv B \cup A, \quad A \cap B \equiv B \cap A.$$

1.4.5 Теорема сложения вероятностей

Портреты случайных событий наглядно иллюстрируют доказательство одной из ключевых теорем теории вероятностей: так называемой *теоремы сложения*. Для этого изобразим на Рис. 14 пару произвольных случайных событий A и B , а также их разбиение на три несовместных события $A \cap \bar{B}$, $A \cap B$ и $\bar{A} \cap B$. Для доказательства теоремы сложения достаточно оперировать лишь двумя из них: $A \cap B$ и $\bar{A} \cap B$.

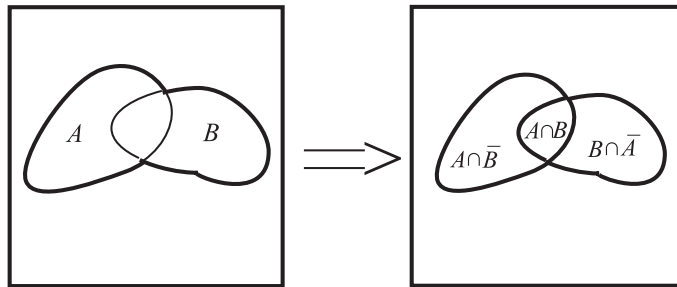


Рис. 1.14

Иллюстрация к доказательству теоремы сложения.

Из рисунков видно, что справедливы равенства теории множеств

$$A \cup B = A \cup (B \cap \bar{A}), \quad B = (B \cap A) \cup (B \cap \bar{A}). \quad (6)$$

Поскольку их правые части представляют собой объединения несовместных событий, то из аксиомы аддитивности (3.14) вытекают соответствующие равенства для вероятностей

$$P(A \cup B) = P(A) + P(B \cap \bar{A}), \quad P(B) = P(B \cap A) + P(B \cap \bar{A}).$$

Исключив отсюда $P(B \cap \bar{A})$, приходим к равенству, математически выражающему содержание теоремы сложения

$$P(A \cup B) = P(A) + P(B) - P(B \cap A). \quad (7)$$

Таким образом вероятность объединения двух событий равна сумме их вероятностей, минус вероятность пересечения этих событий.¹

Замечание: Мы могли бы еще быстрее убедиться в справедливости равенства (7), трактуя вероятности как площади областей единичного квадрата. Пусть образом события A служит область A , а ее площадь

¹Общие формулы вероятности объединения произвольного числа событий приведены в разделе (1.7.4).

S_A численно равна вероятности этого события. Аналогично, вероятность события B равна площади S_B соответствующей области, а S_{AB} – площадь и вероятность пересечения событий $A \cap B$. Из Рис. 14 видно, что площадь объединенной области $A \cup B$ равна сумме их площадей, минус площадь области пересечения $S_{A \cap B}$, дважды учтенной суммой $S_A + S_B$. Другими словами, геометрически очевидно соотношение

$$S_{A \cup B} = S_A + S_B - S_{A \cap B},$$

эквивалентное равенству (7). ■

На практике, при вычислении вероятности объединения событий часто опираются на формулу, вытекающую из закона де Моргана:

$$\overline{\bigcup_{k=1}^n A_k} = \bigcap_{k=1}^n \bar{A}_k. \quad (8)$$

Закон де Моргана применительно к трем событиям A , B и C иллюстрирует диаграмма Венна, данная на Рис. 15. На ней закрашено событие, противоположное объединению событий $A \cup B \cup C$. С другой стороны, закрашенную область можно, очевидно, трактовать как пересечение (общую часть) противоположных A , B и C областей.

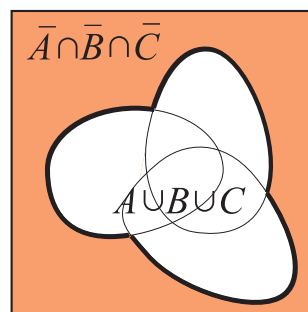


Рис. 1.15

Иллюстрация закона де Моргана.

Из равенства (идентичности) событий (8) следует равенство их вероятностей

$$P\left(\overline{\bigcup_{k=1}^n A_k}\right) = P\left(\bigcap_{k=1}^n \bar{A}_k\right).$$

Глядя на Рис. 15, легко понять, что событие, противоположное событию (8), есть объединение событий $\{A_1, \dots, A_n\}$. Поэтому из (3.15) имеем формулу

$$P\left(\bigcup_{k=1}^n A_k\right) = 1 - P\left(\bigcap_{k=1}^n \bar{A}_k\right), \quad (9)$$

выражающую вероятность объединения событий через вероятность пересечения противоположных событий. В частности, для случая двух событий отсюда следует эквивалентное (7) соотношение

$$P(A \cup B) = 1 - P(\bar{A} \cap \bar{B}). \quad (10)$$

Применим формулу (9) для вычисления вероятности, что при бросании n монет хотя бы одна из них упадет орлом вверх. Пусть A_k – выпадение

орла на k -й монете. Тогда \overline{A}_k – выпадение на этой монете решки. В данном случае легко сосчитать вероятность пересечения противоположных событий с помощью схемы шансов. Действительно, пересечению противоположных событий $\bigcap_{k=1}^n \overline{A}_k$, то есть выпадению решек на всех n монетах, благоприятствует один шанс из 2^n . Следовательно, вероятность выпадения хотя бы одного орла, согласно (9), равна

$$P = 1 - 2^{-n}.$$

1.5 Вероятностное пространство

Мы уже вводили, в разделе 1.2.5, ключевое для описания свойств вероятностной функции $P(A)$, понятие алгебры событий \mathcal{F} . Она необходима для реализации главной цели теории вероятностей: По известным вероятностям некоторой группы событий $\{A, B, \dots\}$ требуется найти вероятности других событий, равных всевозможным объединениям, пересечениям и разностям исходных событий.

В разделе 1.2.5 фундаментом алгебры \mathcal{F} служило пространство Ω , содержащее конечный набор n шансов. Соответственно, алгебра их всевозможных объединений и пересечений содержала конечное число элементов. Ситуация драматически меняется, если число шансов бесконечно (как набор натуральных чисел \mathbb{N}), а тем более несчетно (подобно точкам прямой \mathbb{R} , или “всего лишь” единичного отрезка $x \in [0, 1]$). В последних ситуациях невозможно вычислить вероятности подавляющего большинства потенциально допустимых событий. Утешением здесь служит то, что подобные, не поддающиеся анализу, события крайне редко востребованы в приложениях. Чтобы иметь дело лишь с событиями, вероятности которых можно вычислить, сознательно ограничивают круг исследуемых событий. Возникающие при этом понятия, свойства событий и вероятностей, кратко обсуждаются ниже.

1.5.1 Минимальные алгебры событий

Алгебры событий, содержащие *все* подмножества пространства Ω элементарных событий, называют *максимальными алгебрами*. Мы уже знаем, например, что максимальная алгебра, построенная на пространстве из n шансов, имеет 2^n элементов. Но даже здесь максимальная алгебра подчас избыточна, и полезно выделить лишь часть событий \mathcal{G} максимальной алгебры ($\mathcal{G} \subset \mathcal{F}$), представляющих, по тем или иным резонам,

насуточный интерес. Если множеству \mathcal{G} принадлежат Ω и \emptyset , а также всевозможные объединения, пересечения и разности элементов множества \mathcal{G} , то есть если для любых элементов A, B множества \mathcal{G}

$$A \cup B \in \mathcal{G}, \quad A \cap B \in \mathcal{G}, \quad A/B \in \mathcal{G}, \quad (1)$$

то \mathcal{G} называют просто *алгеброй*. Простейшим примером служит алгебра из двух, достоверного и невозможного, событий $\mathcal{G} = \{\emptyset, \Omega\}$. Однако она слишком бедна. Поэтому стараются выбрать алгебру более богатой элементами, чтобы удалось решать содержательные вероятностные задачи. Для этого выделяют некоторое множество событий $\{A, B, \dots\}$, включающее в себя невозможное и достоверное события, и дополняют его объединениями, пересечениями и разностями входящих в него событий. В итоге приходят к некоторой *минимальной* алгебре \mathcal{G} , совсем не обязательно совпадающей с максимальной.

Описанную только-что алгебру называют минимальной, поскольку любая другая алгебра, элементами которой служат события $\{A, B, \dots\}$, обязательно содержит все остальные события минимальной алгебры \mathcal{G} .

1.5.2 Алгебра Бореля

Дадим пример минимальной алгебры, для чего вновь будем хаотически тыкать иглой в единичный квадрат. Имеется невообразимое множество событий – самых замысловатых фигур, которые можно (а порой чересчур сложно) нарисовать внутри квадрата. Чтобы не утруждать себя описанием всех их, поступим как при игре в морской бой. Покроем квадрат сеткой из отстоящих друг от друга, на расстоянии $1/n$, горизонтальных и вертикальных прямых (см. Рис. 16) и будем говорить, что произошло событие A_{kl} , если игла попадет внутрь kl -го квадратика. На рисунке $n = 10$, так что минимальная алгебра событий, построенная на $n^2 = 100$ квадратиках, содержит конечное, хотя и громадное, число $2^{n^2} = 2^{100}$ событий.² Слева на Рис. 16 изображено одно из них: попадание в стилизованную букву P (Probability), вероятность которого нетрудно подсчитать, опираясь на указанную минимальную алгебру событий. Действительно, вероятность попадания в произвольный квадратик равна $1/100$. Всего квадратиков, составляющих букву P , пятнадцать, а значит,

²Чтобы хотя бы отдаленно почувствовать громадность этого числа, вообразим что некий СуперГейтс заработал десять триллионов долларов ($\$10^{13}$) и разменял их на центы. При этом число элементов указанной алгебры событий окажется примерно во столько раз больше числа центов, во сколько капитал СуперГейтса больше одного цента. И все равно этого “богатства” событий не хватает, чтобы более или менее точно вычислить вероятность попадания в изображенный на Рис. 16 круг.

по аксиоме конечной аддитивности, вероятность интересующего события равна $P(P) = 0.15$.

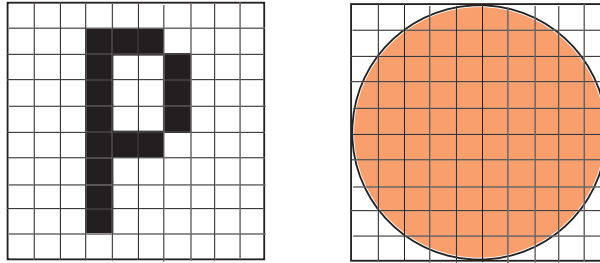


Рис. 1.16

События в разлинованном квадрате.

Однако основанная на квадратиках минимальная алгебра не обладает достаточной “разрешающей способностью”, чтобы с ее помощью аккуратно вычислить вероятность попадания, в изображенный справа на Рис. 16, вписанный в квадрат круг. Если считать, что квадратик, полностью или частично накрытый кругом, благоприятствует событию: попаданию в круг, то “вероятность” попадания в круг равна $P' = 0.88$. С другой стороны, геометрический способ исчисления вероятностей подсказывает правильный ответ

$$P = \frac{\pi}{4} \simeq 0.785$$

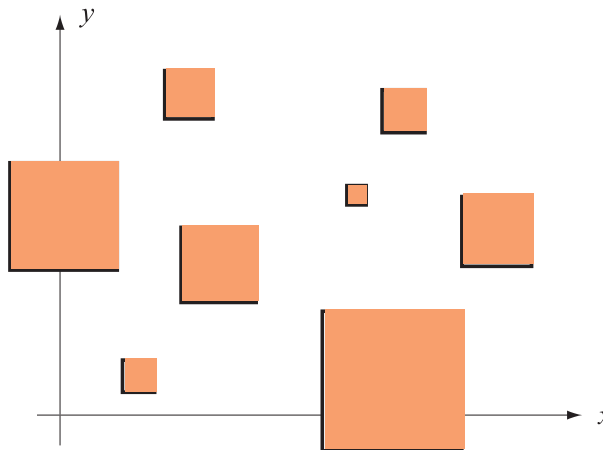


Рис. 1.17

Мозаика полуоткрытых квадратов, входящих в алгебру Бореля.

Выражаясь геометрическим языком, мы столкнулись со своеобразной квадратурой круга: проблемой вычисления площади круга подсчетом площади накрывающих его квадратиков. Эта проблема давно решена в теории интегрирования, согласно которой площадь круга можно

вычислить сколь угодно точно, как суммарную площадь все большего числа все более мелких прямоугольников (или квадратов, как частного случая последних). Чтобы точно вычислить площадь круга, необходимо взять счетное число³ квадратов (прямоугольников). Аналогично, чтобы точно вычислить вероятность попадания иглой в круг, необходимо построить минимальную алгебру, *счетного* множества элементов которого достаточно для точного нахождения площади круга. В качестве “фундамента” такой алгебры на плоскости с декартовыми координатами (x, y) удобно взять набор всевозможных, произвольно расположенных, прямоугольников произвольных размеров

$$\{ [x, x + \Delta x), [x, y + \Delta y) \} \quad (2)$$

Алгебру \mathcal{F} , состоящую из подобных всевозможных прямоугольников, вместе с их конечными и счетными объединениями и пересечениями,

$$\bigcup_{m=1}^{\infty} A_m \subset \mathcal{F}, \quad \bigcap_{m=1}^{\infty} A_m \subset \mathcal{F} \quad (3)$$

называют σ -алгеброй или *алгеброй Бореля*. Некоторые элементы борелевской алгебры изображены на Рис. 17.



Рис. 1.18

Элементы борелевской алгебры в одномерном пространстве \mathbb{R} .

Замечание: Понятие борелевской алгебры легко обобщить на случай D -мерных пространств, заменив прямоугольники D -мерными параллелепипедами. Но пожалуй наибольшее значение для приложений имеет одномерная борелевская алгебра, элементами которой служат всевозможные полуоткрытые интервалы $[x, x + \Delta x)$ на оси x . Они как нельзя лучше подходят для количественного описания измерений случайной величины X (силы тока, магнитного поля, индекса Доу-Джонса и других), поскольку в реальных измерениях обычно интересуются попаданием значений величины в некоторый заданный интервал. Пример элементов одномерной Борелевской алгебры приведен на Рис. 18. ■

Чтобы воспользоваться преимуществами алгебры Бореля при исчислении вероятностей, аксиому конечной аддитивности (3.14) заменяют *аксиомой счетной аддитивности*, согласно которой для любого счетного

³Счетным называют бесконечное множество элементов, которые удастся пронумеровать натуральными числами.

набора взаимно несовместных событий $\{A_1, A_2, \dots, A_n, \dots\}$ справедливо равенство

$$P\left(\bigcup_{m=1}^{\infty} A_m\right) = \sum_{m=1}^{\infty} P(A_m). \quad (4)$$

Именно оно, в частности, дает возможность точно вычислить вероятность попадания в круг. Функции множеств, обладающие свойством (4), называют *счетно-аддитивными*.

Читатель уже достаточно подготовлен, чтобы воспринять ключевое понятие теории вероятностей: понятие *вероятностного пространства*:

Тройку (Ω, \mathcal{F}, P) , где Ω пространство элементарных событий, \mathcal{F} заданная на нем алгебра событий, а P счетно-аддитивная функция, определенная на алгебре \mathcal{F} , отображающая множества на отрезок $[0, 1]$ и такая, что $P(\Omega) = 1$, называют вероятностным пространством.

1.6 Условные вероятности

Часто в испытании интересуются не одним, а сразу несколькими случайными событиями. Например событиями A и B . Будем полагать, что нам известны относящиеся к ним вероятности $P(A)$, $P(B)$ и $P(A \cap B)$.

Прогнозируя последствия тех или иных событий, да и просто в обыденной жизни, иногда полагают, что одно из событий, например B , реализуется в испытании. Естественно выяснить: как это повлияет на вероятность события A . К примеру, рыбаки спрашивают: Если завтра пойдет дождь (событие B), будет ли рыба лучше клевать (событие A)? Подобные вопросы закономерны и в куда более серьезных ситуациях. Так испытателям самолетов жизненно необходимо знать, как увеличится вероятность катастрофы при поломке одной из турбин.

Количественный ответ на эти и многие другие важные для приложений вопросы дает *условная вероятность* $P(A|B)$ события A , при условии, что событие B имело место. Вертикальную черту ставят для отделения случайного события, вероятностью которого интересуются, от условий, чье выполнение в испытании подразумевается.

Имеется формула, выражающая условную вероятность $P(A|B)$ через *безусловные* вероятности $P(A)$, $P(B)$ и $P(A \cap B)$ (в дальнейшем, как и ранее, будем называть безусловные вероятности просто вероятностями). В теории вероятностей эта формула постулируется. Однако, чтобы лучше понять ее суть, вновь прибегнем к услугам схемы шансов, где обещанная формула связи условной и безусловных вероятностей, как и другие аксиомы теории, может быть легко проверена.

1.6.1 Скачки на ипподроме

Зритель сидит на дальней от финиша трибуне ипподрома и наблюдает заезд десяти лошадей под номерами от 1 до 10. Лошади одинаково сильны, а жокеи искусны, так что любой из них с равной возможностью может выиграть заезд. Зритель заключил пари, что победит лошадь с номером, бóльшим 5 (то есть с номерами 6, 7, 8, 9, 10). Обозначим это событие буквой A .

Зритель знает, что жокеи на первых семи лошадях в синих кепках, а на остальных в красных. Обозначим событие, состоящее в том, что в заезде победит жокей в синей кепке, за B .

Вычислим вначале обычные, безусловные, вероятности событий A , B и $A \cap B$. Всего шансов в данном испытании (заезде) $n = 10$, по числу участвующих в заезде лошадей. Причем событиям A и B благоприятствуют $m_A = 5$ и $m_B = 7$ шансов (см. Рис. 19). Кроме того, пересечению событий $A \cap B$ благоприятствует $m_{AB} = 2$ шанса. Таким образом, по формуле (1.3) исчисления вероятностей на основе шансов, получаем

$$P(A) = \frac{m_A}{n} = \frac{1}{2}, \quad P(B) = \frac{m_B}{n} = \frac{7}{10}, \quad P(A \cap B) = \frac{m_{AB}}{n} = \frac{1}{5}. \quad (1)$$

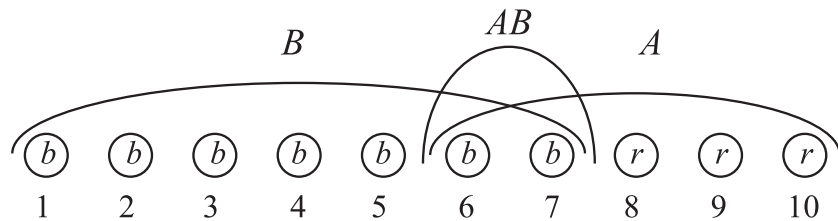


Рис. 1.19

Иллюстрация к подсчету шансов выиграть пари.

Сидя на трибуне, далеко от финиша, зритель увидел, что победил жокей в синей кепке. Иными словами, произошло событие B . Однако он не может издали разглядеть номер победившего жокея. Сосчитаем вероятность $P(A|B)$ зрителю выиграть пари с учетом имеющейся информации. Для этого надо из исходных шансов $n = 10$ оставить лишь число шансов m_B , благоприятствующих событию B , а из шансов события A взять лишь m_{AB} шансов новой полной группы шансов m_B . Соответственно, условную вероятность выиграть пари вычисляют по формуле

$$P(A|B) = \frac{m_{AB}}{m_B}. \quad (2)$$

Подставив в правую часть равенства указанные шансы, будем иметь

$$P(A|B) = \frac{2}{7}.$$

Как заметил зритель, выигрыш жокея в синей кепке сильно снизил его шансы, а вероятность выигрыша уменьшилась:

$$P(A|B) = \frac{2}{7} < P(A) = \frac{1}{2}.$$

Замечание: Условная вероятность может и превышать безусловную вероятность. Так если бы зритель поставил на первых пяти лошадях (событию A благоприятствуют шансы 1, 2, 3, 4, 5), то победа жокея в синей кепке (событие B) делает выигрыш вероятнее: $P(A|B) = 5/7$. ■

Выразим полученный результат через известные безусловные вероятности (1), для чего поделим числитель и знаменатель в (2) на исходное число шансов n . В итоге получим

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

В дальнейшем будем обозначать пересечение событий более кратко:

$$A \cap B \Leftrightarrow AB.$$

В этих обозначениях формула условной вероятности примет вид

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

В общем случае это соотношение служит *определением* условной вероятности. Им с успехом пользуются и тогда, когда схема шансов заведомо неприменима.

Поскольку все равно, какое из событий брать за A , а какое за B , наряду с (3) справедлива формула

$$P(B|A) = \frac{P(AB)}{P(A)}. \quad (4)$$

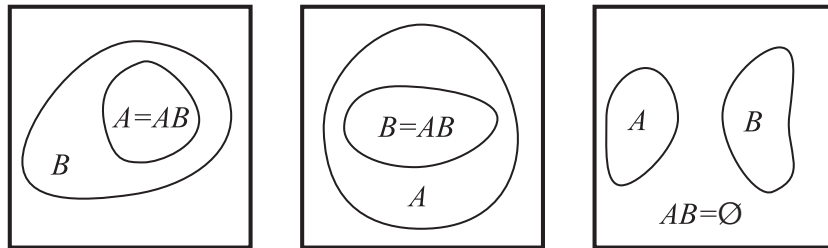


Рис. 1.20

Диаграммы Венна, иллюстрирующие происхождение формул (5) и (6).

1.6.2 Свойства условных вероятностей

Понятие условной вероятности служит мощным инструментом анализа случайных событий. Поэтому остановимся на нем подробнее.

Частные случаи условной вероятности

Укажем несколько частных следствий определения условной вероятности. Так если $A \subset B$ (все элементы множества A заодно принадлежат и множеству B), то $AB = A$ и условная вероятность удовлетворяет неравенству

$$P(A|B) = \frac{P(A)}{P(B)} \geq P(A) \quad (A \subset B).$$

Если, напротив, $B \subset A$, то $AB = B$ и условная вероятность оказывается равной единице

$$P(A|B) = \frac{P(B)}{P(B)} = 1 \quad (B \subset A). \quad (5)$$

Если же события A и B несовместны, то $AB = \emptyset$, а условная вероятность равна нулю

$$P(A|B) = \frac{P(\emptyset)}{P(B)} = \frac{0}{P(B)} = 0 \quad (AB = \emptyset). \quad (6)$$

Диаграммы Венна, поясняющие приведенные формулы, даны на Рис. 20.

Подчас свойство условных вероятностей (6) формулируют проще: Если $P(AB) = 0$, то $P(A|B) = 0$. Однако такое упрощение может обернуться грубыми ошибками. Дело в том, что в указанной формулировке упускают из вида принципиальное отличие невозможного события ($AB = \emptyset$) от событий, реализующихся с нулевой вероятностью ($P(AB) = 0$). А мы уже знаем, что если невозможное событие не происходит никогда, то события с нулевой вероятностью реализуются сплошь и рядом. Поэтому может случиться так, что $P(AB) = 0$ и $P(B) = 0$, а условная вероятность $P(A|B) \neq 0$. Убедимся в этом, вернувшись к задаче Бюффона. Вычислим условную вероятность пересечения иглы и прямой (событие A) при условии, что игла упала под углом φ_0 к прямым (событие B). Геометрическим образом пересечения указанных событий служит вертикальный отрезок, изображенный на Рис. 21.

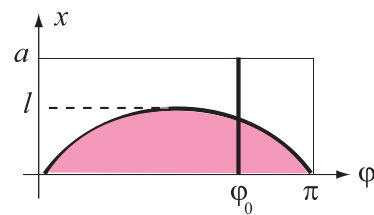


Рис. 1.21
Угол падения иглы φ_0 .

Его площадь равна нулю. Следовательно $P(AB) = 0$. С другой стороны, если игла действительно упала под углом φ_0 , то в качестве нового пространства элементарных событий надо брать совокупность точек отрезка, а вероятность попадания в интервал на нем вычислять как отношение *длины* интервала к *длине* отрезка. Подсчитав таким способом искомую условную вероятность, получим:

$$P(A|\varphi_0) = \frac{l}{a} \sin \varphi_0. \quad (7)$$

Здесь B заменено на параметр φ_0 , явно выражающий суть условия.

Общие свойства условной вероятности

Выясним общие свойства условной вероятности. При этом важно четко понимать, что последняя по сути ничем не отличается от обычных, безусловных вероятностей. Поясним сказанное. Для этого напомним, что как и в физическом эксперименте, под статистическим испытанием подразумевают целый комплекс условий, которые надо выполнить, чтобы испытание признали состоявшимся. Так бросая монету, надо ее сильно раскрутить, а вначале убедиться в симметричности монеты. В обговоренные заранее условия испытания вполне может быть включена и реализация некоторого случайного условия (чтобы увидеть метеоритный дождь, небо должно быть безоблачным). Другое дело, что упоминая о случайном событии, зачастую не указывают всех условий его наблюдения. Таким образом, условная вероятность отличается от “безусловной” лишь “более честным” подходом: Тем, что явно отмечены некоторые специфические условия, при выполнении которых испытание считают имевшим место.

Следовательно, все общие свойства обычных вероятностей без изменения переносятся на условные вероятности. Так областью изменения условной вероятности служит единичный отрезок $P(A|B) \in [0, 1]$. Причем вероятность достоверного события равна единице, а невозможного нулю.

На условную вероятность распространяются и аксиомы конечной (счетной) аддитивности, а также теорема сложения вероятностей (3.14). Только применительно к условным вероятностям она имеет вид

$$P(A \cup B|C) = P(A|C) + P(B|C) - P(AB|C),$$

где C случайное событие, при условии реализации которого вычисляется вероятность объединения событий A и B .

1.6.3 Формула умножения вероятностей

Домножив обе части (3) на $P(B)$ а (4) на $P(A)$, придем к равенствам

$$P(AB) = P(B)P(A|B) = P(A)P(B|A), \quad (8)$$

выражающим содержание теоремы умножения вероятностей. Согласно ей *вероятность произведения двух событий равна произведению вероятности одного из них на условную вероятность другого, при условии что первое произошло.*

Иногда требуется более общая формула умножения, связывающая три и более случайных события. Выведем ее для трех событий A , B и C . Используем предыдущую формулу, представив вероятность их пересечения как вероятность пересечения двух событий $D = AB$ и C . Согласно (8) имеем

$$P(ABC) = P(DC) = P(D) P(C|D) = P(AB) P(C|AB).$$

Подставив сюда, вместо $P(AB)$, последнюю часть формулы (8), получим окончательно

$$P(ABC) = P(A) P(B|A) P(C|AB). \quad (9)$$

Таким образом: *вероятность пересечения трех событий равна вероятности первого из них, умноженной на вероятность второго события, при условии, что первое произошло, и на вероятность третьего события, при условии, что произошли два первых.*

Нетрудно обобщить доказанную формулу на случай произвольного числа n событий A_1, A_2, \dots, A_n . Общая формула умножения вероятностей такова:

$$P(A_1 \dots A_n) = P(A_1) \times P(A_2|A_1) \times P(A_3|A_1A_2) \times \dots \times P(A_n|A_1A_2 \dots A_{n-1}). \quad (10)$$

1.6.4 Формула полной вероятности

Выведем еще одну важную, имеющую многочисленные применения, *формулу полной вероятности*. Перед этим введем понятие *полной группы несовместных событий*:

Группу событий $\{H_1, H_2, \dots, H_n\}$ называют *полной группой несовместных событий*, если все события группы попарно несовместны:

$$H_k H_l = \emptyset \quad (k, l = 1, 2, \dots, n, \quad k \neq l), \quad (11)$$

а их объединение образует достоверное событие:

$$\bigcup_{k=1}^n H_k = \Omega. \quad (12)$$

Другими словами, в каждом испытании реализуется одно и только одно из событий группы. Элементы полной группы несовместных событий называют *гипотезами*.

Возьмем некоторое произвольное случайное событие A и заметим, что выполнено равенство

$$A \equiv A\Omega \equiv A \bigcup_{k=1}^n H_k = \bigcup_{k=1}^n AH_k. \quad (13)$$

Справедливость его наглядно демонстрирует диаграмма Венна на Рис. 22.

Поскольку все объединяемые события AH_k в правой части равенства (13) несовместны, по аксиоме аддитивности имеем

$$P(A) = P\left(\bigcup_{k=1}^n AH_k\right) = \sum_{k=1}^n P(AH_k).$$

С другой стороны, согласно формуле умножения вероятностей (8), каждое слагаемое последней суммы может быть записано в виде

$$P(AH_k) = P(H_k)P(A|H_k).$$

Подставив это равенство в предыдущее, придем к обещанной формуле полной вероятности

$$P(A) = \sum_{k=1}^n P(H_k)P(A|H_k) \quad \left(\sum_{k=1}^n P(H_k) = 1 \right). \quad (14)$$

Как показывает опыт, эта формула очень удобна на практике, так как обычно проще вычислить вероятности гипотез $P(H_k)$ и отвечающие им условные вероятности $P(A|H_k)$, а уж затем, по формуле (14), – найти вероятность “сложного” события A .

Пример: Продемонстрируем использование формулы полной вероятности примером из студенческой жизни. Экзаменационную работу, написанную группой студентов, проверяют три преподавателя. Статистическая обработка архивов факультета показала, что вероятность студенту

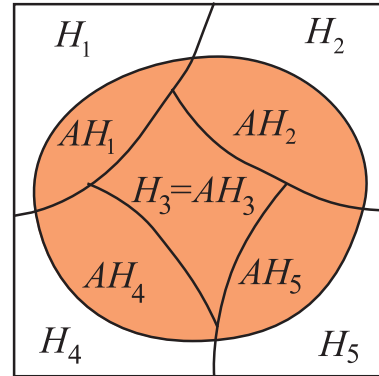


Рис. 1.22

Событие A (закрашенная область) составлено, как детская мозаика, из пересечений AH_1 - AH_5 .

получить неуд у первого из них равна $1/2$, у второго $1/3$, а у третьего $1/4$. Работа Пети с равной возможностью может попасть любому преподавателю. Какова вероятность, что Петя получит неуд (произойдет событие A)?

Полную группу шансов здесь образуют события H_1, H_2, H_3 , что работа Пети попала, соответственно, первому, второму или третьему преподавателю. Равновозможность перечисленных событий означает, что их вероятности равны

$$P(H_1) = P(H_2) = P(H_3) = \frac{1}{3}. \quad (15)$$

Из сказанного также следует, что условные вероятности получить неуд таковы

$$P(A|H_1) = \frac{1}{2}, \quad P(A|H_2) = \frac{1}{3}, \quad P(A|H_3) = \frac{1}{4}. \quad (16)$$

Следовательно, вероятность Пете получить неуд, согласно формуле полной вероятности равна

$$P(A) = P(H_1)P(A|H_1) + P(H_2)P(A|H_2) + P(H_3)P(A|H_3) = \\ \frac{1}{3} \times \frac{1}{2} + \frac{1}{3} \times \frac{1}{3} + \frac{1}{3} \times \frac{1}{4} = \frac{13}{36} \simeq 0.361 \quad \star$$

1.6.5 Формула Байеса

Чтобы лучше понять, о чем пойдет речь, продолжим обсуждение предыдущей задачи. Петя узнал, что ему поставлен неуд. Может ли он оценить, какому преподавателю попала его работа. Для этого Пете надо вычислить вероятности $P(H_k|A)$ что работа попала k -му преподавателю, при условии что получен неуд. Подобные вероятности называют *апостериорными*, чтобы отличить их от *априорных* вероятностей, заданных формулами (15), (16), и известными до того как работы были проверены.

Апостериорные вероятности ценны тем, что корректируют наши знания о случайных событиях на основе поступившей информации. Вычисление апостериорных вероятностей важно практически везде, где используют вероятностные методы, от предсказания погоды и обнаружения сигналов, до игры на бирже и расследования уголовных дел. Судя о людях по их поступкам, мы подсознательно пытаемся оценить апостериорные вероятности их истинных намерений.

Сформулируем и решим общую задачу вычисления апостериорных вероятностей. Пусть, как и прежде, имеется группа гипотез H_1, \dots, H_n ,

известны их априорные вероятности

$$P(H_1), \dots, P(H_n) \quad \left(\sum_{k=1}^n P(H_k) = 1 \right)$$

и условные априорные вероятности

$$P(A|H_1), \dots, P(A|H_n).$$

Стало известно, что в испытании произошло событие A . Спрашивается, чему равны апостериорные вероятности

$$P(H_1|A), \dots, P(H_n|A).$$

Поставленная задача решается с помощью формул условной вероятности (4), умножения вероятностей (8) и полной вероятности (14). Положив в (4) $B = H_k$ и привлекая формулу (8), будем иметь

$$P(H_k|A) = \frac{P(A|H_k) P(H_k)}{P(A)}. \quad (17)$$

Преобразуя знаменатель с помощью формулы полной вероятности (14), получим окончательно

$$P(H_k|A) = \frac{P(A|H_k) P(H_k)}{\sum_{k=1}^n P(H_k) P(A|H_k)}. \quad (18)$$

Это и есть обещанная *формула Байеса*, выражающая апостериорные условные вероятности через априорные вероятности.

Применительно к Пете, получившему неуд, апостериорные вероятности того, что его работа попала, соответственно, первому, второму и третьему преподавателям, равны

$$\begin{aligned} P(H_1|A) &= \frac{6}{13} \simeq 0.462 > \frac{1}{3} \\ P(H_2|A) &= \frac{4}{13} \simeq 0.308 \lesssim \frac{1}{3} \\ P(H_3|A) &= \frac{3}{13} \simeq 0.231 < \frac{1}{3} \end{aligned}$$

Замечание: Заметим, что апостериорные вероятности гипотез, как и априорные вероятности, в сумме дают единицу:

$$\sum_{k=1}^n P(H_k|A) = 1. \quad (19)$$

Это равенство выражает, независящее от итогов испытаний, следствие свойства гипотез в совокупности охватывать все пространство элементарных событий Ω . Равенства, подобные (19), называют условием нормировки. В дальнейшем мы не раз встретимся с ним при анализе статистических свойств случайных величин. ■

1.7 Статистическая независимость

На формуле условной вероятности (6.3)

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

базируется еще одно ключевое понятие теории вероятностей – понятие *статистической независимости* случайных событий. В общем случае информация о том, что в данном испытании произошло событие B , меняет вероятность $P(A|B)$ наступления в том же испытании события A . Однако если события никак не связаны между собой, то вероятность события A остается неизменной, не зависит от B . Математически, факт статистической независимости событий A и B выражается равенством

$$P(A|B) = P(A). \quad (1)$$

Отсюда и из формулы умножения вероятностей (6.8) видно, что вероятность пересечения статистически независимых (будем называть их ниже просто независимыми) событий равна произведению их вероятностей

$$P(AB) = P(A)P(B). \quad (2)$$

Подчеркнем, свойство независимости событий взаимно: если событие A не зависит от B , то и B не зависит от A :

$$P(A|B) = P(A) \quad \Longleftrightarrow \quad P(B|A) = P(B).$$

Формально, чтобы убедиться в независимости случайных событий A и B , надо найти вероятности $P(A)$, $P(B)$ и $P(AB)$, подставить их в (6.3) и проверить, выполняется ли равенство (2). Однако чаще всего независимость событий заранее очевидна, и тогда формулой (2) пользуются для вычисления вероятности их пересечения.

Пример: Примером очевидно независимых событий могут служить выпадение орла на монете и шестерки на кости при одновременном их бросании. Поэтому вероятность описанного события равна

$$P(\text{орел} \cap 6) = \frac{1}{2} \times \frac{1}{6} = \frac{1}{12} \simeq 0.083 \quad \star$$

Если заранее очевидна взаимная независимость n событий $\{A_1, \dots, A_n\}$, то вероятность их пересечения вычисляют по аналогичной (2) формуле

$$P(A_1 A_2 \dots A_n) = P(A_1) \times P(A_2) \times \dots \times P(A_n). \quad (3)$$

Так вероятность, что при бросании трех костей выпадет максимальное число очков (три шестерки) равна

$$P = \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} \simeq 0.0046$$



Рис. 1.23

Последовательное соединение лампочек в гирлянде.

Пример: Другой, более “инженерный” пример независимых событий дает задача о елочной гирлянде (последовательном соединении лампочек). Ток в гирлянде проходит через n последовательно соединенных лампочек (см. Рис. 23). Если хотя бы одна из них перегорит, ток прекращается, а гирлянда перестает светить. Пусть вероятность, что k -я лампочка останется исправной в течении всего рождества, равна p_k . Найти вероятность, что гирлянда будет светить все рождество.

Обозначим за B_k событие, состоящее в том, что k -я лампочка все рождество исправна, а за L , что исправна вся гирлянда. Указанные события связаны между собой равенствами $L = B_1 \cap B_2 \cap \dots \cap B_n$. Если вероятность лампочке оставаться исправной не зависит от состояния остальных лампочек, то вероятность искомого события равна

$$P(L) = P(B_1) \times P(B_2) \times \dots \times P(B_n) = p_1 p_2 \dots p_n. \quad \star$$

Замечание: Отсюда следует важный вывод теории надежности, что приборы с последовательным соединением многих элементов, когда поломка одного выводит из строя весь прибор, очень ненадежны. Пусть, к примеру, гирлянда объединяет сто лампочек, имеющих довольно высокую надежность ($p = 0.99$). Тогда вероятность гирлянде светить все рождество равна

$$P(L) = 0.99^{100} \simeq 0.366$$

Иными словами, к концу рождества погаснут почти две трети таких гирлянд. ■

1.7.1 Независимость и несовместность

На будущее полезно уяснить взаимосвязь свойств независимости и несовместности. Для несовместных событий $P(AB) = P(\emptyset) = 0$. Если к тому же события A и B независимы, то должно выполняться равенство

$$P(A)P(B) = 0.$$

Иначе говоря, несовместные события заодно и независимы, лишь если вероятность хотя бы одного из них равна нулю. В противном случае независимость и несовместность взаимно исключают друг друга.

Заметим однако, что выполнение равенств

$$P(AB) = P(A)P(B) = 0 \tag{4}$$

не гарантирует несовместности случайных событий, а сами они могут оказаться зависимыми. В сказанном убеждает условная вероятность (6.7) пересечения иглой и прямых (событие B) под определенным углом φ_0 (событие A). Вероятности указанных событий подчиняются равенствам (4), однако условная вероятность (6.7) зависит от φ_0 , тем самым свидетельствуя о зависимости событий.

Единственным событием, одновременно независимым и несовместным с любыми другими событиями, является невозможное событие \emptyset . Ситуация здесь напоминает поговорку о черной кошке в темной комнате, которую никто не поймает, потому что ее там нет.

1.7.2 Парадокс Бернштейна

Признание “очевидным” независимости некоторого набора случайных событий, например событий A , B и C , как правило апеллирует к “здравому смыслу”. Иногда он дает осечку. Поэтому полезно тщательно анализировать предпосылки вывода о независимости тех или иных случайных событий.

Зачастую, обосновывая независимость случайных событий, ограничиваются проверкой их *попарной независимости*. То есть, применительно к событиям A , B и C , выясняют, справедливы ли равенства

$$P(AB) = P(A)P(B), \quad P(BC) = P(B)P(C), \quad P(AC) = P(A)P(C). \quad (5)$$

Оказывается, одной попарной независимости недостаточно, чтобы три события можно было считать взаимно независимыми. Покажем это на примере *парадокса Бернштейна*:

Бросим на стол тетраэдр, три грани которого выкрашены, соответственно, в красный, синий и зеленый цвета, а четвертая грань разноцветная, раскрашена всеми тремя цветами. Событие R состоит в том, что тетраэдр упал на грань, содержащий красный цвет. Аналогично определяются события B и G . Согласно схеме шансов, вероятности этих событий равны

$$P(R) = P(B) = P(G) = \frac{2}{4} = \frac{1}{2}, \quad (6)$$

поскольку всего шансов 4, по числу граней тетраэдра, а каждому событию благоприятствует две, одноцветная и разноцветная, грани.

Аналогично легко найти вероятности пересечения любого из двух перечисленных событий. Так как каждому пересечению благоприятствует одна (разноцветная) грань, то справедливы равенства

$$P(RB) = P(BG) = P(RG) = \frac{1}{4}. \quad (7)$$

Из (6), (7) видно, что в данном случае условия попарной независимости событий R , B и G выполнены. Однако нарушено равенство (3). В самом деле, пересечению событий RBG благоприятствует падение тетраэдра на разноцветную грань, а значит

$$P(RBG) = \frac{1}{4} \neq P(R)P(B)P(G) = \frac{1}{8}.$$

Следовательно, указанные события статистически зависимы.

В общем случае, установление независимости n событий $\{A_1, \dots, A_n\}$ требует проверки $2^n - n - 1$ равенств типа (5), (3).

1.7.3 Вероятность объединения независимых событий

Вернемся к теореме сложения вероятностей (4.7). Для независимых событий последнее слагаемое в (4.7) расщепляется на произведение вероятностей отдельных событий, и (4.7) приобретает вид

$$P(A \cup B) = P(A) + P(B) - P(A)P(B). \quad (8)$$

Другую, полезную при расчетах, формулу вероятности объединения независимых событий получим, заметив, что независимость $\{A_1, \dots, A_n\}$ влечет за собой независимость противоположных событий $\{\bar{A}_1, \dots, \bar{A}_n\}$. А значит, согласно формулам (4.9), (4.10),

$$P\left(\bigcup_{k=1}^n A_k\right) = 1 - \prod_{k=1}^n P(\bar{A}_k) \quad (9)$$

и в частности

$$P(A \cup B) = 1 - P(\bar{A})P(\bar{B}). \quad (10)$$

Разумеется, соотношения (8) и (10) эквивалентны. Чтобы убедиться в этом, достаточно вспомнить, что вероятности событий и противоположных им связаны равенством (3.15). Следовательно (10) можно переписать так:

$$P(A \cup B) = 1 - [1 - P(A)][1 - P(B)].$$

Раскрыв квадратные скобки, вернемся к выражению (8).

Аналогично, выразив в (9) вероятности противоположных событий через вероятности исходных событий, для вероятности объединения трех событий A , B и C будем иметь

$$P(A \cup B \cup C) = 1 - [1 - P(A)][1 - P(B)][1 - P(C)]$$

или

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A)P(B) - P(A)P(C) - P(B)P(C) + P(A)P(B)P(C). \quad (11)$$

Формула (9) особенно удобна при исследовании вероятностей объединения большого числа событий. Для иллюстрации решим задачу о люстре с n лампочками:

Пример: Лампочки подсоединены к электрической сети параллельно, так что если перегорит одна из них, остальные лампочки будут светить. Пусть вероятность каждой лампочке оставаться исправной в течении всего вечера равна p и не зависит от состояния остальных лампочек. Спрашивается, какова вероятность что к концу вечеринки в люстре будет светить хотя бы одна лампочка (событие L). Подставив в (9)

$$P(\bar{A}_k) = 1 - p,$$

получим удивительно простой ответ

$$P(L) = 1 - (1 - p)^n. \quad (12)$$

Отсюда видно, что чем больше лампочек в люстре, тем вероятнее не остаться в кромешной тьме. Даже если за вечер перегорает каждая вторая лампочка ($p = 1/2$), то при $n = 10$ люстра будет светить с вероятностью $P(L) = 0.999$. Естественно, подобные следствия параллельного соединения элементов важны не только для устроителей вечеринок, но находят многочисленные применения при конструировании надежных приборов и конструкций. ★

1.7.4 Мнемоническое правило

Формула (11) и подобные ей дают мнемоническое правило реконструкции общих формул для вероятностей объединения событий. Так можно показать, что заменив в (11) произведения вероятностей на вероятности пересечения событий, получим правильную общую формулу, справедливую для произвольных (не только независимых) событий

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC). \quad (13)$$

К примеру для несовместных событий, вероятности пересечений которых равны нулю, получим отсюда очевидное следствие аксиомы конечной аддитивности

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) \\ (A \cap B = A \cap C = B \cap C = \emptyset)$$

Воспользуемся мнемоническим правилом, чтобы установить с помощью (9) общую, подобную (13), формулу вероятности объединения n событий $\{A_1, \dots, A_n\}$. Для простоты будем полагать, что вероятности всех событий одинаковы и равны p . Тогда (9) примет уже знакомый вид

$$P\left(\bigcup_{k=1}^n A_k\right) = 1 - (1 - p)^n.$$

Согласно формуле бинома Ньютона

$$(1 - p)^n = \sum_{k=0}^n (-1)^k \binom{n}{k} p^k.$$

Подставив правую часть последнего равенства в (13), будем иметь

$$P\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n (-1)^{k+1} P_k, \quad (14)$$

n	1	2	3	4	5	6	7	8
$P(W)$	0	0.5	0.3333	0.3750	0.3667	0.3681	0.3679	0.3679

Таблица 1.3: Вероятности всем зрителям ошибиться местами.

где

$$P_k = \binom{n}{k} p^k.$$

Напомним, входящий сюда биномиальный коэффициент (2.11) равен числу сочетаний k элементов из n , то есть, применительно к нашему случаю, совпадает с числом всех различных комбинаций k событий из n без учета порядка их следования в комбинации. Заменяя в (14) P_k суммой $\binom{n}{k}$ вероятностей $P(A_{i_1} \dots A_{i_k})$ пересечений всех сочетаний k событий из n , получим общую формулу вероятности объединения n событий.

Часто, особенно в инженерных приложениях, обсуждаемые события однородны в том смысле, что все вероятности $P(A_{i_1} \dots A_{i_k})$ одинаковы. Тогда формула (14) переходит в равенство

$$P\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n (-1)^{k+1} \binom{n}{k} P(A_1 \dots A_k). \quad (15)$$

Вычитая его правую часть из единицы, придем к не менее полезной формуле вероятности пересечения однородных событий $\{\bar{A}_1, \dots, \bar{A}_n\}$

$$P(\bar{A}_1 \dots \bar{A}_n) = \sum_{k=0}^n (-1)^k \binom{n}{k} P(A_1, \dots, A_k). \quad (16)$$

Пример: Решим задачу о недисциплинированных зрителях: В зрительном зале n мест и билеты на все места в зале проданы. Зрители же рассаживаются совершенно случайно. Какова вероятность, что ни один из них не сядет на “законное” место (событие $W = \bar{A}_1 \dots \bar{A}_n$, где A_k означает, что k -й зритель занял свое место).

Для решения прибегнем к формуле (16), поскольку здесь сравнительно легко найти вероятности $P(A_1 \dots A_k)$, что все первые k зрителей сядут на свои места. По теореме умножения вероятностей (6.10)

$$P(A_1 \dots A_k) = P(A_1)P(A_2|A_1) \dots P(A_k|A_1 \dots A_{k-1}).$$

Очевидно, вероятность m -му зрителю совершенно случайно занять свое место, после того как предыдущие зрители расселись по своим местам,

равна

$$P(A_m | A_1 \dots A_{m-1}) = \frac{1}{n - m + 1}.$$

Следовательно

$$P(A_1 \dots A_k) = \frac{1}{n} \times \frac{1}{n-1} \times \dots \times \frac{1}{n-k+1} = \frac{(n-k)!}{n!}.$$

Подставив это выражение в (16) и вспомнив, чему равен биномиальный коэффициент (2.11), получим окончательно

$$P(W) = \sum_{k=0}^n \frac{(-1)^k}{k!}.$$

Значения $P(W)$ для нескольких значений n приведены в Таблице 3. Из нее видно, что вероятность $P(W)$, как функция n , быстро сходится, и начиная с $n = 7$ практически совпадает с пределом при $n \rightarrow \infty$: $e^{-1} \simeq 0.3679$.

Значение предела довольно неожиданно с позиций “здорового смысла”. Последний подсказывает, что если не в зрительном зале, то на трибунах стадиона кто-то все же займет купленное место. На самом деле, камерный зал на 7 зрителей и стадион на 70 тысяч болельщиков в этом смысле практически эквивалентны: Вероятность что тут и там никто не займет своего места по сути одинакова и равна e^{-1} .

Самые дотошные обратят внимание на более тонкий эффект “несчастливых” четных чисел: Вероятность, что никому не посчастливится сесть на свое место в зале с четным количеством кресел, больше чем та же вероятность с нечетным, на единицу меньшим, числом мест. ★

Глава 2

Случайные величины

В первой главе излагались основы вероятностного анализа случайных явлений. В подавляющем большинстве прикладных задач случайные явления выражают случайными величинами. Так прогноз погоды сводят к предсказанию температуры, влажности, давления. Изучение случайных величин требует специфических понятий и методов. Их последовательному изложению посвящена данная глава.

2.1 Дискретные случайные величины

Прежде всего обсудим *дискретные случайные величины*, принимающие в испытании конечное или счетное число значений.

2.1.1 Распределения дискретных величин

Случайную величину X с ограниченным или счетным числом возможных значений $\{x_1, x_2, \dots, x_n, \dots\}$ называют дискретной. Все ее свойства определяются вероятностями значений

$$P(X = x_1) = p_1, \quad P(X = x_2) = p_2, \quad \dots \quad P(X = x_n) = p_n \quad \dots \quad (1)$$

Если набор возможных значений случайной величины невелик, сведения о ней сводят в таблицу, первая строка которой включает все значения X , а вторая их вероятности (см. Таблицу 1). При этом говорят, что задан *закон распределения* дискретной случайной величины.

Пример: Дискретной является сумма очков, выпадающих при бросании двух игральных костей. По Таблице 1.1 легко установить ее закон распределения. Он приведен в Таблице 2. ★

X	x_1	x_2	\dots	x_n
P	p_1	p_2	\dots	p_n

Таблица 2.1: Таблица, схематически изображающая закон распределения дискретной случайной величины X , имеющей n возможных значений.

В любом испытании реализуется лишь одно из возможных значений случайной величины. Поэтому события $\{A_1, A_2, \dots, A_n, \dots\}$, где A_k состоит в том, что $X = x_k$, образуют полную группу несовместных событий. Их объединение есть достоверное событие

$$\bigcup_k A_k = \Omega,$$

так как в любом испытании X обязательно примет какое-либо значение. Вероятность достоверного события равна единице, а значит вероятности возможных значений подчиняются условию нормировки

$$P\left(\bigcup_k A_k\right) = \sum_k p_k = 1. \quad (2)$$

Пример: Типичной для самых разнообразных приложений дискретной величиной N является номер независимого испытания, в котором впервые происходит некое событие B . К примеру, первый раз выпадает решка. Очевидно, вероятность того, что для реализации события B придется провести $N = n$ испытаний, совпадает с вероятностью пересечения случайных событий

$$B_n = \overline{B^1} \cap \overline{B^2} \cap \dots \cap \overline{B^{n-1}} \cap B^n. \quad (3)$$

Здесь событие B^k состоит в том, что B реализовалось в k -м испытании, а $\overline{B^k}$ противоположное ему событие. Пусть вероятность события B в каждом отдельном испытании равна p . Из независимости испытаний следует, что события в правой части равенства (3) взаимно независимы. Кроме того, если B^k реализуются с одинаковой вероятностью p , то и противоположные события имеют одинаковую вероятность $q = 1 - p$. А значит, по теореме умножения вероятностей независимых событий

$$p_n = P(B_n) = P(N = n) = q^{n-1} p \quad (p + q = 1, \quad n \geq 1). \quad \star \quad (4)$$

X	2	3	4	5	6	7	8	9	10	11	12
P	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$

Таблица 2.2: Распределение суммарного числа очков, выпадающих при бросании двух игральных костей.

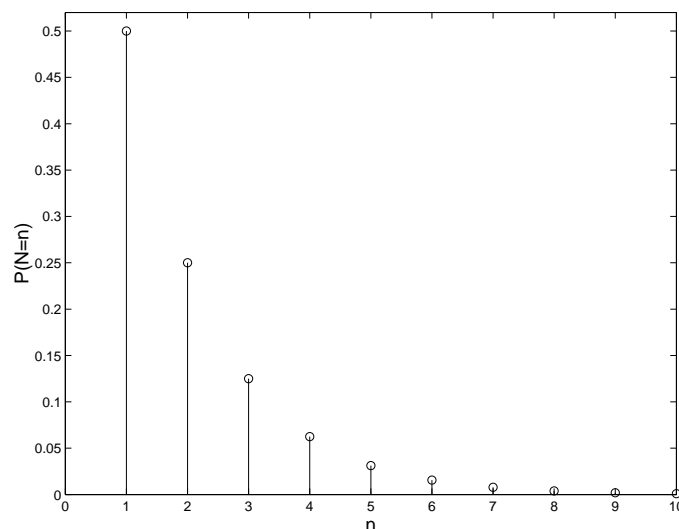


Рис. 2.1

График геометрического распределения числа бросаний до первого выпадения орла. Из графика видно, что хотя число возможных значений N бесконечно, уже вероятность того, что до первого выпадения орла придется сделать 10 бросков, совсем мала.

Формула (4) задает *геометрическое распределение*. Возможные значения распределенной так случайной величины составляют все натуральные числа \mathbb{N} . Применительно к испытаниям это значит, что как бы ни было велико M , имеется, пусть очень малая, но отличная от нуля вероятность, что до первой реализации события B придется сделать больше M испытаний ($N > M$). График геометрического распределения числа бросаний монеты до появления орла ($p = q = 1/2$) дан на Рис. 1.

Замечание: В качестве первого теста на правильность полученного распределения обычно используют условие нормировки (2). Убедимся, что оно выполнено для геометрического распределения (4). В данном

случае сумма вероятностей всех возможных значений равна

$$S = p \sum_{k=0}^{\infty} q^k$$

Суммируя входящую сюда геометрическую прогрессию и учитывая что $q = 1 - p$, имеем

$$S = \frac{p}{1 - q} = \frac{p}{p} = 1. \quad \blacksquare$$

2.1.2 Интегральная функция распределения

Запись распределения случайных величин в виде таблиц неудобна в аналитических расчетах. К ним лучше приспособлены *функции распределения*. Первой из них обсудим *интегральную функцию распределения*

$$F(x) = P(X < x), \quad (5)$$

равную вероятности того, что случайная величина X принимает в испытании значение, меньшее наперед заданного x .

Отметим принципиальное отличие $F(x)$ от вероятности $P(X < x)$. Аргументом последней служат случайные события, а аргумент x функции распределения $F(x)$ точка числовой оси ($x \in \mathbb{R}$). Это позволяет исследовать свойства $F(x)$ стандартными методами теории функций числового аргумента. Так из (5) видно, что $F(x)$ неубывающая функция x , определенная на всей оси $x \in (-\infty, \infty)$ и принимающая значения на отрезке $[0, 1]$. Ее наименьшее значение достигается при $x = -\infty$, а наибольшее при $x = \infty$:

$$F(-\infty) = 0, \quad F(\infty) = 1. \quad (6)$$

Применительно к дискретным случайным величинам, $F(x)$ представляет собой кусочно-постоянную функцию

$$F(x) = \sum_k p_k \theta(x - x_k), \quad (7)$$

терпящую скачки в точках разрешенных значений x_k случайной величины X . При этом величина скачков равна вероятностям разрешенных значений p_k . В (7) использована единичная функция

$$\theta(x) = \begin{cases} 0, & x \leq 0; \\ 1, & x > 0. \end{cases} \quad (8)$$

Типичный график интегральной функции распределения случайного числа бросаний игральной кости до первого выпадения двойки (или любого другого загаданного числа очков) приведен на Рис. 2.

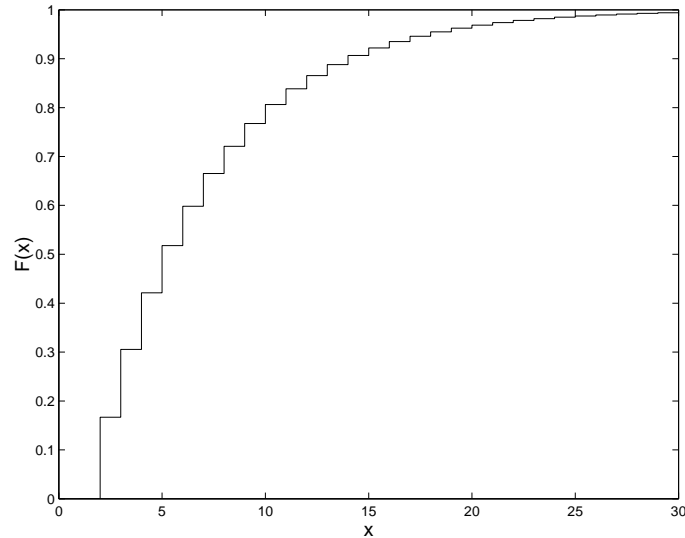


Рис. 2.2

График интегральной функции распределения числа бросаний игральной кости до первого выпадения загаданного числа очков.

Подчеркнем, функция $F(x)$ несет всю информацию о свойствах X . Так с ее помощью нетрудно вычислить вероятность события, которым чаще всего интересуются в приложениях. А именно, вероятность попадания случайной величины X в любой наперед заданный интервал $[a, b)$. Докажем это, для чего обозначим данное событие буквой B ($B : a \leq X < b$), а за A обозначим событие, состоящее в том, что $X < a$. Эти события несовместны. Следовательно, по аксиоме аддитивности,

$$P(A \cup B) = P(A) + P(B).$$

Отсюда имеем

$$P(B) = P(a \leq X < b) = P(A \cup B) - P(A).$$

Кроме того, по определению интегральной функции распределения (5)

$$P(A) = P(X < a) = F(a), \quad P(A \cup B) = P(X < b) = F(b).$$

Отсюда и из предыдущего равенства получаем

$$P(a \leq X < b) = F(b) - F(a). \tag{9}$$

Аналогично вычисляют вероятность попадания X в произвольный набор интервалов на числовой оси, вроде тех, что изображены на Рис. 1.18. Таким образом, интегральная функция распределения позволяет вычислить вероятность любого события из борелевской алгебры событий.

Замечание: Поясняя вероятностные свойства случайных величин, подчас прибегают к наглядной механической интерпретации интегральной функции распределения. А именно, отождествляют вероятность с массой “вероятностного вещества”, распределенного вдоль оси x , а функцию $F(x)$ полагают вероятностной массой слева от точки x . Поскольку $F(\infty) = 1$, то полная масса вероятностного вещества всегда равна единице. Соответственно, предельно ясен и механический смысл правой части равенства (9). Это масса вероятности, сосредоточенная на интервале $[a, b)$. ■

2.1.3 Плотность вероятностей

Распределение вещества вдоль материальной прямой x в механике предпочитают описывать не массой слева $F(x)$, а *плотностью* распределения вещества. В одномерном случае плотность равна

$$f(x) = \frac{dF(x)}{dx}. \quad (10)$$

Функцию $f(x)$ называют *дифференциальной функцией распределения* или *плотностью вероятностей* величины X . Она еще удобней, чем $F(x)$, при анализе свойств случайных величин. Поэтому, говоря о *распределении* случайной величины X (без добавления интегральная или дифференциальная) имеют ввиду именно плотность вероятностей $f(x)$.

С помощью плотности вероятностей, формулу (9) записывают в виде

$$P(a \leq X < b) = \int_a^b f(x) dx. \quad (11)$$

В частности отсюда вытекает явное выражение интегральной функции через плотность вероятностей

$$F(x) = \int_{-\infty}^x f(u) du. \quad (12)$$

Перечислим общие свойства плотностей вероятностей. Устремив в (12) $x \rightarrow \infty$, приходим к условию нормировки, выраженному на языке

плотности вероятностей

$$\int_{-\infty}^{\infty} f(x) dx = 1. \quad (13a)$$

Второе общее свойство плотности вероятностей состоит в том, что $f(x)$, как и любая плотность, неотрицательна:

$$f(x) \geq 0. \quad (13b)$$

Дифференцируя обе части равенства (7) по x и вспомнив, что производная единичной функции равна дельта-функции:

$$\theta'(x) = \delta(x),$$

найдем плотность вероятностей дискретной случайной величины

$$f(x) = \sum_k p_k \delta(x - x_k). \quad (14)$$

Она также имеет наглядную механическую интерпретацию. Образно говоря, это сингулярная плотность набора бусинок (материальных точек) массами $p_1, p_2, \dots, p_n, \dots$, нанизанных на ось x .

Замечание: Если величина C не случайна, ее называют *детерминированной*. В любом испытании детерминированные величины принимают одинаковое значение. По этой причине их называют еще *постоянными* величинами. Они имеют сингулярную плотность вероятностей

$$f(x) = \delta(x - C). \quad \blacksquare \quad (15)$$

Замечание: Обратим внимание, что подобно плотности вещества, плотность вероятностей имеет размерность. Это полезно иметь ввиду, особенно в инженерных приложениях. Напомним, размерность одномерной плотности вещества равна размерности вещества, деленной на размерность длины. Так как вероятностная масса безразмерна, плотность вероятностей имеет размерность, обратную размерности аргумента

$$[f(x)] = [x]^{-1}.$$

Здесь квадратные скобки символизируют размерность заключенных в них величин. К примеру плотность вероятностей случайного напряжения имеет размерность единица на Вольт. \blacksquare

2.1.4 Связь дискретности и непрерывности

Выше обсуждались дискретные величины. Еще чаще встречаются *непрерывные случайные величины*. Типичным примером непрерывной величины служит случайная фаза принятого приемником локационного сигнала

$$X(t) = A \sin(\omega t + \Phi). \quad (16)$$



Рис. 2.3

Иллюстрация к нахождению интегральной функции распределения фазы, равномерно распределенной в интервале $[0, 2\pi]$.

Обычно положение источника сигнала и момент его включения заранее неизвестны. Поэтому полагают, что фаза Φ с равной возможностью принимает любые значения из интервала $\Phi \in [0, 2\pi]$. Это позволяет найти интегральную функцию распределения фазы чисто геометрически: Она равна длине общей части интервала $\Phi < \varphi$ и области возможных значений фазы $\Phi \in [0, 2\pi]$, отнесенной к ее длине 2π (см. Рис. 3).

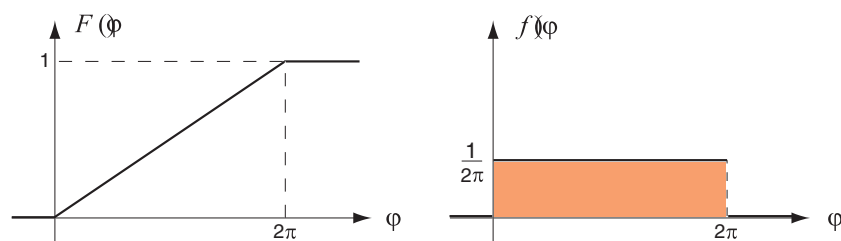


Рис. 2.4

Графики интегральной функции распределения и плотности вероятностей случайной фазы, равномерно распределенной на интервале $\Phi \in [0, 2\pi]$. Площадь закрашенной области под графиком плотности вероятностей равна единице.

График полученной так интегральной функции распределения Φ дан на Рис. 4. Соответствующее аналитическое выражение имеет вид:

$$F(\varphi) = P(\Phi < \varphi) = \begin{cases} 0, & \varphi < 0; \\ \frac{\varphi}{2\pi}, & 0 \leq \varphi < 2\pi; \\ 1 & 2\pi \leq \varphi. \end{cases} \quad (17)$$

Дифференцируя $F(\varphi)$, найдем плотность вероятностей случайной фазы:

$$f(\varphi) = \begin{cases} 0, & \varphi < 0; \\ \frac{1}{2\pi}, & 0 \leq \varphi < 2\pi; \\ 0 & 2\pi \leq \varphi. \end{cases} \quad (18)$$

Она постоянна внутри $\varphi \in (0, 2\pi)$, в полном согласии с равновозможностью значений фазы внутри этого интервала.

Мы еще познакомимся с самыми разнообразными непрерывными случайными величинами. Заметим однако, что ограниченная точность измерений делает все, и в том числе непрерывные величины, дискретными. Так рост человека указывают до сантиметров, а температуру воздуха до градусов. Поэтому, с позиций экспериментаторов, достаточно уметь работать с дискретными случайными величинами, и не обязательно вводить непрерывные. Напротив, теоретикам удобнее непрерывные величины, вероятностные свойства которых более приспособлены для аналитических расчетов. Мы же должны понимать, что выведенные для дискретных величин общие формулы (9)-(13) справедливы практически для всех, в том числе и непрерывных случайных величин: Подобно схеме шансов, теория дискретных величин часто выполняет эвристическую роль, делая наглядным вывод самых общих соотношений.

Подчеркнем еще, что в связи с интенсивным использованием цифровой техники обработки сигналов, первостепенное значение имеет процедура *квантования*. Применительно к случайным величинам она сводится к замене X , могущей принимать любые значения, дискретной случайной величиной \hat{X} . Опишем типичную процедуру квантования:

$$\hat{X} = \Delta \left\lfloor \frac{X}{\Delta} \right\rfloor, \quad (19)$$

где $\lfloor x \rfloor$ означает целую часть x . Пусть X обладает плотностью вероятностей $f(x)$. Тогда \hat{X} дискретная величина с плотностью вероятностей

$$\hat{f}(x) = \sum_k p_k \delta(x - k\Delta), \quad p_k = \int_{k\Delta}^{(k+1)\Delta} f(x) dx. \quad (20)$$

Пример: В математических программах имеются генераторы случайных чисел, равномерно распределенных в интервале $(0, 1)$. В программе MATLAB они вызываются командой

```
>> rand
```

Ниже приведен набор чисел, вызванных этой командой:

```
>> rand(1,12)
ans =
Columns 1 through 5
0.9218    0.7382    0.1763    0.4057    0.9355
Columns 6 through 10
0.9169    0.4103    0.8936    0.0579    0.3529
Columns 11 through 12
0.8132    0.0099
```

Видно, что на деле команда выдает не непрерывно распределенные числа, а их проквантованный аналог (19) с $\Delta = 10^{-4}$. Конечно, реальный шаг квантования гораздо меньше. Задав `long` формат, получим очередную реализацию

```
0.13889088195695
```

Но факт остается фактом: генератор случайных чисел в принципе не может генерировать истинно равномерно распределенные числа. Всегда есть, пусть малый, шаг квантования. Внутри компьютера он равен $\Delta = 2^{-r}$, а для внешнего потребителя, привыкшего к десятичной системе счисления, $\Delta = 10^{-s}$. Здесь $r, s \in \mathbb{N}$.

Выпишем распределение равномерно распределенной в интервале $(0, 1)$ случайной величины, проквантованной с шагом $\Delta = 1/n$. Оно равно

$$\hat{f}(x) = \frac{1}{n} \sum_{k=0}^{n-1} \delta \left(x - \frac{k}{n} \right). \quad (21)$$

2.2 Статистические средние

Функции распределения несут всю информацию о случайных величинах. Зачастую эта информация чересчур подробна. Поэтому вводят более простые числовые характеристики случайных величин. Ниже описаны некоторые из них, чаще всего встречающиеся в приложениях.

2.2.1 Математическое ожидание

Главной числовой характеристикой случайной величины является *математическое ожидание*. Обсудим иллюстрирующую его житейскую

ситуацию. Пусть X_i выручка таксиста в i -й день. Доходность бизнеса таксиста определяется *средней* выручкой

$$\bar{X}_N = \frac{1}{N} \sum_{i=1}^N X_i, \quad (1)$$

где N количество дней, по которым находят среднее. Вычислим предел арифметического среднего (1) при безграничном увеличении N , полагая что выручка не испытывает сезонных колебаний. В таком случае X_i есть исход i -го независимого испытания дискретной случайной величины X . Для подсчета суммы (1) перегруппируем слагаемые. Пусть $\{x_1 < x_2 < x_3 \dots x_k \dots\}$ возможные значения X . Тогда

$$\bar{X}_N = \sum_k x_k \frac{M_k}{N}.$$

Здесь M_k число испытаний из полного числа N , в которых случайная величина приняла значение x_k . В силу свойства статистической устойчивости (1.1.3), каждая дробь под знаком суммы имеет предел

$$\lim_{N \rightarrow \infty} \frac{M_k}{N} = p_k,$$

где p_k вероятность, что случайная величина X примет значение x_k . А значит предел арифметического среднего (1) равен

$$\lim_{N \rightarrow \infty} \bar{X}_N = \sum_k x_k p_k. \quad (2)$$

Выражение справа и есть математическое ожидание случайной величины X . Его еще называют *средним значением* (mean value), тем самым подчеркивая тесную связь математического ожидания с арифметическим средним (1). В дальнейшем будем использовать оба термина как синонимы. В разных отраслях науки и инженерии среднее значение обозначают по-разному. Одно из наиболее распространенных таково

$$E[X] = \sum_k x_k p_k.$$

Здесь E первая буква фразы “Expected value of X ”. В инженерных приложениях среднее значение часто символизируют угловыми скобками:

$$\langle X \rangle = \sum_k x_k p_k. \quad (3)$$

Именно их мы и будем преимущественно применять в дальнейшем.

Пример: Вычислим среднее число очков при бросании игральной кости. Так как вероятности выпадения любого числа очков, от одного до шести, одинаковы и равны $1/6$, сумма (3) принимает вид

$$\langle X \rangle = \frac{1}{6} \sum_{k=1}^6 k.$$

Суммируя арифметическую прогрессию, получаем

$$\langle X \rangle = \frac{1}{6} \times \frac{7 \times 6}{2} = \frac{7}{2} = 3.5 \quad \star \quad (4)$$

Замечание: Обратим внимание, что среднее не совпало ни с одним разрешенным значением. Это довольно типичная ситуация. Так статистика утверждает, что жители некоего городка в среднем съедают $2/3$ яйца в день. Это не значит, что они аккуратно срезают верхушку и выкидывают в мусорный бак. И никто, зная что там же в среднем в неделю венчаются 14.6 пар, не отпускает шуток в адрес неполной 0.6 пары. ■

Пример: Найдем математическое ожидание геометрического распределения. Согласно (3), (1.4) оно равно

$$\langle N \rangle = p \sum_{n=1}^{\infty} n q^{n-1}.$$

Выразим входящий сюда ряд через геометрическую прогрессию

$$\sum_{n=1}^{\infty} n q^{n-1} = \frac{d}{dq} \sum_{n=0}^{\infty} q^n = \frac{d}{dq} \left(\frac{1}{1-q} \right) = \frac{1}{(1-q)^2} = \frac{1}{p^2}.$$

Отсюда и из предыдущего равенства имеем

$$\langle N \rangle = p^{-1}.$$

Так среднее число бросков монеты до выпадения орла равно двум, а до выпадения единицы на игральной кости ее в среднем бросают 6 раз. ★

Дадим интегральную форму записи математического ожидания:

$$\langle X \rangle = \int_{-\infty}^{\infty} x f(x) dx. \quad (5)$$

Подставив сюда (1.14), приходим к выражению (3) для среднего дискретной величины. Сама же формула (5) справедлива всегда, когда имеет смысл интеграл в правой части.

Пример: Вычислим среднее значение случайной фазы Φ , равномерно распределенной в интервале $[0, 2\pi]$. Подставив в (5) плотность вероятностей (1.18), найдем искомое среднее

$$\langle \Phi \rangle = \frac{1}{2\pi} \int_0^{2\pi} \varphi d\varphi = \pi.$$

Результат очевиден: среднее попадает в центр интервала равновероятных значений. Заметим еще, что пределы последнего интеграла совпадают с границами интервала возможных значений, так как вне его плотность вероятностей (1.18) тождественно равна нулю. ★

2.2.2 Среднее по ансамблю

Напомним, математическое ожидание равно пределу арифметического среднего:

$$\langle X \rangle = \lim_{N \rightarrow \infty} \bar{X}_N. \quad (6)$$

Сам предел иногда трактуют как арифметическое среднее по всем, уже сделанным и только предстоящим, испытаниям случайной величины X . Эти (лишь богу известные) итоги всех испытаний образуют *статистический ансамбль*. Поэтому математическое ожидание называют еще *средним по ансамблю* или *статистическим средним* случайной величины X , обозначая его, как и арифметическое среднее, чертой, но без индекса N :

$$\langle X \rangle \equiv \bar{X}. \quad (7)$$

Язык статистических ансамблей вскрывает суть понятия математического ожидания: Так как арифметическое среднее постоянной – одинаковой во всех испытаниях – величины C , равно себе самой:

$$\langle C \rangle \equiv \bar{C} = C,$$

то математическое ожидание часто интерпретируют как *постоянную составляющую*, а разность случайной величины X и ее математического ожидания

$$\tilde{X} = X - \bar{X} \quad (\langle \tilde{X} \rangle \equiv 0) \quad (8)$$

как *случайную компоненту* величины X .

С помощью средних по ансамблю можно выразить любые статистические характеристики случайных величин. Покажем это на примере интегральной функции распределения. Напомним: арифметическое среднее единичной функции $\theta(x - X)$, взятое по N реализациям случайной

величины X , равно относительной частоте испытаний, при которых выполнено неравенство $X < x$. Следовательно, статистическое среднее от $\theta(x - X)$ совпадает с интегральной функцией распределения (1.5):

$$F(x) = \langle \theta(x - X) \rangle . \quad (9)$$

Концепция статистического ансамбля высвечивает замечательное свойство статистических средних: Вслед за арифметическим средним, *среднее по ансамблю линейной комбинации случайных величин равно линейной комбинации средних*. Поясним это свойство на примере величины

$$Z = aX + bY + c ,$$

где a, b, c детерминированные (постоянные), а X и Y произвольные случайные величины. Согласно указанному свойству

$$\bar{Z} = \overline{aX + bY + c} = a\bar{X} + b\bar{Y} + c . \quad (10)$$

В частности, постоянный множитель можно выносить за знак среднего:

$$\langle CX \rangle = C \langle X \rangle . \quad (10a)$$

Пример: Найдем среднее число очков, выпадающих при бросании двух игральных костей. Прямой способ основан на формуле (4) и Таблице 2, и сводится к подсчету суммы двенадцати слагаемых

$$\begin{aligned} \langle Z \rangle = & \frac{1}{36} \times (2 \times 1 + 3 \times 2 + 4 \times 3 + 5 \times 4 + 6 \times 5 + \\ & 7 \times 6 + 8 \times 5 + 9 \times 4 + 10 \times 3 + 11 \times 2 + 12 \times 1) . \end{aligned}$$

Мы скорее придем к ответу, представив число очков на двух костях суммой $X = Y + Z$, где Y и Z , соответственно, число очков на первой и второй костях. Пользуясь свойством (10), имеем

$$\langle X \rangle = \langle Y \rangle + \langle Z \rangle = \frac{7}{2} + \frac{7}{2} = 7 ,$$

поскольку среднее число очков на каждой кости нам уже известно: $\langle Y \rangle = \langle Z \rangle = 7/2$. Сомневающиеся в правильности “слишком быстрого” ответа пусть проверят его, сосчитав предыдущую сумму. ★

Упомянутое замечательное свойство статистических средних, выраженное равенством (10), можно сформулировать еще и так: *Статистическое усреднение и линейные операции коммутируют*. Проиллюстрируем это свойство, дифференцируя обе части равенства (9) по x :

$$\frac{dF(x)}{dx} = f(x) = \frac{d}{dx} \langle \theta(x - X) \rangle .$$

Операция дифференцирования – это линейная операция, и ее можно внести под знак среднего. Вспомнив, что производная единичной функции равна дельта функции, приходим к соотношению

$$f(x) = \langle \delta(X - x) \rangle, \quad (11)$$

согласно которому *плотность вероятностей равна статистическому среднему от дельта функции $\delta(X - x)$.*

Замечание: Напомним, угловые скобки означают усреднение по статистическому ансамблю. Подразумевается, что соответствующие статистические средние могут быть найдены как предел (6) арифметических средних. Это действительно так для большинства средних. Например, арифметическое среднее функции $\theta(x - X)$ по N реализациям X есть относительная частота реализации события $X < x$. Последняя, по свойству статистической устойчивости, сходится к вероятности $P(X < x)$. Применительно к равенству (11) подобная эмпирическая трактовка дает осечку. Действительно, $\delta(x - X)$ равна нулю при всех X кроме $X = x$, где дельта-функция равна бесконечности. Поэтому ни о каком пределе арифметических средних дельта-функции говорить не приходится. Тем не менее соотношение (11), не давая рецепта вычисления плотности вероятностей методом статистических испытаний, оказывается важным инструментом теории, в чем мы в дальнейшем не раз убедимся. ■

2.2.3 Средние функций случайных величин

Часто важна не исходная случайная величина X , а другая: $Y = g(X)$, где $y = g(x)$ известная функция. Если, к примеру, по условию игры, выигрыш растет как квадрат выпавших на кости очков $Y = X^2$, то игроку любопытно знать не среднее число очков, а средний выигрыш $\langle Y \rangle$. Аналогично, подводя к нелинейному детектору с вольт-амперной характеристикой $u = g(i)$ случайный ток I , на выходе имеют случайное напряжение $U = g(I)$. Спрашивается, можно ли найти среднее напряжение по функции распределения тока? Ответ на подобные вопросы в который раз подсказывает обращение к дискретным случайным величинам.

Пусть X дискретная величина, свойства которой заданы Таблицей 1. Тогда $Y = g(X)$ также дискретная величина со значениями

$$y_1 = g(x_1), \quad y_2 = g(x_2), \dots, y_n = g(x_n), \dots$$

реализующимися с вероятностями $p_1, p_2 \dots$ значений аргумента функции $Y = g(X)$. Следовательно, согласно (4), среднее Y равно

$$\langle Y \rangle = \sum_k y_k p_k = \sum_k g(x_k) p_k.$$

Заменив, как при переходе от (3) к (5), сумму интегралом, получим общую формулу среднего от функции случайной величины

$$\langle Y \rangle = \langle g(X) \rangle = \int_{-\infty}^{\infty} g(x) f(x) dx. \quad (12)$$

Пример: Найдём средний выигрыш упомянутого выше игрока в кости:

$$\langle X^2 \rangle = \frac{1}{6} \sum_{k=1}^6 k^2$$

Сосчитав сумму в лоб, или вспомнив известную формулу

$$\sum_{k=1}^n k^2 = \frac{n(n+1)(2n+1)}{6}, \quad (13)$$

получим ответ

$$\langle X^2 \rangle = \frac{91}{6} \simeq 15.17 \quad \star \quad (14)$$

Пример: Вычислим среднее радиолокационного сигнала (1.16). Согласно (12), (1.18) имеем

$$\langle X(t) \rangle = \frac{A}{2\pi} \int_0^{2\pi} \sin(\omega t + \varphi) d\varphi = \frac{A}{2\pi} \int_0^{2\pi} \sin \varphi d\varphi = 0. \quad \star$$

Замечание: При упрощении интеграла здесь использован прием, который не раз пригодится в дальнейшем: Пусть $g(x)$ периодическая функция с периодом L . Тогда интеграл по периоду от $g(x+a)$ не зависит от произвольного сдвига a :

$$\int_0^L g(x+a) dx = \int_0^L g(x) dx. \quad \blacksquare \quad (15)$$

Пример: Для передачи радиосигнала используют колебания высокой частоты. Чтобы его услышать, сигнал детектируют. С этой целью применяют детекторы, нелинейно преобразующие входной сигнал и выделяющие его постоянную составляющую. Например *линейный детектор*, описываемый кусочно-линейной функцией

$$g(x) = x \theta(x) = \begin{cases} x, & x \geq 0; \\ 0, & x < 0; \end{cases} \quad (16)$$

График линейного детектора и сигналов на его входе и выходе изображены на Рис. 4. Спрашивается, чему равно статистическое среднее пропущенного через линейный детектор гармонического сигнала (1.16)?

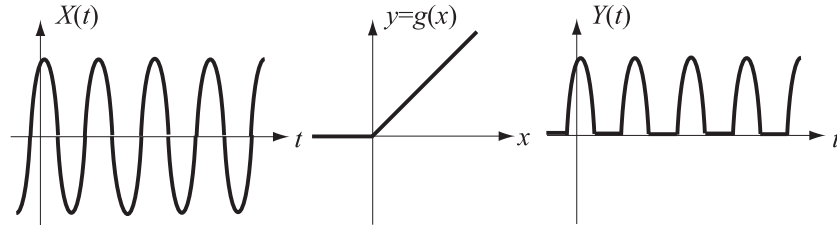


Рис. 2.5

В середине график характеристики линейного детектора, а слева и справа схематические графики входного и выходного сигналов.

Найдем среднее, пользуясь приемом (15). Здесь он позволяет заметить процесс (1.16) случайной величиной $X = A \sin \Phi$. Линейный детектор преобразует ее в $Y = A\theta(\sin \Phi) \sin \Phi$. Таким образом

$$\langle Y \rangle = A \langle \theta(\sin \varphi) \sin \varphi \rangle = \frac{A}{2\pi} \int_0^\pi \sin \varphi d\varphi = \frac{A}{\pi}. \quad \star$$

2.2.4 Моменты и дисперсия

Функции распределения, несущие всю информацию о случайных величинах, довольно трудно измерить. Поэтому зачастую ограничиваются измерением более простых средних, предпочитая оперировать парой-тройкой чисел, а не функциями распределения. Прежде всего измеряют математическое ожидание случайной величины, дополняя его несколькими *моментами*. Моментом n -го порядка называют среднее

$$\mu_n = \int_{-\infty}^{\infty} x^n f(x) dx \quad (17)$$

степенной функции X^n ($n \in \mathbb{N}$). По этой терминологии математическое ожидание есть момент первого порядка. Кроме него обычно измеряют *средний квадрат* случайной величины

$$\mu_2 = \langle X^2 \rangle = \int_{-\infty}^{\infty} x^2 f(x) dx. \quad (18)$$

Но пожалуй наиболее информативным средним, помимо математического ожидания, является *дисперсия* случайной величины X . Так называют средний квадрат ее случайной составляющей (8):

$$D[X] = \langle \widetilde{X}^2 \rangle = \langle (X - \langle X \rangle)^2 \rangle. \quad (19)$$

Если математическое ожидание есть постоянная составляющая X , то дисперсия служит *количественной мерой случайности* – разброса значений X вокруг ее среднего. В частности для детерминированной величины, совпадающей со своим средним, дисперсия равна нулю.

Выразим дисперсию произвольной случайной величины X через ее первые два момента. Для этого раскроем квадратные скобки под знаком среднего в (19) и используем свойства (10), (10а) статистических средних:

$$D[X] = \langle (X^2 - 2 \langle X \rangle X + \langle X \rangle^2) \rangle = \langle X^2 \rangle - 2 \langle X \rangle \langle X \rangle + \langle X \rangle^2.$$

Частично сократив последние два слагаемых, получим окончательно

$$D[X] = \langle X^2 \rangle - \langle X \rangle^2 = \mu_2 - \mu_1^2. \quad (20)$$

Таким образом, *дисперсия случайной величины равна ее среднему квадрату минус квадрат среднего*.

Поскольку дисперсия равна среднему квадрата \widetilde{X}^2 , неотрицательного в любых испытаниях, дисперсия также неотрицательна

$$D[X] \geq 0 \quad (21)$$

и равна нулю лишь для постоянных величин. Кроме того, из (21) и (20) вытекает еще одно полезное неравенство

$$\langle X^2 \rangle \geq \langle X \rangle^2. \quad (22)$$

Напомним: дисперсия есть мера разброса случайной величины вокруг ее среднего. Однако эта мера не всегда удобна в инженерных приложениях, где имеют дело с *размерными* величинами: током, напряжением, давлением и т.д. Мету случайности размерных величин лучше характеризуют величины той же размерности. Поэтому инженеры предпочитают не дисперсию, а *среднеквадратичное отклонение* случайной величины от ее среднего:

$$\sigma_x = \sqrt{D[X]} \iff D[X] = \sigma_x^2. \quad (23)$$

Его еще называют *стандартным отклонением*, или просто *стандартом* случайной величины X .

Если нас интересуют свойства лишь случайной компоненты \widetilde{X} (8), разумно, вместо моментов μ_n (17), вычислять *центральные моменты* случайной величины X . По определению они равны

$$\nu_n = \langle \widetilde{X}^n \rangle = \langle (X - \langle X \rangle)^n \rangle = \int_{-\infty}^{\infty} (x - \mu_1)^n f(x) dx \quad (n \in \mathbb{N}). \quad (24)$$

В частности 1-й центральный момент всегда равен нулю, а второй совпадает с дисперсией случайной величины ($D[X] \equiv \nu_2$).

Разберем несколько примеров, позволяющих лучше освоить введенные выше ключевые понятия теории случайных величин.

Пример: Вычислим дисперсию и среднеквадратичное отклонение числа очков, выпадающих на игральной кости. Привлекая уже имеющиеся в нашем распоряжении первые два момента (4) и (14), будем иметь

$$D[X] = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12} \simeq 2.9167, \quad \sigma_x \simeq 1.7078 \quad \star$$

Куда важней для теории вероятностей и ее приложений другой, более простой пример, рассмотренный ниже.

Пример: Пусть событие A реализуется в испытании с вероятностью p . Зададим случайную величину (так называемый *индикатор события*) $X(A)$, равную единице, если A происходит, и нулю в противном случае. Найти плотность вероятностей, дисперсию и стандарт индикатора события.

Очевидно, плотность вероятностей такова

$$f(x) = q \delta(x) + p \delta(x - 1) \quad (q = 1 - p). \quad (25)$$

Это так называемое *распределение Бернулли*. Его первые два момента одинаковы:

$$\langle X \rangle = 0 \times q + 1 \times p = p, \quad \langle X^2 \rangle = 0^2 \times q + 1^2 \times p = p,$$

а дисперсия и стандартное отклонение равны

$$D[X] = p - p^2 = p(1 - p) = pq, \quad \sigma_x = \sqrt{p(1 - p)}. \quad (26)$$

График стандарта индикатора события, выражающий зависимость случайности события A от вероятности p , дан на Рис. 4. Из него видно, что максимум стандартного отклонения приходится на $p = 1/2$. При $p \rightarrow 1$ или при $p \rightarrow 0$, когда событие A становится все более достоверным или все более невозможным — так или иначе все менее случайным — стандарт стремится к нулю. ★

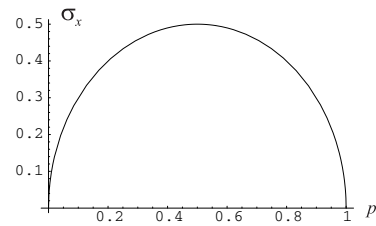


Рис. 2.6
График стандартного отклонения индикатора события.

Пример: Найдём плотность вероятностей, среднее, центральные моменты и стандарт отклонения случайной величины X , равномерно распределенной в интервале (a, b) .

Равномерность распределения означает, что $f(x)$ постоянна внутри интервала разрешенных значений (a, b) и обращается в нуль вне его. Амплитуду $f(x)$ находят из условия нормировки. В итоге равномерное распределение примет вид

$$f(x) = \frac{1}{b-a} [\theta(x-a) - \theta(x-b)] . \quad (27)$$

Для лучшего понимания дальнейшего, явно введем координату центра интервала возможных значений и его полуширину:

$$c = \frac{a+b}{2}, \quad \ell = \frac{b-a}{2} . \quad (27a)$$

Среднее плотности вероятностей (27) совпадает с центром интервала возможных значений

$$\langle X \rangle = \frac{1}{b-a} \int_a^b x dx = \frac{1}{2} \frac{b^2 - a^2}{b-a} = c ,$$

а центральные моменты

$$\nu_n = \frac{1}{b-a} \int_a^b (x-c)^n dx = \frac{1}{2\ell} \int_{-\ell}^{\ell} y^n dy = \begin{cases} \frac{\ell^n}{(n+1)}, & n = 2m ; \\ 0, & n = 2m + 1 ; \end{cases}$$

не зависят от положения центра интервала возможных значений. Кроме того, из-за симметрии равномерного распределения вокруг его центра,

все нечетные центральные моменты обращаются в нуль. В частности, дисперсия и стандарт равномерного распределения равны

$$D[X] = \frac{\ell^2}{3}, \quad \sigma_x = \frac{\ell}{\sqrt{3}} \simeq 0.577 \ell.$$

График равномерного распределения приведен на Рис. 7 слева. Там же справа изображено 500 реализаций равномерно распределенной в интервале $(0, 1)$ случайной величины X , сгенерированные программой MATLAB:

```
>> x=rand(1,500);
>> plot(x)
```

★

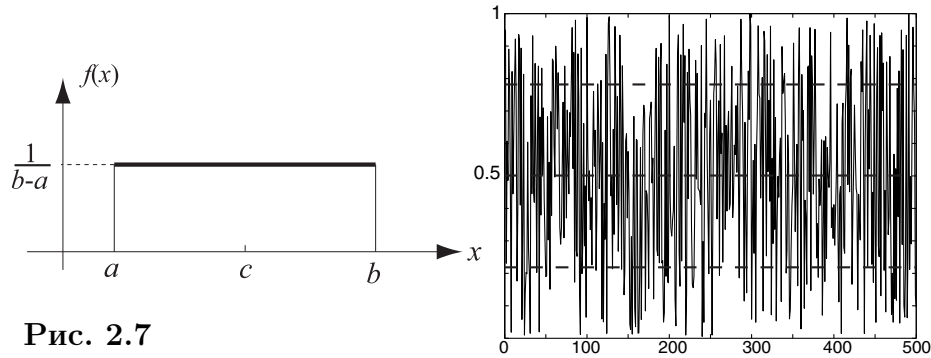


Рис. 2.7

График плотности вероятностей равномерно распределенной случайной величины (слева) и 500 реализаций случайной величины, равномерно распределенной в единичном интервале $(0, 1)$ (справа). Там же пунктиром указаны ее среднее значение $\langle X \rangle = 0.5$ и границы интервала $(\langle X \rangle - \sigma_x, \langle X \rangle + \sigma_x)$.

2.2.5 Характеристическая функция

Помимо функций распределения, для анализа свойств случайных величин привлекают *характеристическую функцию*. Мы придем к ней, положив в (12) $g(x) = e^{iux}$, где i мнимая единица:

$$\Theta(u) = \langle e^{iuX} \rangle = \int_{-\infty}^{\infty} f(x) e^{iux} dx. \quad (28)$$

Впоследствии обнаружатся многие замечательные свойства $\Theta(u)$, делающие ее ценным инструментом исследования случайных величин. Здесь же заметим, что $\Theta(u)$ сохраняет полную информацию о плотности вероятностей $f(x)$. В самом деле, правая часть равенства (28) есть *интеграл*

Фурье от плотности вероятностей. Следовательно, зная характеристическую функцию случайной величины X , можно восстановить ее плотность вероятностей обратным преобразованием Фурье

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Theta(u) e^{-iux} dx. \quad (29)$$

Одно из полезных свойств $\Theta(u)$ состоит в том, что моменты случайной величины можно вычислить не интегрированием, как в формулах (17), (24), а *дифференцированием* характеристической функции. Это важно, так как дифференцирование проще операции интегрирования. Дифференцируя обе части равенства (28) n раз по u , будем иметь

$$\frac{d^n \Theta(u)}{du^n} = i^n \int_{-\infty}^{\infty} f(x) x^n e^{iux} dx.$$

Положив затем $u = 0$, после тривиальных выкладок получим

$$\mu_n = \int_{-\infty}^{\infty} x^n f(x) dx = \frac{1}{i^n} \left. \frac{d^n \Theta(u)}{du^n} \right|_{u=0}. \quad (30)$$

Иначе говоря, моменты случайной величины X определяют разложение ее характеристической функции в ряд Тейлора

$$\Theta(u) = \sum_{n=0}^{\infty} \mu_n \frac{(iu)^n}{n!}. \quad (31)$$

Дифференцированием удастся вычислить и центральные моменты X . Покажем это, заметив, что характеристическая функция случайной компоненты \tilde{X} (8) равна

$$\tilde{\Theta}(u) = \langle e^{iu\tilde{X}} \rangle = \langle e^{iuX - iu\langle X \rangle} \rangle.$$

Поскольку постоянный множитель выносится за знак среднего, то

$$\tilde{\Theta}(u) = \Theta(u) e^{-iu\langle X \rangle}. \quad (32)$$

Отсюда, по аналогии с (31), имеем

$$\nu_n = \frac{1}{i^n} \left. \frac{d^n \tilde{\Theta}(u)}{du^n} \right|_{u=0} = \frac{1}{i^n} \frac{d^n}{du^n} \left[\Theta(u) e^{-iu\langle X \rangle} \right] \Big|_{u=0}. \quad (33)$$

Дадим несколько примеров вычисления и использования характеристических функций.

Пример: Первой найдем характеристическую функцию распределения Бернулли (25). Из (25), (28) имеем

$$\Theta(u) = q + p e^{iu}. \quad \star \quad (34)$$

Пример: Второй вычислим характеристическую функцию геометрического распределения (1.4). Для этого вначале запишем его плотность вероятностей

$$f(x) = \frac{p}{q} \sum_{n=1}^{\infty} q^n \delta(x - n). \quad (35)$$

Подставив (35) в (29) и суммируя геометрическую прогрессию, найдем

$$\Theta(u) = \frac{p e^{iu}}{1 - q e^{iu}}. \quad (36)$$

Дифференцируя ее “в лоб”, можно вычислить моменты геометрического распределения. Мы же воспользуемся мощностью современных программ. В данном случае прибегнем к услугам программы МАТЕМАТИКА. Для этого вспомним, что математическое ожидание геометрического распределения нами уже вычислено: $\langle X \rangle = 1/p$. Домножив (36) на $\exp(-iu/p)$, перейдем к характеристической функции случайной компоненты \tilde{X} (8) геометрически распределенной величины. Обозначим ее `Theta[u]`. После чего выполним команды

```
Series[Theta[u], {u, 0, 3}]
```

```
FullSimplify[%]
```

В итоге получим

$$\text{Theta}[u] = 1 + \frac{q}{p^2} \frac{(iu)^2}{2} + \frac{(1+q)q}{p^3} \frac{(iu)^3}{6} + \dots$$

В частности, коэффициент при $(iu)^2/2$ равен дисперсии геометрического распределения, а его стандарт

$$\sigma_x = \frac{\sqrt{1-p}}{p} = \langle X \rangle \sqrt{1 - \frac{1}{\langle X \rangle}}. \quad \star \quad (37)$$

Пример: Следующей найдем характеристическую функцию равномерно распределенной величины (27). В обозначениях (27а) будем иметь

$$\Theta(u) = \frac{1}{2\ell} \int_a^b e^{iux} dx = e^{iuc} \frac{\sin u\ell}{u\ell}. \quad (38)$$

Первый множитель здесь равен $e^{iu\langle X \rangle}$, а значит второй есть характеристическая функция случайной компоненты \tilde{X} :

$$\tilde{\Theta}(u) = \frac{\sin ul}{ul}. \quad (39)$$

Она действительна из-за симметрии равномерного распределения относительно его среднего значения. ★

Пример: Найдём характеристическую функцию случайного гармонического радиолокационного сигнала $X = A \sin \Phi$. Она равна

$$\Theta(u) = \langle e^{iuA \sin \Phi} \rangle.$$

Вычислив среднее с помощью плотности вероятностей (1.18), получим

$$\Theta(u) = \frac{1}{2\pi} \int_0^{2\pi} e^{iuA \sin \varphi} d\varphi = J_0(Au). \quad (40)$$

Здесь $J_0(z)$ функция Бесселя нулевого порядка. Хорошо известно ее разложение в ряд Тейлора:

$$J_0(z) = 1 - \left(\frac{z}{2}\right)^2 + \frac{1}{2^2} \left(\frac{z}{2}\right)^4 - \frac{1}{2^2 3^2} \left(\frac{z}{2}\right)^6 + \dots \quad (41)$$

Отсюда и из формулы (31) легко найти любой момент случайной величины $X = A \sin \Phi$. В частности, дисперсия и стандарт случайного радиолокационного сигнала таковы

$$D[X] = \frac{1}{2}A^2, \quad \sigma_x = \frac{A}{\sqrt{2}}.$$

Обратим внимание, что мы вычислили их, даже не зная плотности вероятностей X (ее мы отыщем и подробно обсудим в будущем). ★

2.2.6 Свойства характеристической функции

Укажем общие свойства характеристической функции. Прежде всего, из условия нормировки плотности вероятностей $f(x)$ и из определения характеристической функции (29) следует, что

$$\Theta(u=0) \equiv 1. \quad (42)$$

Кроме того из интегрального неравенства

$$|\Theta(u)| = \left| \int_{-\infty}^{\infty} f(x) e^{iux} dx \right| \leq \int_{-\infty}^{\infty} f(x) dx = 1$$

вытекает, что

$$|\Theta(u)| \leq 1. \quad (43)$$

Условия (42), (43) необходимы но недостаточны для того, чтобы $\Theta(u)$ была характеристической функцией некоторой случайной величины X . Для этого $\Theta(u)$ должна быть еще *положительно определенной*, то есть чтобы ее обратное преобразование Фурье (30) было неотрицательно.

Порой теоретики моделируют характеристическую функцию случайной величины X некоторой функцией $\Theta(u)$. Выводы теории имеют вероятностный смысл лишь если $\Theta(u)$ обладает перечисленными свойствами. Труднее всего проверить наличие положительной определенности. Однако если она нарушена, это как правило легко показать, и исключить негодную функцию $\Theta(u)$ из рассмотрения. Действенным критерием отсева ложных характеристических функций служит проверка значения ее второй производной в нуле. Для характеристических функций случайных величин это значение должно быть строго положительным

$$\gamma = - \left. \frac{d^2\Theta(u)}{du^2} \right|_{u=0} > 0, \quad (44)$$

так как, согласно (31), γ равна среднему квадрату, положительному для случайных величин.

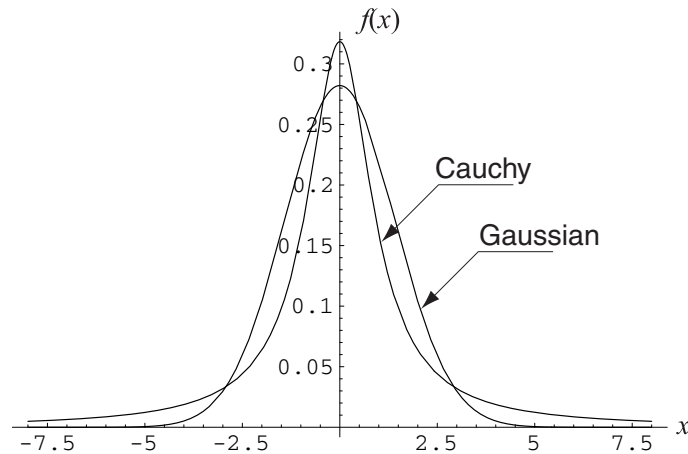


Рис. 2.8

Графики Коши и Гауссова плотностей вероятностей. Видно, что при больших x распределение Коши стремится к нулю медленнее, чем распределение Гаусса. Этим и объясняется отсутствие среднего квадрата у случайных величин с распределением Коши.

Пример: Важную роль в теории играют характеристические функции

$$\Theta(u) = e^{-|u|^\alpha} \quad (\alpha \geq 0). \quad (45)$$

Однако не все степени $\alpha \geq 0$ разрешены. Критерий (44) отсекает запрещенные значения α . Мы укажем их, но сперва обсудим несколько частных случаев.

Возьмем $\alpha = 1$. Тогда производная функции (45) при $u = 0$ не существует (равна бесконечности), что не противоречит критерию (45). Действительно, вычислив обратное преобразование Фурье (30), имеем

$$\Theta(u) = e^{-|u|} \iff f(x) = \frac{1}{\pi} \frac{1}{1+x^2}. \quad (46)$$

Это знаменитое *распределение Коши*. Его график приведен на Рис. 8. Значение $\gamma = \infty$ здесь связано с медленным (как $f(x) \sim x^{-2}$ при $x \rightarrow \pm\infty$) спаданием хвостов распределения Коши. Само же оно всюду больше нуля, гарантируя тем самым положительную определенность $\Theta(u)$.

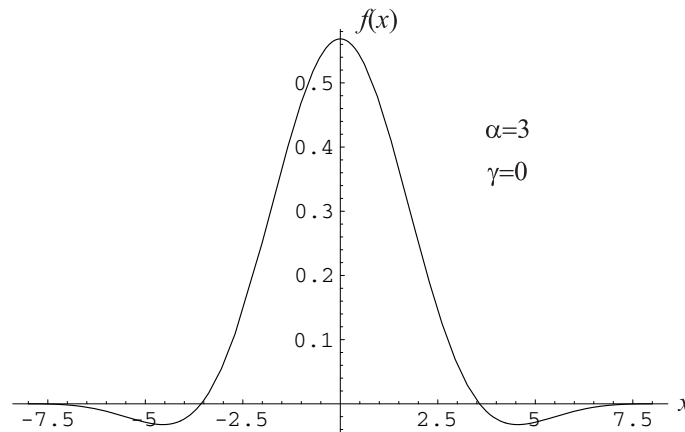


Рис. 2.9

График $f(x)$, отвечающей функции $\Theta(u)$ (45) при $\alpha = 3$. Видно, что $f(x)$ знакопеременная и не может быть плотностью вероятностей.

Следующим обсудим случай $\alpha = 2$. Для него $\gamma = 2 > 0$, а значит функция (45) прошла тест (44) на звание характеристической функции. В самом деле, обратное преобразование Фурье (30) дает

$$\Theta(u) = e^{-u^2} \iff f(x) = \frac{1}{2\sqrt{\pi}} \exp\left(-\frac{x^2}{4}\right). \quad (47)$$

Это еще более знаменитое *распределение Гаусса* со стандартным отклонением $\sigma_x = \sqrt{2}$. График его также приведен на Рис.8. В дальнейшем мы часто будем сталкиваться с этим замечательным распределением.

Выпишем напоследок явное выражение второй производной функции (45). При $\alpha \geq 2$ она существует при любых u и равна

$$\Theta^{(2)}(u) = e^{-|u|^\alpha} \alpha \left[\alpha |u|^{2\alpha-1} - (\alpha - 1) |u|^{\alpha-2} \right],$$

а при $\alpha > 2$ и $u = 0$ обращается в нуль. Это значит, что при $\alpha > 2$ функция (45) не может быть характеристической. Сказанное подтверждает, построенный численным интегрированием, график обратного преобразования Фурье от функции $e^{-|u|^3}$, изображенной на Рис. 9. С другой стороны, можно показать, что для всех $0 < \alpha \leq 2$ функция $\Theta(u)$ (45) положительно определена. ★

2.3 Испытания Бернулли

Неоднократно встречавшееся понятие независимых испытаний обретает глубокий смысл, тесно связанный с идеологией научного эксперимента, при обсуждении *испытаний Бернулли*. Так называют серию независимых испытаний, в которых с одинаковой вероятностью может реализоваться или нет некое случайное событие. Ниже мы подробно обсудим испытания Бернулли и их вероятностные свойства.

2.3.1 Независимые испытания

Напомним главную идею, лежащую в основе научных экспериментов. Она гласит, что если тщательно воспроизвести условия опыта, например заново собрать электрическую цепь и подключить к ней заданный источник питания, то исход будет тот же, что и в предыдущих опытах: по цепи потечет ток предсказанной силы. В испытаниях случайных событий, например события A , как бы тщательно ни воспроизводились условия испытаний, полная повторяемость отсутствует: событие A или произойдет или нет. Тем не менее, выполняя серию испытаний, экспериментатор должен свято блюсти *идентичность условий* их проведения. Это накладывает принципиальное ограничение на вероятностные свойства случайных событий в разных испытаниях: Предыдущие испытания не должны влиять на возможные исходы последующих. Иными словами, случайные события в разных испытаниях должны быть *статистически независимы*.

Пример: Классический пример независимых испытаний дает бросание монеты: Хорошо согласуется с опытом тот факт, что если каждый раз высоко подбрасывать монету и сильно ее раскручивать, то вероятность выпадения герба при n -ом бросании не зависит от того, сколько раз выпал орел в предыдущих $n - 1$ бросаниях: Даже если случилось так, что в первых десяти испытаниях выпала решка, вероятность выпадения орла в 11-м испытании та же, что и в первом. А именно $p = 1/2$. ★

Замечание: Данный пример служит хорошим тестом на понимание сути случайных явлений. Многие его не выдерживают, искренне полагая, что если орел не выпал десять раз, то вероятность ему выпасть в одиннадцатый больше одной второй. Вообще наша психика устроена так, что подсознательно ищет причинно-следственные связи всюду, даже где их нет. Поэтому человек не в силах написать длинную последовательность нулей и единиц, которую объективные статистические тесты признали бы совершенно случайной. К примеру, пытаясь составить случайную последовательность, человек старательно чередует нули и единицы. В то же время известно, что в истинно случайной последовательности вероятность встретить набор только нулей (единиц) длиной $\log_2 n$ стремится к единице при $n \rightarrow \infty$. Говорят, существует компьютерная игра, состоящая в том, что человек вводит плюсы и минусы, а компьютер прогнозирует знак следующего хода. И чем больше ходов сделано, тем чаще компьютер угадывает, опираясь на скрытую закономерность, присущую человеческой психологии. ■

2.3.2 Биномиальное распределение

Утверждение о независимости исходов последовательных испытаний нельзя логически обосновать или надежно подтвердить испытаниями. Поэтому независимость исходов входит в теорию как постулат, определяющий структуру вероятностных свойств серии испытаний. Опишем ее, для чего пометим события индексами. Например A_i, B_k , где индекс указывает номер испытания. Так B_k есть событие B в k -м испытании. Соответственно, $P(B_k)$ вероятность, что B произойдет в k -м испытании.

Сосредоточимся на одном событии A и противоположном ему \bar{A} . Пусть A может реализоваться в каждом испытании с одинаковой вероятностью p . Тогда вероятность противоположного события \bar{A} равна $1 - p$:

$$P(A) = p, \quad P(\bar{A}) = q = 1 - p, \quad p + q = 1. \quad (1)$$

События A и \bar{A} несовместны в одном испытании, а в разных независимы.

Поэтому справедливы равенства

$$\begin{aligned} P(A_i A_i) = p, \quad P(\bar{A}_i \bar{A}_i) = q, \quad P(A_i \bar{A}_i) = 0, \\ P(A_i A_k) = p^2, \quad P(\bar{A}_i \bar{A}_k) = q^2, \quad P(A_i \bar{A}_k) = pq \\ (i \neq k). \end{aligned} \quad (2)$$

Пока совершенно неважно, что собой представляет событие A . Поэтому назовем его S (success – успех), а противоположное F (failure – неудача). Теперь мы готовы дать окончательное определение испытаний Бернулли:

Серия n независимых испытаний, имеющих два исхода: успех S или неудача F , с вероятностями p и q , неизменными для всех испытаний, называют испытаниями Бернулли.

Так как испытания независимы, вероятность любой конкретной последовательности успехов и неудач, например

$$\underbrace{S_1 S_2 F_3 S_4 F_5 \dots F_{n-2} F_{n-1} S_n}_n \quad (3)$$

равна произведению их вероятностей p и q :

$$P(S_1 S_2 F_3 S_4 F_5 \dots F_{n-2} F_{n-1} S_n) = p p q p q \dots q q p = p^k q^{n-k}, \quad (4)$$

где k количество успехов в n испытаниях, а $n - k$ число неудач.



Рис. 2.10

Шары, полученные как награда за успехи в испытаниях, расставленные в порядке (5) проведения испытаний.

Обычно более существенно общее число успехов k , а не то, как чередуются успехи и неудачи в n испытаниях. К примеру говорят, что бизнесмен успешно вел дело, если ему сопутствовало больше успехов, чем неудач ($k > n/2$). Поэтому найдем вероятности $P_n(k)$, что в n испытаниях было k успехов. Начнем с подсчета всех комбинаций успехов и неудач (3), содержащих k успехов. Для этого выпишем из (3) последовательность номеров успешных испытаний

$$j_1 < j_2 < \dots < j_{k-1} < j_k. \quad (5)$$

Будем трактовать их как номера k шаров, вынутых из урны с n пронумерованными шарами (см. Рис. 10). Напомним — всего способов извлечь k шаров из n без возвращения равно числу размещений

$$n \times (n - 1) \times \dots (n - k + 1) = \frac{n!}{(n - k)!}$$

Мы найдем число комбинаций успехов с возрастающей последовательностью номеров (5), оставив из всех размещений лишь выборки с возрастающими номерами. Для этого надо поделить число размещений на число $k!$ перестановок k шаров. В итоге оказывается, что число последовательностей (3) с k успехами равно биномиальному коэффициенту $\binom{n}{k}$ (1.2.11).

В каждой серии из n испытаний конкретная последовательность (3) k успехов и $n - k$ неудач *несовместна* с любой другой, хотя бы и содержащей то же число успехов. Следовательно, по аксиоме аддитивности, вероятность объединения всех разных последовательностей (3), содержащих k успехов, равна сумме их вероятностей (4):

$$P_n(k) = \binom{n}{k} p^k q^{n-k}. \quad (6)$$

Формула (6) для вероятности k успехов из n испытаний носит название *биномиального распределения*. Графики биномиального распределения для $n = 12$ и разных значений вероятности успеха p даны на Рис. 11.

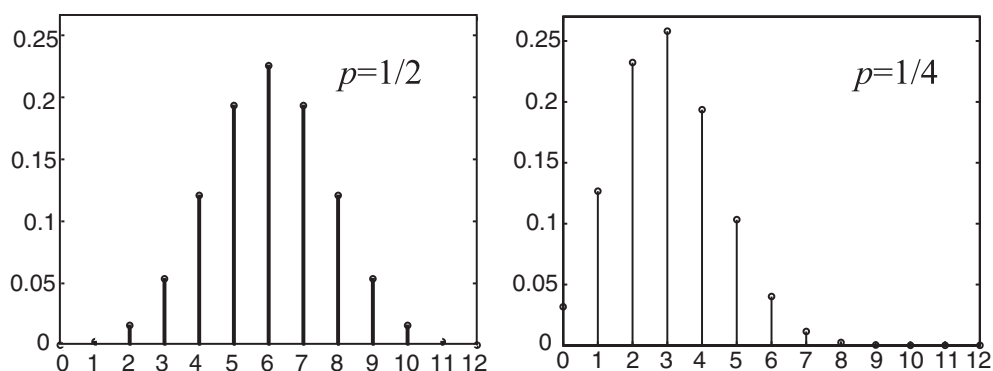


Рис. 2.11

Биномиальные распределения для разных вероятностей успеха.

Видно, что при меньшем p вероятности числа успехов группируются левее, при меньших значениях k .

Формула (6) дает вероятность события H_k , что в серии из n испытаний, трактуемых как одно макро-испытание, будет k успехов. Набор всех

таких $n + 1$ событий $\{H_0, H_1, \dots, H_n\}$ образует полную группу несовместных событий.

Часто истинный успех (событие T true success) приходит, лишь если в серии из n испытаний успехи перевалят некоторое пороговое значение m . Например, спортсмена введут в национальную сборную, если в восьми отборочных соревнованиях он победил не менее трех раз. В силу взаимной несовместности H_k , вероятность события T получают суммированием вероятностей всех благоприятствующих T событий H_k :

$$P(T) = P(H_m \cup H_{m+1} \cup \dots \cup H_n) = \sum_{k=m}^n P_n(k).$$

Пример: Некоторое устройство состоит из 6 блоков, причем для нормальной работы устройства (событие W working) необходимо, чтобы исправно функционировало как минимум 5 блоков. Пусть вероятности отказа блоков в критической ситуации одинаковы и равны $q = 0.1$. Какова вероятность, что устройство сработает в такой ситуации?

В данном случае $p = 0.9$, $q = 0.1$, $n = 6$ и $m = 5$. Следовательно

$$P(W) = \binom{6}{5} \times 0.9^5 \times 0.1 + \binom{6}{6} \times 0.9^6 \simeq 6 \times 0.5905 \times 0.1 + 0.5314 \simeq 0.8857 \quad \star$$

Вероятности (6) удобно трактовать еще как вероятности возможных значений случайной величины K — числа успехов в n испытаниях. Ее плотность вероятностей равна

$$f(x) = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} \delta(x - k), \quad (7)$$

а соответствующая характеристическая функция имеет вид

$$\Theta(u) = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} e^{iku}. \quad (8)$$

Сопоставив ее с известной формулой бинома Ньютона

$$(a + b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k},$$

придем к более компактной форме записи характеристической функции случайного числа K успехов в n испытаниях

$$\Theta(u) = (pe^{iu} + q)^n. \quad (9)$$

Найдем среднее значение случайной величины K , опираясь на свойство статистических средних (2.10). Для этого запишем K в виде суммы более простых случайных величин

$$K = X_1 + X_2 + \dots + X_n, \quad (10)$$

где X_m равно единице, если событие A_m имело место, и нулю в противном случае. Каждая из них распределена по закону Бернулли (2.26) со средним $\langle X_k \rangle = p$. Следовательно, по свойству средних (2.10)

$$\langle X \rangle = \sum_{m=1}^n \langle X_m \rangle = \underbrace{p + p + \dots + p}_n = np. \quad (11)$$

Как и следовало ожидать, среднее число успехов линейно зависит как от вероятности успеха в одном испытании, так и от числа испытаний.

Замечание: Советуем читателю вывести соотношение (11), используя формулу вычисления среднего с помощью характеристической функции:

$$\langle K \rangle = -i \frac{d\Theta(u)}{du} \Big|_{u=0}. \quad \blacksquare$$

Пример: Найдем характеристическую функцию случайной компоненты \tilde{K} числа успехов в n испытаниях и вычислим с ее помощью дисперсию и стандарт числа успешных испытаний. Помножив (9) на e^{-ipnu} , получим

$$\tilde{\Theta}(u) = (pe^{iqu} + qe^{-ipu})^n.$$

Разложим $\tilde{\Theta}(u)$ в ряд Тейлора по степеням u до u^2 включительно

$$\tilde{\Theta}(u) = 1 - \frac{n}{2}pq u^2 + O(u^3).$$

Следовательно, дисперсия и стандарт числа успехов в n испытаниях равны

$$D[K] = npq, \quad \sigma_k = \sqrt{npq}. \quad \star \quad (12)$$

Обратим внимание, что дисперсия числа успехов, как и их среднее, линейно растет с числом испытаний n .

2.3.3 Закон редких событий

Часто в физических, инженерных и экономических приложениях число испытаний n велико, а вероятность исследуемого события p мала. Так число испытаний может быть равно числу атомов газа в сосуде или количеству единиц продукции, выпускаемой фабрикой за год, а интересующее событие – распад атома или наличие брака. Вероятности указанных событий как правило очень малы. В подобных случаях пользуются не биномиальным распределением, а более простым *законом редких событий*. Он возникает как предел биномиального распределения при

$$n \rightarrow \infty, \quad np = \lambda \quad (0 < \lambda < \infty). \quad (13)$$

Событие, вероятность которого p подчиняется предельным соотношениям (13), назовем *редким событием*. Входящий в (13) параметр λ количественно характеризует степень редкости события.

Пусть $P_k(\lambda)$ вероятность что редкое событие произойдет k раз. По данному выше определению она равна

$$P_k(\lambda) = \lim_{n \rightarrow \infty} P_n(k) = \lim_{n \rightarrow \infty} \binom{n}{k} p^k q^{n-k}. \quad (14)$$

Найдем вначале вероятность $P_0(\lambda)$, что редкое событие ни разу не произойдет. Согласно биномиальному распределению

$$P_n(0) = q^n = (1 - p)^n = \left(1 - \frac{\lambda}{n}\right)^n.$$

Следовательно, из 2-го замечательного предела имеем

$$P_0(\lambda) = \lim_{n \rightarrow \infty} P_n(0) = \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}. \quad (15)$$

Выведем рекуррентное соотношение, связывающее вероятности $P_k(\lambda)$ и $P_{k+1}(\lambda)$. Заметим, что из биномиального распределения (6) вытекает равенство

$$P_n(k+1) = \frac{n-k}{k+1} \times \frac{p}{q} \times P_n(k).$$

Переходя в его обеих частях к пределу (13), будем иметь

$$\lim_{n \rightarrow \infty} P_n(k+1) = \lim_{n \rightarrow \infty} \left[\frac{n-k}{k+1} \frac{p}{q} \right] \times \lim_{n \rightarrow \infty} P_n(k).$$

С учетом (14) отсюда получаем

$$P_{k+1}(\lambda) = P_k(\lambda) \times \lim_{n \rightarrow \infty} \left[\frac{n-k}{k+1} \frac{p}{q} \right].$$

Вычислим оставшийся предел, заменив p на λ/n , а q на $1 - \lambda/n$:

$$\lim_{n \rightarrow \infty} \left[\frac{n-k}{k+1} \frac{p}{q} \right] = \frac{\lambda}{1+k} \lim_{n \rightarrow \infty} \left[\frac{1 - \frac{k}{n}}{1 - \frac{\lambda}{n}} \right] = \frac{\lambda}{1+k}.$$

Отсюда и из (15) следует искомое рекуррентное соотношение

$$P_{k+1}(\lambda) = \frac{\lambda}{k+1} P_k(\lambda).$$

Положив здесь $k = 0$, с учетом (15) имеем

$$P_1(\lambda) = \lambda P_0(\lambda) = \lambda e^{-\lambda}.$$

Применяя рекуррентную формулу k раз, получим окончательно

$$P_k(\lambda) = \frac{\lambda^k}{k!} e^{-\lambda}. \quad (16)$$

Это так называемый *закон Пуассона*. Его трактуют как закон распределения случайной величины K , имеющей бесконечное число значений $K = 0, 1, 2, \dots$. Соответствующая характеристическая функция равна

$$\Theta(u) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{1}{k!} \lambda^k e^{iku} = \exp [\lambda(e^{iu} - 1)]. \quad (17)$$

Нетрудно найти первые слагаемые ее разложения в ряд Тейлора по u :

$$\Theta(u) = 1 + i\lambda u - \frac{1}{2} \lambda(\lambda + 1) u^2 + \dots$$

откуда

$$\langle K \rangle = \lambda, \quad \langle K^2 \rangle = \lambda(\lambda + 1), \quad D[K] = \lambda, \quad \sigma_k = \sqrt{\lambda}. \quad (18)$$

Конечно, число испытаний n в реальных условиях ограничено. Поэтому на практике закон Пуассона используют так: Малую вероятность p редкого события в одном испытании умножают на число испытаний n . Полученное значение $\lambda = pn$ подставляют в закон Пуассона (16) для приближенного расчета вероятности, что в серии из n испытаний событие произошло k раз. При этом обычно оказывается, что $P_k(\lambda)$ становится исчезающе малым задолго до того, как k приближается к n .

Проиллюстрируем описанную схему применения закона Пуассона примером, заимствованным из технических приложений:

K	0	1	2	3	4	5	6	7
$P_k(\lambda)$	0.223	0.335	0.251	0.126	0.047	0.014	0.004	$7.5 \cdot 10^{-4}$
$P_{100}(k)$	0.221	0.336	0.253	0.126	0.047	0.014	0.003	$6.7 \cdot 10^{-4}$

Таблица 2.3: Вероятности K бракованных шурупов в контрольной партии из $n = 100$ шурупов, сосчитанные по закону Пуассона и с помощью биномиального распределения. Вероятность одному шурупу быть бракованным $p = 0.015$.

Пример: Фабрика выпускает шурупы. Вероятность что шуруп бракованный равна $p = 0.015$. Требуется найти вероятность того, что в контрольной партии из $n = 100$ шурупов будут бракованными k из них.

Применим закон Пуассона, для чего найдем вначале параметр

$$\lambda = p \times n = 0.015 \times 100 = 1.5$$

Следовательно, по закону Пуассона, искомые вероятности равны

$$P_k(1.5) = \frac{(1.5)^k}{k!} e^{-1.5}.$$

В Таблице 3 даны значения вероятностей числа бракованных шурупов, вычисленные по закону Пуассона, и для сравнения приведены вероятности, найденные с помощью биномиального распределения (6). Из таблицы видно, что вероятности, вычисленные по закону Пуассона, незначительно отличаются от биномиальных вероятностей. Кроме того и те и другие вероятности практически равны нулю уже при $k \geq 7 \ll n$. ★

Рассмотрим типичную процедуру выборочной отбраковки изделий: Из партии наугад отбирают n изделий и тщательно проверяют их. Если число бракованных среди проверенных изделий превышает критическое значение m , то бракуют всю партию. Обычно $m \ll n$. Поэтому, вычисляя вероятность отбраковки партии, пользуются формулой (1.3.15) для вероятности противоположного события. Согласно ей

$$P(K > m) = 1 - P(K \leq m) = 1 - \sum_{k=0}^m P_k(\lambda), \quad (19)$$

то есть требуется суммировать не бесконечное число, а всего $m + 1$ слагаемых. Так вероятность, что из ста шурупов хотя бы один будет бракованный

$$P(K > 0) = 1 - P_0(1.5) \simeq 0.777,$$

а что их будет больше двух равна

$$P(K > 2) = 1 - P_0(1.5) - P_1(1.5) - P_2(1.5) \simeq 0.191.$$

Замечание: Закон редких событий опубликован Пуассоном в 1837 году в работе, посвященной приложениям теории вероятностей к судебной практике. Однако долгое время закон Пуассона оставался без должного внимания. Его признание датируют 1898 годом, когда обнаружилось поразительное согласие этого закона со смертностью в 14 кавалерийских корпусах Германии от удара копытом с 1874 по 1894 год. Данные группировались по корпусам и годам, так что полное их число было $N = 14 \times 20 = 280$. Всего погибло 196 солдат, то есть в среднем в год в корпусе гибли $196/280 = 0.7$ человек. Взяв это значение за λ , из закона Пуассона найдем, что вероятность отсутствия смертей в год в корпусе равна

$$P_0(0.7) = e^{-0.7} = 0.497,$$

а вероятности погибнуть от удара копытом одному и двум солдатам корпуса в год такова

$$P_1(0.7) = 0.7 \times e^{-0.7} = 0.348, \quad P_2(0.7) = \frac{1}{2} \times (0.7)^2 \times e^{-0.7} = 0.122.$$

Оценим число случаев отсутствия, одной и двух смертей в корпусах в год, помножив указанные вероятности на число наблюдений:

$$M_0 \simeq 0.497 \times 280 \simeq 139, \quad M_1 \simeq 0.348 \times 280 \simeq 97, \quad M_2 \simeq 0.122 \times 280 \simeq 34.$$

В реальности было $M_0 = 140$, $M_1 = 91$ и $M_2 = 32$, что более чем хорошо согласуется с оценками, основанными на законе Пуассона. ■

Замечание: В начале двадцатого века закономерности редких случайных событий изучались датским математиком Эрлангом. Он выяснил, что вероятности наступления редких событий в ограниченной серии n испытаний аккуратнее описываются усеченным законом Пуассона

$$P_k(\lambda) = C \frac{\lambda^k}{k!} \quad \text{где} \quad C = \left(\sum_{k=0}^n \frac{\lambda^k}{k!} \right)^{-1}. \quad \blacksquare \quad (20)$$

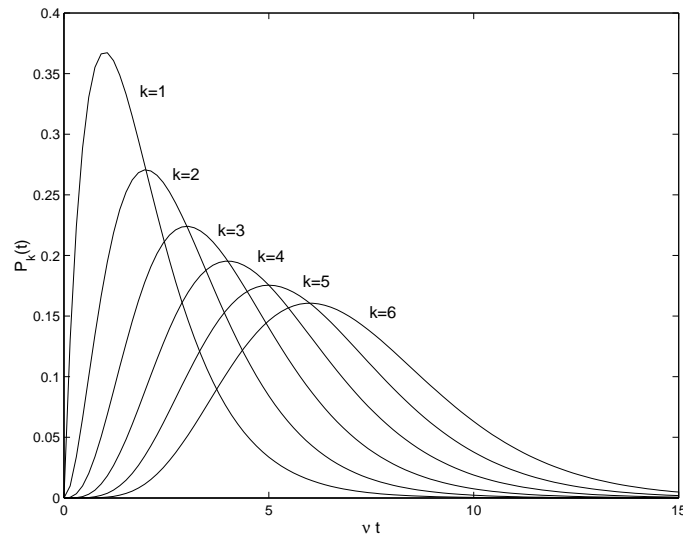


Рис. 2.12

Вероятности регистрации k α частиц в зависимости от νt . С ростом k их максимумы смещаются в сторону больших νt .

2.3.4 Закон Пуассона в физике

Многочисленные применения закона Пуассона в физике иллюстрирует задача о радиоактивном распаде. Пусть счетчик α частиц помещен вблизи источника радиоактивности. Индикатор счетчика фиксирует число α частиц, попавших в него за время t . Найдем вероятность $P_k(t)$, что за это время счетчик зарегистрирует k частиц. Для этого разобьем время наблюдения на n интервалов $\Delta t = t/n$, таких, что вероятность регистрации одной α частицы за Δt очень мала, а возможностью регистрации большего числа частиц и вовсе можно пренебречь. Кроме того, естественно полагать, что вероятность регистрации одной частицы пропорциональна длительности интервала

$$p \sim \nu \Delta t = \nu \frac{t}{n}.$$

Из физики радиоактивности ясно, что испускания α частиц в непрерывающиеся промежутки времени статистически независимы. А значит справедлива схема Бернулли для n независимых испытаний, где роль успеха играет регистрация счетчиком α частицы за время Δt , а неудачу – отсутствие регистрации в этом промежутке. Причем n безгранично велико, вероятность регистрации $p = \nu \Delta t$ исчезающе мала, а

$$\lambda = pn = \nu \frac{t}{n} n = \nu t \tag{21}$$

имеет произвольное значение. Подставив λ (21) в закон Пуассона (16), найдем вероятности, что за время t счетчик зарегистрирует k частиц

$$P_k(t) = \frac{(\nu t)^k}{k!} e^{-\nu t}. \quad (22)$$

Параметр ν характеризует радиоактивность окружающей среды и имеет ясный физический смысл. Он равен *средней частоте* регистрации α частиц (их среднему числу за единицу времени). График вероятностей (22) зарегистрировать одну, две и более частиц, в зависимости от безразмерного времени νt , изображен на Рис. 12.

Осталось добавить, что закон (22) прекрасно согласуется с экспериментом в самых разнообразных физических ситуациях.

Пример: Пусть концентрация молекул газа равна ϱ . Найдем вероятность, что внутри шара радиуса r молекул нет. Здесь

$$\lambda = \frac{4}{3} \pi r^3 \varrho.$$

Подставив это значение λ в (15), получим

$$P_0(r) = \exp\left(-\frac{4}{3} \pi \varrho r^3\right). \quad \star \quad (23)$$

2.4 Законы большого числа испытаний

Помимо обсужденного только-что закона редких событий, имеется еще ряд, связанных с испытаниями Бернулли, важнейших предельных соотношений, полезных при анализе большого числа независимых испытаний $n \gg 1$. Только, в отличие от закона редких событий, они справедливы не при $p \rightarrow 0$, а при любых наперед заданных значениях вероятности p успеха в одном испытании.

2.4.1 Локальная теорема Муавра-Лапласа

Ключевое предельное соотношение, на которое опираются остальные, дает *локальная теорема Муавра-Лапласа*. Муавр доказал ее в 1730 году для случая $p = q = 1/2$. Вслед за Муавром мы подробно обсудим именно этот частный случай, а затем сформулируем общую теорему, справедливую при любых $0 < p < 1$.

Для удобства выкладок примем, что число испытаний четное и равно $n = 2N$. Кроме того введем новый аргумент $m = k - N$, равный отклонению от половины испытаний. Введем еще одно обозначение

$$P(m) = P_{2N}(N + m) = \binom{2N}{N + m} \frac{1}{2^{2N}} \quad (1)$$

для вероятности, что случайная величина $M = K - N$ примет значение m . Запишем $P(m)$ в явной форме

$$P(m) = \frac{(2N)!}{(N + m)!(N - m)!} \left(\frac{1}{2}\right)^{2N}.$$

Помножив числитель и знаменатель на $(N!)^2$, будем иметь

$$P(m) = P_0 \frac{(N!)^2}{(N + m)!(N - m)!}. \quad (2)$$

Здесь выделена вероятность нулевого значения случайной величины M :

$$P_0 = \frac{(2N)!}{(N!)^2} \left(\frac{1}{2}\right)^{2N}. \quad (3)$$

Сократив одинаковые сомножители в числителе и знаменателе (2), получим

$$P(m) = P_0 \frac{N(N - 1) \dots (N - m + 1)}{(N + m)(N + m - 1) \dots (N + 1)}.$$

Разделив затем числитель и знаменатель дроби на N^m , будем иметь

$$P(m) = P_0 \frac{\left(1 - \frac{1}{N}\right) \left(1 - \frac{2}{N}\right) \dots \left(1 - \frac{m-1}{N}\right)}{\left(1 + \frac{1}{N}\right) \left(1 + \frac{2}{N}\right) \dots \left(1 + \frac{m}{N}\right)} = \frac{\prod_{j=1}^{m-1} \left(1 - \frac{j}{N}\right)}{\prod_{j=1}^m \left(1 + \frac{j}{N}\right)}. \quad (4)$$

Будем считать далее, что $m \ll N$, так что все сомножители в числителе и знаменателе близки к единице. Пользуясь этим, заменим произведения в (4) выражениями, асимптотически справедливыми при $m \ll N$, вроде тех, что мы использовали в 1-й главе при решении задачи о совпадении дней рождения. Подробно обсудим произведения в знаменателе. Логарифмируя их, получим

$$\ln \prod_{j=1}^m \left(1 + \frac{j}{N}\right) = \sum_{j=1}^m \ln \left(1 + \frac{j}{N}\right).$$

Логарифмы в сумме заменим известными асимптотиками

$$\ln(1+x) = x + O(x^2) \implies \ln\left(1 + \frac{j}{N}\right) = \frac{j}{N} + O\left(\frac{j^2}{N^2}\right).$$

Подставив их в сумму, придем к асимптотическому соотношению

$$\ln \prod_{j=1}^m \left(1 + \frac{j}{N}\right) = \sum_{j=1}^m \left[\frac{j}{N} + O\left(\frac{j^2}{N^2}\right) \right] = \frac{1}{N} \sum_{j=1}^m j + O\left(\frac{m^3}{N^2}\right).$$

Пользуясь формулой суммы арифметической прогрессии, получаем

$$\ln \left[\prod_{j=1}^m \left(1 + \frac{j}{N}\right) \right] = \frac{m(m+1)}{2N} + O\left(\frac{m^3}{N^2}\right).$$

Отсюда имеем искомую асимптотику произведения в знаменателе (4)

$$\prod_{j=1}^m \left(1 + \frac{j}{N}\right) = \exp\left[\frac{m(m+1)}{2N}\right] \left[1 + O\left(\frac{m^3}{N^2}\right)\right].$$

Точно также находят асимптотику числителя в правой части (4):

$$\prod_{j=1}^{m-1} \left(1 - \frac{j}{N}\right) = \exp\left[-\frac{m(m-1)}{2N}\right] \left[1 + O\left(\frac{m^3}{N^2}\right)\right].$$

Подставив последние два соотношения в (4), получим окончательно

$$P(m) = P_0 \exp\left(-\frac{m^2}{N}\right) \left[1 + O\left(\frac{m^3}{N^2}\right)\right]. \quad (5)$$

Найдем асимптотику вероятности (3) нулевого значения случайной величины M . Для этого вспомним *формулу Стирлинга*

$$N! = \sqrt{2\pi N} N^N e^{-N} \left[1 + O\left(\frac{1}{N}\right)\right] \quad (N \rightarrow \infty). \quad (6)$$

Она надежно работает при любых N , начиная с $N = 1$. В этом убеждает Таблица 4, где даны факториалы и результаты расчетов по формуле Стирлинга. Последняя дает

$$(2N)! \simeq \sqrt{4\pi N} (2N)^{2N} e^{-2N}, \quad (N!)^2 \simeq 2\pi N (N)^{2N} e^{-2N}.$$

Отсюда и из (3) имеем

$$P_0 \simeq \frac{1}{\sqrt{\pi N}}. \quad (7)$$

N	1	2	3	4	5	10
$N!$	1	2	6	24	120	3628800
Стирлинг	1.09	1.91	5.83	23.5	118.02	3598896
%	9	5	3	2	1.7	0.8

Таблица 2.4: Точные значения факториалов и их приближенные значения по формуле Стирлинга. В нижней строчке даны проценты ошибки вычисления факториала по этой формуле.

N	4	10	20	50	100
m^*	2.52	4.64	7.37	13.57	21.54
$\exp(-m^{*2}/N)$	0.20	0.12	0.07	0.03	0.01

Таблица 2.5: Значения экспоненциального множителя в (6) для критического значения m^* и разного числа испытаний.

Подставив это значение в (5), получим окончательно

$$P(m) = P_{2N}(m + N) \simeq \frac{1}{\sqrt{\pi N}} \exp\left(-\frac{m^2}{N}\right) \left[1 + O\left(\frac{m^3}{N^2}\right)\right]. \quad (8)$$

Эта асимптотическая формула и составляет содержание локальной теоремы Муавра Лапласа в частном случае $p = q = 1/2$.

Соотношение (8) означает, что если $m^3 \ll N^2$, то при подсчете вероятностей $P_{2N}(m + N)$ можно пользоваться приближенной формулой

$$P_{2N}(m + N) \simeq \frac{1}{\sqrt{\pi N}} \exp\left(-\frac{m^2}{N}\right) \quad (m \ll m^* = N^{2/3}). \quad (9)$$

Подчеркнем, при $N \gg 1$ эта формула по сути справедлива для любых m , так как входящая в нее экспонента становится малой задолго до того как m достигает границу $m^* = N^{2/3}$ применимости асимптотической формулы (8). Это видно из Таблицы 5, где для разных N даны m^* и соответствующие значения экспоненты в формуле (9).

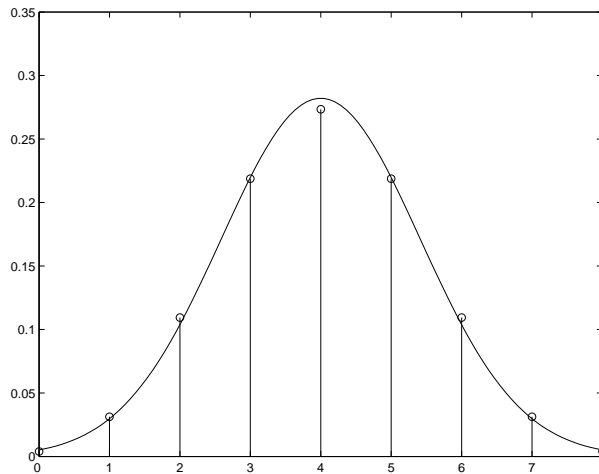


Рис. 2.13

Биномиальное распределение (вертикальные отрезки) и гауссова кривая для случая $n = 8$ ($N = 4$) и $p = 1/2$.

Вернемся к исходным числу успехов k и полному числу испытаний n . Напомним, они связаны с вспомогательными величинами m и N равенствами

$$k = m + N, \quad n = 2N \quad \Leftrightarrow \quad m = k - \frac{n}{2}, \quad N = \frac{n}{2}.$$

С их помощью перепишем (9) в более удобной для вычислений форме

$$P_n(k) \simeq \frac{2}{\sqrt{n}} \varphi(x) \quad (n \gg 1), \quad (10a)$$

где

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \quad (10b)$$

функция Гаусса, а

$$x = \frac{2}{\sqrt{n}} \left(k - \frac{n}{2}\right). \quad (10c)$$

Замечание: Заметим, что n в (10а-с) не обязательно должно быть четным. Четность ($n = 2N$) была нужна нам лишь для упрощения выкладок. ■

Замечание: Хотя приближенная формула (10а-с) вытекает из соотношения (8), асимптотически справедливого при $n \rightarrow \infty$, она хорошо аппроксимирует биномиальное распределение и при не очень больших n . В

этом убеждает график на Рис. 13, где изображено биномиальное распределение и его гауссова аппроксимация (10) (сплошная кривая) для $n = 8$ ($N = 4$). ■

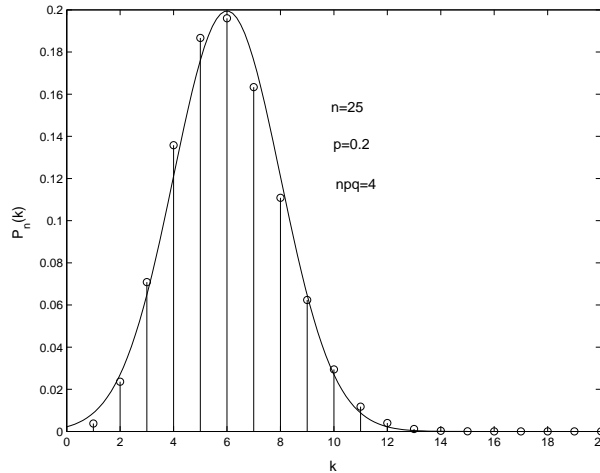


Рис. 2.14

Биномиальное распределение (вертикальные отрезки) и гауссова кривая при $n = 25$, $p = 1/5$ и $npq = 4$. Несмотря на умеренное значение параметра npq , график демонстрирует хорошее согласие биномиальных вероятностей и аппроксимации Муавра-Лапласа.

При произвольных вероятностях успеха p и неудачи q ($p + q = 1$) справедлива аналогичная асимптотическая формула, составляющая содержание общей теоремы Муавра-Лапласа. Опуская громоздкие выкладки, приведем приближенную формулу вычисления биномиального распределения при больших npq :

$$P_n(k) \simeq \frac{1}{\sqrt{npq}} \varphi(x), \quad x = \frac{k - np}{\sqrt{npq}} \quad (npq \gg 1). \quad (11)$$

Здесь $\varphi(x)$ задана равенством (10b).

2.4.2 Интегральная формула Муавра-Лапласа

На практике чаще используют не локальную, а *интегральную формулу Муавра-Лапласа*. Дело в том, что при $n \gg 1$, когда формулы (10), (11) справедливы, вероятности отдельных чисел успехов k , пропорциональные $1/\sqrt{n}$, близки к нулю. Поэтому обычно интересуются вероятностью, что число успехов K попадает в некоторый отрезок $K \in [k_1, k_2]$. Вероятность объединения несовместных событий равна сумме их вероятностей,

а значит вероятность попасть в заданный отрезок равна

$$P(k_1, k_2) = \sum_{k=k_1}^{k_2} P_n(k).$$

Подставив сюда, вместо $P_n(k)$, выражение (11), получим:

$$P(k_1, k_2) \simeq \sum_{k=k_1}^{k_2} \varphi(x_k) \Delta x,$$

где

$$x_k = \frac{k - np}{\sqrt{npq}}, \quad \Delta x = x_{k+1} - x_k = \frac{1}{\sqrt{npq}}.$$

При $n \gg 1$ сумму можно заменить интегралом и пользоваться формулой

$$P(k_1, k_2) \simeq \int_a^b \varphi(x) dx = \frac{1}{\sqrt{2\pi}} \int_a^b \exp\left(-\frac{x^2}{2}\right) dx. \quad (12)$$

Здесь введены обозначения

$$a = \frac{k_1 - np}{\sqrt{npq}}, \quad b = \frac{k_2 - np}{\sqrt{npq}}. \quad (12a)$$

Формула (12) и есть интегральная формула Муавра-Лапласа. Тем не менее, применяя ее, обходятся без вычисления интегралов, а пользуются таблицами функции

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{z^2}{2}\right) dz, \quad (13)$$

и находят требуемые вероятности по формуле

$$P(k_1, k_2) = P\left(a < \frac{K - np}{\sqrt{npq}} < b\right) = \Phi(b) - \Phi(a). \quad (14)$$

Часто вместо $\Phi(x)$ (13) привлекают специальную функцию

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-z^2) dz. \quad (15)$$

Ее обозначением служит аббревиатура полного названия: *Error Function*. Предыдущая функция $\Phi(x)$ выражается через $\operatorname{erf}(x)$ равенством

$$\Phi(x) = \frac{1}{2} \left[1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right) \right], \quad (16)$$

а соотношение (14) приобретает вид

$$P\left(a < \frac{K - np}{\sqrt{npq}} < b\right) = \frac{1}{2} \left[\operatorname{erf}\left(\frac{b}{\sqrt{2}}\right) - \operatorname{erf}\left(\frac{a}{\sqrt{2}}\right) \right]. \quad (17)$$

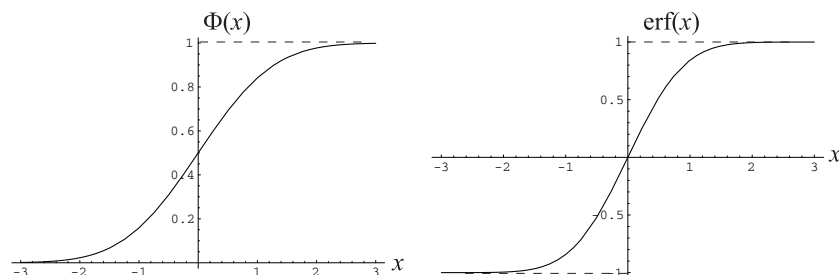


Рис. 2.15

Графики функций $\Phi(x)$ и $\operatorname{erf}(x)$, используемых в интегральной формуле Муавра-Лапласа.

Функция $\operatorname{erf}(x)$ имеется во всех математических программах. Так в программе MATLAB ее вызывают командой

```
>> erf(x)
```

а в программе МАТЕМАТИКА командой

```
Erf[x]
```

Графики функций $\Phi(x)$ и $\operatorname{erf}(x)$ даны на Рис. 15.

2.4.3 Закон Больших Чисел

В первой главе было дано статистическое определение вероятности, согласно которому вероятность события A равна пределу

$$P(A) = \lim_{n \rightarrow \infty} \frac{K}{n}, \quad (18)$$

где K число испытаний, из полного числа n , в которых имело место событие A . До сих пор однако было неясно, в каком смысле здесь понимается сходимость. Теперь мы готовы дать понятию сходимости в (18) строгий математический смысл. Рассмотрим неравенство

$$\left| \frac{K}{n} - P(A) \right| < \varepsilon, \quad (19)$$

где $\varepsilon > 0$ наперед заданное число. При $n \gg 1$ можно рассчитать вероятность выполнения этого неравенства по формуле (17). В данном случае мы интересуемся вероятностью, что число успехов K лежит в интервале

$$np - n\varepsilon < K < np + n\varepsilon. \quad (20)$$

Здесь $p = P(A)$. Соответствующие значения a и b из формул (12) равны

$$a = -\varepsilon \sqrt{\frac{n}{pq}}, \quad b = \varepsilon \sqrt{\frac{n}{pq}}.$$

Подставив их в (17), найдем вероятность выполнения неравенства (19)

$$P\left(\left|\frac{K}{n} - P(A)\right| < \varepsilon\right) = \operatorname{erf}\left(\varepsilon \sqrt{\frac{n}{2pq}}\right). \quad (21)$$

Здесь учтена нечетность функции $\operatorname{erf}(x)$.

Правая часть (21) при $n \rightarrow \infty$ и $\varepsilon > 0$ стремится к единице. Таким образом, мы установили замечательный факт: *Как бы мало ни было $\varepsilon > 0$, вероятность выполнения неравенства (19) стремится к единице.* Это утверждение носит название *Закона Больших Чисел*.

Обычно Закон Больших Чисел формулируют, привлекая понятие *сходимости по вероятности*. В общем случае его определяют так:

Случайную последовательность $\{Z_1, Z_2, \dots, Z_n, \dots\}$ называют сходящейся по вероятности, при $n \rightarrow \infty$, к детерминированной величине z , если при любом $\varepsilon > 0$ вероятность выполнения неравенства $|z - Z_n| > \varepsilon$ стремится к нулю.

Таким образом, статистическая устойчивость события A означает, что его относительная частота сходится по вероятности к $P(A)$.

2.5 Примеры непрерывных величин

Нам уже эпизодически встречались непрерывные случайные величины, принимающие любое значение в некотором интервале (a, b) или даже на всей числовой оси. Так мы освоились с равномерно распределенными величинами, имеем представление о распределениях Гаусса и Коши. Однако непрерывные величины заслуживают более серьезного внимания: с их помощью формулируют важные законы физики, экономики, инженерных приложений. Помимо того они имеют специфические свойства, отсутствующие у дискретных величин. Все это заставляет заняться систематическим изучением непрерывных случайных величин.

Начнем с детального обсуждения конкретных распределений непрерывных величин, часто встречающихся в приложениях.

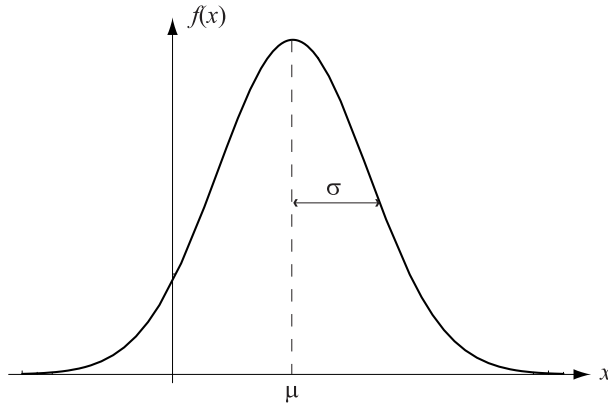


Рис. 2.16

График распределения Гаусса. Его параметр μ задает центр симметрии, а σ определяет характерную ширину $f(x)$.

2.5.1 Распределение Гаусса

Центральное место в теории и приложениях занимают *гауссовы величины*. Распределение гауссовой величины X таково

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad (1)$$

а его интегральная функция распределения имеет вид

$$F(x) = P(X < x) = \Phi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{2} \left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sqrt{2}\sigma}\right) \right]. \quad (2)$$

Гауссово распределение полностью определяется параметрами (μ, σ) , имеющими ясный геометрический смысл: μ есть центр колоколообразного графика распределения (1), а σ равно его характерной ширине (см. Рис. 16). Эти параметры имеют еще и прозрачную вероятностную интерпретацию. Так μ равно средней величине X . Покажем это, записав (1) в форме

$$f(x) = \tilde{f}(x - \mu),$$

где

$$\tilde{f}(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right). \quad (3)$$

В новых обозначениях среднее равно

$$\langle X \rangle = \int_{-\infty}^{\infty} x f(x) dx = \int_{-\infty}^{\infty} x \tilde{f}(x - \mu) dx.$$

Перейдем в последнем интеграле к интегрированию по $y = x - \mu$ ($x = y + \mu$, $dx = dy$). В итоге получим

$$\langle X \rangle = \int_{-\infty}^{\infty} y \tilde{f}(y) dy + \mu \int_{-\infty}^{\infty} \tilde{f}(y) dy .$$

Здесь первый интеграл равен нулю из-за четности $\tilde{f}(y)$, а второй единице в силу условия нормировки. Следовательно $\langle X \rangle = \mu$.

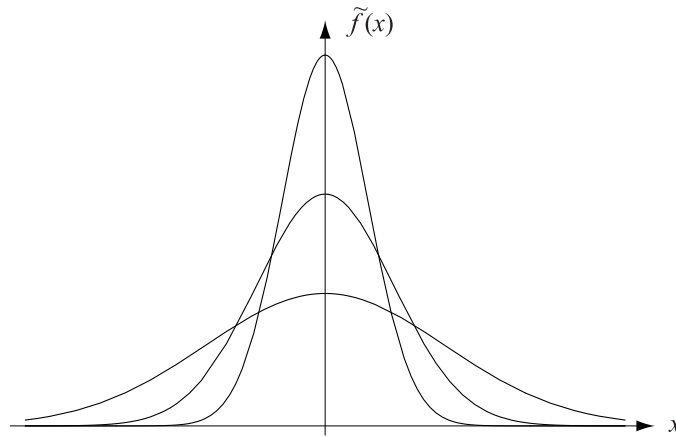


Рис. 2.17

Графики распределения Гаусса при $\mu = 0$ и разных σ . С ростом σ кривые становятся шире, и в то же время ниже, чтобы не нарушалось условие нормировки, согласно которому площадь под кривыми равна единице.

Напомним смысл вспомогательной функции $\tilde{f}(x)$ (3). Это распределение случайной компоненты \tilde{X} гауссовой величины X . Как известно, второй момент \tilde{X} равен дисперсии X :

$$D[X] = \langle \tilde{X}^2 \rangle = \int_{-\infty}^{\infty} x^2 \tilde{f}(x) dx .$$

Подставив сюда явное выражение $\tilde{f}(x)$ (3), будем иметь

$$D[X] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x^2 \exp\left(-\frac{x^2}{2\sigma^2}\right) dx .$$

Сделаем замену переменных

$$z = \frac{x}{\sigma} \quad \Longleftrightarrow \quad x = \sigma z, \quad dx = \sigma dz.$$

Это дает

$$D[X] = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 \exp\left(-\frac{z^2}{2}\right) dz. \quad (4)$$

Сведем оставшийся интеграл к известному *интегралу Пуассона*

$$\int_{-\infty}^{\infty} \exp(-bz^2) dz = \sqrt{\frac{\pi}{b}} \quad (\operatorname{Re} b > 0). \quad (5)$$

Интегрированием по частям находим

$$\begin{aligned} \int_{-\infty}^{\infty} z^2 \exp\left(-\frac{z^2}{2}\right) dz &= - \int_{-\infty}^{\infty} z d\left[\exp\left(-\frac{z^2}{2}\right)\right] = \\ &= \int_{-\infty}^{\infty} \exp\left(-\frac{z^2}{2}\right) dz = \sqrt{2\pi}. \end{aligned}$$

Отсюда и из (4) имеем $D[X] = \sigma^2$. Таким образом, σ^2 есть дисперсия гауссовой величины X , а параметр σ равен ее стандарту.

Замечание: Гауссовы случайные величины наиболее часто встречаются в приложениях. Причины этого мы выясним позже. Сейчас же заметим, что на заре становления теории вероятностей бытовало мнение, что все случайные имеют гауссово распределение, а отклонение распределения от гауссова трактовалось как нечто аномальное. Поэтому у гауссовых случайных величин есть еще одно общепринятое наименование. Их называют *нормальными величинами* и используют для гауссовой величины со средним μ и дисперсией σ^2 сокращенное обозначение $N(\mu, \sigma)$. ■

Гауссово распределение нигде не равно нулю. Это значит, что измеренные значения гауссовой величины X могут быть сколь угодно велики. Однако на практике гауссовы величины редко выходят за некоторый ограниченный интервал. Этот известный экспериментаторам факт выражает знаменитое *правило трех сигм*. Придем к нему, вычислив вероятность, что отклонение гауссовой величины от ее среднего значения не больше 3σ . Выразив искомую вероятность через интегральную функцию распределения (2), будем иметь

$$P(X \in [\mu - 3\sigma, \mu + 3\sigma]) = \operatorname{erf}\left(\frac{3}{\sqrt{2}}\right) \simeq 0.9973.$$

Вероятность противоположного события, что измеренное значение X будет лежать за пределами указанного интервала, равна

$$P(|X - \mu| > 3\sigma) \simeq 0.003.$$

Иными словами, как правило, лишь в трех из тысячи измерений X отклоняется от μ на величину, большую 3σ .

Обсудим в заключение характеристическую функцию гауссовой величины. Нетрудно показать, что она равна

$$\Theta(u) = \exp\left(iu\mu - \frac{1}{2}\sigma^2 u^2\right). \quad (6)$$

Отвечающая ей характеристическая функция случайной компоненты гауссовой величины такова

$$\tilde{\Theta}(u) = \exp\left(-\frac{1}{2}\sigma^2 u^2\right). \quad (7)$$

Сопоставив ее разложение в ряд Тейлора по степеням u

$$\tilde{\Theta}(u) = \sum_{k=0}^{\infty} (-1)^k \frac{\sigma^{2k}}{2^k k!} u^{2k}$$

с родственным (2.32) разложением произвольной $\tilde{\Theta}(u)$

$$\tilde{\Theta}(u) = \sum_{k=0}^{\infty} \frac{\nu_k}{k!} (iu)^k,$$

найдем общую формулу для всех центральных моментов гауссовых величин

$$\nu_{2k} = (2k - 1)!! \sigma^{2k}, \quad \nu_{2k-1} \equiv 0 \quad (k \in \mathbb{N}).$$

Здесь использовано общепринятое обозначение

$$(2k - 1)!! = (2k - 1) \times (2k - 3) \dots 3 \times 1 = \frac{(2k)!}{2^k k!}.$$

2.5.2 Экспоненциальное и родственные распределения

В теории и приложениях важную роль играют распределения, пропорциональные экспоненциальной функции. Сюда относится и обсужденное выше распределение Гаусса. Ниже мы вкратце обсудим другие распределения такого рода, наиболее часто встречающиеся в приложениях.

Экспоненциальное распределение

Чрезвычайно важное для самых разнообразных приложений распределение непрерывных случайных величин вытекает из закона редких событий (3.22). Напомним, $P_k(t)$ (3.22) равно вероятности, что с момента включения счетчика α частиц (выберем его за начало отсчета времени $t = 0$) до текущего момента t в счетчик попало k частиц. Взглянем на указанные вероятности с другой точки зрения, для чего введем случайную величину T , равную длительности интервала от момента включения счетчика до первого попадания в него α частицы. Очевидно, вероятность того, что $T > t$, равна вероятности $P_0(t)$, что за время t счетчик не зарегистрировал частиц:

$$P(T > t) = P_0(t) = e^{-\nu t} \quad (t > 0). \quad (8)$$

Отсюда сразу находим интегральную и дифференциальную функции распределения случайной величины T :

$$\begin{aligned} F(t) &= P(T < t) = 1 - P(T > t) = 1 - e^{-\nu t}, \\ f(t) &= \frac{dF(t)}{dt} = \nu e^{-\nu t} \quad (t > 0). \end{aligned} \quad (9)$$

Плотность вероятностей $f(t)$ (9) называют *экспоненциальным распределением*, по имени функции e^x . Графики функций (9) даны на Рис. 18.

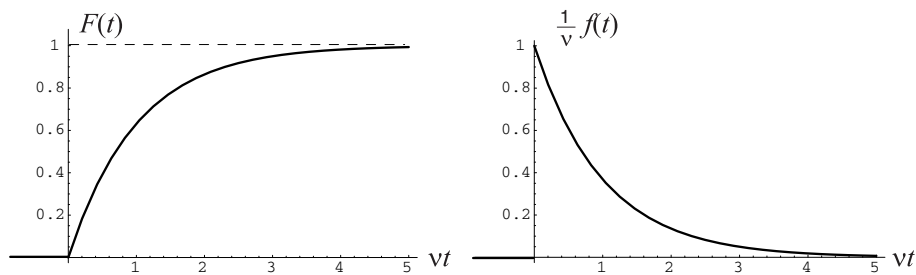


Рис. 2.18

Графики интегральной и дифференциальной функций распределения экспоненциально распределенной случайной величины T .

Легко сосчитать, что среднее значение и стандарт экспоненциального распределения одинаковы и равны

$$\langle T \rangle = \sigma = \frac{1}{\nu}, \quad (10)$$

а его характеристическая функция такова

$$\Theta(u) = \nu \int_0^{\infty} e^{iut - \nu t} dt = \frac{\nu}{\nu - iu}. \quad (11)$$

Замечание: Экспоненциальное распределение представляет собой типичный образчик распределений неотрицательных случайных величин, характерных для самых различных приложений, от физики и биологии до экономики и теории очередей в интернете. ■

Экспоненциальное распределение (9) имеет еще одну важную интерпретацию. Укажем ее, для чего рассмотрим моменты t_k попадания α частиц в счетчик. Из независимости регистрации частиц в неперекрывающиеся промежутки времени следует, что ничего не изменится, если за начало отсчета взять любой момент t_k . Следовательно, распределения (9) заодно являются распределениями *длительностей интервалов между последовательными моментами регистрации* $T = t_{k+1} - t_k$.

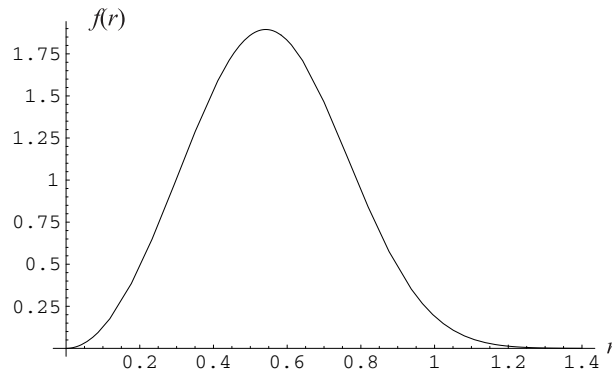


Рис. 2.19

Плотность вероятностей кратчайших расстояний между молекулами газа единичной концентрации ($\varrho = 1$).

Пример: Напомним, выражение (3.23) дает вероятность того, что внутри шара радиуса r нет молекул газа. Пользуясь независимостью положений молекул, совместим центр шара с одной из них. Тогда, согласно (3.23), вероятность, что случайное расстояние R от центральной до ближайшей молекулы больше r , равна правой части равенства (3.23). Соответственно, интегральная и дифференциальная функции распределения кратчайших расстояний между молекулами таковы

$$F(r) = 1 - P_0(r) = 1 - \exp\left(-\frac{4}{3}\pi r^3 \varrho\right),$$

$$f(r) = 4\pi \sigma r^2 \exp\left(-\frac{4}{3}\pi r^3 \varrho\right).$$

а среднее кратчайшее расстояние равно

$$\langle R \rangle = \int_0^{\infty} r f(r) dr = \frac{\Gamma\left(\frac{1}{3}\right)}{\sqrt[3]{36} \pi \varrho} \simeq \frac{0.554}{\sqrt[3]{\varrho}}. \quad (12)$$

График плотности вероятностей кратчайших расстояний между хаотически расположенными молекулами газа с единичной концентрацией ($\varrho = 1$) дан на Рис. 19. ★

Гамма функция

В правую часть равенства (12) входит значение *гамма функции*. Поскольку мы еще не раз еще встретимся с ней в дальнейшем, дадим определение этой функции и краткую справку о ее свойствах. По определению, гамма функция равна

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx. \quad (13)$$

Интеграл справа существует только при $\alpha > 0$. Однако аналитическим продолжением гамма функцию доопределяют на всю комплексную плоскость, за исключением точек $\alpha = 0, -1, -2, \dots$, где она имеет полюсы. График гамма функции вдоль действительной оси изображен на Рис. 20. При расчетах с привлечением гамма функции бывают полезными ее *рекуррентное свойство*, согласно которому

$$\Gamma(\alpha + 1) = \alpha \Gamma(\alpha) \quad (14)$$

и свойство симметрии

$$\Gamma(\alpha) \Gamma(1 - \alpha) = \frac{\pi}{\sin \pi \alpha}. \quad (15)$$

В частности, из рекуррентного соотношения (14) следует, что

$$\Gamma(n + 1) = n!, \quad (16)$$

а свойство симметрии дает

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}. \quad (17)$$

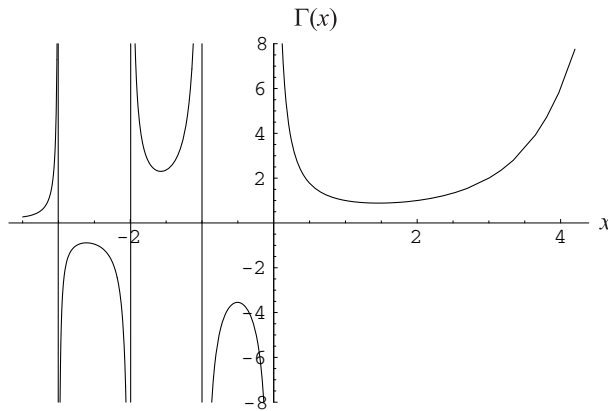


Рис. 2.20

Зависимость гамма функции от вещественного аргумента.

Заметим еще, что в теории гамма функций доказывается справедливость интегрального соотношения

$$\int_0^{\infty} x^{\alpha-1} e^{-\gamma x} dx = \frac{\Gamma(\alpha)}{\gamma^{\alpha}}, \quad (18)$$

справедливого при $\alpha > 0$ и любого комплексного γ с положительной реальной частью $\text{Re } \gamma > 0$.

Гамма распределение

Так называют плотность вероятностей

$$f(x; \gamma, \alpha) = \frac{\gamma^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\gamma x} \quad (x > 0). \quad (19)$$

неотрицательной случайной величины X . Найдем характеристическую функцию гамма распределения:

$$\Theta(u) = \frac{\gamma^{\alpha}}{\Gamma(\alpha)} \int_0^{\infty} x^{\alpha-1} e^{-(\gamma-iu)x} dx \quad (x > 0).$$

Вычислив входящий сюда интеграл с помощью формулы (18), будем иметь

$$\Theta(u) = \frac{1}{\left(1 - \frac{i u}{\gamma}\right)^{\alpha}}. \quad (20)$$

Распределение Вейбулла

Распределения, близкие экспоненциальным, возникают при анализе продолжительности работы технических приборов, процессов старения и других родственных явлений. Обсудим некоторые связанные с ними понятия, для чего введем функцию, дополнительную к интегральному распределению

$$R(t) = P(T \geq t) = 1 - F(t). \quad (21)$$

Применительно к случайной продолжительности работы прибора T , она равна вероятности прибору быть исправным дольше наперед заданного времени t . Поэтому $R(t)$ (21) называют *функцией выживания* (survival function). Это убывающая с ростом t функция. Чем быстрее она спадает, тем больше опасность прибору вскоре выйти из строя. Поэтому отношение скорости убывания функции выживания к ней самой

$$\Lambda(t) = -\frac{1}{R(t)} \frac{dR(t)}{dt} \quad (22)$$

называют функцией опасности (hazard function).

Для экспоненциального распределения (9) функция опасности постоянна: $R(t) = \nu = \text{const}$. Для приборов, животных и людей, для которых характерны процессы старения, эта функция имеет тенденцию к увеличению ростом t . Если функция опасности известна при любых t , по ней можно рассчитать функцию выживаемости. Возьмем к примеру достаточно реалистичную степенную функцию опасности

$$\Lambda(t) = \frac{\alpha}{\beta} t^{\alpha-1} \quad (\alpha > 0, \beta > 0).$$

Тогда функция выживания равна

$$R(t) = \exp\left(-\frac{t^\alpha}{\beta}\right).$$

По $R(t)$ легко определить плотность вероятностей продолжительности работы прибора:

$$f(t; \alpha, \beta) = -\frac{dR(t)}{dt} = \frac{\alpha}{\beta} t^{\alpha-1} \exp\left(-\frac{t^\alpha}{\beta}\right) \quad (t > 0). \quad (23)$$

Последнюю называют *распределением Вейбулла*. При $\alpha = 1$ оно переходит в экспоненциальное распределение (9) с *интенсивностью* $\nu = 1/\beta$.

При $\alpha = 2$ распределение Вейбулла превращается в известное распределение Рэлея

$$f(t) = \frac{2}{\beta} t \exp\left(-\frac{t^2}{\beta}\right). \quad (24)$$

Оно соответствует линейно растущей функции опасности. Мы еще встретимся с распределением Рэлея в связи с физическими и инженерными приложениями теории вероятностей.

2.6 Функции случайных величин

К одной из типичных процедур обработки случайных величин относится нелинейное преобразование. Уже упоминалось к примеру, что приемник электромагнитных сигналов содержит детектор, преобразующий случайный сигнал X по некоторому закону $y = g(x)$. Помимо линейного детектора (2.17), на практике используют квадратичный $y = \gamma x^2$, экспоненциальный $y = a \exp(-bx)$ и другие нелинейные устройства. Они преобразуют входную случайную величину X в выходную $Y = g(X)$, статистические свойства которой требуется выяснить. Потребность в анализе вероятностных свойств функций случайных величин возникает не только в электротехнике. С примерами из других областей мы встретимся ниже. А пока изучим общие закономерности трансформации вероятностных свойств функций случайных величин.

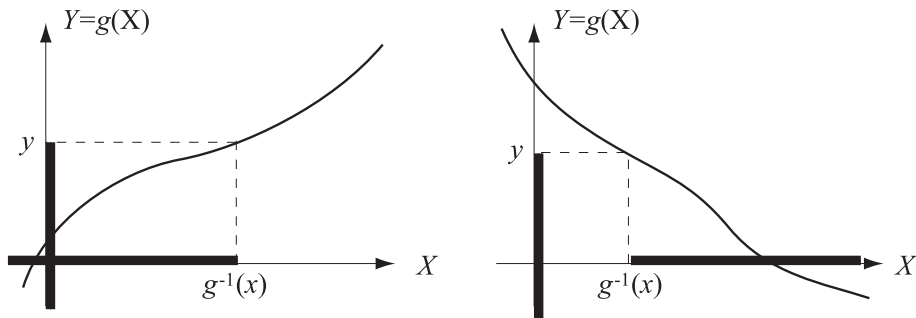


Рис. 2.21

Графики строго монотонных функций, взаимно-однозначно отображающих значения случайных величин X и Y . Жирными отрезками помечены участки, где выполнены неравенства (2) и (5).

2.6.1 Строго монотонные функции

Обсудим вначале как связаны распределения $Y = g(X)$ с распределениями исходной величины X для строго монотонно возрастающей

функции $y = g(x)$. Ее характерный график дан на Рис. 21 слева. Найдем прежде всего интегральную функцию распределения Y . По определению (1.5) она равна¹

$$F_Y(y) = P(Y < y) = P(g(X) < y). \quad (1)$$

Для строго монотонно возрастающей функции $y = g(x)$ эквивалентны неравенства

$$Y < y \iff g(X) < y \iff X < g^{-1}(y). \quad (2)$$

Здесь за $x = g^{-1}(y)$ обозначена обратная к $y = g(x)$ функция. Таким образом, в указанном случае

$$P(g(X) < y) = P(X < g^{-1}(y)) = F_X(g^{-1}(y)).$$

Объединяя конец этого равенства с началом равенства (1), придем к искомой формуле, выражающей интегральную функцию Y через интегральную функцию случайной величины X :

$$F_Y(y) = F_X(g^{-1}(y)). \quad (3)$$

Дифференцируя по y , найдем связь плотностей вероятностей X и Y :

$$f_Y(y) = f_X(g^{-1}(y)) \frac{dg^{-1}(y)}{dy}. \quad (4)$$

Выясним, как преобразуются распределения в случае строго монотонно убывающей функции $Y = g(X)$ (см. график справа на Рис. 21). Для нее эквивалентны неравенства

$$Y < y \iff X > g^{-1}(y), \quad (5)$$

что означает равенство вероятностей

$$P(Y < y) = P(X > g^{-1}(y)) = 1 - P(X < g^{-1}(y)).$$

Выразив их через интегральные функции распределения, получим

$$F_Y(y) = 1 - F_X(g^{-1}(y)). \quad (6)$$

¹Здесь и всюду ниже, где могут быть разночтения, будем использовать индексы. Они помогут различать распределения разных случайных величин. Так обозначим интегральную функцию Y как F_Y , а плотность вероятностей X как f_X .

Дифференцируя (6) по y , придем к аналогичной (4) формуле, связывающей плотности вероятностей Y и X :

$$f_Y(y) = -f_X(g^{-1}(y)) \frac{dg^{-1}(y)}{dy}. \quad (7)$$

Она только знаком отличается от (4). Минус обеспечивает неотрицательность плотности вероятностей, поскольку в данном случае производная обратной функции $g^{-1}(y)$ меньше нуля.

Объединим формулы (4) и (7) в одну, справедливую для любых строго монотонных функций:

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right|. \quad (8)$$

Дадим несколько примеров использования формулы (8).

Пример: Обсудим вначале трансформацию плотности вероятностей для простейшей линейной функции

$$y = ax + b \quad \Longleftrightarrow \quad x = \frac{y - b}{a}.$$

В этом случае равенство (8) принимает вид

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y - b}{a}\right).$$

Множитель $1/|a|$ в правой части равенства обеспечивает выполнение условия нормировки f_Y и корректирует ее размерность, если X и Y величины разной размерности. ★

При моделировании случайных чисел с требуемыми статистическими свойствами, типична следующая проблема: Имеется генератор случайных чисел, выдающий случайные числа с одним распределением, а надо получить случайную последовательность, обладающую другим распределением. Эту проблему легко решить с помощью формул типа (3). Как правило, генерируется последовательность чисел, равномерно распределенных в интервале $(0, 1)$. В программе MATLAB набор n таких чисел $\{X_1, \dots, X_n\}$ вызывают командой

```
>> x=rand(1, n)
```

Найдем функцию $y = g(x)$, преобразующую числа $\{X_1, \dots, X_n\}$ в последовательность

>> $y=g(x)$

чисел с нужной интегральной функцией распределения $F_Y(y)$. Для этого заметим, что для равномерно распределенных в интервале $(0, 1)$ величин X имеем $F_X(x) = x$. Соответственно, равенство (3) сразу дает решение задачи

$$F_Y(y) = g^{-1}(y) \quad \Longleftrightarrow \quad y = F_Y^{-1}(x). \quad (9)$$

Здесь $F_Y^{-1}(x)$ функция, обратная к $F_Y(y)$.

Пример: Найдем функцию $y = g(x)$, преобразующую равномерно распределенное в интервале $(0, 1)$ числа X в неотрицательные числа Y с экспоненциальным распределением

$$F_Y(y) = 1 - e^{-\nu y} \quad (y > 0).$$

В этом случае первое равенство (9) запишется в виде

$$x = g^{-1}(y) = 1 - e^{-\nu y} \quad \Longleftrightarrow \quad y = -\frac{1}{\nu} \ln(1 - x).$$

Заметив далее, что случайные величины X и $1 - X$ одинаково распределены, получим окончательно

$$y = g(x) = -\frac{1}{\nu} \ln x. \quad \star$$

Пример: Найдем функцию $y = g(x)$, преобразующую равномерно распределенные в интервале $(0, 1)$ числа X в гауссовы случайные величины Y с нулевым средним и единичной дисперсией ($\mu = 0, \sigma = 1$). Из (9) с учетом (5.2) имеем

$$x = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{y}{\sqrt{2}} \right) \right] \quad \Longleftrightarrow \quad y = \sqrt{2} \operatorname{erfinv}(2x - 1).$$

Здесь заимствовано обозначение программы MATLAB, где за $x = \operatorname{erfinv}(y)$ обозначена функция, обратная к функции ошибок $y = \operatorname{erf}(x)$. \star

Пример: Пусть корабль береговой охраны стоит в море на удалении h от берега. На корабле вращается прожектор, включаясь в совершенно случайные моменты времени и освещая на мгновение точку береговой линии (см. Рис. 22). Найти распределение случайной координаты X освещаемой точки берега.

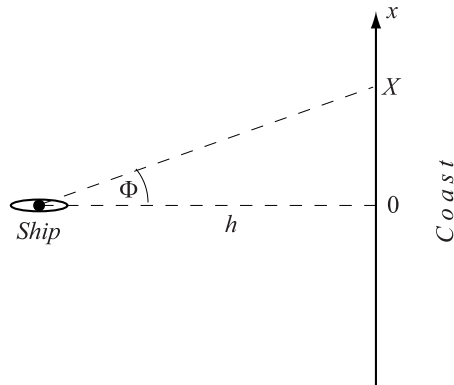


Рис. 2.22

Иллюстрация к задаче о корабле, освещающем берег.

Пусть Φ случайный угол между направлением прожектора в момент включения и отрезком, соединяющим корабль с ближайшей к нему точкой берега. Из условия ясно, что Φ равномерно распределен в интервале $[-\pi/2, \pi/2]$. Мы же пока рассмотрим более общую задачу, считая, что Φ имеет произвольную плотность вероятностей $f_\Phi(\varphi)$, равную нулю вне указанного интервала. Найдем интегральную функцию распределения случайной координаты X , заметив, что X и Φ связаны равенством

$$X = h \operatorname{tg} \Phi \quad \Longleftrightarrow \quad \Phi = \operatorname{arctg} \left(\frac{X}{h} \right).$$

Следовательно, искомая интегральная функция распределения равна

$$F_X(x) = P \left(\Phi < \operatorname{arctg} \frac{x}{h} \right) = \int_{-\pi/2}^{\operatorname{arctg}(x/h)} f_\Phi(\varphi) d\varphi.$$

Подставив сюда равномерное распределение $f_\Phi(\varphi) = 1/\pi$, получим окончательно

$$F_X(x) = \frac{1}{\pi} \operatorname{arctg} \left(\frac{x}{h} \right) + \frac{1}{2}.$$

Дифференцируя его по x , найдем плотность вероятностей случайной координаты освещенной точки берега

$$f_X(x) = \frac{1}{\pi} \frac{h}{x^2 + h^2}.$$

Это уже известное нам *распределение Коши*. Мы лучше поймем его геометрический смысл, заменив прожектор лампой, светящей во все стороны. Тогда освещенность береговой линии, убывающая обратно пропорционально квадрату расстояния, будет пропорциональна распределению Коши.

В общем случае распределением Коши называют

$$f(x) = \frac{1}{\pi} \frac{\sigma}{\sigma^2 + (x - \mu)^2}. \quad (10)$$

У него нет среднего и дисперсии, поскольку соответствующие интегралы расходятся. Параметры же μ и σ , полностью определяющие распределение Коши (10), называют *центром* (location parameter) и *масштабом* (scale parameter) распределения Коши. ★

Пример: Пусть X -случайная величина на входе экспоненциального детектора. Тогда на его выходе наблюдается величина

$$Y = e^{sX+m}. \quad (11)$$

Здесь $s > 0$ и m заданные константы. Пусть кроме того входная величина X гауссова с $\mu = 0$, $\sigma = 1$ и интегральной функцией распределения

$$F_X(x) = \frac{1}{2} + \frac{1}{2} \operatorname{erf} \left(\frac{x}{\sqrt{2}} \right). \quad (12)$$

Подставив сюда обратную функцию

$$x = \frac{\ln y - m}{s},$$

найдем интегральную функцию распределения случайной величины на выходе экспоненциального детектора

$$F_Y(y) = \frac{1}{2} + \frac{1}{2} \operatorname{erf} \left(\frac{\ln y - m}{\sqrt{2}s} \right) \quad (y > 0). \quad (13)$$

Дифференцируя ее по y , придем к плотности вероятностей

$$f_Y(y) = \frac{1}{\sqrt{2\pi}y} \exp \left(-\frac{(\ln y - m)^2}{2s^2} \right). \quad (14)$$

Это распределение называют *логарифмически нормальным*. Оно часто встречается в инженерных приложениях и заслуживает подробного обсуждения. Из (10) видно, что экспоненциальный детектор превращает произвольную величину X в *неотрицательную* Y . Поэтому интегральная функция логарифмически нормального распределения (а вместе с ней и плотность вероятностей) тождественно равна нулю при $y < 0$. Ее график и график логарифмически нормальной плотности вероятностей (14) при $m = 0$ и $s = 1$ изображены на Рис. 23.

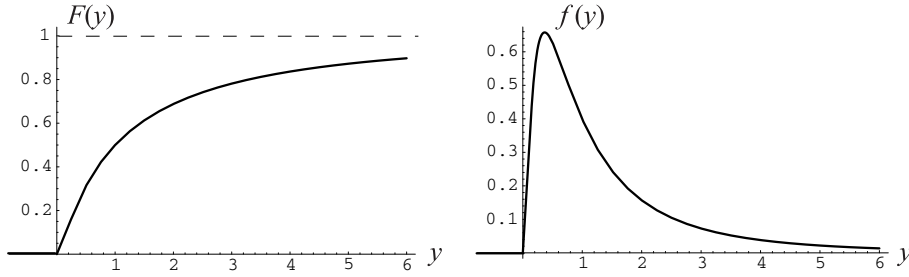


Рис. 2.23

Интегральная функция и плотность вероятностей логарифмически нормально распределенной случайной величины.

Вычислим моменты логарифмически нормальной величины

$$\langle Y^n \rangle = \langle e^{n(sX+m)} \rangle = e^{nm} \langle e^{nsX} \rangle. \quad (15)$$

Из определения характеристической функции (2.29) видно, что среднее экспоненты случайной величины равно

$$\langle e^{aX} \rangle = \Theta(-ia). \quad (16)$$

Применительно к нашему случаю отсюда имеем:

$$\Theta(u) = \exp\left(-\frac{1}{2}u^2\right) \iff \langle e^{nsX} \rangle = \exp\left(\frac{1}{2}n^2s^2\right). \quad (17)$$

Подставив последнее равенство в (15), получим окончательно

$$\langle Y^n \rangle = \exp\left(\frac{1}{2}n(ns^2 + 2m)\right).$$

В частности, для первых двух моментов отсюда имеем

$$\langle Y \rangle = \exp\left(\frac{1}{2}s^2 + m\right), \quad \langle Y^2 \rangle = \exp\left(2(s^2 + m)\right). \quad \star \quad (18)$$

Замечание: Моменты полезны для количественной оценки свойств случайных величин. Кроме того, их сравнительно легко измерить в эксперименте. Поэтому многие склонны игнорировать функции распределения и всецело доверяются выводам, основанным на анализе моментов. Порой, однако, статистические выводы, апеллирующие к моментам, могут вводить в заблуждение. Продемонстрируем этот удручающий факт на примере логарифмически нормального распределения с параметром $m = -s^2/2$. Тогда среднее Y постоянно, а второй момент и дисперсия Y экспоненциально растут с ростом s :

$$\langle Y \rangle \equiv 1, \quad \langle Y^2 \rangle = \exp(s^2), \quad D[Y] = \exp(s^2) - 1.$$

Казалось бы отсюда следует, что с ростом s случайная величина Y все сильнее отклоняется от среднего в сторону больших значений $Y \gg 1$. На деле, как видно из (13), вероятность случайной величине Y превысить среднее значение быстро спадает к нулю

$$P(Y > 1) = 1 - F_Y(1) = \frac{1}{2} \operatorname{erfc} \left(\frac{s}{2\sqrt{2}} \right). \quad (19)$$

Здесь использована дополнительная к erf функция

$$\operatorname{erfc}(z) = 1 - \operatorname{erf}(z). \quad (20)$$

Иными словами, чем больше s , тем больше “вероятностная масса” скапливается в области малых $y < 1$, а значит все реже и реже Y будет превышать в испытаниях уровень $\langle Y \rangle = 1$. Экспоненциальный же рост моментов логарифмически нормального распределения (14) обусловлен все более медленным спаданием его хвоста с увеличением s . График зависимости от s вероятности превышения среднего уровня (19) дан на Рис. 24. ■

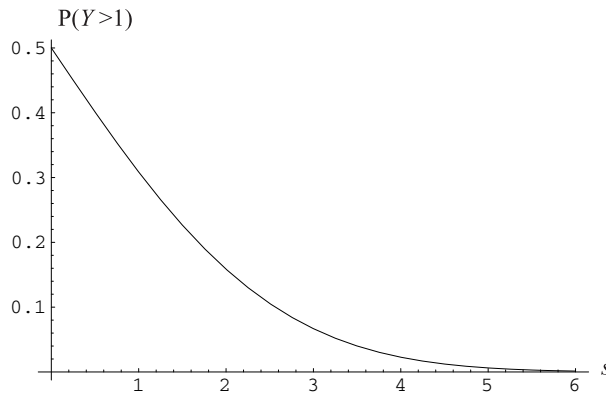


Рис. 2.24

График вероятности, что логарифмически распределенная случайная величина Y превысит среднее значение, в зависимости от s (при $m = -s^2/2$).

Замечание: Пользуясь удобным случаем, определим понятие *медианы*. Применительно к произвольной случайной величине Y , медианой называют такой уровень y_{med} , что с равной возможностью Y принимает значения как большие, так и меньшие y_{med} . Из определения вытекает уравнение для медианы:

$$F(y_{\text{med}}) = \frac{1}{2}.$$

В частности, согласно (13), медиана логарифмически нормального распределения равна $y_{\text{med}} = e^m$. В обсужденном выше случае $m = -s^2/2$ медиана быстро стремится к нулю с ростом s . Последнее означает, что с ростом s “вероятностная масса” сосредотачивается во все более малой окрестности начала координат $y = 0$. ■

Замечание: Помимо медианы, при обсуждении вероятностных свойств случайных величин привлекают понятие моды y_{mod} . Так называют координаты локальных максимумов плотности вероятностей. Для логарифмически нормального распределения медиана и мода совпадают. ■

Замечание: Укажем одну из причин, по которым логарифмически нормальное распределение играет важную роль в инженерных и физических приложениях. Оно возникает при анализе поведения систем со случайными параметрами. В простейшем случае поведение подобной системы с течением времени описывается дифференциальным уравнением

$$\frac{dY}{dt} = (aX + b)Y. \quad Y(t = 0) = 1.$$

параметр которого (коэффициент при Y в правой части) случаен. Решение данного уравнения таково

$$Y(t) = e^{atX+bt}$$

и с точностью до замены at на s , bt на m , совпадает с логарифмически нормальной величиной (11). ■

2.6.2 Нарушение строгой монотонности

Вернемся к обсуждению формулы (8). Она связывает распределение величин $Y = g(X)$ и X , если функция $g(x)$ строго монотонна (и желательнее всюду дифференцируема). Если строгая монотонность нарушается, формула (8) иногда теряет смысл, а плотность вероятностей $f_Y(y)$ приходится находить из первых принципов, обращаясь к равенствам для вероятностей случайным величинам принять те или иные значения.

Пример: Проиллюстрируем сказанное на примере линейного детектора ($y = x\theta(x)$), принцип действия которого иллюстрируется Рис. 5. Для величины на выходе линейного детектора выполнение неравенства $Y < 0$ невозможно, а значит $F_Y(y) \equiv 0$ при $y < 0$. С другой стороны, при любом $y > 0$ выполнено равенство $F_Y(y) = F_X(y)$, так как при $y > 0$ неравенство $Y < y$ эквивалентно неравенству $X < y$. Следовательно, при любых

y интегральная функция распределения Y описывается формулой

$$F_Y(y) = \begin{cases} 0, & y < 0; \\ F_X(y) & y \geq 0. \end{cases} \quad (21)$$

Если X знакопеременная, а $F_X(0) \neq 0$, то интегральная функция (21) терпит скачок при $y = 0$. К примеру, если X гауссова с равным нулю средним ($\mu = 0$), то скачок $F_Y(y)$ при переходе через точку $y = 0$ равен $1/2$. Известно, что производная функции $F(y)$, терпящей скачок в некоторой точке y_0 , содержит сингулярное слагаемое $[F(y_0)]\delta(y - y_0)$, где квадратные скобки символизируют величину скачка $F(y)$. Имея это ввиду, выпишем плотность вероятностей случайной величины на выходе линейного детектора

$$f_Y(y) = p\delta(y) + \theta(y)f_X(y). \quad (22)$$

Здесь коэффициент

$$p = F_X(0) = \int_{-\infty}^0 f_X(x) dx$$

равен вероятности входной величине X быть отрицательной. Графики $F_Y(y)$ и соответствующей плотности вероятностей $f_Y(y)$ в случае, когда X гауссова с $\mu = 0$ и $\sigma = 1$, дан на Рис. 25.

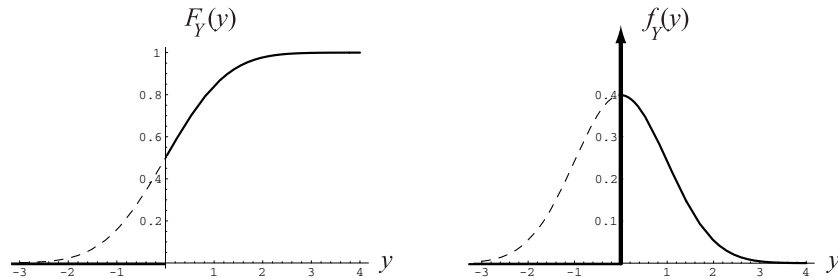


Рис. 2.25

Распределения случайной величины Y на выходе линейного детектора. Штрихом изображены распределения входной гауссовой величины X . Жирная стрелка на правом графике символизирует дельта-функцию.

2.6.3 Немонотонные функции

Помимо обсужденных выше монотонных функций случайных величин, на практике встречаются и немонотонные функции $Y = g(X)$. Типичным примером здесь служит квадратичный детектор. Если на его

вход подают X , то на выходе имеют $Y = \gamma X^2$. Выясним как в этом и других подобных ситуациях связаны плотности вероятностей X и Y . Для этого найдем вероятность, что случайная величина Y принимает значения, заключенные в интервале $[y, y + dy)$. Для наглядности будем иллюстрировать рассуждения на примере немонотонной функции $Y = g(X)$, изображенной на Рис. 26.

Для (бесконечно) малого dy и непрерывной случайной величины Y вероятность ее попадания в указанный интервал вычисляют по асимптотической формуле

$$P(Y \in [y, y + dy)) = f_Y(y) dy. \quad (23)$$

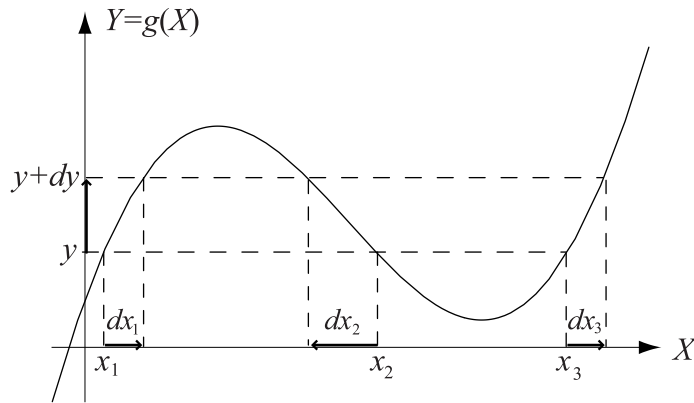


Рис. 2.26

Иллюстрация к расчету плотности вероятностей случайной величины Y , заданной немонотонной функцией $Y = g(X)$. Жирными стрелками указана ориентация интервалов dy и dx_k .

Будем трактовать равенство $y = g(X)$ как уравнение относительно X . В общем случае оно имеет несколько корней $x_k = g_k^{-1}(y)$ (на Рис. 26 три корня). При этом случайная величина Y окажется в интервале $[y, y + dy)$, если X попадет в любой из k интервалов, примыкающих к корням уравнения $y = g(X)$ (см. Рис. 26). Длины этих интервалов равны

$$|dx_k| = \left| \frac{dg_k^{-1}(y)}{dy} \right| dy. \quad (24)$$

Соответственно, вероятность случайной величине Y попасть в интервал $[y, y + dy)$ равна сумме вероятностей несовместных событий: попаданий X в k упомянутых интервалов на оси X

$$f_Y(y) dy = \sum_k f_X(x_k) |dx_k|$$

Отсюда с учетом (24) получаем

$$f_Y(y) = \sum_k f_X(g_k^{-1}(y)) \left| \frac{dg_k^{-1}(y)}{dy} \right|. \quad (25)$$

Здесь суммирование ведется по всем корням уравнения $y = g(x)$ относительно x .

Пример: С помощью формулы (25) найдем плотность вероятностей случайной величины $Y = \gamma X^2$ ($\gamma > 0$) на выходе квадратичного детектора. Так как при $y < 0$ уравнение $y = \gamma x^2$ корней не имеет, для отрицательных y имеем $f_Y(y) \equiv 0$. При $y > 0$ существует два корня

$$x_{1,2} = \mp \sqrt{\frac{y}{\gamma}} \quad \Longrightarrow \quad \left| \frac{dx_{1,2}}{dy} \right| = \frac{1}{2\sqrt{\gamma y}} \quad (y > 0)$$

Подставив их вместе с производными в (25), получим окончательно

$$f_Y(y) = \begin{cases} 0, & y < 0; \\ \frac{1}{2\sqrt{\gamma y}} \left[f_X\left(-\sqrt{\frac{y}{\gamma}}\right) + f_X\left(\sqrt{\frac{y}{\gamma}}\right) \right] & y > 0. \end{cases} \quad \star \quad (26)$$

Замечание: Обратим внимание, что если $f_X(0) \neq 0$, то $f_Y(y)$ стремится к бесконечности при $y \rightarrow 0$. Это связано с аномальным сжатием интервалов dy , по сравнению с интервалами dx , в окрестности точки $x = 0$, $y = 0$. Чтобы вероятности попадания в эти интервалы были равны, плотность вероятностей $f_Y(y)$ вынуждена стремиться к бесконечности. С подобными бесконечными сжатиями интервалов, порождающими стремление к бесконечности плотностей вероятностей, мы еще раз встретимся в следующем примере. ■

Пример: Мы уже исследовали статистические свойства случайного гармонического сигнала $X = A \sin \Phi$, где Φ равномерно распределенная в интервале $[0, 2\pi]$ фаза. В частности, уже найдена его характеристическая функция (2.40). Вычислим теперь плотность вероятностей этой случайной величины.

Очевидно, возможные значения X лежат в интервале $X \in [-A, A]$, вне которого плотность вероятностей $f_X(x)$ равна нулю. Внутри этого интервала уравнение $x = A \sin \Phi$ имеет два корня $\varphi_1(x)$ и $\varphi_2(x)$, показанные на Рис. 27 слева. Таким образом, согласно (25), искомая плотность

вероятностей равна

$$f_X(x) = f(\varphi_1(x)) \left| \frac{d\varphi_1(x)}{dx} \right| + f(\varphi_2(x)) \left| \frac{d\varphi_2(x)}{dx} \right|. \quad (27)$$

Найдем производную функции $x = A \sin \varphi$ в точках корней

$$\left. \frac{dx}{d\varphi} \right|_{\varphi_k} = A \cos \varphi_k = \pm A \sqrt{1 - \sin^2 \varphi_k} = \pm \sqrt{A^2 - x^2}. \quad (28)$$

Здесь плюс отвечает положительным, а минус отрицательным значениям $\cos \varphi_k$.

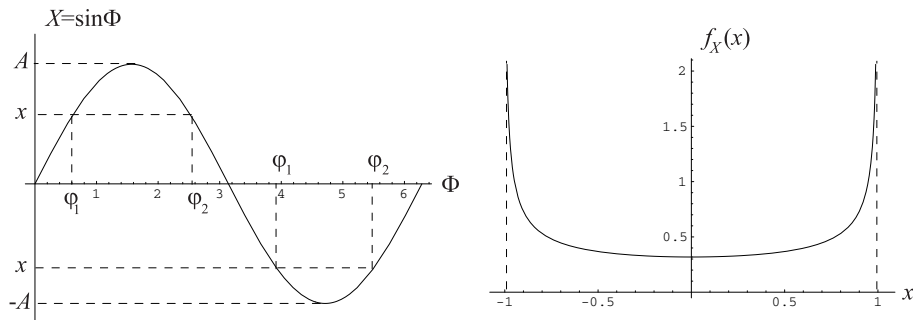


Рис. 2.27

Иллюстрация к построению плотности вероятностей случайной величины $X = A \sin \Phi$.

Так как модули производных (28) одинаковы, а плотность равномерно распределенной фазы постоянна и равна $1/2\pi$, из (27) имеем

$$f_X(x) = \frac{1}{\pi} \frac{1}{\sqrt{A^2 - x^2}} \quad (|x| < A). \quad (29)$$

График этого распределения для случая $A = 1$ изображен на Рис. 27 справа. Его стремление к бесконечности при $x \rightarrow \pm A$ обусловлено замедлением функции $x = A \sin \varphi$ в окрестности экстремальных точек $\varphi = \pi/2$ и $\varphi = 3\pi/2$.

Глава 3

Совокупности случайных величин

Случайные явления обычно описываются целой совокупностью взаимосвязанных случайных величин $\{X_1, \dots, X_n\}$, и возникает потребность в исследовании их совместных вероятностных свойств. В данной главе последовательно излагаются методы анализа подобных совокупностей. Как правило, идеи упомянутых методов удастся наглядно донести на примере пары случайных величин X и Y . Поэтому подробные выкладки будут проводиться преимущественно для этого случая.

3.1 Совместные вероятностные свойства

Прежде всего ознакомимся с основными понятиями и способами вероятностного описания совокупностей случайных величин.

3.1.1 Интегральная функция распределения

Как и для одной случайной величины, статистические свойства совокупности $\{X_1, \dots, X_n\}$ полностью определяются *совместной n -мерной интегральной функцией распределения*

$$F(x_1, x_2, \dots, x_n) = P(X_1 < x_1 \cap X_2 < x_2 \cap \dots \cap X_n < x_n), \quad (1)$$

равной вероятности пересечения событий, состоящих в том, что первая случайная величина X_1 меньше x_1 , вторая x_2 и наконец последняя меньше x_n . Зачастую достаточно знать вероятностные свойства всего двух

случайных величин X и Y . Вся информация о них содержится в интегральной функции распределения

$$F(x, y) = P(X < x \cap Y < y), \quad (2)$$

равной вероятности пересечения событий $X < x$ и $Y < y$.

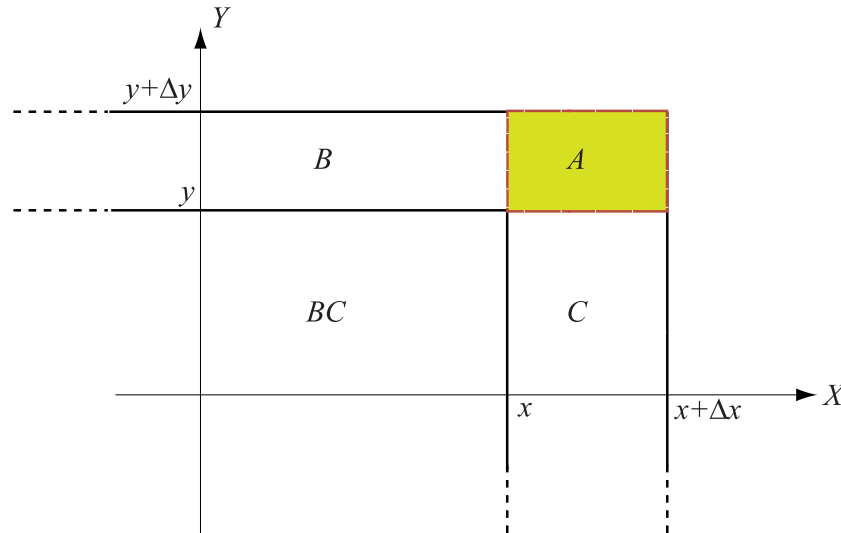


Рис. 3.1

Иллюстрация к выводу формулы для плотности вероятностей.

Выразим через $F(x, y)$ вероятность, что в одном и том же испытании случайные величины X и Y попадают в интервалы $X \in [x, x + \Delta x)$, $Y \in [y, y + \Delta y)$. Для этого введем три случайных события

$$A : \{x \leq X < x + \Delta x \cap y \leq Y < y + \Delta y\}, \\ B : \{X < x \cap Y < y + \Delta y\}, \quad C : \{X < x + \Delta x \cap Y < y\}.$$

Они изображены на Рис. 1. Очевидно, искомая вероятность равна вероятности события A , а вероятности событий B , C и их пересечения BC напрямую связаны с $F(x, y)$ (2):

$$P(B) = F(x, y + \Delta y), \quad P(C) = F(x + \Delta x, y), \quad P(B \cap C) = F(x, y). \quad (3)$$

Найдем вероятность объединения перечисленных событий. Из рисунка видно, что она равна

$$P(A \cup B \cup C) = F(x + \Delta x, y + \Delta y). \quad (4)$$

С другой стороны, по аксиоме сложения вероятностей, с учетом несовместности событий A и $B \cup C$, имеем

$$P(A \cup B \cup C) = P(A) + P(B \cup C).$$

Далее, согласно теореме сложения вероятностей,

$$P(B \cup C) = P(B) + P(C) - P(B \cap C).$$

Следовательно

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(B \cap C),$$

или

$$P(A) = P(A \cup B \cup C) - P(B) - P(C) + P(B \cap C).$$

Выразив вероятности в правой части равенства через функции распределения (3), (4), придем к искомой формуле

$$P(A) = P(x \leq X < x + \Delta x \cap y \leq Y < y + \Delta y) = F(x + \Delta x, y + \Delta y) - F(x, y + \Delta y) - F(x + \Delta x, y) + F(x, y). \quad (5)$$

3.1.2 Совместная плотность вероятностей

Найденная вероятность (5) попадания в прямоугольник A (см. Рис. 1) позволяет естественным образом определить *совместную плотность вероятностей* совокупности двух случайных величин. Действительно, плотность “вероятностной массы” на плоскости (X, Y) , как и плотность любого вещества, равна пределу при $\Delta x \rightarrow 0$, $\Delta y \rightarrow 0$, вероятности попадания (X, Y) в прямоугольник A , деленной на его площадь $\Delta x \Delta y$:

$$f(x, y) = \lim_{\Delta x, \Delta y \rightarrow 0} \frac{P(A)}{\Delta x \Delta y} = \lim_{\Delta x, \Delta y \rightarrow 0} \frac{F(x + \Delta x, y + \Delta y) - F(x, y + \Delta y) - F(x + \Delta x, y) + F(x, y)}{\Delta x \Delta y}.$$

Из математического анализа известно, что этот предел равен второй смешанной производной $F(x, y)$ по x и y . Следовательно, *совместная плотность вероятностей случайных величин* $\{X, Y\}$ равна

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}. \quad (6)$$

Ее еще называют *двумерной плотностью вероятностей*.

3.1.3 Статистические средние

Выразим интегральную функцию распределения совокупности $\{X, Y\}$ на языке средних по ансамблю. По аналогии с (2.2.9), $F(x, y)$ равна среднему от произведения единичных функций

$$F(x, y) = \langle \theta(x - X) \theta(y - Y) \rangle . \quad (7)$$

Дифференцируя обе части равенства по x и y , а затем поменяв порядок дифференцирования и статистического усреднения, получим среднее по ансамблю, равное двумерной плотности вероятностей

$$f(x, y) = \langle \delta(X - x) \delta(Y - y) \rangle . \quad (8)$$

Очевидное обобщение этого соотношения на совокупность n случайных величин $\{X_1, X_2, \dots, X_n\}$ дает их n -мерную плотность вероятностей

$$f(x_1, x_2, \dots, x_n) = \left\langle \prod_{k=1}^n \delta(X_k - x_k) \right\rangle . \quad (9)$$

Запись функций распределения с помощью статистических средних удобна для вывода многих важных свойств совокупностей случайных величин. Найдем, к примеру, как выражается с помощью плотности вероятностей среднее по ансамблю функции $g(X_1, X_2, \dots, X_n)$. Согласно выкалывающему свойству дельта-функции

$$g(X_1, X_2, \dots, X_n) = \int \cdots \int_{\mathbb{R}^n} g(x_1, x_2, \dots, x_n) \prod_{k=1}^n \delta(X_k - x_k) d^n x .$$

Усреднив это равенство по ансамблю $\{X_1, X_2, \dots, X_n\}$ и поменяв местами операции интегрирования и статистического усреднения, будем иметь

$$\begin{aligned} \langle g(X_1, X_2, \dots, X_n) \rangle &= \\ &= \int \cdots \int_{\mathbb{R}^n} g(x_1, x_2, \dots, x_n) \left\langle \prod_{k=1}^n \delta(X_k - x_k) \right\rangle d^n x = \\ &= \int \cdots \int_{\mathbb{R}^n} g(x_1, x_2, \dots, x_n) f(x_1, x_2, \dots, x_n) d^n x . \end{aligned} \quad (10)$$

В частности, среднее от функции двух случайных величин равно

$$\langle g(X, Y) \rangle = \iint_{-\infty}^{\infty} g(x, y) f(x, y) dx dy . \quad (11)$$

Замечание: Формулы (10), (11) демонстрируют главное преимущество плотностей вероятностей перед интегральными функциями распределения: Зная плотность вероятностей, можно найти среднее любой функции случайных величин, вычисляя упомянутые интегралы. ■

3.1.4 Общие свойства совместных функций распределения

Укажем общие свойства функций распределения совокупности случайных величин. Прежде всего отметим, что как и любая вероятность, $F(x, y)$ подчиняется неравенствам

$$0 \leq F(x, y) \leq 1. \quad (12a)$$

Заметим далее, что $Y < \infty$ есть достоверное событие. Поэтому, положив в (2) $y = \infty$, получим одномерную функцию распределения случайной величины X

$$F(x, y = \infty) \equiv P(X < x \cap \Omega) \equiv P(X < x) \equiv F_X(x). \quad (12b)$$

Вследствие равноправия X и Y , справедливо аналогичное тождество

$$F(x = \infty, y) \equiv P(Y < y) \equiv F_Y(y), \quad (12b')$$

выражающее, вместе с (12b), *свойство согласованности интегральных функций распределения*. Естественно, эти свойства вытекают и из определения $F(x, y)$ (7) как статистического среднего. В самом деле

$$\theta(\infty - X) \equiv 1.$$

Отсюда и из (7) вновь приходим к тождеству (12b').

Положив оба аргумента $F(x, y)$ равными бесконечности, придем к еще одному общему свойству $F(x, y)$:

$$F(x = \infty, y = \infty) \equiv P(\Omega) \equiv 1. \quad (12c)$$

Наконец заметив, что $(X < -\infty)$ и $(Y < -\infty)$ невозможные события, будем иметь

$$F(-\infty, -\infty) \equiv P(\emptyset) \equiv 0. \quad (12d)$$

Перечислим общие свойства совместной плотности вероятностей $f(x, y)$, опираясь на статистическое среднее (11).

Заметим во-первых, что среднее по ансамблю значений любой неотрицательной функции $g(X, Y) \geq 0$ не может быть меньше нуля. С другой стороны, как видно из (11), для этого необходимо и достаточно, чтобы была всюду неотрицательной совместная плотность вероятностей

$$f(x, y) \geq 0. \quad (13a)$$

Статистическое среднее функции одного случайного аргумента $g(X)$ выражается равенством

$$\langle g(X) \rangle = \int_{-\infty}^{\infty} g(x) f_X(x) dx.$$

В то же время, согласно (11), это среднее равно

$$\langle g(X) \rangle = \iint_{-\infty}^{\infty} g(x) f(x, y) dx dy.$$

Из последних двух равенств и произвольности $g(x)$ следует, что

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy. \quad (13b)$$

Это и аналогичное ему равенство

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx \quad (13b')$$

выражают *свойство согласованности* совместных плотностей вероятностей. Наконец, подставив в (11) $g(X, Y) \equiv 1$, придем к свойству нормировки совместной плотности вероятностей

$$\langle 1 \rangle = \iint_{-\infty}^{\infty} f(x, y) dx dy \equiv 1. \quad (13c)$$

3.1.5 Условные функции распределения

Напомним, условная вероятность $P(A|B)$ произвольных случайных событий A и B равна

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \quad (14)$$

Отсюда, как частный случай, вытекает определение условных функций распределения. В самом деле, возьмем условную вероятность

$$P(X < x | y \leq Y < y + \Delta y)$$

того, что в данном испытании случайная величина X будет меньше x , при условии что другая случайная величина Y в том же испытании попадет в интервал $[y, y + \Delta y)$. Согласно (14) имеем

$$P(X < x | y \leq Y < y + \Delta y) = \frac{P(X < x \cap y \leq Y < y + \Delta y)}{P(y \leq Y < y + \Delta y)}. \quad (15)$$

Знаменатель в правой части этой формулы выражается через интегральную функцию распределения случайной величины Y :

$$P(y \leq Y < y + \Delta y) = F_Y(y + \Delta y) - F_Y(y).$$

Аналогично и числитель (15) равен разности совместных интегральных функций

$$P(X < x \cap y \leq Y < y + \Delta y) = F(x, y + \Delta y) - F(x, y).$$

Подставив правые части последних двух равенств в (15), будем иметь

$$P(X < x | y \leq Y < y + \Delta y) = \frac{F(x, y + \Delta y) - F(x, y)}{F_Y(y + \Delta y) - F_Y(y)}.$$

Предел этого выражения при $\Delta y \rightarrow 0$ дает условную функцию распределения случайной величины X при условии, что $Y = y$:

$$F_X(x|y) = \lim_{\Delta y \rightarrow 0} \frac{F(x, y + \Delta y) - F(x, y)}{F_Y(y + \Delta y) - F_Y(y)}.$$

Найдем предел правой части, для чего поделим числитель и знаменатель на Δy

$$F_X(x|y) = \lim_{\Delta y \rightarrow 0} \frac{\frac{F(x, y + \Delta y) - F(x, y)}{\Delta y}}{\frac{F_Y(y + \Delta y) - F_Y(y)}{\Delta y}}.$$

Пользуясь тем, что предел частного равен частному пределов, и заметив, что последние равны производным интегральных функций распределения

$$\begin{aligned} \lim_{\Delta y \rightarrow 0} \frac{F(x, y + \Delta y) - F(x, y)}{\Delta y} &= \frac{\partial F(x, y)}{\partial y}, \\ \lim_{\Delta y \rightarrow 0} \frac{F_Y(y + \Delta y) - F_Y(y)}{\Delta y} &= \frac{dF_Y(y)}{dy} = f_Y(y), \end{aligned}$$

придем к окончательному выражению для условной интегральной функции распределения случайной величины X

$$F_X(x|y) = \frac{1}{f_Y(y)} \frac{\partial F(x, y)}{\partial y}. \quad (16)$$

Найдем условную плотность вероятностей случайной величины X , при условии, что Y равно y . Дифференцируя (16) по x и вспомнив выражение (6) для совместной плотности вероятностей, будем иметь

$$f_X(x|y) = \frac{f(x, y)}{f_Y(y)}. \quad (17)$$

Подобно тому как условная вероятность (14) обладает всеми общими свойствами обычной, безусловной, вероятности, условная плотность вероятностей (17) удовлетворяет всем свойствам обычных плотностей вероятностей. А именно, $f_X(x|y)$ неотрицательная функция, подчиняющаяся условию нормировки по x . В последнем нетрудно убедиться, интегрируя обе части равенства (17) по x и вспомнив свойство согласованности (13b') совместной плотности вероятностей

$$\int_{-\infty}^{\infty} f_X(x|y) dx = \frac{\int_{-\infty}^{\infty} f_X(x, y) dx}{f_Y(y)} = \frac{f_Y(y)}{f_Y(y)} \equiv 1.$$

Замечание: Обратим внимание, что условная плотность вероятностей $f(x|y)$ зависит от двух переменных x и y . Они имеют принципиально разный смысл. Чтобы подчеркнуть это, x называют аргументом, а y параметром $f(x|y)$. Условная плотность вероятностей, как функция аргумента, подчиняется общим свойствам плотностей вероятностей. Параметры же условной плотности вероятностей выражают условия, при которых она определена, и по ним условная плотность вероятностей не обязана удовлетворять указанным свойствам. ■

Пример: В начале раздела 1.4, посвященного геометрическому исчислению вероятностей, мы обсудили задачу о посредственном стрелке, который никогда не промахивается, стреляя по круглой мишени радиусом R , но с равной возможностью попадает в любую точку мишени. Совместим центр мишени с центром декартовой системы координат и найдем совместную плотность вероятностей случайных координат (X, Y) точки, куда попадает при очередном выстреле (испытании) посредственный стрелок. Из равной возможности попадания в любую точку мишени вытекает, что совместная плотность вероятностей координат (X, Y) постоянна внутри круга радиуса R и равна нулю вне его. Причем из условия нормировки (13с) следует, что внутри круга $f(x, y)$ обратно пропорционально его площади πR^2 . Таким образом

$$f(x, y) = \frac{1}{\pi R^2} \theta(R^2 - x^2 - y^2). \quad (18)$$

Найдем одномерные плотности вероятностей координат X и Y . В силу симметрии совместной плотности вероятностей (18) по X и Y , они одинаковы. Поэтому ограничимся построением лишь плотности вероят-

ностей $f_X(x)$ координаты X . Согласно (13b), (18)

$$f_X(x) = \frac{1}{\pi R^2} \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} dy = \frac{2}{\pi R^2} \sqrt{R^2-x^2} \quad (|x| \leq R). \quad (19)$$

Таким образом, координата X (а вместе с ней и Y), неравномерно распределена внутри интервала $|x| < R$.

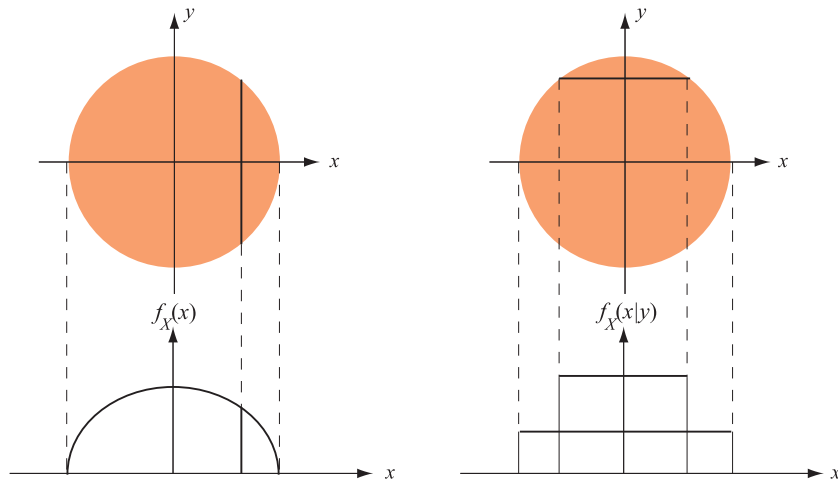


Рис. 3.2

Построение одномерной (слева) и условной (справа) плотностей вероятностей координаты X точки попадания в мишень посредством стрелком. Высота первой из них пропорциональна длине вертикального отрезка внутри круга, проектирующегося в точку x , а вторая постоянна внутри интервала – проекции на ось x заключенного внутри круга горизонтального отрезка с вертикальной координатой y .

Напротив, условная плотность вероятностей $f_X(x|y)$ (17) постоянна внутри интервала разрешенных значений $|x| \leq \sqrt{R^2 - y^2}$, а сам этот интервал сужается с ростом параметра y :

$$f_X(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{\theta(R^2 - x^2 - y^2)}{2\sqrt{R^2 - y^2}}. \quad (20)$$

Графики, поясняющие построение одномерной $f_X(x)$ и условной $f_X(x|y)$ плотностей вероятностей координаты X даны на Рис. 2. ★

3.1.6 Независимые случайные величины

Пусть теперь X и Y статистически независимы. По теореме умножения для вероятности независимых событий, их совместная интегральная функция распределения распадается на произведение функций распределения независимых случайных величин:

$$F(x, y) = P(X < x \cap Y < y) = P(X < x)P(Y < y) = F_X(x)F_Y(y).$$

Вслед за интегральной функцией, произведению одномерных плотностей вероятностей равна и совместная плотность вероятностей независимых случайных величин. Таким образом, если X и Y независимы, то

$$F(x, y) = F_X(x)F_Y(y), \quad f(x, y) = f_X(x)f_Y(y). \quad (21)$$

В частности, как видно отсюда и из (17), если X и Y независимы, то их одномерная и условная плотности вероятностей совпадают:

$$f_X(x|y) = f_X(x) \quad (\text{для независимых } X \text{ и } Y). \quad (22)$$

Если статистическая независимость случайных величин заранее очевидна, по формуле (21) находят их совместные функции распределения. Если же есть сомнения в независимости X и Y , то (22) используют как тест на их независимость. В частности, рассмотренные выше координаты X и Y точки попадания в мишень посредством стрелком статистически зависимы, поскольку их одномерная (19) и условная (20) плотности вероятностей различны. Примеры совместных плотностей вероятностей независимых случайных величин даны ниже.

Пример: Пусть X и Y равномерно распределены в интервалах $X \in [a, b]$ и $Y \in [c, d]$. Это значит, что их одномерные плотности вероятностей равны

$$f_X(x) = \frac{1}{b-a}[\theta(x-a) - \theta(x-b)], \\ f_Y(y) = \frac{1}{d-c}[\theta(y-c) - \theta(y-d)].$$

Если к тому же X и Y статистически независимы, то их совместная плотность вероятностей равна произведению указанных одномерных вероятностей. Другими словами, $f(x, y)$ постоянна внутри прямоугольника ($x \in [a, b]$, $y \in [c, d]$) и равна нулю вне его. ★

Пример: Пусть (V_x, V_y, V_z) проекции, на оси (x, y, z) декартовой системы координат, случайной скорости движения молекулы газа. Из физических соображений ясно, что в покоем газе указанные проекции скорости статистически независимы и обладают одинаковой плотностью вероятностей $f(v)$. А значит совместная плотность вероятностей проекций скорости (обозначим ее $\varphi(v_x, v_y, v_z)$) равна

$$\varphi(v_x, v_y, v_z) = f(v_x) f(v_y) f(v_z).$$

С другой стороны, в газе нет выделенных направлений и $\varphi(v_x, v_y, v_z)$ должна быть инвариантной относительно поворотов, то есть должна зависеть лишь от величины скорости молекулы $v = \sqrt{v_x^2 + v_y^2 + v_z^2}$:

$$\varphi(v_x, v_y, v_z) = \varphi(\sqrt{v_x^2 + v_y^2 + v_z^2}).$$

Следовательно, совместная плотность вероятностей проекций скорости молекулы обязана подчиняться функциональному уравнению

$$f(v_x) f(v_y) f(v_z) = \varphi(\sqrt{v_x^2 + v_y^2 + v_z^2}).$$

Легко проверить что решением данного уравнения является гауссова плотность вероятностей

$$f(v) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{v^2}{2\sigma^2}\right) \quad (23a)$$

с произвольной дисперсией σ^2 . Из статистической физики известна средняя кинетическая энергия молекул газа

$$\frac{m \langle V^2 \rangle}{2} = \frac{m}{2} (\langle V_x^2 \rangle + \langle V_y^2 \rangle + \langle V_z^2 \rangle) = \frac{3}{2} kT. \quad (23b)$$

Здесь k - постоянная Больцмана, m - масса молекулы, а T абсолютная температура газа. Таким образом, входящая в (23a) дисперсия и стандарт проекций скорости молекул газа равны

$$\sigma^2 = \frac{kT}{2m}, \quad \sigma = \sqrt{\frac{kT}{2m}}. \quad (23c)$$

Следовательно, окончательное выражение для совместной плотности вероятностей проекций скорости молекул газа таково

$$\varphi(v_x, v_y, v_z) = \left(\frac{m}{\pi kT}\right)^{3/2} \exp\left(-\frac{m}{kT}(v_x^2 + v_y^2 + v_z^2)\right). \quad (24)$$

Оно известно в физике под названием *распределения Максвелла*. ★

Пример: Полезным для дальнейшего примером совместной плотности вероятностей совокупности независимых случайных величин является плотность вероятностей гауссовой совокупности $\{X_1, \dots, X_n\}$ с нулевыми средними и одинаковыми дисперсиями σ^2 . Плотность вероятностей каждой из них по отдельности задается выражением (2.5.3). Их произведение дает искомую совместную плотность вероятностей

$$f(x_1, \dots, x_n) = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n \exp \left(-\frac{1}{2\sigma^2} \sum_{k=1}^n x_k^2 \right). \quad (25)$$

Как и трехмерное распределение Максвелла (24), оно инвариантно относительно поворотов декартовой системы координат $\{x_1, \dots, x_n\}$ в n -мерном пространстве \mathbb{R}^n . Чтобы подчеркнуть этот факт, перепишем (25) в виде

$$f(x_1, \dots, x_n) = \varphi_n(r), \quad (26)$$

$$\varphi_n(r) = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n \exp \left(-\frac{r^2}{2\sigma^2} \right), \quad r = \sqrt{x_1^2 + \dots + x_n^2}. \quad \star$$

3.2 Функции совокупностей

В этом разделе мы установим связи статистических свойств функций совокупностей случайных величин с вероятностными свойствами исходной совокупности.

3.2.1 Интегральная функция распределения

Часто в прикладных задачах интересуются не всей совокупностью случайных величин, а лишь некоторой функцией совокупности

$$Z = g(X_1, X_2, \dots, X_n). \quad (1)$$

К примеру, в статистической физике важны не столько совместные статистические свойства проекций скорости молекул газа (V_x, V_y, V_z) , сколько величина скорости

$$V = \sqrt{V_x^2 + V_y^2 + V_z^2}. \quad (2)$$

Покажем как найти интегральную функцию распределения функции случайных величин на примере функции двух аргументов

$$Z = g(X, Y). \quad (3)$$

Из определения (2.1.5) следует, что интегральная функция распределения случайной величины Z (3) равна

$$F(z) = P(Z < z) = \langle \theta(g(X, Y) - z) \rangle . \quad (4)$$

Выразим входящее сюда статистическое среднее через совместную плотность вероятностей аргументов X и Y . Согласно (1.11), (4) имеем

$$F(z) = \iint_{\mathcal{D}(z)} f(x, y) dx dy . \quad (5)$$

Здесь $\mathcal{D}(z)$ зависящая от z область интегрирования в плоскости (x, y) , где выполнено неравенство

$$\mathcal{D} : g(x, y) < z .$$

Пример: Распределением Максвелла в физике называют не только трехмерное распределение (1.24) проекций скорости молекул газа вдоль осей декартовой системы координат, но и одномерное распределение величины скорости молекулы (2). Найдем его, опираясь на трехмерный аналог формулы (5). Согласно ей, интегральная функция распределения скорости молекулы равна

$$F(v) = \iiint_{\mathcal{S}_v} \varphi(v_x, v_y, v_z) dv_x dv_y dv_z \quad (6)$$

интегралу по внутренности шара \mathcal{S}_v радиуса v , центр которого совпадает с центром декартовой системы координат $\{v_x, v_y, v_z\}$. Подставив сюда совместную плотность вероятностей (1.24) и перейдя к интегрированию по сферическим координатам, сведем интеграл к однократному

$$F(v) = 4\pi \left(\frac{m}{\pi kT} \right)^{3/2} \int_0^v \exp \left(-\frac{m}{kT} v^2 \right) v^2 dv .$$

Дифференцируя обе части равенства по v , найдем плотность вероятностей величины скорости молекул газа

$$f(u) = \sqrt{\frac{2}{\pi}} u^2 \exp \left(-\frac{u^2}{2} \right) . \quad (7)$$

Здесь, для удобства анализа, дана плотность вероятностей безразмерной скорости, нормированной на стандартное отклонение (1.23с) проекции скорости в произвольном направлении

$$U = V \sqrt{\frac{2m}{kT}} . \quad (8)$$

График распределения Максвелла (7) безразмерной скорости (8) молекулы газа дан на Рис. 3. Физики характеризуют распределение Максвелла (7) тремя разными величинами. Это, во-первых, средняя скорость

$$\langle U \rangle = \sqrt{\frac{2}{\pi}} \int_0^{\infty} u^3 \exp\left(-\frac{u^2}{2}\right) du = \sqrt{\frac{8}{\pi}} \simeq 1.596.$$

Во-вторых, наиболее вероятная скорость (мода распределения Максвелла)

$$u_{\text{mod}} = \sqrt{2} \simeq 1.414$$

Наконец, распределение (7) характеризуют *среднеквадратичной скоростью*

$$\bar{U}_{\text{sq}} = \sqrt{\langle U^2 \rangle} = \sqrt{3} \simeq 1.732 \quad \star$$

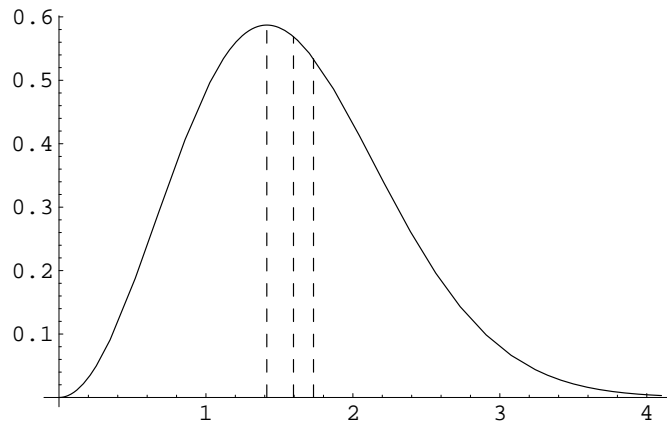


Рис. 3.3

Распределение Максвелла (7) безразмерной скорости (8) молекул. Вертикальными линиями указаны (в порядке возрастания) мода, средняя величина и среднеквадратичная скорость молекул газа. Видно, что перечисленные числовые характеристики случайной скорости не сильно отличаются друг от друга.

Пример: Найдём интегральную функцию распределения максимума совокупности случайных величин $\{X_1, \dots, X_n\}$:

$$Z = \max\{X_1, \dots, X_n\}.$$

Событие $Z < z$ эквивалентно тому, что в данном испытании все величины $\{X_1, \dots, X_n\}$ оказываются меньше z . Следовательно, искомая интегральная функция распределения равна

$$F_{\max}(z) = P(X_1 < z, \dots, X_n < z) = F(\underbrace{z, \dots, z}_n),$$

где $F(x_1, \dots, x_n)$ совместная интегральная функция распределения случайных величин $\{X_1, \dots, X_n\}$. В частности, если они статистически независимы и равномерно распределены в интервале $(0, 1)$, то

$$F_{\max}(z) = z^n, \quad f_{\max}(z) = n z^{n-1} \quad (0 < z < 1), \quad (9)$$

а среднее значение наибольшей из n равномерно распределенных величин

$$\langle Z \rangle = \frac{n}{n+1}$$

с ростом n приближается к единице. ★

3.2.2 χ^2 -распределение

Естественное обобщение формулы (5) на n -мерный случай имеет вид

$$F(z) = \int_{\mathcal{D}(z)} f(x_1, \dots, x_n) d^n x. \quad (10)$$

Знак интеграла здесь символизирует интегрирование по n -мерной области $\mathcal{D}(z)$, где выполнено неравенство

$$g(x_1, \dots, x_n) < z. \quad (11)$$

Применим формулу (10) для вычисления интегральной функции распределения и плотности вероятностей случайной величины

$$Y = \sum_{k=1}^n X_k^2, \quad (12)$$

где $\{X_1, \dots, X_n\}$ независимые гауссовы величины, совместная плотность вероятностей которых задана равенством (1.26). В данном случае \mathcal{D} – внутренность n -мерного шара с центром в начале декартовой системы координат $\{x_1, \dots, x_n\}$ и радиусом \sqrt{y} . Пользуясь симметрией плотности вероятностей (1.26), сведем интеграл (10) к однократному

$$F_n(y) = \int_0^{\sqrt{y}} \varphi_n(r) S_n(r) dr. \quad (13)$$

Здесь интегрирование ведется по концентрическим сферическим слоям, элементарный объем которых равен $dV = S_n(r) dr$, где S_n площадь гиперповерхности n -мерной сферы радиуса r . Известно, что она равна

$$S_n(r) = 2 \frac{\pi^{n/2}}{\Gamma(n/2)} r^{n-1}. \quad (14)$$

В частности, с учетом свойств гамма функции отсюда следуют знакомые формулы площади сферы в 1D, 2D и 3D пространствах

$$S_1(r) = 2, \quad S_2(r) = 2\pi r, \quad S_3(r) = 4\pi r^2.$$

Продифференцировав обе части равенства (13) по y , найдем плотность вероятностей случайной величины Y (12)

$$f_n(y) = \frac{1}{2\sqrt{y}} \varphi(\sqrt{y}) S_n(\sqrt{y}). \quad (15)$$

Подставив сюда явные выражения $\varphi_n(r)$ (1.26) и $S_n(r)$ (14), получим окончательно

$$f_n(y) = \frac{1}{2^{n/2}\Gamma(n/2)\sigma^n} y^{\frac{n}{2}-1} \exp\left(-\frac{y}{2\sigma^2}\right) \quad (y > 0). \quad (16)$$

Это распределение широко используется в математической статистике, где оно носит название χ^2 -распределения с n степенями свободы. Заметим еще, что оно является частным случаем уже встречавшегося нам гамма распределения (2.5.19). Графики плотности вероятностей (16) при $\sigma = 1$ и разных степеней свободы изображены на Рис. 4

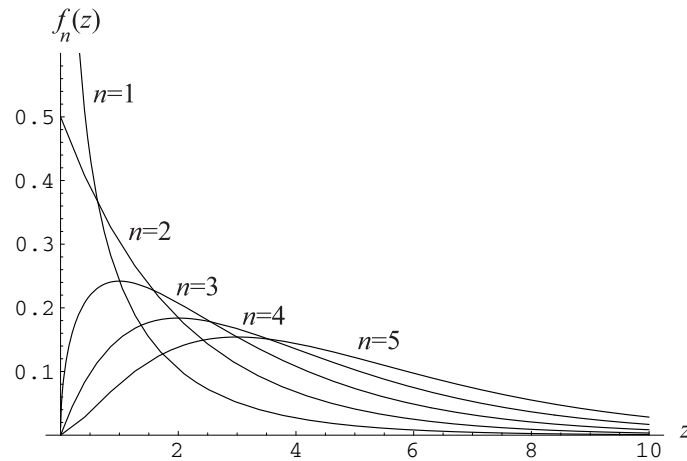


Рис. 3.4

Графики χ^2 -распределения для разных степеней свободы и при $\sigma = 1$.

Частный случай плотности вероятностей (16) при $n = 1$

$$f_1(y) = \frac{1}{\sqrt{2\pi y} \sigma} \exp\left(-\frac{y}{2\sigma^2}\right) \quad (17)$$

можно трактовать как плотность вероятностей гауссова шума, пропущенного через квадратичный детектор ($Y = X^2$). Конечно, значение распределения (17) не ограничивается чисто инженерными приложениями. К нему приходят, например, при изучении квантовомеханических хаотических состояний. В этом случае X играет роль случайной величины волновой функции, а $Y = X^2$ ее плотности. Распределение (17) применительно к случайной плотности хаотической волновой функции называют *распределением Портера-Томаса*.

Замечание: Обратим внимание, что в случае $n = 1$ плотность вероятностей $f_n(y)$ имеет бесконечную особенность при $y = 0$. Для $n \geq 2$ эта особенность пропадает. Более того, с ростом n плотность вероятностей $f_n(y)$ все сильнее прижимается к нулю в окрестности $y = 0$. Это связано со свойством геометрии n -мерного пространства, состоящим в том, что чем больше его размерность n , тем быстрее уменьшается объем областей с уменьшением их линейных размеров. Продемонстрируем это свойство, исследовав вопрос: в пространстве какой размерности выгоднее всего покупать апельсины. Пусть в любом пространстве толщина кожуры апельсина составляет одну десятую радиуса апельсина. Известно, что объем апельсина радиуса R в n -мерном пространстве равен

$$V_n(R) = \frac{\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2} + 1\right)} R^n. \quad (18)$$

Найдем отношение объема мякоти к полному объему апельсина. Оно равно

$$R_n = \frac{V_n(9R/10)}{V_n(R)} = \left(\frac{9}{10}\right)^n$$

и максимально в 1D пространстве $R_1 = 0.9$. В родном 3D пространстве мякоти существенно меньше: $R_3 = 0.729$. Наконец в 10D пространстве $R_{10} = 0.348$, а значит отношение объема мякоти к объему всего апельсина практически в два раза меньше, чем в земных магазинах. Читателю полезно запомнить на будущее, что в пространствах очень большой размерности почти весь объем апельсина приходится на его кожуру. ■

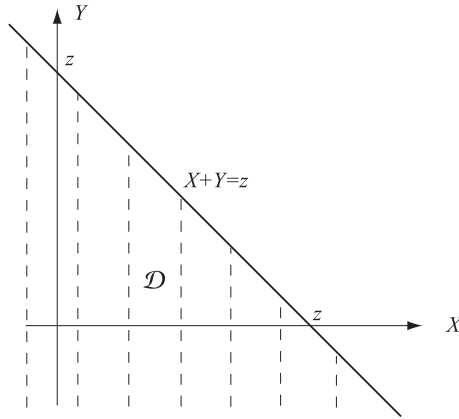


Рис. 3.5

Изображение области \mathcal{D} , вероятностная масса которой равна интегральной функции распределения $F_Z(z)$ суммы $Z = X + Y$.

3.2.3 Распределение суммы

Применим формулу (5) к нахождению функций распределения суммы случайных величин

$$Z = X + Y .$$

В данном случае область \mathcal{D} , изображенная на Рис. 5, лежит ниже прямой $z = X + Y$. Соответственно, двойной интеграл (29) сводится к повторному интегрированию

$$F_Z(z) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{z-x} f(x, y) dy \right] dx .$$

Дифференцируя обе части равенства по z , получим плотность вероятностей суммарной величины Z :

$$f_Z(z) = \int_{-\infty}^{\infty} f(x, z - x) dx . \quad (19)$$

В частности, для независимых слагаемых, совместная плотность вероятностей которых распадается на произведение, отсюда имеем

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x) dx . \quad (20)$$

Это знаменитый *интеграл свертки*. Операция свертки имеет много общего с умножением. Поэтому будем использовать форму записи свертки

плотностей вероятностей

$$f_Z(z) = f_X(z) \otimes f_Y(z), \quad (20a)$$

где, сходный с символом умножения \times , знак \otimes символизирует свертку.

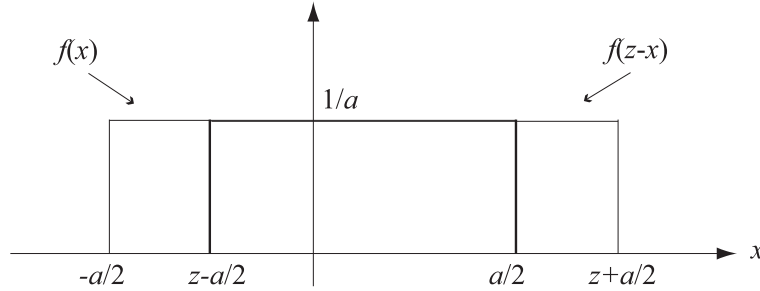


Рис. 3.6

Иллюстрация к построению распределения Симпсона (21).

Пример: Найдём по формуле (20) плотность вероятностей суммы $Z = X_1 + X_2$ независимых и равномерно распределённых в интервале $(-a/2, a/2)$ величин. Их плотности вероятностей равны

$$f(x) = \begin{cases} \frac{1}{a}, & |x| < \frac{a}{2}; \\ 0, & |x| \geq \frac{a}{2}. \end{cases}$$

Интеграл свертки (20) в данном случае удобно вычислять графически. На Рис. 6 изображены графики $f(x)$ и $f(z-x)$. Видно, что произведение этих функций, в случае $0 < |z| < a$, образует прямоугольник высотой $1/a^2$ и длиной $a - |z|$. Его площадь равна искомой $f_Z(z)$. При $|z| > a$ прямоугольники $f(x)$ и $f(z-x)$ не перекрываются, а $f_Z(z) = 0$. Таким образом, плотность вероятностей суммы имеет треугольную форму:

$$f_Z(z) = \begin{cases} \frac{1}{a^2}(a - |z|), & |z| < a; \\ 0 & |z| \geq a. \end{cases} \quad (21)$$

Её иногда называют *распределением Симпсона* ★.

Пример: Вычислим, опираясь на формулы (20), плотность вероятностей суммы независимых слагаемых

$$T_n = \sum_{k=1}^n \tau_k, \quad (22)$$

каждое из которых имеет экспоненциальное распределение (2.5.9). Обозначим за $f_n(t)$ плотность вероятностей суммы (22) n слагаемых. Очевидно, она подчиняется рекуррентному соотношению

$$f_{n+1}(t) = f_n(t) \otimes f_1(t), \quad f_1(t) = \nu e^{-\nu t} \quad (t > 0).$$

Ввиду неотрицательности случайных величин τ_k , свертка принимает вид

$$f_{n+1}(t) = \int_0^t f_n(x) f_1(t-x) dx.$$

Подставив сюда экспоненциальное распределение $f_1(t)$, будем иметь

$$f_{n+1}(t) = \nu e^{-\nu t} \int_0^t f_n(x) e^{\nu x} dx.$$

Поочередно беря $n = 1, 2, \dots$, получим окончательно

$$f_n(t) = \frac{(\nu t)^{n-1}}{(n-1)!} e^{-\nu t} \quad (t > 0). \quad (23)$$

Замечание: Обратим внимание, что правая часть равенства (23), с точностью до замены k на $n-1$, совпадает с правой частью соотношения (2.3.22). Таким образом, графики на Рис. 2.12 заодно являются графиками плотностей вероятностей (23). ■

3.2.4 Преобразование распределений с помощью дельта-функций

При нахождении плотностей вероятностей функций случайных величин часто удобно пользоваться определением плотности вероятностей как статистического среднего от дельта-функций, и свойствами дельта-функции. Например известной формулой

$$\delta(kx - a) = \frac{1}{|k|} \delta\left(x - \frac{a}{k}\right). \quad (24)$$

Ниже даны примеры, иллюстрирующие сказанное.

Пример: Отыщем плотность вероятностей линейной комбинации

$$Z = aX + bY + c$$

случайных величин X и Y с плотностью вероятностей $f(x, y)$. Она равна

$$f_Z(z) = \langle \delta(z - aX - bY - c) \rangle = \iint_{-\infty}^{\infty} \delta(z - ax - by - c) f(x, y) dx dy.$$

Избавимся с помощью дельта-функции от интеграла по y . Используя равенство

$$\delta(z - ax - by - c) = \frac{1}{|b|} \delta\left(y - \frac{z - ax - c}{b}\right)$$

и выкалывающее свойство дельта-функции, получим окончательно

$$f_Z(z) = \frac{1}{|b|} \int_{-\infty}^{\infty} f\left(x, \frac{z - ax - c}{b}\right) dx.$$

При $a = b = 1$, $c = 0$ имеем, как частный случай, формулу (20). ★

Найдем аналогичным способом плотность вероятностей произведения случайных величин $Z = XY$:

$$f_Z(z) = \iint_{-\infty}^{\infty} \delta(z - xy) f(x, y) dx dy = \iint_{-\infty}^{\infty} \frac{1}{|x|} \delta\left(y - \frac{z}{x}\right) f(x, y) dx dy.$$

Пользуясь выкалывающим свойством, получим окончательно

$$f_Z(z) = \int_{-\infty}^{\infty} \frac{1}{|x|} f\left(x, \frac{z}{x}\right) dx. \quad (25)$$

Пример: Найдем с помощью (25) плотность вероятностей произведения независимых гауссовых величин X и Y с нулевыми средними и одинаковой дисперсией σ^2 . В этом случае

$$f(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right),$$

а интеграл (25) принимает вид:

$$f_Z(z) = \frac{1}{\pi\sigma^2} \int_0^{\infty} \exp\left[-\frac{1}{2\sigma^2} \left(x^2 + \frac{z^2}{x^2}\right)\right] \frac{dx}{x}.$$

Заглянув в справочник интегралов или прибегнув к услугам программы МАТНЕМАТИСА, получим

$$f_Z(z) = \frac{1}{\pi\sigma^2} K_0\left(\frac{|z|}{\sigma^2}\right). \quad (26)$$

Здесь $K_0(z)$ модифицированная функция Бесселя второго рода нулевого порядка. Для справки укажем ее асимптотики:

$$K_0(z) \sim \ln\left(\frac{2}{z}\right) \quad (z \rightarrow 0); \quad K_0(z) \sim \sqrt{\frac{\pi}{2z}} e^{-z} \quad (z \rightarrow \infty).$$

График плотности вероятностей $f_Z(z)$ (26) при $\sigma = 1$ изображен на Рис. 7.

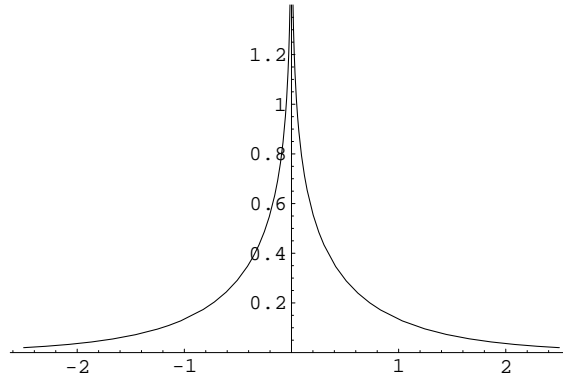


Рис. 3.7

Плотность вероятностей произведения гауссовых случайных величин с нулевым средним и единичной дисперсией.

Для понимания характерных особенностей плотности вероятностей произведения $Z = XY$, удобно трактовать Z линейной функцией Y со случайным коэффициентом X . Малые значения коэффициента X порождают бесконечную логарифмическую особенность плотности вероятностей (26) в окрестности $z = 0$, а большие X ответственны за медленное, в сравнении с гауссовым распределением, спадание к нулю $f_Z(z)$. ★

3.2.5 Распределение Стьюдента

Приведем, для полноты картины, еще и плотность вероятностей частного случайных величин $Z = X/Y$:

$$f_Z(z) = \left\langle \delta\left(z - \frac{X}{Y}\right) \right\rangle = \langle |Y| \delta(X - Yz) \rangle.$$

Выразив последнее среднее через совместную плотность вероятностей $f(x, y)$ и используя выкалывающее свойство дельта-функции, получим

$$f_Z(z) = \int_{-\infty}^{\infty} |y| f(yz, y) dy. \quad (27)$$

В статистике большую роль играет *распределение Стьюдента*. Так называют плотность вероятностей частного

$$Z = \frac{X_0}{U}, \quad (28)$$

где

$$U = \sqrt{\frac{1}{n} \sum_{k=1}^n X_k^2}, \quad (29)$$

а $\{X_0, X_1, \dots, X_n\}$ совокупность $n + 1$ независимых гауссовых величин с нулевыми средними и одинаковыми дисперсиями σ^2 . Найдем искомую плотность вероятностей по формуле (27). Имея ввиду, что числитель в (28) распределен по гауссу, а также что U неотрицательна, получим

$$f_n(z) = \frac{1}{\sqrt{2\pi}\sigma} \int_0^\infty u \exp\left(-\frac{u^2 z^2}{2\sigma^2}\right) f_U(u) du. \quad (30)$$

Здесь $f_U(u)$ плотность вероятностей случайной величины U (29). Соответствующую интегральную функцию распределения найдем, заменив верхний предел интеграла справа в (13) на \sqrt{nu} :

$$F_U(u) = \int_0^{\sqrt{nu}} \varphi_n(r) S_n(r) dr.$$

Дифференцируя равенство по u , найдем плотность вероятностей

$$f_U(u) = \sqrt{n} \varphi_n(\sqrt{n}u) S_n(\sqrt{n}u).$$

Подставив сюда $\varphi_n(r)$ (1.26) и $S_n(r)$ (14), получим окончательно

$$f_U(u) = \frac{2 \left(\frac{n}{2}\right)^{n/2}}{\sigma^n \Gamma\left(\frac{n}{2}\right)} u^{n-1} \exp\left(-\frac{nu^2}{2\sigma^2}\right) \quad (u > 0). \quad (31)$$

Поместив правую часть (31) в (30), будем иметь

$$f_n(z) = \sqrt{\frac{2}{\pi}} \frac{\left(\frac{n}{2}\right)^{n/2}}{\sigma^{n+1} \Gamma\left(\frac{n}{2}\right)} \int_0^\infty u^n \exp\left(-\frac{u^2(n+z^2)}{2\sigma^2}\right) du. \quad (32)$$

Зависимость $f_n(z)$ от аргумента z легко установить уже сейчас, не вычисляя интеграла в правой части равенства. Покажем это, перейдя к новой переменной интегрирования

$$v = \sqrt{\frac{n+z^2}{2\sigma^2}} u.$$

В итоге получим

$$f_n(z) = \frac{C_n}{\sqrt{\pi n} \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{z^2}{n}\right)^{-\frac{n+1}{2}}.$$

Здесь

$$C_n = 2 \int_0^\infty v^n e^{-v^2} dv.$$

Последний интеграл выражается через гамма функцию:

$$C_n = \Gamma\left(\frac{n+1}{2}\right).$$

Подставив ее в предыдущее равенство, получим окончательно

$$f_n(z) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n} \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{z^2}{n}\right)^{-\frac{n+1}{2}}. \quad (33)$$

Это так называемое *распределение Стьюдента с n степенями свободы*.

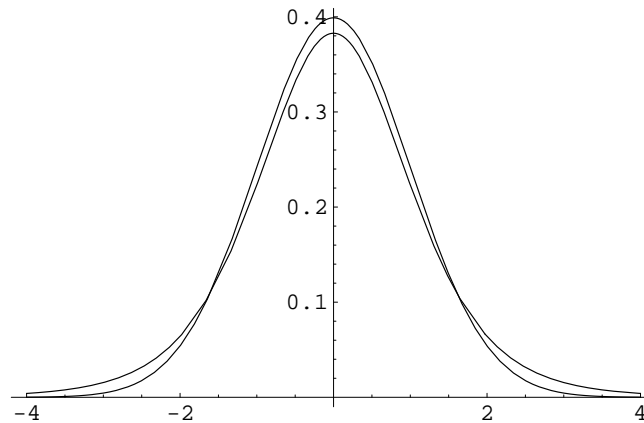


Рис. 3.8

Распределение Стьюдента при $n = 6$ и предельное распределение Гаусса (34). Видно что графики очень близки. Существенное отличие пожалуй лишь в более медленно спадающих хвостах распределения Стьюдента.

Укажем некоторые характерные особенности распределения Стьюдента (33). Оно универсально в том смысле, что не зависит от дисперсии σ^2 компонент $\{X_0, X_1, \dots, X_n\}$, формирующих величину Z (28).

При $n = 1$ распределение Стьюдента совпадает с распределением Коши (2.2.46), а с ростом n асимптотически стремится к гауссовой плотности вероятностей:

$$\lim_{n \rightarrow \infty} f_n(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right). \quad (34)$$

График распределения Стьюдента при $n = 6$ и предельного гауссова распределения (34) изображен на Рис. 8.

3.3 Отображения совокупностей

Выше обсуждены свойства одной случайной величины Z (2.1), равной функции набора величин $\{X_1, \dots, X_n\}$. В приложениях бывает необходимо исследовать трансформацию статистических свойств совокупности случайных величин при их отображении в другую совокупность. Некоторые возникающие при этом полезные соотношения обсуждаются ниже.

3.3.1 Взаимно-однозначное отображение

Для удобства выкладок перейдем к векторным обозначениям

$$\{X_1, \dots, X_n\} = \mathbf{X}, \quad \{Y_1, \dots, Y_n\} = \mathbf{Y}. \quad (1)$$

Пусть дано отображение случайного вектора \mathbf{X} в \mathbf{Y}

$$\mathbf{Y} = \boldsymbol{\alpha}(\mathbf{X}), \quad (2)$$

заданное детерминированной векторной функцией

$$\mathbf{y} = \boldsymbol{\alpha}(\mathbf{x}) : \quad \{y_1 = \alpha_1(x_1, \dots, x_n), \dots, y_n = \alpha_n(x_1, \dots, x_n)\}. \quad (3)$$

Будем для определенности считать, что она всюду дифференцируема, а равенство (2) осуществляет взаимно-однозначное отображение области Ω_x возможных значений \mathbf{X} на область Ω_y возможных значений \mathbf{Y} , такое, что существует обратное дифференцируемое отображение $\Omega_y \mapsto \Omega_x$

$$\mathbf{X} = \boldsymbol{\beta}(\mathbf{Y}). \quad (4)$$

Для этого необходимо и достаточно, чтобы всюду в Ω_y был отличен от нуля и ограничен якобиан отображения:

$$0 < |J(\mathbf{y})| < \infty, \quad (5)$$

где

$$J(\mathbf{y}) = \det \begin{vmatrix} \frac{\partial \beta_1}{\partial y_1} & \cdots & \frac{\partial \beta_1}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial \beta_n}{\partial y_1} & \cdots & \frac{\partial \beta_n}{\partial y_n} \end{vmatrix}. \quad (6)$$

Выясним как выражается совместная плотность вероятностей $f_{\mathbf{Y}}(\mathbf{y})$ случайного вектора \mathbf{Y} через совместную плотность вероятностей $f_{\mathbf{X}}(\mathbf{x})$, которую будем считать известной. Для этого найдем статистическое среднее от функции $g(\alpha(\mathbf{X}))$, где $g(\mathbf{y})$ произвольная функция векторного аргумента \mathbf{y} . Согласно формуле (1.10) имеем

$$\langle g(\alpha(\mathbf{X})) \rangle = \langle g(\mathbf{Y}) \rangle = \int_{\Omega_x} g(\alpha(\mathbf{x})) f_{\mathbf{X}}(\mathbf{x}) d^n x. \quad (7)$$

Перейдя к новым переменным интегрирования

$$\mathbf{y} = \alpha(\mathbf{x}) \quad \Longleftrightarrow \quad \mathbf{x} = \beta(\mathbf{y})$$

и заметив, что элементарные объемы преобразуются по правилу

$$d^n x = |J(\mathbf{y})| d^n y, \quad (8)$$

будем иметь

$$\langle g(\mathbf{Y}) \rangle = \int_{\Omega_y} g(\mathbf{y}) f_{\mathbf{X}}(\beta(\mathbf{y})) |J(\mathbf{y})| d^n y. \quad (9)$$

С другой стороны, то же самое среднее выражается через совместную плотность вероятностей $f_{\mathbf{Y}}(\mathbf{y})$ случайной совокупности \mathbf{Y} равенством

$$\langle g(\mathbf{Y}) \rangle = \int_{\Omega_y} g(\mathbf{y}) f_{\mathbf{Y}}(\mathbf{y}) d^n y.$$

Сравнивая его с предыдущим и учитывая произвольность $g(\mathbf{y})$, приходим к искомой формуле, связывающей плотности вероятностей случайных векторов \mathbf{Y} и \mathbf{X} :

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(\beta(\mathbf{y})) |J(\mathbf{y})|. \quad (10)$$

Пример: Поставим типичную для инженерных приложений статистическую задачу. Пусть $Z = U + iV$ комплексное число, реальная и мнимая части которого равны

$$U = X + a, \quad V = Y, \quad (11)$$

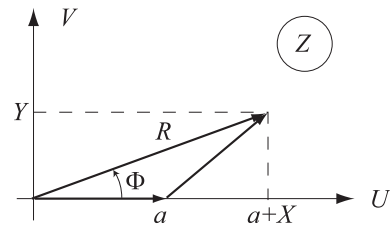


Рис. 3.9

Иллюстрация к исследованию статистики комплексного числа Z .

где X и Y гауссовы величины с нулевыми средними и одинаковой дисперсией σ^2 , а a детерминированная величина. По тем или иным физическим или инженерным соображениям, обычно удобно представлять комплексное число в экспоненциальной форме

$$Z = R e^{i\Phi},$$

где R абсолютная величина, а Φ фаза случайного числа Z . Требуется, по известной статистике совокупности $\{X, Y\}$, определить плотность вероятностей R и Φ .

В данном случае отображение плоскости (X, Y) на область возможных значений $\{R \in (0, \infty), \Phi \in (-\pi, \pi)\}$ задается парой функций

$$x = r \cos \varphi - a, \quad y = r \sin \varphi \quad (12)$$

с якобианом

$$J = \det \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \varphi} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \varphi} \end{vmatrix} = \det \begin{vmatrix} \cos \varphi & -r \sin \varphi \\ \sin \varphi & r \cos \varphi \end{vmatrix} = r. \quad (13)$$

Пусть $f(x, y)$ совместная плотность вероятностей величин $\{X, Y\}$. Обозначим плотность вероятностей совокупности $\{R, \Phi\}$ за $w(r, \varphi)$. Из (10), (12), (13) следует, что они связаны равенством

$$w(r, \varphi) = r f(r \cos \varphi - a, r \sin \varphi). \quad (14)$$

Подставив сюда плотность вероятностей независимых гауссовых величин $\{X, Y\}$

$$f(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right),$$

получим окончательно

$$w(r, \varphi) = \frac{r}{2\pi\sigma^2} \exp\left(-\frac{r^2 + a^2 - 2ar \cos \varphi}{2\sigma^2}\right), \quad (15)$$

$$r \in (0, \infty), \quad \varphi \in (-\pi, \pi).$$

Обсудим свойства найденного распределения. При $a = 0$ оно распадается на произведение *рэлеевского распределения* R

$$w_R(r) = \frac{r}{\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right), \quad r > 0, \quad (16)$$

и равномерной плотности вероятностей Φ :

$$w_\Phi(\varphi) = \frac{1}{2\pi}, \quad \varphi \in (-\pi, \pi). \quad (17)$$

Таким образом, при $a = 0$ абсолютная величина и фаза комплексного числа Z (11) статистически независимы.

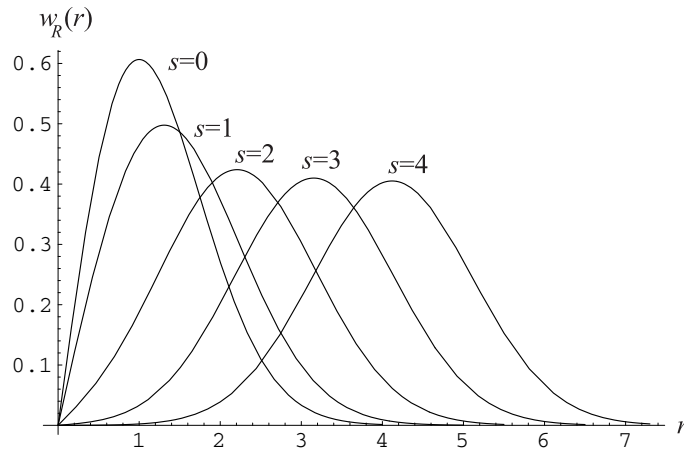


Рис. 3.10

Графики обобщенного распределения Рэля при $\sigma = 1$ и разных отношениях сигнала к шуму.

Если $a \neq 0$, то R и Φ оказываются зависимыми, а их плотности вероятностей существенно трансформируются. Покажем это, вычислив одномерные плотности вероятностей R и Φ при $a \neq 0$. Интегрируя совместную плотность вероятностей (15) по разрешенным значениям фазы и пользуясь табличным интегралом

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{z \cos \varphi} d\varphi = I_0(z), \quad (18)$$

где $I_0(z)$ модифицированная функция Бесселя нулевого порядка, будем иметь

$$w_R(r) = \frac{r}{\sigma^2} \exp\left(-\frac{r^2 + a^2}{2\sigma^2}\right) I_0\left(\frac{ar}{\sigma^2}\right). \quad (19)$$

Эту плотность вероятностей называют *обобщенным распределением Рэля*. Его форма зависит от безразмерного параметра $s = a/\sigma$, который в инженерных приложениях трактуют как *отношение сигнала к шуму*. Графики плотности вероятностей (19) при разных s даны на Рис. 10.

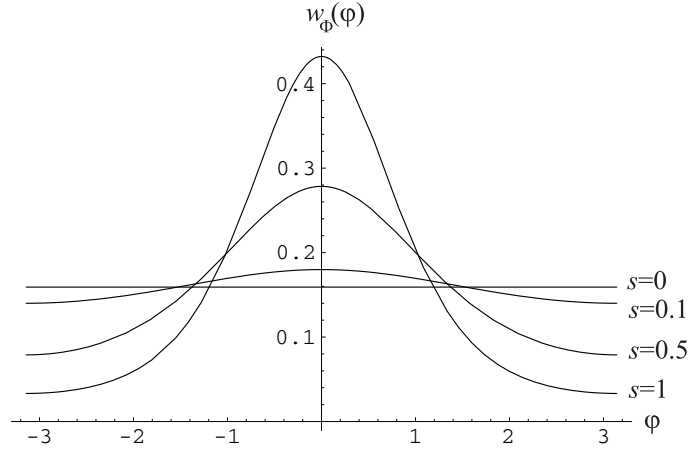


Рис. 3.11

Плотность вероятностей фазы комплексного числа Z для разных отношений сигнала к шуму $s = a/\sigma$. При $s = 0$ оно вырождается в равномерное, а с ростом $|s|$ все более сосредотачивается в окрестности $\varphi = 0$.

Проследим теперь, как меняется с ростом s плотность вероятностей случайной фазы Φ . Интегрируя (15) по всем $r > 0$, получим

$$w_{\Phi}(\varphi) = \frac{1}{2\pi} e^{-\frac{s^2}{2}} \left[1 + \sqrt{\frac{\pi}{2}} z e^{\frac{z^2}{2}} \left(1 + \operatorname{erf} \left(\frac{z}{\sqrt{2}} \right) \right) \right] \quad (z = s \cos \varphi). \quad (20)$$

Графики этого распределения приведены на Рис. 11. ★

Пример: Во многих приложениях важно знать, как преобразуются плотности вероятностей при линейных отображениях случайных величин

$$Y_i = \sum_{j=1}^n \alpha_{ij} X_j$$

или в векторно-матричных обозначениях

$$\mathbf{Y} = \mathcal{A}\mathbf{X} \quad (\mathcal{A} = \|\alpha_{ij}\|). \quad (21)$$

Если детерминант матрицы \mathcal{A} не равен нулю, то имеется обратное отображение

$$\mathbf{X} = \mathcal{B}\mathbf{Y}, \quad (22)$$

где \mathcal{B} обратная к \mathcal{A} матрица. При этом детерминант обратного отображения (6) всюду одинаков и равен детерминанту матрицы \mathcal{B} , а формула (10) приобретает вид

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(\mathcal{B}\mathbf{y}) |\det \mathcal{B}|. \quad \star \quad (23)$$

3.4 Корреляции случайных величин

Пусть X и Y зависимые случайные величины. Типичный для многочисленных приложений вопрос состоит в следующем: Можно ли, по измеренной величине X , предсказать, какое значение в том же испытании примет Y , используя зависимость между ними. К сожалению, обычно не существует детерминированной связи между исследуемыми случайными величинами. Зависимость между ними проявляется *статистически* и выражается тем, что их совместная плотность вероятностей $f(x, y)$ не распадается на произведение плотностей вероятностей каждой из величин по отдельности. Тем не менее иногда удается подобрать такую детерминированную функцию $\varphi(x)$, что вспомогательная случайная величина $Y' = \varphi(X)$ довольно хорошо прогнозирует искомую величину Y . Выбор подходящей функции $\varphi(x)$ сильно зависит от конкретных совместных статистических свойств совокупности (X, Y) . Поэтому подробно обсудим здесь лишь возможность *линейного прогнозирования*, когда искомую случайную величину Y аппроксимируют линейной функцией

$$Y' = aX + b, \quad (1)$$

выбирая a и b так, чтобы стала минимальной *среднеквадратичная ошибка*

$$\varepsilon^2 = \langle (Y - Y')^2 \rangle. \quad (2)$$

Возникающие при решении задачи линейного прогнозирования важные понятия теории совокупностей случайных величин обсуждаются ниже.

3.4.1 Корреляция и ковариация

Прежде всего подставим правую часть равенства (1) в (2) и раскроем полученное среднее

$$\begin{aligned} \varepsilon^2 &= \langle (Y - aX - b)^2 \rangle = \\ &= \langle Y^2 \rangle + a^2 \langle X^2 \rangle + b^2 - 2a \langle XY \rangle - 2b \langle Y \rangle + 2ab \langle X \rangle. \end{aligned} \quad (3)$$

Помимо уже известных первых двух моментов случайных величин X и Y , сюда входит статистическое среднее

$$\langle XY \rangle = \iint_{-\infty}^{\infty} xy f(x, y) dx dy \quad (4)$$

произведения X и Y . Его называют *корреляцией* случайных величин.

Минимальное значение среднеквадратичной ошибки (2) найдем, решив пару уравнений относительно a и b , вытекающих из условий экстремума функции (3):

$$\left\{ \frac{\partial \varepsilon^2}{\partial a} = 0, \quad \frac{\partial \varepsilon^2}{\partial b} = 0 \right\} \iff \begin{cases} a \langle X^2 \rangle + b \langle X \rangle = \langle XY \rangle, \\ a \langle X \rangle + b = \langle Y \rangle. \end{cases}$$

Отсюда имеем:

$$a = \frac{\langle XY \rangle - \langle X \rangle \langle Y \rangle}{\sigma_x^2}, \quad b = \langle Y \rangle - a \langle X \rangle. \quad (5)$$

Здесь использовано уже знакомое обозначение дисперсии X

$$\sigma_x^2 = \langle X^2 \rangle - \langle X \rangle^2 = \langle \tilde{X}^2 \rangle.$$

Особого внимания заслуживает числитель выражения для a . Подобно дисперсии, он равен корреляции случайных компонент X и Y

$$\text{Cov}(X, Y) = \langle XY \rangle - \langle X \rangle \langle Y \rangle \equiv \langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle \equiv \langle \tilde{X} \tilde{Y} \rangle. \quad (6)$$

Поэтому, по аналогии с дисперсией (variance), $\text{Cov}(X, Y)$ называют *ковариацией* (covariance) случайных величин X и Y .

Подставив (5) и (6) в правую часть (3), найдем минимальное значение среднеквадратичной ошибки.

$$\min \varepsilon^2 = \sigma_x^2 \sigma_y^2 (1 - \rho^2). \quad (7)$$

Сюда входит безразмерная величина

$$\rho = \frac{\text{Cov}(X, Y)}{\sigma_x \sigma_y} = \frac{\langle \tilde{X} \tilde{Y} \rangle}{\sqrt{\langle \tilde{X}^2 \rangle \langle \tilde{Y}^2 \rangle}}, \quad (8)$$

называемая *коэффициентом корреляции* X и Y . Из неотрицательности ε^2 следует, что коэффициент корреляции лежит в пределах

$$-1 \leq \rho \leq 1. \quad (9)$$

Из (7) видно также, что чем больше по абсолютной величине коэффициент корреляции, тем меньше, при фиксированных σ_x и σ_y , ошибка линейного прогнозирования (1). Таким образом, коэффициент корреляции количественно выражает *степень линейной зависимости* величин X и Y . При $\rho = 0$ она пропадает вовсе, а $|\rho| = 1$ означает, что между X и Y имеется детерминированная линейная зависимость.

Добавим еще, что если $\rho = 0$, то случайные величины X и Y называются *некоррелированными*. В противном случае они являются *коррелированными* величинами. Очевидно, статистически независимые случайные величины заодно и некоррелированы. Обратное вообще говоря не верно.

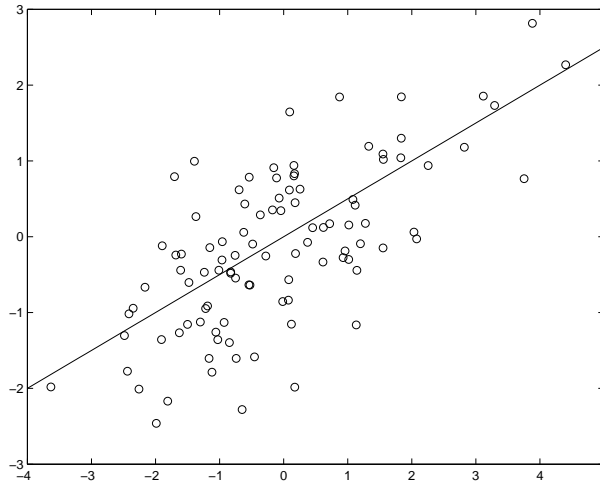


Рис. 3.12

Сто пар реализаций (X, Y) и график линии регрессии (11) в случае равных дисперсий сигнала и шума.

Подставив (5) в (1) и учитывая определение коэффициента корреляции (8), запишем уравнение оптимального линейного прогнозирования в окончательной форме, наглядно выражающей его суть:

$$Y' = \langle Y \rangle + \frac{\sigma_y}{\sigma_x} \rho (X - \langle X \rangle) = \langle Y \rangle + \frac{\sigma_y}{\sigma_x} \rho \tilde{X}. \quad (10)$$

Первое слагаемое справа представляет собой как бы “нулевую итерацию” линейного прогнозирования, согласно которой прогнозируемое значение Y' принимается равным среднему оцениваемой случайной величины Y . Второй член корректирует нулевую итерацию с учетом корреляции между \tilde{X} и \tilde{Y} . Прямую, заданную уравнением (10), называют *линией регрессии* Y на X .

Пример: Сформулируем характерную для инженерных приложений задачу линейного прогнозирования. Пусть Y (случайный) сигнал, который надо измерить. В процессе приема сигнал искажается, так что в действительности измеряют смесь $X = Y + N$ сигнала и шума N (noise). Требуется оценить значение сигнала по измеренной величине X . Обычно источники сигнала и шума различны, и их можно считать некоррелированными. Кроме того, не ограничивая общности, примем средние сигнала и шума равными нулю. Тогда уравнение оптимального линейного прогнозирования примет вид

$$Y' = \frac{\sigma^2}{\sigma^2 + \rho^2} X, \quad (11)$$

где σ^2 дисперсия сигнала Y , а ρ^2 дисперсия шума N . На Рис. 12 кружками изображены точки на плоскости (X, Y) , полученные генерацией ста гауссовых случайных чисел $\{Y, N\}$ с нулевым средним и единичной дисперсией. Соответствующая программа MATLAB такова:

```
>> y=randn(1,100);
>> n=randn(1,100);
>> plot(y+n,y,'o')
>> hold on
>> y=linspace(-4,5);
>> plot(y,y/2)
```

Там же проведена прямая линейного прогноза, заданная в данном случае уравнением $Y' = X/2$. ★

Пример: Возьмем случайные величины

$$X = \sin \Phi, \quad Y = \cos \Phi. \quad (12)$$

Они жестко связаны уравнением $X^2 + Y^2 = A^2$. Тем не менее

$$\langle XY \rangle = \frac{A^2}{2\pi} \int_0^{2\pi} \sin \varphi \cos \varphi d\varphi = 0 \quad \implies \quad \rho = 0,$$

а линейный прогноз, сводящийся здесь к равенству $Y' = 0$, полностью неэффективен. Этот факт констатируют, говоря, что случайные величины (12) *статистически зависимы, но некоррелированы*. ★

3.4.2 Ковариационная матрица

Мы придем к еще одному ключевому понятию теории вероятностей, исследуя дисперсию линейной комбинации случайных величин

$$Y = \sum_{i=1}^n \alpha_i X_i. \quad (13)$$

Здесь и далее $\{\alpha_1, \dots, \alpha_n\}$ набор детерминированных коэффициентов. Дисперсия Y равна среднему квадрату линейной комбинации (13) случайных составляющих X_i :

$$D[Y] = \left\langle \left(\sum_{i=1}^n \alpha_i \tilde{X}_i \right)^2 \right\rangle. \quad (14)$$

Пользуясь тем, что среднее суммы равно сумме средних, а также определением ковариации (6), запишем (14) в виде двойной суммы

$$D[Y] = \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(X_i, X_j) \alpha_i \alpha_j. \quad (15)$$

В итоге мы пришли к квадратичной форме коэффициентов $\{\alpha_1, \dots, \alpha_n\}$. Ее называют *положительно определенной*, так как она неотрицательна для любых $\{\alpha_1, \dots, \alpha_n\}$. Коэффициенты квадратичной формы (15) образуют *ковариационную матрицу*

$$\mathcal{C} = \left\| c_{ij} \right\|, \quad c_{ij} = \text{Cov}(X_i, X_j) = \langle \widetilde{X}_i \widetilde{X}_j \rangle. \quad (16)$$

На ее диагонали расположены дисперсии величин $\{X_1, \dots, X_n\}$, а сама матрица симметрична в силу очевидного свойства ее элементов (16)

$$c_{ij} = c_{ji}. \quad (17)$$

Вслед за квадратичной формой (15), ковариационную матрицу (16) также называют *положительно определенной*.

Заметим в заключение, что если компоненты вектора \mathbf{X} некоррелированы, то есть если все

$$\text{Cov}(X_i, X_j) = 0 \quad (i \neq j),$$

то дисперсия суммы оказывается равной

$$D[Y] = \sum_{i=1}^n \alpha_i^2 D[X_i].$$

Укажем, важное для понимания дальнейшего, следствие этого соотношения, вытекающее при всех $\alpha_i \equiv 1$: *Дисперсия суммы некоррелированных случайных величин равна сумме дисперсий слагаемых*:

$$D \left[\sum_{i=1}^n X_i \right] = \sum_{i=1}^n D[X_i]. \quad (18)$$

Ранее мы установили, что среднее суммы случайных величин равно сумме их средних. Теперь выяснилось, что подобным замечательным свойством обладает и дисперсия суммы, но лишь независимых слагаемых.

3.5 Еще раз о Законе Больших Чисел

В разделе (2.4.3) уже обсуждался Закон Больших Чисел применительно к вероятности произвольного события A . В более широком понимании, этот закон является одним из фундаментальных законов, объясняющих самые разнообразные явления. Так он снимает противоречие между хаосом микроскопических движений молекул окружающей среды, и возможностью описать ее макроскопическое поведение динамическими законами, например уравнениями газодинамики: Хотя отдельные частицы газа имеют случайные скорости, даже очень точные измерения давления P газа дают вполне определенные, не случайные, значения P . Заметим, на заре становления физики, возможность измерения давления использовалась как неопровержимый аргумент против молекулярной теории строения вещества. И лишь Закон Больших Чисел расставил все на свои места, примирив молекулярную теорию с газодинамикой и другими физическими теориями, изучающими закономерности поведения сплошных сред.

В предыдущем разделе мы выяснили, что дисперсия суммы независимых случайных величин равна сумме дисперсий слагаемых. Этот факт и доказанное ниже *неравенство Чебышева* позволят дать более общую формулировку Закона Больших Чисел.

3.5.1 Неравенство Чебышева

Между дисперсией случайной величины и ее плотностью вероятностей имеется важная связь, выраженная неравенством Чебышева

$$P(|X| \geq \varepsilon) \leq \frac{\langle X^2 \rangle}{\varepsilon^2}. \quad (1)$$

Докажем его, считая X непрерывной случайной величиной с интегрируемой в обычном смысле плотностью вероятностей $f(x)$. Заметим однако, что доказательство легко обобщить на произвольный случай, лишь бы средний квадрат X был ограничен: $\langle X^2 \rangle < \infty$.

По определению, средний квадрат случайной величины равен

$$\langle X^2 \rangle = \int_{-\infty}^{\infty} x^2 f(x) dx.$$

Разобьем ось интегрирования на интервал $|x| < \varepsilon$ и области, где $|x| \geq \varepsilon$:

$$\langle X^2 \rangle = \int_{|x| \geq \varepsilon} x^2 f(x) dx + \int_{-\varepsilon}^{\varepsilon} x^2 f(x) dx.$$

Заметим далее, что из неотрицательности $f(x)$ и x^2 следует, что

$$\int_{-\varepsilon}^{\varepsilon} x^2 f(x) dx \geq 0,$$

а значит

$$\langle X^2 \rangle \geq \int_{|x| \geq \varepsilon} x^2 f(x) dx.$$

В свою очередь оставшийся интеграл удовлетворяет неравенству

$$\int_{|x| \geq \varepsilon} x^2 f(x) dx \geq \varepsilon^2 \int_{|x| \geq \varepsilon} f(x) dx,$$

поскольку при $|x| \geq \varepsilon$ и $x^2 \geq \varepsilon^2$ (см. Рис. 13). Таким образом, мы пришли к неравенству

$$\langle X^2 \rangle \geq \varepsilon^2 \int_{|x| \geq \varepsilon} f(x) dx.$$

Обратим внимание на вероятностный смысл оставшегося интеграла:

$$\int_{|x| \geq \varepsilon} f(x) dx = P(|X| \geq \varepsilon).$$

Он равен вероятности, что модуль X больше или равен ε . Следовательно, предыдущее неравенство эквивалентно неравенству Чебышева (1).

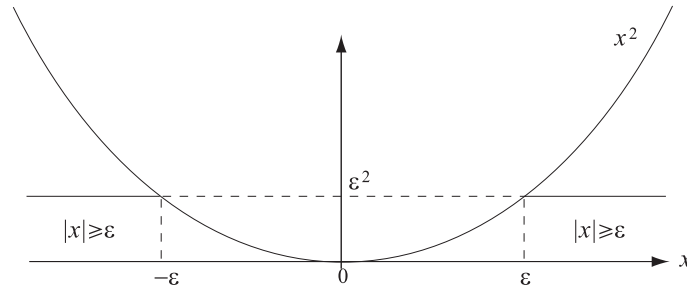


Рис. 3.13

Иллюстрация к доказательству неравенства Чебышева.

В теории вероятностей широко используют неравенство Чебышева в иной форме. А именно, от X переходят к случайной компоненте $\tilde{X} = X - \langle X \rangle$. Средний квадрат последней равен дисперсии X , а значит неравенство Чебышева можно переписать в виде

$$P(|X - \langle X \rangle| \geq \varepsilon) \leq \frac{\sigma_x^2}{\varepsilon^2}. \quad (2)$$

Следовательно: *вероятность отклонения произвольной случайной величины X от ее среднего на величину, большую ε , не превышает ее дисперсии, деленной на ε^2 .*

3.5.2 Закон Больших Чисел в форме Чебышева

Неравенство Чебышева позволяет дать простое и довольно универсальное доказательство Закона Больших Чисел, с частным случаем которого мы уже сталкивались в рамках схемы независимых испытаний Бернулли. Сформулируем Закон Больших Чисел в форме теоремы, известной как *теорема Чебышева*:

Теорема: Если $\{X_1, \dots, X_n\}$ последовательность взаимно-независимых случайных величин с одинаковыми дисперсиями $\sigma_x^2 < \infty$ и средними $\langle X \rangle$, то для любого $\varepsilon > 0$ справедливо предельное неравенство

$$\lim_{n \rightarrow \infty} P \left(\left| \frac{1}{n} \sum_{k=1}^n X_k - \langle X \rangle \right| < \varepsilon \right) = 1. \quad (3)$$

Для доказательства введем вспомогательную случайную величину

$$Y = \frac{1}{n} \sum_{k=1}^n X_k. \quad (4)$$

Очевидно, ее статистическое среднее равно $\langle X \rangle$. Дисперсия же Y равна сумме дисперсий слагаемых

$$\sigma_y^2 = \sum_{k=1}^n \frac{\sigma_x^2}{n^2} = \frac{1}{n} \sigma_x^2. \quad (5)$$

Таким образом, согласно неравенству Чебышева применительно к случайной величине Y (4), имеем

$$P(|Y - \langle X \rangle| \geq \varepsilon) \leq \frac{\sigma_y^2}{\varepsilon^2} = \frac{\sigma_x^2}{n\varepsilon^2}. \quad (6)$$

Напомним, вероятность противоположного событию $|Y - \langle X \rangle| \geq \varepsilon$ события $|Y - \langle X \rangle| < \varepsilon$ равна

$$P(|Y - \langle X \rangle| < \varepsilon) = 1 - P(|Y - \langle X \rangle| \geq \varepsilon) \geq 1 - \frac{\sigma_x^2}{n\varepsilon^2}.$$

Заменив Y правой частью равенства (4), в пределе $n \rightarrow \infty$ получим

$$\lim_{n \rightarrow \infty} P \left(\left| \frac{1}{n} \sum_{k=1}^n X_k - \langle X \rangle \right| < \varepsilon \right) \geq 1.$$

Поскольку вероятность любого события не превышает единицы, отсюда следует утверждение теоремы Чебышева (3).

3.5.3 Построение гистограмм

Закон Больших Чисел математически выражает фундаментальное свойство статистической устойчивости случайных явлений, изучаемых теорией вероятностей. Он обосновывает возможность измерения статистических характеристик случайных величин по итогам большой серии испытаний. Сама же процедура измерения тех или иных статистических характеристик, а также оценка точности соответствующих измерений, требует разработки своих специфических идей и методов. Проиллюстрируем сказанное описанием процедуры измерения плотности вероятностей случайных величин.

Пусть итогом n независимых измерений случайной величины Z стал вектор $\mathbf{Z} = \{Z_1, \dots, Z_n\}$ измеренных значений. Эмпирическая оценка формы плотности вероятностей Z по вектору \mathbf{Z} сводится к построению *гистограммы*. Так называют набор примыкающих столбиков одинаковой ширины Δ . Их высота N_k равна числу измеренных значений случайной величины Z , попадающих в основание k -го столбика

$$[z_{\min} + k\Delta - \Delta, z_{\min} + k\Delta) \quad (7)$$

В программе MATLAB построение гистограммы выполняет команда

```
>> hist(z,m)
```

Здесь \mathbf{z} вектор измеренных значений, а m число столбиков гистограммы. Причем за z_{\min} берется минимальное измеренное значение, а ширина столбиков равна

$$\Delta = \frac{z_{\max} - z_{\min}}{m},$$

где z_{\max} максимальное из измеренных значений. Команда

```
>> hist(z)
```

с опущенным необязательным аргументом m генерирует гистограмму из десяти прямоугольников. Приведем несколько примеров гистограмм.

Пример: Первой построим гистограмму равномерно распределенной в интервале $(0, 1)$ величины Z по тысяче ее значений:

```
>> z=rand(1,1000);  
>> hist(z)
```

Первая строка формирует вектор из $n=1000$ реализаций Z , а вторая строит гистограмму. Результат изображен на Рис. 14. ★

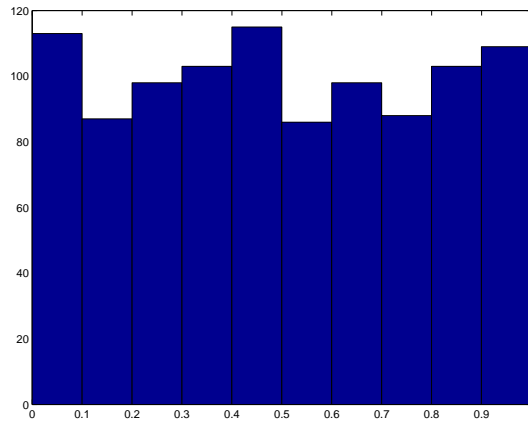


Рис. 3.14

Гистограмма равномерно распределенной величины Z , построенная по тысяче реализаций. Вдоль горизонтали автоматически отложены возможные значения Z , а по вертикали число реализаций, попавших в основание соответствующего столбика.

Построенная гистограмма (Рис. 14) может вызвать недовольство экспериментатора, так как высоты столбиков довольно сильно колеблются вокруг истинного значения 100, равного умноженной на $n = 1000$ вероятности попадания Z в интервал длиной $\Delta = 0.1$. Чтобы гистограмма была ближе по форме к исследуемой плотности вероятностей, надо увеличивать число n . Кроме того, при измерении плотностей вероятностей $f(z)$ неравномерно распределенных случайных величин, стараются улучшить “разрешающую способность”. Последняя определяется числом m столбцов гистограммы.

Убедимся на примерах, что рост числа испытаний, в согласии с Законом Больших Чисел, повышает точность измерения плотности вероятностей.

Пример: Построим гистограмму случайной величины, имеющей распределение Симпсона (2.21) при $a = 1$. Сформируем два вектора одинаковой длины из реализаций равномерно распределенных величин

```
>> x=rand(1,10000);
>> y=rand(1,10000);
```

а затем построим гистограмму центрированной суммы

```
>> z=x+y-1;
>> hist(z,30)
```

Здесь мы в десять раз увеличили число реализаций, но взяли в три раза большее число столбиков. Итоговая гистограмма изображена на Рис. 15.

Видно, что она довольно хорошо повторяет треугольную форму распределения Симпсона. ★

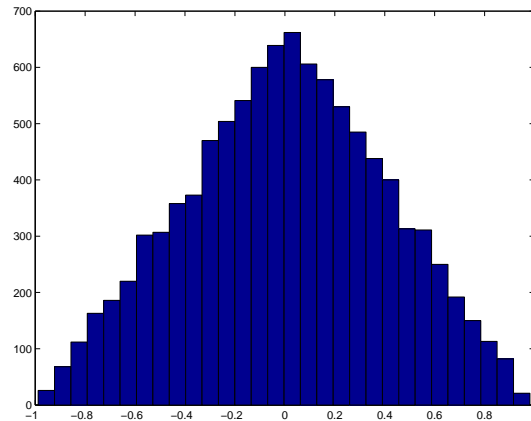


Рис. 3.15

Гистограмма распределения Симпсона, построенная по $n = 10,000$ значениям, рассортированным по $m = 30$ столбикам.

Пример: Построим гистограмму максимума $U = \max(X, Y, Z)$ из трех независимых величин, равномерно распределенных в интервале $(0, 1)$. Гистограмма строится программой

```
>> x=rand(3,50000);  
>> u=max(x);  
>> hist(u,50)
```

Как и следовало ожидать, соответствующая гистограмма, построенная на Рис. 16, близка к параболической форме распределения U : $f(u) = 3u^2$ (2.9). ★

Помимо плотности вероятностей, на практике может понадобиться интегральная функция распределения исследуемых случайных величин. Составим программу построения интегральной функции распределения гауссовой величины. Создадим вектор из 5,000 реализаций:

```
>> n=5000;  
>> x=rand(n,1);
```

Затем построим гистограмму:

```
>> subplot(2,1,1)  
>> hist(x,50)
```

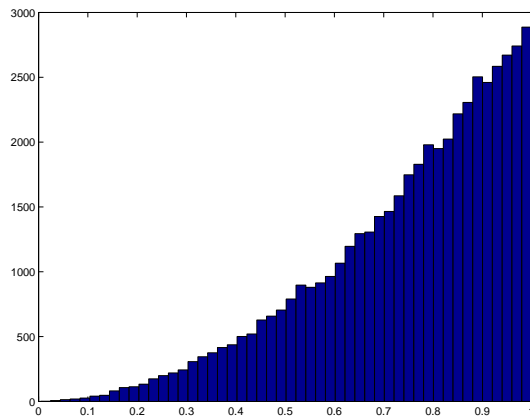


Рис. 3.16

Гистограмма максимума из трех равномерно распределенных величин.

Построим эмпирическую интегральную функцию распределения. Для этого сформируем векторы высот столбиков и координат их центров:

```
>> [f,xc]=hist(x,50);
```

а затем сформируем вектор эмпирической интегральной функции распределения, и построим график

```
>> P=cumsum(f)/n;
```

```
>> subplot(2,1,2)
```

```
>> plot(xc,P)
```

После чего сравним ее с теоретической интегральной функцией распределения гауссовой величины $N(0,1)$:

```
>> F=(1+erf(xc/sqrt(2)))/2;
```

```
>> hold on
```

```
>> plot(xc,F)
```

Замечание: Эмпирическая интегральная функция распределения лучше согласуется с теоретической, чем гистограмма с гауссовой плотностью вероятностей. Это связано с тем, что численные методы лучше приспособлены к вычислению интегральных характеристик, чем дифференциальных (в данном случае плотности вероятностей). ■

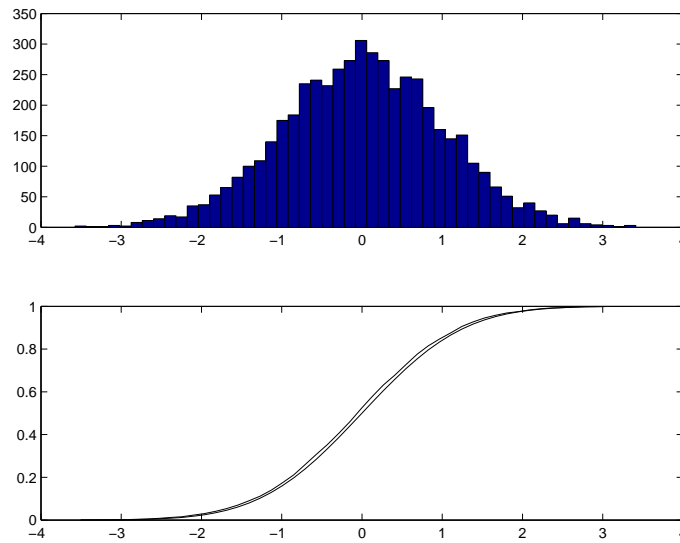


Рис. 3.17

Гистограмма и эмпирическая интегральная функция распределения гауссовой случайной величины $X = \text{randn}$, генерируемой датчиком случайных чисел программы MATLAB.

3.6 Характеристическая функция совокупности случайных величин

Наиболее ценные качества характеристических функций, как инструмента теории вероятностей, проявляются при анализе совокупностей случайных величин. Этот раздел посвящен обсуждению свойств случайных совокупностей с помощью характеристических функций.

3.6.1 Многомерные характеристические функции

По определению, характеристической функцией совокупности случайных величин $\mathbf{X} = \{X_1, \dots, X_n\}$ называют функцию векторного аргумента $\mathbf{u} = \{u_1, \dots, u_n\}$, равную

$$\Theta(\mathbf{u}) = \langle e^{i(\mathbf{u} \cdot \mathbf{X})} \rangle = \left\langle \exp \left(i \sum_{k=1}^n u_k X_k \right) \right\rangle. \quad (1)$$

Как и в одномерном случае, n -мерная характеристическая функция (1) обращается, в силу условия нормировки, в единицу при $\mathbf{u} = 0$. Если же положить равными нулю лишь некоторые компоненты вектора \mathbf{u} , то

в итоге получим характеристическую функцию меньшей совокупности случайных величин, для которых $u_k \neq 0$.

Заметим еще, что как и в случае характеристической функции одной случайной величины, моментные функции компонент вектора \mathbf{X} можно вычислить дифференцированием совместной характеристической функции. Так корреляцию случайных величин X_k и X_m находят по формуле

$$\langle X_k X_m \rangle = - \left. \frac{\partial^2 \Theta(\mathbf{u})}{\partial u_k \partial u_m} \right|_{\mathbf{u}=0}. \quad (2)$$

Зная n -мерную характеристическую функцию (1), легко найти характеристическую функцию любой линейной комбинации совокупности, например характеристическую функцию суммы случайных величин

$$Y = \sum_{k=1}^n X_k. \quad (3)$$

Для этого надо лишь взять компоненты вектора \mathbf{u} одинаковыми и равными u :

$$\Theta_Y(u) = \Theta(\underbrace{u, \dots, u}_n) = \left\langle \exp \left(iu \sum_{k=1}^n X_k \right) \right\rangle. \quad (4)$$

Если все компоненты случайного вектора $\mathbf{X} = \{X_1, \dots, X_n\}$ статистически независимы, среднее в правой части (1), а вместе с ним и n -мерная характеристическая функция $\Theta(\mathbf{u})$, распадается на произведение одномерных характеристических функций каждого из слагаемых

$$\Theta(\mathbf{u}) = \prod_{k=1}^n \Theta_k(u_k), \quad (5)$$

где $\Theta_k(u)$ характеристическая функция случайной величины X_k . В частности, характеристическая функция суммы (3) одинаково распределенных независимых случайных величин равна

$$\Theta_Y(u) = \Theta^n(u) \quad (6)$$

n -й степени характеристической функции слагаемых суммы.

Пример: Найдем с помощью (6) характеристическую функцию (2.3.9) биномиального распределения (2.3.7). Напомним, биномиально распределенная величина K представима в виде суммы (2.3.10) независимых

величин, подчиняющихся закону Бернулли (2.2.26). Следовательно, характеристическая функция K равна n -й степени характеристической функции (2.2.34) распределения Бернулли

$$\Theta_K(u) = (q + p e^{iu})^n.$$

Заметим, что если бы мы знали формулу (6) ранее, то могли бы вывести биномиальное распределение обратным преобразованием Фурье от правой части последнего равенства. ★

Пример: Получим характеристическую функцию плотности вероятностей (2.23) суммы (2.22) n экспоненциально распределенных независимых случайных величин. Для этого напомним, что характеристическая функция каждого слагаемого равна

$$\Theta_\tau(u) = \int_0^\infty \nu e^{-\nu t + iut} dt = \frac{\nu}{\nu - iu}.$$

Возведя ее в n -ю степень, получим характеристическую функцию суммы (2.22)

$$\Theta_T(u) = \left(\frac{\nu}{\nu - iu} \right)^n. \quad \star$$

Пример: Определим характеристическую функцию χ^2 распределения (2.16). Напомним, оно является плотностью вероятностей суммы (2.12) квадратов независимых гауссовых случайных величин. Найдем характеристическую функцию отдельного слагаемого

$$\Theta(u) = \langle e^{iux^2} \rangle = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2\sigma^2} + iux^2\right) dx = \frac{1}{\sqrt{1 - 2iu\sigma^2}}.$$

Ее n -я степень дает характеристическую функцию χ^2 -распределения

$$\Theta_Y(u) = (1 - 2iu\sigma^2)^{-n/2}.$$

Заметим, это частный случай характеристической функции (2.5.20) гамма распределения (2.5.19). ★

3.6.2 Кумулянты

Мы уже знаем, что дисперсия суммы независимых величин равна сумме дисперсий слагаемых. Характеристическая функция позволяет

указать бесконечное множество других статистических характеристик, обладающих подобным *свойством аддитивности*. Покажем это, записав характеристическую функцию $\Theta(u)$ произвольной случайной величины X в виде

$$\Theta(u) = e^{\Psi(u)}, \quad (7)$$

где

$$\Psi(u) = \ln \Theta(u). \quad (8)$$

Разложим вспомогательную функцию $\Psi(u)$ в ряд по степеням u

$$\Psi(u) = \sum_{m=1}^{\infty} \frac{\varkappa_m}{m!} (iu)^m.$$

Коэффициенты \varkappa_m этого разложения называют *кумулянтами* (или *семиинвариантами*) случайной величины X . С их помощью характеристическая функция записывается в форме

$$\Theta(u) = \exp \left(\sum_{m=1}^{\infty} \frac{\varkappa_m}{m!} (iu)^m \right). \quad (9)$$

Из свойства (5) характеристических функций совокупностей независимых величин следует, что m -й кумулянт суммы равен сумме кумулянтов слагаемых. В частности, подставив (9) в (6), получим

$$\Theta_y(u) = \exp \left(\sum_{m=1}^{\infty} \frac{n\varkappa_m}{m!} (iu)^m \right). \quad (10)$$

Отсюда видно, что кумулянты суммы независимых и одинаково распределенных случайных величин равны кумулянтам слагаемых, умноженным на их число n .

Дифференцируя (10) n раз по iu , положив $u = 0$ и приравняв результат к μ_n , нетрудно получить соотношения, связывающие кумулянты с моментами. Так для первых четырех кумулянтов имеем

$$\begin{aligned} \varkappa_1 &= \mu_1, \\ \varkappa_2 &= \mu_2 - \mu_1^2 = \sigma_x^2, \\ \varkappa_3 &= \mu_3 - 3\mu_1^2\mu_2 + 2\mu_1^3, \\ \varkappa_4 &= \mu_4 - 3\mu_2^2 - 4\mu_1\mu_3 + 12\mu_1^2\mu_2 - 6\mu_1^4. \end{aligned}$$

Обратим внимание: первые два кумулянта совпадают со средним и дисперсией X . Остальные характеризуют форму плотности вероятностей $f(x)$. Поскольку форма $f(x)$ более тесно связана не с моментами, а центральными моментами случайной величины, дадим формулы связи высших (при $k > 2$) кумулянтов с центральными моментами. Для этого

достаточно в приведенных формулах положить $\mu_1 = 0$, а остальные моменты заменить центральными моментами. В итоге получим

$$\varkappa_3 = \nu_3, \quad \varkappa_4 = \nu_4 - 3\nu_2^2 = \nu_4 - 3\sigma_x^4. \quad (11)$$

Обсудим связь кумулянтов с формой плотности вероятностей. Если у случайной величины X не равен нулю лишь первый кумулянт \varkappa_1 , то ее характеристическая функция и плотность вероятностей таковы

$$\Theta(u) = e^{i\varkappa_1 u} \iff f(x) = \delta(x - \varkappa_1).$$

Иными словами, величина X , у которой отличен от нуля лишь первый кумулянт, детерминирована, а ее значение равно $X = \varkappa_1$. Подчеркнем, как и в данной вырожденной ситуации, первый кумулянт ответственен, в общем случае, за смещение плотности вероятностей $f(x)$ вдоль оси x , и не влияет на форму $f(x)$.

Пусть теперь отличны от нуля первые два кумулянта $\varkappa_1 \neq 0$ и $\varkappa_2 \neq 0$, а остальные равны нулю. Тогда плотность вероятностей оказывается гауссовой:

$$\Theta(u) = \exp\left(i\varkappa_1 u - \frac{\varkappa_2}{2} u^2\right) \iff f(x) = \frac{1}{\sqrt{2\pi\varkappa_2}} \exp\left(-\frac{(x-\varkappa_1)^2}{2\varkappa_2}\right). \quad (12)$$

Причем, если первый кумулянт учитывает смещение плотности вероятностей, то второй определяет ее эффективную ширину. Точнее, эффективная ширина определяется стандартом, равным корню второго кумулянта:

$$\sigma = \sqrt{\varkappa_2}.$$

Этот же смысл второго кумулянта, как квадрата эффективной ширины плотности вероятностей, сохраняется за ним и в общем случае. Поэтому, анализируя свойства плотностей вероятностей, не зависящие от их ширины, переходят, как правило, от исходной случайной величины X к вспомогательной безразмерной величине

$$\hat{X} = \frac{X}{\sigma} = \frac{X}{\sqrt{\varkappa_2}}. \quad (13)$$

Она нормирована в том смысле, что дисперсия \hat{X} равна единице.

У безразмерной нормированной случайной величины \hat{X} (13) все кумулянты безразмерны и равны

$$\kappa_m = \frac{\varkappa_m}{\sigma^m} = \frac{\varkappa_m}{\varkappa_2^{m/2}} \quad (m = 3, 4, \dots). \quad (14)$$

Эти безразмерные кумулянты называют *кумулянтными коэффициентами* исходной случайной величины X .

Мы выяснили, что характеристическая функция, у которой все кумулянты, кроме первых двух, равны нулю, отвечает гауссовой плотности вероятностей. Оказывается, функция $\Theta(u)$ с любым другим ограниченным набором не равных нулю кумулянт теряет свойство положительной определенности. Следовательно, истинная характеристическая функция может иметь или первые два, или бесконечное число не равных нулю кумулянт. Все же чем меньше кумулянтные коэффициенты (14), тем ближе плотность вероятностей случайной величины к гауссовой. Поэтому кумулянтные коэффициенты (14) используют как количественную меру близости формы плотности вероятностей к гауссовой. На практике для этой цели привлекают лишь два первых кумулянтных коэффициента, имеющие специальные названия: *Асимметрия*

$$\kappa_3 = \frac{\varkappa_3}{\sigma^3} = \frac{\nu_3}{\sigma^3} \quad (15a)$$

и *эксцесс*

$$\kappa_4 = \frac{\varkappa_4}{\sigma^4} = \frac{\varkappa_4}{\varkappa_2^2}. \quad (15b)$$

Названия отражают смысл этих величин. Асимметрия ответственна за отклонение формы плотности вероятностей $f(x)$ от симметричной, а эксцесс указывает степень отклонения (симметричной) формы плотности вероятностей от гауссовой. Хотя это не всегда так, при объяснении смысла эксцесса обычно говорят, что при $\kappa_4 > 0$ вершина плотности вероятностей $f(x)$ более острая, чем у гауссова распределения с той же дисперсией, а при $\kappa_4 < 0$ вершина более тупая.

Приведем несколько примеров плотностей вероятностей и соответствующих им кумулянтных коэффициентов.

Пример: Пусть X равномерно распределена в интервале $X \in (a, b)$. Ее плотность вероятностей задана выражением (2.2.27), а характеристическая функция равна (2.2.38). Кумулянты порядка выше первого удобно находить из разложения в ряд Тейлора логарифма характеристической функции (2.2.39) случайной компоненты \tilde{X} :

$$\tilde{\Psi}(u) = \ln \tilde{\Theta}(u) = \ln \left(\frac{\sin u\ell}{u\ell} \right).$$

Здесь $\ell = (b-a)/2$. Разлагая $\tilde{\Psi}(u)$ по степеням u вручную или с помощью математических программ, получим

$$\tilde{\Psi}(u) = -\frac{\ell^2}{6} u^2 - \frac{\ell^4}{180} u^4 - \frac{\ell^6}{2835} u^6 + O(u^8).$$

Как и следовало ожидать, все нечетные кумулянты случайной компоненты равномерно распределенной случайной величины равны нулю, а первые три четных кумулянта таковы:

$$\kappa_2 = \sigma^2 = \frac{1}{3}\ell^2, \quad \kappa_4 = -\frac{2}{15}\ell^4, \quad \kappa_6 = \frac{16}{63}\ell^6.$$

Поделив четвертый кумулянт на квадрат второго, найдем эксцесс равномерной плотности вероятностей

$$\kappa_4 = -\frac{6}{5}. \quad \star$$

Пример: Найдем кумулянты экспоненциального распределения (2.5.9)

$$f(t) = \nu e^{-\nu t} \quad \Longleftrightarrow \quad \Theta(u) = \frac{\nu}{\nu - iu}.$$

Разложение соответствующей характеристической функции в ряд по степеням u таково

$$\ln\left(\frac{\nu}{\nu - iu}\right) = \sum_{m=1}^{\infty} \frac{(iu)^m}{m \nu^m}.$$

Отсюда следует, что кумулянты экспоненциального распределения, в том числе его асимметрия и эксцесс, равны

$$\kappa_m = \frac{(m-1)!}{\nu^m}, \quad \kappa_3 = 2, \quad \kappa_4 = 6. \quad \star$$

3.6.3 Центральная предельная теорема

Аппарат характеристических функций и обсужденное выше аддитивное свойство кумулянтов суммы независимых величин позволяет довольно легко убедиться в справедливости главной теоремы теории вероятностей, так называемой *Центральной Предельной Теоремы* (ЦПТ), находящей многочисленные применения не только в физике, а пожалуй всюду, где наблюдаются случайные явления.

Центральная предельная теорема оперирует суммами большого числа случайных величин

$$X_1 + X_2 + \dots + X_n \quad (n \gg 1). \quad (16)$$

Чтобы не утонуть в несущественных технических деталях, будем считать $\{X_1, X_2, \dots, X_n\}$ независимыми и одинаково распределенными. Кроме того, положим для простоты, что ограничены все моменты, а вместе с ними и кумулянты слагаемых.

Прежде чем формулировать Центральную Предельную Теорему, укажем ее отношение к другому универсальному закону теории вероятностей – Закону Больших Чисел. Последний утверждает, что арифметическое среднее суммы независимых случайных величин

$$\bar{X}_n = \frac{1}{n}(X_1 + X_2 + \dots + X_n) \quad (17)$$

становится с ростом n все менее случайным. В самом деле, если дисперсии слагаемых в (17) равны σ_x^2 , то стандарт суммы (17), характеризующий отклонение от статистического среднего, стремится к нулю как

$$\sigma_n = \frac{\sigma_x}{\sqrt{n}}.$$

Для многих приложений, однако, отклонения арифметического от статистического среднего принципиальны, и важно знать их вероятностные свойства. Грубый Закон Больших Чисел не дает ответа на этот вопрос. Зато на него отвечает более тонкая Центральная Предельная Теорема.

Поскольку ЦПТ описывает отклонения от среднего, исключим из суммы (16) статистические средние слагаемых и будем интересоваться плотностью вероятностей суммы случайных компонент:

$$X^{(n)} = \frac{1}{\sqrt{n}} \sum_{k=1}^n \tilde{X}_k. \quad (18)$$

Она нормирована на \sqrt{n} , чтобы дисперсия $X^{(n)}$ не менялась с ростом n , а оставалась равной дисперсии слагаемых $D[X^{(n)}] = \sigma_x^2$. Обсудим характеристическую функцию случайной величины $X^{(n)}$ (18):

$$\Theta_n(u) = \left\langle \exp \left(i \frac{u}{\sqrt{n}} \sum_{k=1}^n \tilde{X}_k \right) \right\rangle = \Theta^n \left(\frac{u}{\sqrt{n}} \right). \quad (19)$$

Здесь $\Theta(u)$ характеристическая функция слагаемых \tilde{X}_k . Выразим ее через кумулянты $\tilde{\chi}_k$ и подставим в (19). В итоге получим

$$\Theta_n(u) = \exp \left[n \sum_{m=2}^{\infty} \frac{\tilde{\chi}_m}{m!} \left(\frac{i u}{\sqrt{n}} \right)^m \right].$$

Суммирование здесь начинается с $m = 2$, поскольку мы заранее исключили первый кумулянт, перейдя к суммированию случайных компонент \tilde{X}_k . Еще раз перепишем $\Theta_n(u)$, выделив первый сомножитель:

$$\Theta_n(u) = \exp \left(-\frac{\tilde{\chi}_2}{2} u^2 \right) \exp \left(-i \frac{\tilde{\chi}_3}{6\sqrt{n}} u^3 + \frac{\tilde{\chi}_4}{24n} u^4 + \dots \right). \quad (20)$$

Отсюда видно, что при любом заданном $|u| < \infty$ предел характеристической функции нормированной на \sqrt{n} суммы случайных компонент (18) равен

$$\lim_{n \rightarrow \infty} \Theta_n(u) = \Theta_\infty(u) = \exp\left(-\frac{\sigma_x^2}{2} u^2\right) \quad (21)$$

— характеристической функции гауссовой случайной величины. А значит и плотность вероятностей суммы (18) сходится, при $n \rightarrow \infty$, к гауссовой плотности вероятностей

$$f_\infty(x) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left(-\frac{x^2}{2\sigma_x^2}\right). \quad (22)$$

Это утверждение и составляет содержание Центральной Предельной Теоремы.

Путь, приведший нас к Центральной Предельной Теореме, может показаться кому-то чересчур формальным. Чтобы наглядно убедиться в ее справедливости, дадим несколько примеров сумм случайных величин, и проследим эволюцию их плотностей вероятностей с ростом n .

Пример: Прежде всего отметим, что обсужденные в разделе 2.4 формулы Муавра-Лапласа, описывающие асимптотику биномиального распределения $P_n(k)$ (2.3.6) при $n \rightarrow \infty$, выражают ЦПТ для сумм независимых величин, имеющих распределение Бернулли (2.2.25). Соответственно, графики на Рис. 2.13 и 2.14 иллюстрируют справедливость ЦПТ. То, что биномиально распределенная величина остается дискретной при любом сколь угодно большом n , в то время как предельная гауссова величина непрерывна, не меняет сути дела. Ситуация напоминает отношения между классической и квантовой механикой. Согласно последней, физические величины, к примеру энергия элементарных частиц, принимает лишь дискретный ряд значений. С позиций же классической механики энергия может быть какой угодно. Классическая и квантовая механика прекрасно уживаются в макроскопическом пределе, когда разница между соседними уровнями энергии очень мала, и дискретностью можно пренебречь. Переход от “квантового” к “классическому” описанию, в случае биномиального распределения, осуществляет интегральная формула Муавра-Лапласа (2.4.12). ★

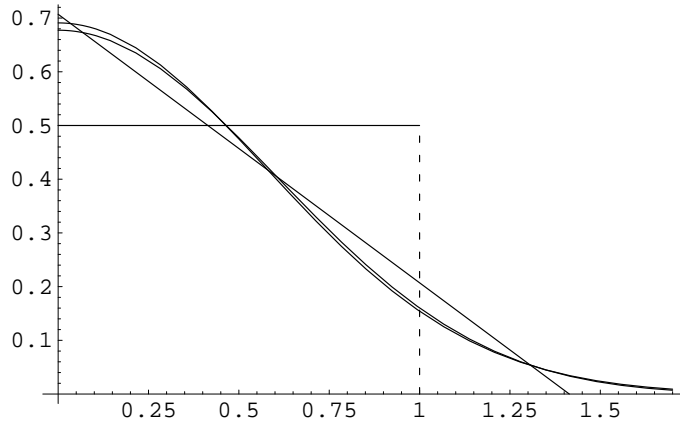


Рис. 3.18

Графики плотностей вероятностей суммы n равномерно распределенных слагаемых при $n = 1$ (прямоугольник), $n = 2$ (треугольник) и $n = 8$. Последняя плотность вероятностей почти повторяет график изображенного здесь же гауссова распределения (24).

Пример: Возьмем сумму равномерно распределенных в интервале $(-1, 1)$ случайных величин. Характеристическая функция их нормированной суммы (18) равна

$$\Theta_z(u) = \left[\frac{\sqrt{n} \sin\left(\frac{u}{\sqrt{n}}\right)}{u} \right]^n. \quad (23)$$

На Рис. 18 изображена плотность вероятностей, полученная численно обратным преобразованием Фурье (23) при $n = 8$, а также предельное гауссово распределение

$$f_\infty(x) = \sqrt{\frac{3}{2\pi}} \exp\left(-\frac{3x^2}{2}\right). \quad (24)$$

Видно, что графики гауссова распределения и плотности вероятностей суммы восьми равномерно распределенных слагаемых практически совпадают. ★

Замечание: Рассуждая на “физическом уровне строгости” оценим число слагаемых n , при котором плотность вероятностей суммы (18) становится практически гауссовой. Заметим, что “тело” плотности вероятностей суммы (18) фактически определяется первым сомножителем в правой части (20). Оно существенно не равно нулю при $|u| \lesssim 1/\sqrt{\varkappa_2}$. Если на границах этого интервала оставшийся сомножитель

$$\exp\left(-i\frac{\varkappa_3}{6\sqrt{n}}u^3 + \frac{\varkappa_4}{24n}u^4 + \dots\right)\Big|_{u=1/\sqrt{\varkappa_2}} \simeq 1, \quad (23)$$

то можно ожидать, что распределение суммы близко к гауссову. Как видно из (20), для этого необходимо, чтобы асимметрия и эксцесс слагаемых подчинялись неравенствам

$$|\kappa_3| \ll 6\sqrt{n}, \quad |\kappa_4| \ll 24n. \quad (24)$$

Во всех приведенных выше примерах эти неравенства с лихвой выполняются уже при $n \sim 10$. ■

Замечание: Мы пришли к утверждению Центральной Предельной Теоремы в простейшем случае суммы одинаково распределенных независимых величин. Можно показать однако, что она справедлива при гораздо более общих предположениях. Грубо говоря, она “работает”, пока дисперсии слагаемых ограничены и не слишком сильно отличаются друг от друга, а статистическая зависимость слагаемых X_m и X_l суммы (16) довольно быстро пропадает с увеличением $|m - l|$. ■

3.7 Гауссовы совокупности

Обсужденная выше Центральная Предельная Теорема высвечивает ключевое место распределенных по Гауссу случайных величин как в самой теории вероятностей, так и в прикладных статистических задачах. Еще большую роль в теории и приложениях играют *гауссовы совокупности*, свойства которых исследуются ниже.

3.7.1 Определение гауссовой совокупности

Прежде чем дать определение гауссовой совокупности, сделаем важное замечание: Возьмем две последовательности независимых величин $\{X_1, \dots, X_n\}$ и $\{Y_1, \dots, Y_n\}$. Из ЦПТ следует, что если суммы

$$\sum_{k=1}^n X_k \xrightarrow[n \rightarrow \infty]{} X, \quad \sum_{k=1}^n Y_k \xrightarrow[n \rightarrow \infty]{} Y,$$

сходятся к гауссовым величинам X и Y , то и сумма

$$\sum_{k=1}^n Z_k \quad (Z_k = aX_k + bY_k)$$

сходится при $n \rightarrow \infty$ к гауссовой величине $Z = aX + bY$. Иными словами, вместе с элементами совокупности $\{X, Y\}$, гауссовой будет любая

их линейная комбинация. Это, вытекающее из ЦПТ, свойство *инвариантности относительно линейных преобразований* берут за определение гауссовых совокупностей. В общем случае набор случайных величин $\{Z_1, \dots, Z_n\}$ называют *гауссовой совокупностью*, если любая их линейная комбинация

$$\sum_{k=1}^n \alpha_k Z_k \quad (1)$$

имеет гауссово распределение.

3.7.2 Необходимые сведения из теории матриц

Изучение гауссовых совокупностей требует знания основ теории матриц. Напомним их. Набор величин $\mathbf{x} = \{x_1, \dots, x_n\}$ называют *вектором*, а прямоугольную таблицу $A = \|a_{ij}\|$, где i номер строки, а j столбца, *матрицей*. Обсудим лишь квадратные матрицы размером $n \times n$. Их примером служит ковариационная матрица \mathcal{C} (4.16). Поменяв местами строки и столбцы, получим *транспонированную матрицу*

$$A = \|a_{ij}\| \quad \Longrightarrow \quad A^T = \|a_{ji}\|. \quad (2)$$

Напомним, ковариационная матрица симметрична, и значит $\mathcal{C} \equiv \mathcal{C}^T$.

Умножение вектора на матрицу определяют равенством

$$\mathbf{y} = A\mathbf{x} \quad \Longleftrightarrow \quad y_i = \sum_{j=1}^n a_{ij} x_j. \quad (3)$$

Аналогично вводят *скалярное произведение*

$$(\mathbf{y} \cdot \mathbf{x}) = \sum_{j=1}^n y_j x_j \quad (4)$$

и операцию *умножения матриц*

$$D = AB \quad \Longleftrightarrow \quad d_{ij} = \sum_{m=1}^n a_{im} b_{mj}. \quad (5)$$

Применяя умножение несколько раз, находят произведение любого числа матриц. Легко доказать, что умножение матриц *ассоциативно*. Для трех матриц это значит, что

$$D E F = (D E) F = D (E F). \quad (6)$$

Если произведение матриц равно *единичной матрице*

$$\mathcal{A}\mathcal{B} = \mathcal{I} = \|\delta_{ij}\| ,$$

то \mathcal{B} называют *обратной* к матрице \mathcal{A} и пользуются обозначением $\mathcal{B} = \mathcal{A}^{-1}$. При этом выполняются соотношения

$$\mathcal{B} = \mathcal{A}^{-1} \iff \mathcal{B}^{-1} = \mathcal{A}, \quad (\mathcal{A}\mathcal{B})^{-1} = \mathcal{B}^{-1}\mathcal{A}^{-1}. \quad (7)$$

Отсюда и из (6) следует, в частности, что

$$(\mathcal{D}\mathcal{E}\mathcal{F})^{-1} = \mathcal{F}^{-1}\mathcal{E}^{-1}\mathcal{D}^{-1}. \quad (8)$$

Квадратной матрице ставят в соответствие число, так называемый *детерминант*, обладающий свойствами:

$$\det(\mathcal{A}\mathcal{B}) = \det \mathcal{A} \det \mathcal{B}, \quad \det(\mathcal{A}^T) = \det \mathcal{A}. \quad (9)$$

Подробнее обсудим свойства ковариационной матрицы \mathcal{C} (4.16). Как и любая положительно определенная матрица, она имеет n ортонормированных *собственных векторов*

$$\{\gamma_1, \gamma_2, \dots, \gamma_n\}, \quad (\gamma_m \cdot \gamma_l) = \delta_{ml}, \quad (10)$$

удовлетворяющих уравнениям

$$\mathcal{C}\gamma_m = \varrho_m^2 \gamma_m, \quad (11)$$

где $\varrho_m^2 \geq 0$ *собственные числа* матрицы \mathcal{C} . Ниже будем считать, что $\det \mathcal{C} > 0$. Тогда собственные числа строго положительны: $\varrho_m^2 > 0$.

Возьмем матрицу

$$\mathcal{G} = \|\gamma_1, \dots, \gamma_n\|, \quad (12)$$

столбцами которой служат собственные векторы (10). Из их ортогональности следует, что \mathcal{G}^T обратна к матрице \mathcal{G} :

$$\mathcal{G}^T \mathcal{G} = \|(\gamma_i \cdot \gamma_j)\| = \|\delta_{ij}\| \implies \mathcal{G}^T = \mathcal{G}^{-1}. \quad (13)$$

Матрица \mathcal{G} , для которой справедливы соотношения (13), называют *ортогональной*. Вместе с \mathcal{G} ортогональна и матрица \mathcal{G}^T , так что

$$\mathcal{G}\mathcal{G}^T = \|\delta_{ij}\|. \quad (13')$$

Известно, что произведение трех матриц

$$\mathcal{G}^T \mathcal{C} \mathcal{G} = \mathcal{L} = \|\varrho_i^2 \delta_{ij}\| \quad (14)$$

дает *диагональную матрицу* \mathcal{L} . Ее боковые элементы равны нулю, а диагональные совпадают с собственными числами ковариационной матрицы. Умножив (14) слева на \mathcal{G} , а справа на \mathcal{G}^T и учитывая свойства ортогональных матриц (13) и (13'), придем к знаменитому представлению ковариационной матрицы в виде произведения диагональной и ортогональных матриц

$$\mathcal{C} = \mathcal{G} \mathcal{L} \mathcal{G}^T. \quad (14')$$

Из (14), (9) и ортогональности матриц \mathcal{G} и \mathcal{G}^T следует, что детерминант матрицы \mathcal{L} равен детерминанту ковариационной матрицы

$$\det \mathcal{C} = \det \mathcal{L} = \prod_{k=1}^n \varrho_k^2. \quad (15)$$

Кроме того из (14) и (8) имеем

$$\mathcal{L}^{-1} = \left\| \frac{\delta_{ij}}{\varrho_i^2} \right\| = \mathcal{G}^T \mathcal{C}^{-1} \mathcal{G} \quad \Longleftrightarrow \quad \mathcal{C}^{-1} = \mathcal{G} \mathcal{L}^{-1} \mathcal{G}^T. \quad (16)$$

Рассмотрим произвольную квадратичную форму

$$\sum_{k=1}^n q_{ij} x_i x_j = (\mathbf{x} \cdot \mathcal{Q} \mathbf{x}). \quad (18)$$

Из определения скалярного произведения (4) и умножения вектора на матрицу (3) следует справедливость равенства

$$(\mathbf{x} \cdot \mathcal{Q} \mathbf{x}) = (\mathcal{Q}^T \mathbf{x} \cdot \mathbf{x}). \quad (19)$$

Пример: Протестируем приведенные формулы на примере матрицы

$$\mathcal{C} = \left\| \begin{array}{ccc} 5 & 3 & 1 \\ 3 & 4 & 2 \\ 1 & 2 & 3 \end{array} \right\|. \quad (20)$$

Найдем ортогональную матрицу \mathcal{G} с помощью программы MATLAB. Для этого сформируем ковариационную матрицу

```
>> C=[5 3 1; 3 4 2; 1 2 3];
```

Затем выполним команду, генерирующую ортогональную матрицу и собственные числа ковариационной матрицы

```
>> [G,L]=eig(C)
```

В итоге получим

$$\mathcal{G} = \begin{vmatrix} 0.4226 & -0.5999 & 0.6793 \\ -0.7461 & 0.1953 & 0.6366 \\ 0.5145 & 0.7759 & 0.3651 \end{vmatrix}, \quad \mathcal{L} = \begin{vmatrix} 0.9213 & 0 & 0 \\ 0 & 2.7302 & 0 \\ 0 & 0 & 8.3485 \end{vmatrix}.$$

Команды

>> $\mathbf{G}*\mathbf{G}'$

и

>> $\mathbf{G}'*\mathbf{G}$

иллюстрируют свойства ортогональности (13) и (13'). Выполнив команду

>> $\mathbf{G}'*\mathbf{C}*\mathbf{G}$

убедимся в справедливости соотношения (14). Команда

>> $\mathbf{G}*\mathbf{L}*\mathbf{G}'$

возвращает нас к матрице (20). Аналогично, выполнив команду

>> $\mathbf{G}'*\text{inv}(\mathbf{C})*\mathbf{G}$

получим, в соответствии с равенством (16), матрицу

$$\mathcal{L}^{-1} = \begin{vmatrix} 1.0854 & 0 & 0 \\ 0 & 0.3663 & 0 \\ 0 & 0 & 0.1198 \end{vmatrix}. \quad \star$$

3.7.3 Плотность вероятностей гауссовой совокупности

Найдем плотность вероятностей гауссовой совокупности $\tilde{\mathbf{Y}}$ с нулевыми средними и заданной ковариационной матрицей \mathbf{C} . Для этого представим $\tilde{\mathbf{Y}}$ в виде линейной комбинации

$$\tilde{\mathbf{Y}} = \mathbf{A}\mathbf{X} \quad (\mathbf{A} = \mathcal{G}) \quad (21)$$

статистически независимых гауссовых случайных величин

$$\mathbf{X} = \{X_1, \dots, X_n\}, \quad \langle X_m X_l \rangle = \varrho_m^2 \delta_{ml} \quad (m, l = 1, 2, \dots, n), \quad (22)$$

ковариационная матрица которых равна диагональной матрице \mathcal{L} (14).

Убедимся вначале, что гауссова совокупность $\tilde{\mathbf{Y}}$ (21) действительно обладает требуемой ковариационной матрицей. Для этого необходимо и достаточно, чтобы выполнялись равенства

$$\langle \tilde{Y}_i \tilde{Y}_j \rangle = \sum_{m=1}^n \sum_{l=1}^n a_{im} a_{jl} \langle X_m X_l \rangle = c_{ij}.$$

Последние, в силу статистической независимости компонент \mathbf{X} , принимают вид

$$c_{ij} = \sum_{m=1}^n a_{im} a_{jm} \varrho_m^2.$$

На матричном языке это значит, что выполнено равенство

$$\mathcal{C} = \mathcal{A} \mathcal{L} \mathcal{A}^T. \quad (23)$$

Подставив сюда $\mathcal{A} = \mathcal{G}$, и вспомнив соотношение (14'), убедимся, что $\widetilde{\mathbf{Y}}$ гауссова совокупность с требуемой ковариационной матрицей.

Приступим к отысканию плотности вероятностей $f_{\widetilde{\mathbf{Y}}}(\mathbf{y})$ гауссовой совокупности $\widetilde{\mathbf{Y}}$ (21). Согласно (3.23)

$$f_{\widetilde{\mathbf{Y}}}(\mathbf{y}) = f_{\mathbf{X}}(\mathcal{B}\mathbf{y}) |\det \mathcal{B}|. \quad (24)$$

где $\mathcal{B} = \mathcal{G}^T$ матрица, обратная к \mathcal{A} (21). В (24) входит плотность вероятностей независимых величин \mathbf{X} (22):

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(\sqrt{2\pi})^n \sqrt{\det \mathcal{C}}} \exp \left[-\frac{1}{2} \sum_{k=1}^n \frac{x_k^2}{\varrho_k^2} \right]. \quad (25)$$

Здесь учтено равенство (15), выражающее произведение дисперсий случайных компонент \mathbf{X} через детерминант ковариационной матрицы. Выразим показатель экспоненты в (25) на языке скалярных произведений

$$\sum_{k=1}^n \frac{x_k^2}{\varrho_k^2} = (\mathbf{x} \cdot \mathcal{L}^{-1} \mathbf{x}).$$

Кроме того, в силу ортогональности матрицы \mathcal{G}^T , имеем $|\det \mathcal{B}| = 1$. Следовательно

$$f_{\widetilde{\mathbf{Y}}}(\mathbf{y}) = f_{\mathbf{X}}(\mathcal{B}\mathbf{y}) = f_{\mathbf{X}}(\mathcal{G}^T \mathbf{y}) = \frac{1}{(\sqrt{2\pi})^n \sqrt{\det \mathcal{C}}} \exp \left[-\frac{1}{2} (\mathcal{G}^T \mathbf{y} \cdot \mathcal{L}^{-1} \mathcal{G}^T \mathbf{y}) \right].$$

Преобразуем входящее сюда скалярное произведение с помощью (19) и второго равенства (16):

$$(\mathcal{G}^T \mathbf{y} \cdot \mathcal{L}^{-1} \mathcal{G}^T \mathbf{y}) = (\mathbf{y} \cdot \mathcal{G} \mathcal{L}^{-1} \mathcal{G}^T \mathbf{y}) = (\mathbf{y} \cdot \mathcal{C}^{-1} \mathbf{y}).$$

Таким образом

$$f_{\widetilde{\mathbf{Y}}}(\mathbf{y}) = \frac{1}{(\sqrt{2\pi})^n \sqrt{\det \mathcal{C}}} \exp \left[-\frac{1}{2} (\mathbf{y} \cdot \mathcal{C}^{-1} \mathbf{y}) \right]. \quad (26)$$

До сих пор мы считали, для удобства выкладок, что средние значения компонент $\tilde{\mathbf{Y}}$ (21) равны нулю. Возьмем теперь самый общий случай гауссовой совокупности $\mathbf{Y} = \tilde{\mathbf{Y}} + \boldsymbol{\mu}$, где $\boldsymbol{\mu} = \langle \mathbf{Y} \rangle$ вектор средних значений. Из (26) и очевидного равенства

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\tilde{\mathbf{Y}}}(\mathbf{y} - \boldsymbol{\mu})$$

имеем окончательно

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{(\sqrt{2\pi})^n \sqrt{\det \mathcal{C}}} \exp \left[-\frac{1}{2} \left((\mathbf{y} - \boldsymbol{\mu}) \cdot \mathcal{C}^{-1} (\mathbf{y} - \boldsymbol{\mu}) \right) \right]. \quad (27)$$

3.7.4 Эллипсоиды рассеяния

Выясним геометрическую структуру плотности вероятностей (27), исследовав *поверхности равного уровня*, где $f_{\mathbf{Y}}(\mathbf{y})$ имеет одинаковые значения. Из (27) видно, что уравнение указанных поверхностей таково

$$\left((\mathbf{y} - \boldsymbol{\mu}) \cdot \mathcal{C}^{-1} (\mathbf{y} - \boldsymbol{\mu}) \right) = \kappa^2. \quad (28)$$

На них $f_{\mathbf{Y}}(\mathbf{y})$ одинакова и равна

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{(\sqrt{2\pi})^n \sqrt{\det \mathcal{C}}} \exp \left(-\frac{\kappa^2}{2} \right). \quad (29)$$

Уравнение (28) записано в декартовой системе координат \mathbf{y} с расположенными вдоль осей базисными векторами

$$\{\mathbf{b}_1, \dots, \mathbf{b}_n\}, \quad \mathbf{b}_k = \underbrace{\{0, \dots, 0\}}_{k-1}, \underbrace{\{1, 0, \dots, 0\}}_{n-k}.$$

Введем новые декартовы координаты. Поместим их центр в точку $\boldsymbol{\mu}$, а затем повернем жестко скрепленные базисные векторы $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$. В итоге получим новый базис $\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$. Связь между старыми $\{y_1, \dots, y_n\}$ и новыми $\{x_1, \dots, x_n\}$ координатами описывается равенствами

$$\mathbf{y} = \mathcal{R}\mathbf{x} + \boldsymbol{\mu} \quad \Longleftrightarrow \quad \mathbf{x} = \mathcal{R}^T(\mathbf{y} - \boldsymbol{\mu}), \quad (30)$$

где \mathcal{R} ортогональная матрица, столбцами которой служат векторы нового базиса $\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$, разложенные по старому базису $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$: $r_{ij} = (\mathbf{b}_i \cdot \mathbf{a}_j)$. Подставив (30) в (28), запишем уравнение поверхности равного уровня в новой системе координат

$$\left(\mathcal{R}\mathbf{x} \cdot \mathcal{C}^{-1} \mathcal{R}\mathbf{x} \right) = \kappa^2 \quad \Longrightarrow \quad \left(\mathbf{x} \cdot \mathcal{R}^T \mathcal{C}^{-1} \mathcal{R}\mathbf{x} \right) = \kappa^2. \quad (31)$$

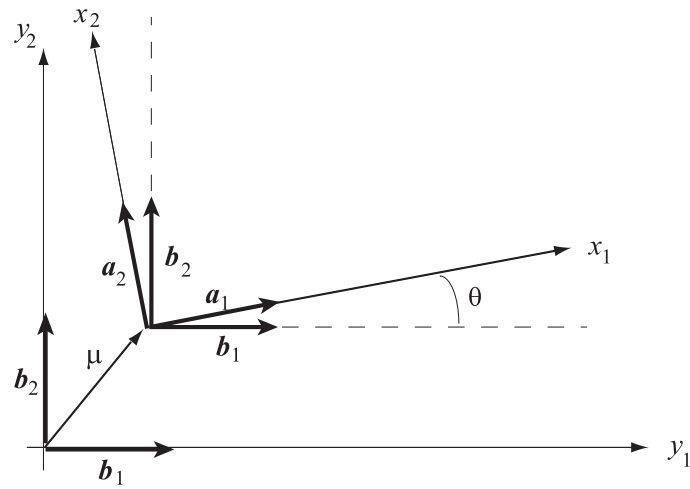


Рис. 3.19

Переход от старой $\{y_1, y_2\}$ к новой $\{x_1, x_2\}$ декартовой системе координат, повернутой относительно старой на угол θ .

Пример: Проиллюстрируем переход от старых к новым координатам в наиболее наглядном случае плоскости ($n = 2$). Старая \mathbf{y} и повернутая относительно нее на угол θ новая система координат \mathbf{x} , вместе с базисными векторами, изображены на Рис. 19. Ортогональная матрица перехода от старых координат к новым полностью определяется единственным в данном случае углом поворота θ и равна

$$\mathcal{R}(\theta) = \|\mathbf{a}_1, \mathbf{a}_2\| = \begin{vmatrix} (\mathbf{b}_1 \cdot \mathbf{a}_1) & (\mathbf{b}_1 \cdot \mathbf{a}_2) \\ (\mathbf{b}_2 \cdot \mathbf{a}_1) & (\mathbf{b}_2 \cdot \mathbf{a}_2) \end{vmatrix} = \begin{vmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{vmatrix}. \quad (32)$$

Советуем читателю проверить геометрически очевидное свойство матрицы $\mathcal{R}(\theta)$

$$\mathcal{R}(\theta_1) \mathcal{R}(\theta_2) = \mathcal{R}(\theta_1 + \theta_2),$$

а также справедливость общих свойств ортогональных матриц, в двумерном случае сводящихся к равенствам

$$\mathcal{R}^{-1}(\theta) = \mathcal{R}(-\theta) = \mathcal{R}^T(\theta). \quad \star$$

Положив в (31) $\mathcal{R} = \mathcal{G}$, где \mathcal{G} ортогональная матрица (12), составленная из собственных векторов ковариационной матрицы, и используя первое соотношение (16), преобразуем уравнение (31) к виду

$$(\mathbf{x} \cdot \mathcal{L}^{-1} \mathbf{x}) = \frac{x_1^2}{\varrho_1^2} + \frac{x_2^2}{\varrho_2^2} + \dots + \frac{x_n^2}{\varrho_n^2} = \kappa^2. \quad (32)$$

Если, как мы условились, все $\varrho_i^2 > 0$, то поверхности равного уровня имеют вид вложенных друг в друга подобных эллипсоидов с длинами полуосей $\kappa\varrho_i$. Их называют *эллипсоидами рассеяния*.

Замечание: Приведенные построения раскрывают геометрический смысл ортогональной матрицы \mathcal{G} (12): Она поворачивает декартову систему координат $\mathbf{y} \mapsto \mathbf{x}$ так, что новые оси совпадают с осями симметрии эллипсоидов рассеяния (32). Очевиден также геометрический смысл собственных чисел ковариационной матрицы. Их корни $\{\varrho_1, \dots, \varrho_n\}$ определяют эффективную ширину плотности вероятностей $f_{\mathbf{Y}}(\mathbf{y})$ в направлениях осей симметрии эллипсоидов рассеяния. В связи с этим эллипсоид симметрии, отвечающий значению $\kappa = 1$, иногда называют *главным эллипсоидом рассеяния*. Полезно также иметь ввиду статистическую интерпретацию собственных чисел ковариационной матрицы. Она исходит из того, что ортогональное преобразование

$$\mathbf{Y} = \mathcal{G} \mathbf{X} + \boldsymbol{\mu} \quad (33)$$

выражает произвольную гауссову совокупность \mathbf{Y} через независимые гауссовы величины \mathbf{X} с дисперсиями $\{\varrho_1^2, \dots, \varrho_n^2\}$. Поэтому $\{\varrho_1, \dots, \varrho_n\}$ называют еще *главными среднеквадратичными отклонениями плотности вероятностей $f_{\mathbf{Y}}(\mathbf{y})$* . ■

Напомним, при анализе одной гауссовой величины часто интересуются вероятностью попадания ее значений в симметрично расположенный относительно среднего μ интервал $(\mu - s, \mu + s)$, где s полуширина интервала. В частности, при $s = 3\sigma$ приходят к знаменитому правилу трех σ , обсужденному в разделе 2.5.1. Внутренность эллипсоидов рассеяния служит естественным обобщением упомянутых интервалов на случай n -мерных гауссовых совокупностей. Найдем вероятность попадания внутрь эллипсоида рассеяния с заданной κ . Заметим, что правая часть выражения (25) есть исследуемая плотность вероятностей $f_{\mathbf{y}}(\mathbf{y})$, записанная в системе координат, оси которой совпадают с осями симметрии эллипсоида рассеяния. Следовательно, искомая вероятность равна

$$P(\kappa) = \frac{1}{(\sqrt{2\pi})^n \prod_{k=1}^n \varrho_k} \int_{\mathcal{E}_\kappa} \exp \left[-\frac{1}{2} \sum_{k=1}^n \frac{x_k^2}{\varrho_k^2} \right] dx^n.$$

Здесь \mathcal{E}_κ внутренность эллипсоида. Сделав замену переменных

$$z_1 = \frac{x_1}{\varrho_1}, \quad \dots \quad z_n = \frac{x_n}{\varrho_n}$$

перейдем к интегрированию по внутренности сферы

$$\mathcal{S}_\kappa : \quad z_1^2 + z_2^2 + \dots + z_n^2 \leq \kappa^2$$

в n -мерном пространстве с декартовыми координатами $\{z_1, \dots, z_n\}$:

$$P(\kappa) = \frac{1}{(\sqrt{2\pi})^n} \int_{S_\kappa} \exp\left[-\frac{1}{2}(z_1^2 + \dots + z_n^2)\right] dz^n.$$

В силу сферической симметрии подынтегральной функции, последний интеграл сводится к однократному

$$P(\kappa) = \frac{1}{(\sqrt{2\pi})^n} \int_0^\kappa \exp\left(-\frac{r^2}{2}\right) S_n(r) dr,$$

где $S_n(r)$ площадь n -мерной сферы радиуса r , заданная равенством (2.14). Подставив ее в последнее выражение, получим окончательно

$$P(\kappa) = \frac{2^{(2-n)/2}}{\Gamma(n/2)} \int_0^\kappa \exp\left(-\frac{r^2}{2}\right) r^{n-1} dr. \quad (34)$$

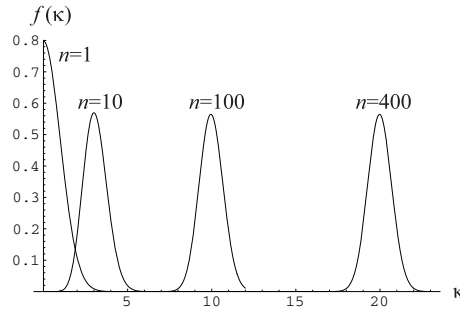


Рис. 3.20

Графики распределения $f(\kappa)$. Видно, что чем больше n , тем меньше относительная (отнесенная к размеру κ_n) ширина $f(\kappa)$.

Замечание: Заметим, что параметр κ в (34) задает линейные размеры эллипсоида рассеяния. Дифференцируя (34) по κ , получим функцию

$$f(\kappa) = \frac{2^{(2-n)/2}}{\Gamma(n/2)} \kappa^{n-1} \exp\left(-\frac{\kappa^2}{2}\right).$$

Ее естественно трактовать как распределение вероятностной массы гауссовой совокупности \mathbf{Y} вдоль оси линейных размеров эллипсоидов рассеяния. В самом деле, $f(\kappa) d\kappa$ равно вероятностной массе, заключенной между эллипсоидами с размерами κ и $\kappa + d\kappa$. На Рис. 20 даны графики $f(\kappa)$ для разных n . Видно, что при $n \gg 1$ вероятностная масса сосредоточена в окрестности эллипсоида размером $\kappa_n \approx \sqrt{n}$. Этот факт

широко используют в статистической физике, где n невообразимо велико (пропорционально числу молекул вещества). При этом, как правило, пренебрегают шириной $f(\kappa)$ и заменяют ее дельта-функцией

$$f(\kappa) \approx \delta(\kappa - \kappa_n).$$

Эффект концентрации вероятностной массы в малой окрестности некоторого эллипсоида обусловлен конкуренцией, описанного в разделе 2.3, “эффекта кожуры апельсина” и экспоненциально быстрого спада к нулю гауссовых распределений. ■

3.7.5 Двумерное гауссово распределение

Из всех многомерных гауссовых плотностей вероятностей в приложениях наиболее употребительна двумерная плотность вероятностей. Изучим ее подробнее. Пусть $\{X, Y\}$ гауссова совокупность, статистические свойства которой полностью определены пятью моментами

$$\langle X \rangle = \mu_x, \quad \langle Y \rangle = \mu_y, \quad \langle \tilde{X}^2 \rangle = \sigma_x^2, \quad \langle \tilde{Y}^2 \rangle = \sigma_y^2, \quad \langle \tilde{X}\tilde{Y} \rangle = \sigma_x\sigma_y\rho. \quad (35)$$

Последние три составляют элементы ковариационной матрицы

$$\mathcal{C} = \begin{vmatrix} \sigma_x^2 & \sigma_x\sigma_y\rho \\ \sigma_x\sigma_y\rho & \sigma_y^2 \end{vmatrix} \implies \det \mathcal{C} = \sigma_x^2\sigma_y^2(1 - \rho^2). \quad (36)$$

Соответственно обратная матрица такова

$$\mathcal{C}^{-1} = \begin{vmatrix} \frac{1}{\sigma_x^2(1 - \rho^2)} & \frac{-\rho}{\sigma_x\sigma_y(1 - \rho^2)} \\ \frac{-\rho}{\sigma_x\sigma_y(1 - \rho^2)} & \frac{1}{\sigma_y^2(1 - \rho^2)} \end{vmatrix}.$$

Подставив ее в двумерный вариант выражения (27), будем иметь:

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1 - \rho^2}} \times \exp \left[-\frac{1}{2(1 - \rho^2)} \left(\frac{(x - \mu_x)^2}{\sigma_x^2} - \frac{2\rho(x - \mu_x)(y - \mu_y)}{\sigma_x\sigma_y} + \frac{(y - \mu_y)^2}{\sigma_y^2} \right) \right]. \quad (37)$$

Если коэффициент корреляции ρ равен нулю, то совместная плотность вероятностей (32) распадается на произведение плотностей вероятностей величин X и Y

$$f(x, y) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp \left[-\frac{(x - \mu_x)^2}{2\sigma_x^2} \right] \times \frac{1}{\sqrt{2\pi}\sigma_y} \exp \left[-\frac{(y - \mu_y)^2}{2\sigma_y^2} \right].$$

Мы установили замечательный факт, сильно облегчающий анализ свойств гауссовых совокупностей: *Некоррелированность компонент гауссовой совокупности равносильна их статистической независимости.*

Дальнейший анализ статистических свойств двумерной гауссовой совокупности проведем, положив средние ее компонент равными нулю: $\mu_x = \mu_y = 0$. Это допущение сильно облегчит выкладки. С другой стороны, при необходимости, легко восстановить μ_x и μ_y в их правах, заменив в совместной плотности вероятностей

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp \left[-\frac{1}{2(1-\rho^2)} \left(\frac{x^2}{\sigma_x^2} - \frac{2\rho xy}{\sigma_x\sigma_y} + \frac{y^2}{\sigma_y^2} \right) \right] \quad (38)$$

и других подобных соотношениях x на $x - \mu_x$ и y на $y - \mu_y$.

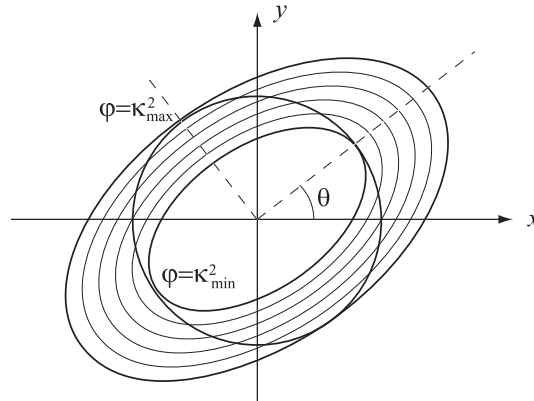


Рис. 3.21

Иллюстрация к определению ориентации эллипса рассеяния: Квадратичная форма φ (39) имеет на окружности те же значения κ^2 , что и пересекающие окружность эллипсы рассеяния. При этом минимум достигается в точках касания окружности и внутреннего, а максимум – внешнего эллипса.

Найдем ориентацию эллипсов рассеяния плотности вероятностей (38). Иначе говоря, определим угол θ , под которым одна из осей симметрии наклонена к оси x . Отыщем θ , пользуясь геометрически наглядными построениями. Введем квадратичную форму

$$\varphi(x, y) = \frac{1}{(1-\rho^2)} \left(\frac{x^2}{\sigma_x^2} - \frac{2\rho xy}{\sigma_x\sigma_y} + \frac{y^2}{\sigma_y^2} \right), \quad (39)$$

принимаящую на эллипсах рассеяния одинаковые значения κ^2 , и изучим ее поведение вдоль единичной окружности ($x = \cos \theta$, $y = \sin \theta$).

Из Рис. 21 видно, что минимальное κ_{\min}^2 и максимальное κ_{\max}^2 значения $\varphi(x, y)$ достигаются при углах θ наклона к оси x осей симметрии эллипсов рассеяния. Геометрические соображения подсказывают также, что

$$\kappa_{\min}^2 = \frac{1}{\varrho_{\max}^2}, \quad \kappa_{\max}^2 = \frac{1}{\varrho_{\min}^2}. \quad (40)$$

Найдем угол θ наклона осей симметрии, исследовав функцию

$$\psi(\theta) = (1 - \rho^2) \varphi(\cos \theta, \sin \theta)$$

на экстремум по θ . Запишем $\psi(\theta)$ в явном виде:

$$\psi(\theta) = \left(\frac{1}{\sigma_x^2} + \frac{1}{\sigma_y^2} \right) + \left(\frac{1}{\sigma_x^2} - \frac{1}{\sigma_y^2} \right) \cos 2\theta - \frac{\rho}{\sigma_x \sigma_y} \sin 2\theta.$$

Дифференцируя ее по θ и приравняв производную нулю, получим уравнение углов θ наклона осей симметрии эллипсов рассеяния

$$\tan 2\theta = \frac{2\rho\sigma_x\sigma_y}{\sigma_x^2 - \sigma_y^2}. \quad (41)$$

Подробнее обсудим случай равных дисперсий компонент $\{X, Y\}$ гауссовой совокупности: $\sigma_x = \sigma_y = \sigma$. Тогда правая часть равенства (41) обращается в бесконечность. Это означает всего лишь, что оси симметрии развернуты под углом в 45° к осям x и y . Найдем соответствующие собственные числа ковариационной матрицы. Подставив в (39)

$$\{x, y\} = \left\{ \pm \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right\}, \quad \sigma_x = \sigma_y = \sigma,$$

и учитывая соотношения (40), будем иметь

$$\varrho_{\min}^2 = \sigma^2(1 - \rho), \quad \varrho_{\max}^2 = \sigma^2(1 + \rho). \quad (42)$$

Как и следовало ожидать, с ростом модуля коэффициента корреляции $|\rho|$ эллипсы рассеяния становятся все более сплюснутыми.

Напомним, главным эллипсом рассеяния называют эллипс с полуосями $(\varrho_{\min}, \varrho_{\max})$. Наряду с ним вводят понятие *полного эллипса рассеяния*, внутрь которого попадают практически все точки с координатами (X, Y) . Укажем эллипс, достойный звания полного эллипса рассеяния. Для этого заметим, что, согласно (34), вероятность попадания в эллипс с полуосями $(\kappa\varrho_{\min}, \kappa\varrho_{\max})$ равна

$$P(\kappa) = 1 - \exp\left(-\frac{\kappa^2}{2}\right) \quad (n = 2). \quad (43)$$

Положив здесь $\kappa = 3$, получим вероятность

$$P(3) \simeq 0.9889. \quad (44)$$

Она близка к единице, а значит на двумерный случай уместно распространить правило трех σ : *Почти всегда точки (X, Y) попадают внутрь полного эллипса рассеяния с полуосями $(3\sigma_{\min}, 3\sigma_{\max})$.*

Замечание: Мы обосновали правило трех σ для пары случайных величин тем, что вероятность (44) попадания в полный эллипс близка к единице, как и вероятность $P \simeq 0.9973$ одной гауссовой величине отклониться от среднего на расстояние, меньшее 3σ . Все же двумерное правило трех σ обладает гораздо меньшим “запасом прочности”. В самом деле, противоположная вероятность при $n = 2$ примерно в четыре раза больше аналогичной вероятности для стандартного правила трех σ . Это значит, что экспериментатор, измеряющий гауссову совокупность $\{X, Y\}$, в четыре раза чаще столкнется с нарушением правила трех σ , чем измеряющий одну гауссову величину. ■

Замечание: Очевидно, правило трех σ не распространяется на случай больших n , где, как отмечено выше, вероятностная масса сосредоточена в малой окрестности эллипса с линейным размером $\kappa \simeq \sqrt{n}$. ■

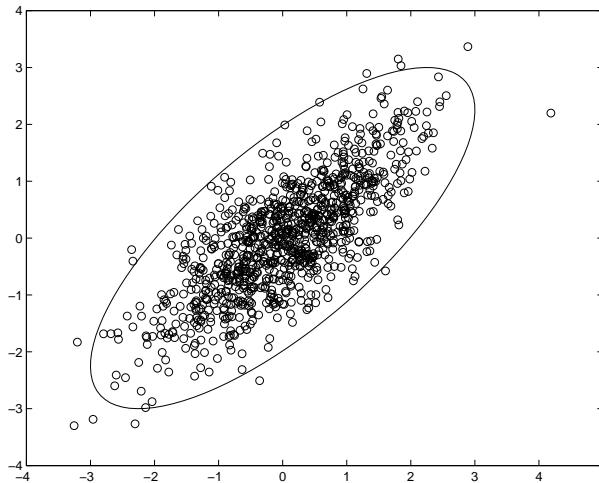


Рис. 3.22

Тысяча реализаций пары гауссовых величин $\{X, Y\}$ с коэффициентом корреляции $\rho = 3/4$. Здесь же изображен полный эллипс рассеяния. Вне его лежит 13 точек, что вполне согласуется с вероятностью $Q \simeq 0.011$ не попасть в полный эллипс рассеяния.

Пример: С помощью программы MATLAB смоделируем тысячу пар X, Y значений гауссовой совокупности случайных величин с нулевым средним, единичной дисперсией и коэффициентом корреляции $\rho = 3/4$. Введем коэффициент корреляции, диагональный элемент ортогональной матрицы \mathcal{G} и корни собственных чисел ковариационной матрицы:

```
>> r=3/4;
>> g=1/sqrt(2); a=sqrt(1+r); b=sqrt(1-r);
```

Затем сформируем реализации независимых гауссовых величин

```
>> u=a*randn(1,1000); v=b*randn(1,1000);
```

и образуем из них реализации исследуемых величин $\{X, Y\}$

```
>> x=g*(u+v); y=g*(u-v);
```

Изобразим результат графически

```
>> plot(x,y,'o')
```

Нанесем на тот же график полный эллипс рассеяния:

```
>> hold on
>> t=linspace(0,2*pi);
>> c=a*cos(t); s=b*sin(t);
>> x=g*(c+s); y=g*(c-s);
>> plot(3*x,3*y)
```

Результат моделирования представлен на Рис. 22. ★

3.7.6 Условные гауссовы распределения

Помимо обычных плотностей вероятностей, в приложениях важно знать условные плотности вероятностей. Например $f_X(x|y)$ – плотность вероятностей X при условии, что другая случайная величина Y примет в испытании значение y . Регулярный способ нахождения условных плотностей вероятностей опирается на общую формулу (1.17). В случае гауссовых совокупностей, однако, удобно находить их, пользуясь эквивалентностью независимости и некоррелированности компонент гауссовых совокупностей. Покажем это на примере гауссовой совокупности $\{X, Y\}$. Для большей наглядности выкладок положим пока их средние равными нулю. Представим X и Y в форме

$$X = aU + bV, \quad Y = cV, \quad (45)$$

где U и V независимые гауссовы величины с нулевыми средними и единичными дисперсиями. Коэффициенты (a, b, c) в (45) определим из усло-

вия, что $\{X, Y\}$ обладает ковариационной матрицей (36). Для этого должны выполняться равенства

$$\sigma_x^2 = a^2 + b^2, \quad \sigma_y^2 = c^2, \quad \sigma_x \sigma_y \rho = bc.$$

Отсюда имеем

$$c = \sigma_y, \quad b = \sigma_x \rho, \quad a = \sigma_x \sqrt{1 - \rho^2}. \quad (46)$$

Найдем статистические свойства X при условии, что $Y = y$. Из второго равенства (45) и из (46) видно, что оно эквивалентно условию $V = y/\sigma_y$. Подставив правую часть этого равенства и коэффициенты a, b (46) в первое равенство (45), будем иметь

$$X|_y = \sigma_x \sqrt{1 - \rho^2} U + \frac{\sigma_x}{\sigma_y} \rho y.$$

Индекс ' $|y$ ' означает, что выражение справа описывает свойства X при условии $Y = y \iff V = y/c$. Из независимости U и V следует, что U , как и прежде, гауссова с нулевым средним и единичной дисперсией. Следовательно, $X|_y$ также гауссова со средним и дисперсией

$$\langle X \rangle|_y = \mu_{x|y} = \frac{\sigma_x}{\sigma_y} \rho y, \quad \sigma_{x|y}^2 = \sigma_x^2 (1 - \rho^2). \quad (47)$$

Их называют *условным средним* и *условной дисперсией* случайной величины X . Подставив (47) в гауссово распределение (2.5.1), найдем искомую условную плотность вероятностей

$$f_X(x|y) = \frac{1}{\sqrt{2\pi(1 - \rho^2)}\sigma_x} \exp \left[-\frac{1}{2(1 - \rho^2)} \left(x - \frac{\sigma_x}{\sigma_y} \rho y \right)^2 \right]. \quad (48)$$

Пусть теперь X и Y имеют ненулевые средние μ_x и μ_y . Чтобы получить выражение для условной вероятности в этом общем случае, достаточно заменить в (48) (x, y) на $(x - \mu_x, y - \mu_y)$. В итоге имеем

$$f_X(x|y) = \frac{1}{\sqrt{2\pi(1 - \rho^2)}\sigma_x} \times \exp \left[-\frac{1}{2(1 - \rho^2)\sigma_x} \left(x - \mu_x - \frac{\sigma_x}{\sigma_y} \rho (y - \mu_y) \right)^2 \right]. \quad (49)$$

Поменяв x на y , придем к условной плотности вероятностей

$$f_Y(y|x) = \frac{1}{\sqrt{2\pi(1 - \rho^2)}\sigma_y} \times \exp \left[-\frac{1}{2(1 - \rho^2)\sigma_y} \left(y - \mu_y - \frac{\sigma_y}{\sigma_x} \rho (x - \mu_x) \right)^2 \right]. \quad (50)$$

Замечание: Из (50) видно, что условное среднее случайной величины Y , при условии, что $X = x$, равно

$$\mu_{y|x} = \mu_y + \frac{\sigma_y}{\sigma_x} \rho (x - \mu_x).$$

С точностью до замены $\mu_{y|x}$ на Y' и x на X , это равенство совпадает с уравнением линии регрессии (4.10). Таким образом, мы раскрыли еще одно свойство линии регрессии: Для гауссовых совокупностей она повторяет условное среднее. ■

При $\rho = 0$ условные плотности вероятностей (49), (50) вырождаются в обычные одномерные плотности вероятностей гауссовых величин X и Y , а с ростом $|\rho|$ все более концентрируются вокруг условных средних. Это свойство условных плотностей вероятностей иллюстрирует Рис. 23.

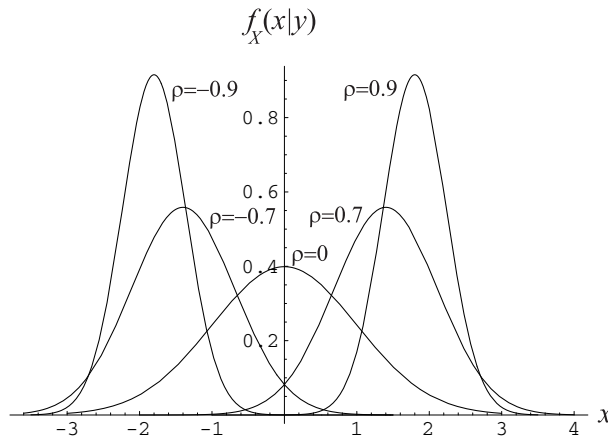


Рис. 3.23

Графики условной плотности вероятностей (48) при $\sigma_x = \sigma_y = 1$, $y = 2$ и разных значениях коэффициента корреляции.

При $|\rho| \rightarrow 1$ условные плотности вероятностей слабо сходятся к дельта-функции. К примеру

$$\lim_{\rho \rightarrow 1} f_Y(y|x) = \delta \left(y - \mu_y - \frac{\sigma_y}{\sigma_x} (x - \mu_x) \right).$$

Помножив обе части равенства на $f_X(x)$, найдем предельное сингулярное выражение для совместной плотности вероятностей

$$\lim_{\rho \rightarrow 1} f(x, y) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp \left(-\frac{(x - \mu_x)^2}{2\sigma_x^2} \right) \delta \left(y - \mu_y - \frac{\sigma_y}{\sigma_x} (x - \mu_x) \right).$$

Пользуясь свойством дельта-функции (2.24), его можно переписать в другой эквивалентной форме

$$\lim_{\rho \rightarrow 1} f(x, y) = \frac{1}{\sqrt{2\pi}\sigma_y} \exp\left(-\frac{(y - \mu_y)^2}{2\sigma_y^2}\right) \delta\left(x - \mu_x - \frac{\sigma_x}{\sigma_y}(y - \mu_y)\right).$$

Замечание: Равенства (45), (46) представляют собой частный случай применения разложения Холецкого (Cholesky factorization). Так называют разложение ковариационной матрицы

$$\mathcal{C} = \mathcal{T}^T \mathcal{T} \quad (51)$$

на верхнюю \mathcal{T} и нижнюю \mathcal{T}^T треугольные матрицы. Известно, что для любой положительно определенной матрицы \mathcal{C} существует единственное разложение вида (51). Рассмотрим равенство

$$\mathbf{Y} = \mathcal{T}^T \mathbf{U} + \boldsymbol{\mu}, \quad (52)$$

где $\boldsymbol{\mu}$ вектор средних значений гауссовой совокупности \mathbf{Y} , а \mathbf{U} независимые гауссовы величины с единичными дисперсиями. При этом элементы ковариационной матрицы совокупности \mathbf{Y} равны:

$$\langle \tilde{Y}_i \tilde{Y}_j \rangle = \sum_m \sum_{l=1}^n \tau_{mi} \tau_{lj} \langle U_m U_l \rangle = \sum_{m=1}^n \tau_{mi} \tau_{mj}.$$

На матричном языке последнее равенство означает, что ковариационная матрица совокупности \mathbf{Y} равна правой части равенства (51), то есть заданной ковариационной матрице. Таким образом, (52) есть еще одно, альтернативное к (6.33), представление гауссовой совокупности в виде линейной комбинации независимых гауссовых величин. ■

Пример: В программе MATLAB предусмотрена операция вычисления треугольной матрицы \mathcal{T} по заданной матрице \mathcal{C} . Пользуясь соответствующей командой MATLAB, найдем разложение Холецкого ковариационной матрицы (20). Выполним команды

```
>> C=[5 3 1; 3 4 2; 1 2 3];
>> T=chol(C)
```

В итоге получим

$$\mathcal{T} = \begin{vmatrix} 2.2361 & 1.3416 & 0.4472 \\ 0 & 1.4832 & 0.9439 \\ 0 & 0 & 1.3817 \end{vmatrix}.$$

Убедимся в справедливости равенства (51):

```
>> T'*T
```

Читатель легко проверит, что эта команда возвращает к матрице \mathcal{C} . ★

3.7.7 Характеристическая функция гауссовой совокупности

Найдем характеристическую функцию n -мерной гауссовой совокупности \mathbf{Y} с заданными средними $\langle \mathbf{Y} \rangle = \boldsymbol{\mu}$ и ковариационной матрицей \mathcal{C} . Для этого введем вспомогательную гауссову величину $Z = (\mathbf{u} \cdot \mathbf{Y})$, где \mathbf{u} произвольный детерминированный n -мерный вектор. Искомая характеристическая функция равна

$$\Theta_{\mathbf{Y}}(\mathbf{u}) = \langle e^{i(\mathbf{u} \cdot \mathbf{Y})} \rangle = \langle e^{iZ} \rangle .$$

С другой стороны, в силу гауссовости Z , имеем

$$\langle e^{iZ} \rangle = e^{i(\mathbf{u} \cdot \boldsymbol{\mu})} \exp\left(-\frac{1}{2}D[Z]\right) .$$

Подставив сюда, записанный в матричных обозначениях, аналог равенства (4.15)

$$D[Z] = (\mathbf{u} \cdot \mathcal{C} \mathbf{u}) ,$$

получим окончательно

$$\Theta_{\mathbf{Y}}(\mathbf{u}) = \langle e^{iZ} \rangle = \exp\left[i(\boldsymbol{\mu} \cdot \mathbf{u}) - \frac{1}{2}(\mathbf{u} \cdot \mathcal{C} \mathbf{u})\right] . \quad (53)$$

В частности, двумерная характеристическая функция

$$\Theta(u, v) = \langle e^{iuX+ivY} \rangle \quad (54)$$

гауссовой совокупности $\{X, Y\}$ со средними $\langle X \rangle = \mu_x$, $\langle Y \rangle = \mu_y$ и ковариационной матрицей (36) равна

$$\Theta(u, v) = \exp\left[i\mu_x u + i\mu_y v - \frac{1}{2}(\sigma_x^2 u^2 + \sigma_y^2 v^2 + 2\sigma_x \sigma_y \rho uv)\right] . \quad (55)$$

В типичном для приложений случае нулевых средних и одинаковых дисперсий, (55) принимает особенно простой вид

$$\Theta(u, v) = \exp\left[-\frac{\sigma^2}{2}(u^2 + v^2 + 2\rho uv)\right] . \quad (56)$$

Как уже отмечалось, характеристические функции служат удобным инструментом анализа свойств случайных величин. Применительно же к гауссовым совокупностям, аппарат характеристических функций способен буквально творить чудеса, помогая вычислять статистические характеристики, которые трудно извлечь непосредственно из плотностей вероятностей. Ниже разобрано несколько примеров, иллюстрирующих некоторые возможности аппарата характеристических функций.

Пример: Пусть $\{X, Y\}$ гауссова совокупность с нулевыми средними и одинаковыми дисперсиями σ^2 . Найдем плотность вероятностей разности $Z = X - Y$. При этом нет необходимости вычислять интегралы типа (2.19), а достаточно положить в (56) $v = -u$. В итоге получим характеристическую функцию гауссовой величины

$$\Theta_Z(u) = \exp\left[-\sigma^2(1 - \rho)u^2\right]$$

с плотностью вероятностей

$$f_Z(z) = \frac{1}{2\sqrt{\pi\sigma^2(1 - \rho)}} \exp\left[-\frac{z^2}{4\sigma^2(1 - \rho)}\right].$$

При $\rho \rightarrow 1$ она слабо сходится к дельта-функции $\delta(z)$, а при $\rho = -1$ имеет максимальную дисперсию $\sigma_z^2 = 2\sqrt{2}\sigma^2$. Последнее означает, что разность антикоррелированных случайных величин только увеличивает случайный разброс. ★

Пример: В разделе (2.4) найдена плотность вероятностей (2.26) произведения $Z = XY$ гауссовых величин. Аппарат характеристических функций позволяет сравнительно легко решать подобные задачи. Возьмем, к примеру, случайную величину

$$Z = XU + YV, \quad (58)$$

где $\{X, Y, U, V\}$ гауссова совокупность независимых величин с нулевыми средними и одинаковой дисперсией σ^2 . Найдем вначале условную характеристическую функцию Z , полагая X и Y заданными. При этом $Z|_{XY}$ оказывается линейной комбинацией оставшихся случайных величин $\{U, V\}$ с гауссовой характеристической функцией

$$\Theta_{Z|_{X,Y}}(u) = \langle e^{iuZ|_{XY}} \rangle = \exp\left[-\frac{\sigma^2 u^2}{2} (X^2 + Y^2)\right].$$

Усредним ее по X и Y . В силу независимости X и Y среднее распадается на произведение одинаковых средних. Вычислим одно из них

$$\left\langle \exp\left(-\frac{\sigma^2 u^2}{2} X^2\right) \right\rangle = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2} \left(\sigma^2 u^2 + \frac{1}{\sigma^2}\right) x^2\right].$$

Последний интеграл равен

$$\left\langle \exp\left(-\frac{\sigma^2 u^2}{2} X^2\right) \right\rangle = \sqrt{\frac{1}{1 + \sigma^4 u^2}}. \quad (59)$$

Возведя правую часть равенства в квадрат, найдем искомую характеристическую функцию случайной величины Z (58):

$$\Theta_Z(u) = \frac{1}{1 + \sigma^4 u^2}.$$

Отсюда, с помощью обратного преобразования Фурье, имеем

$$f_Z(z) = \frac{1}{2\sigma^2} \exp\left(-\frac{|z|}{\sigma^2}\right) \quad (60)$$

известное *распределение Лапласа*. Построим его гистограмму, численно смоделировав 50,000 реализаций случайной величины Z (58):

```
>> XU=prod(randn(2,50000)); YV=prod(randn(2,50000));  
>> Z=XU+YV;
```

Построим гистограмму массива реализаций Z :

```
>> z=linspace(-7,7);  
>> hist(Z,z)
```

Полученная гистограмма изображена на Рис. 24. ★

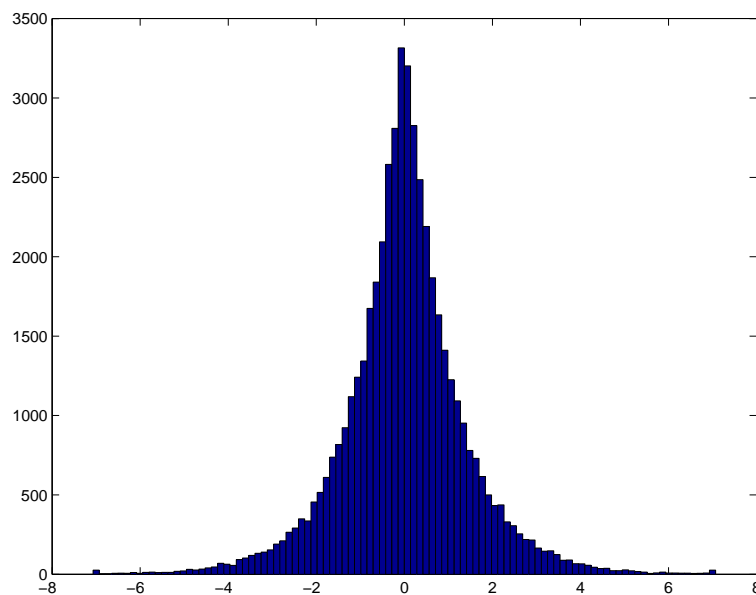


Рис. 3.24

Гистограмма распределения Лапласа, построенная численным моделированием случайной величины Z (58).

Замечание : Заметим, что в качестве побочного продукта мы нашли характеристическую функцию (59) плотности вероятностей (2.26). ■

В приложениях важно уметь вычислять корреляции функций случайных величин

$$B(\rho) = \langle g(X)g(Y) \rangle = \iint_{-\infty}^{\infty} g(x)g(y)f(x, y) dx dy, \quad (61)$$

где $g(x)$ заданная функция. Выведем, опираясь на характеристическую функцию (56), полезное соотношение для $B(\rho)$. Дифференцируя (56) n раз по ρ , будем иметь

$$\frac{\partial^n}{\partial \rho^n} \Theta(u, v) = (-1)^n (uv)^n \sigma^{2n} \Theta(u, v).$$

Обращая это равенство по Фурье, запишем его на языке плотности вероятностей $f(x, y)$:

$$\frac{\partial^n f}{\partial \rho^n} = \sigma^{2n} \frac{\partial^{2n} f}{\partial x^n \partial y^n}. \quad (62)$$

Дифференцируя равенство (61) n раз по ρ , с учетом (62) получим

$$\frac{d^n B(\rho)}{d\rho^n} = \sigma^{2n} \iint_{-\infty}^{\infty} g(x)g(y) \frac{\partial^{2n} f}{\partial x^n \partial y^n} dx dy.$$

После интегрирования по частям, придем к уравнению относительно среднего (61)

$$\frac{d^n B(\rho)}{d\rho^n} = \sigma^{2n} \langle g^{(n)}(x)g^{(n)}(y) \rangle. \quad (63)$$

Дадим несколько примеров использования соотношения (63).

Пример: Пусть $g(x) = x$. Взяв $n = 1$, из (63) имеем:

$$\frac{dB(\rho)}{d\rho} = \sigma^2. \quad (64)$$

Это простейшее дифференциальное уравнение относительно $B(\rho)$. Найдем единственное интересующее нас решение уравнения (64), дополнив его “начальным” условием. Последнее получим, заметив, что при $\rho = 0$ величины X и Y статистически независимы, а искомое среднее равно

$$B(\rho)|_{\rho=0} = \langle X \rangle \langle Y \rangle = 0.$$

Отсюда и из (64) имеем

$$B(\rho) = \sigma^2 \rho. \quad \star$$

Ответ очевиден. Сам пример был нужен для демонстрации техники работы с соотношением (63). Возьмем пример посложнее:

Пример: Пусть $g(x) = x^2$. Здесь естественно положить $n = 2$:

$$\frac{d^2 B(\rho)}{d\rho^2} = 4\sigma^4.$$

Дополнив уравнение очевидными условиями

$$B(\rho)|_{\rho=0} = \langle X^2 \rangle \langle Y^2 \rangle = \sigma^4, \quad \left. \frac{dB(\rho)}{d\rho} \right|_{\rho=0} = 4\sigma^2 \langle X \rangle \langle Y \rangle = 0,$$

получим окончательно

$$B(\rho) = \sigma^4(1 + 2\rho^2). \quad \star$$

Пример: Многие технические устройства преобразуют непрерывные величины в дискретные. Рассмотрим простейший пример подобного рода. Пусть

$$g(x) = \text{sign}(x) = \begin{cases} 1, & x > 0, \\ -1, & x < 0, \end{cases}$$

функция знака, принимающая значение 1, если x положительна, и -1 , если отрицательна. Производная этой функции равна $\text{sign}'(x) = 2\delta(x)$. Следовательно, положив в (63) $n = 1$, получим:

$$\frac{dB(\rho)}{d\rho} = 4\sigma^2 \langle \delta(X)\delta(Y) \rangle = 4\sigma^2 f(x=0, y=0).$$

Найдем $f(x=0, y=0)$, подставив в (38) $x=0, y=0$ и $\sigma_x = \sigma_y = \sigma$. Добавив кроме того очевидное начальное условие, будем иметь

$$\frac{dB(\rho)}{d\rho} = \frac{2}{\pi\sqrt{1-\rho^2}}, \quad B(\rho)|_{\rho=0} = 0.$$

Таким образом

$$B(\rho) = \frac{2}{\pi} \int_0^\rho \frac{d\rho}{\sqrt{1-\rho^2}} = \frac{2}{\pi} \arcsin(\rho). \quad \star$$

Пример: Найдем корреляцию амплитуд гауссовых случайных величин

$$B(\rho) = \langle |X| |Y| \rangle.$$

Из (63) и ответа предыдущего примера получим

$$\frac{dB(\rho)}{d\rho} = \sigma^2 \langle \text{sign}(X) \text{sign}(Y) \rangle = \sigma^2 \frac{2}{\pi} \arcsin(\rho).$$

Дополним это уравнение условием, которое в данном случае удобнее брать при $\rho = 1$:

$$B(\rho)|_{\rho=1} = \langle X^2 \rangle = \sigma^2.$$

В итоге имеем

$$B(\rho) = \sigma^2 \left[1 - \frac{2}{\pi} \int_{\rho}^1 \arcsin(\rho) d\rho \right].$$

Вычислив интеграл, получим окончательно

$$B(\rho) = \sigma^2 \frac{2}{\pi} \left[\sqrt{1 - \rho^2} + \rho \arcsin(\rho) \right]. \quad \star$$

Вдумчивый читатель уже сообразил, что имеется множество полезных соотношений, родственных (63), и справедливых для произвольных гауссовых совокупностей. Так читатель сам легко выведет формулу

$$\frac{d \langle g(X, Y) \rangle}{d \text{Cov}(X, Y)} = \left\langle \frac{\partial^2 g(X, Y)}{\partial X \partial Y} \right\rangle, \quad (65)$$

справедливую для произвольной гауссовой совокупности $\{X, Y\}$. Из нее, в частности, следует известная *формула расщепления корреляций*

$$\langle g(X) Y \rangle = \langle g(X) \rangle \langle Y \rangle + \left\langle \frac{dg(X)}{dX} \right\rangle \text{Cov}(X, Y). \quad (66)$$