Что было, а также чего не было, но что вполне могло бы быть прочитано в курсе лекций под названием

ТЕОРИЯ ВЕРОЯТНОСТЕЙ

Чернова Н.И.

— Знаете что, милый Арамис? — сказал д'Артаньян, ненавидевший стихи почти так же сильно, как латынь. — Добавьте к достоинству трудности достоинство краткости, и вы сможете быть уверены в том, что ваша поэма будет иметь никак не менее двух достоинств.

Содержание

Введение

Студентам первого курса ЭФ читать введение строго воспрещается!

Учебное пособие практически дословно повторяет курс лекций по теории вероятностей, читаемый автором студентам первого курса отделения экономики экономического факультета НГУ.

Курс теории вероятностей продолжается далее полугодовым курсом математической статистики. Затем студентам предстоит полугодовой курс регрессионного анализа, полугодовой курс теории временных рядов (в рамках курса эконометрии), знакомство в ряде дальнейших курсов с основами теории игр и теории принятия решений.

Объём курса ограничен рамками не более чем пятнадцати лекций короткого весеннего семестра и слабой подготовленностью слушателей, за плечами которых к моменту начала изучения предмета имеется лишь один семестр математического анализа и линейной алгебры.

Несмотря на это, читаемый автором курс не избегает, в том числе, таких абсолютно не знакомых студентам абстрактных понятий, как сигма-алгебры и меры, и вообще стремится быть корректным, полным и доказательным, в отличие от чисто рецептурных курсов, читаемых на экономических факультетах и отделениях остальных вузов.

Такое содержание курса сложилось в последние пять-шесть лет, и автор пока не видит необходимости в упрощении материала. Оправданием сложности курса могут служить два обстоятельства: во-первых, постоянный семестровый контроль работы студентов приводит к тому, что более четверти слушателей усваивают материал полностью в течение семестра на отличном или близком к нему уровне. Ещё примерно половина студентов вполне справляется с материалом после дополнительных летних месяцев подготовки. Во-вторых, студенты первого курса, не будучи ещё расслаблены «лёгкими» предметами, способны воспринять как должное курс лекций практически любой (разумной) сложности и насыщенности.

Основная проблема, которую читатель отметит для себя в данном пособии, заключается в сжатости материала. Несмотря на стремление к строгости изложения в целом, математическое ожидание излагается так, как это принято на нематематических факультетах — в дискретном и непрерывном случаях, без изложения общей теории интеграла Лебега. Не только недостаточный в сравнении с механико-математическим факультетом объём курса математического анализа тому причиной, но и глубокая уверенность автора, что во всём — в том числе и в уровне серьёзности материала — следует знать меру.

С нежеланием перегрузить студентов неподъёмным для их возраста и опыта материалом связано и отсутствие в курсе важной для экономистов темы про условные распределения и условные математические ожидания. В 2004/5 уч. г. этой теме была посвящена последняя лекция «для любителей», но в пособие она не вошла. И напротив, в тексте присутствует ряд утверждений и примеров, которые не входят обычно в курс лекций,— например, теорема 13, доказательство теоремы 5, пример 13.

Читателю, желающему освоить курс, стоит выполнять все содержащиеся в тексте упражнения и отвечать на заданные вопросы. В конце имеется список полезных задач по тем разделам курса, которые не вполне покрываются практическими занятиями, либо дополняющих (но не заменяющих) материал практических занятий.

Автор искренне признателен своим коллегам по кафедре теории вероятностей и математической статистики ММФ НГУ, в течение многих лет вынужденным терпеть рассказы автора о высоком уровне обучения математике на ЭФ. Автор снимает шляпу перед самоотверженным трудом своих друзей и ассистентов Е. А. Бакланова и В. В. Милосердова, по зову души и долгу службы этот уровень обеспечивающих за счёт своего времени, сил и нервов.

Н. И. Чернова

ГЛАВА 1

Классическая вероятностная схема

> ...Да, первые страницы рассказа обнаруживают, что я очень плохо думаю о публике. Я употребил обыкновенную хитрость романистов: начал повесть эффектными сценами, вырванными из средины или конца её, прикрыл их туманом. Ты, публика, добра, очень добра, а потому ты неразборчива и недогадлива. На тебя нельзя положиться, что ты с первых страниц можешь различить, будет ли содержание повести стоить того, чтобы прочесть её, у тебя плохое чутьё, оно нуждается в пособии, а пособий этих два: или имя автора, или эффектность манеры.
>
> *Н.Г.Чернышевский, Что делать?*

§ 1. Основные формулы комбинаторики

В данном разделе мы займёмся подсчётом числа «шансов». О числе шансов говорят, когда возможно несколько результатов какого-либо действия (извлечение карты из колоды, подбрасывание кубика или монетки). Число шансов — это число способов проделать это действие или, что то же самое, число возможных результатов этого действия.

Теорема о перемножении шансов. Пусть одно действие можно проделать пятью способами, а другое — двумя. Каким числом способов можно проделать пару этих действий?

Теорема 1. *Пусть множество A состоит из k элементов: $A = \{a_1, \ldots, a_k\}$, а множество B — из m элементов: $B = \{b_1, \ldots, b_m\}$. Тогда можно образовать ровно km пар (a_i, b_j), взяв первый элемент из множества A, а второй — из множества B.*

Замечание 1. Можно сформулировать утверждение теоремы 1 так: если первый элемент можно выбрать k способами, а второй элемент — m способами, то пару элементов можно выбрать km способами.

Доказательство. С элементом a_1 мы можем образовать m пар: (a_1, b_1), (a_1, b_2), $\ldots$, (a_1, b_m). Столько же пар можно составить с элементом a_2, столько же — с элементом a_3 и с любым другим из k элементов множества A. Т. е. всего возможно km пар, в которых первый элемент выбран из множества A, а второй — из множества B. □

Упражнение. С помощью теоремы 1 доказать, что:

а) при подбрасывании трёх монет возможно $2 \cdot 2 \cdot 2 = 8$ различных результатов;

б) бросая дважды игральную кость, получим $6 \cdot 6 = 36$ различных результатов;

в) трёхзначных чисел бывает $9 \cdot 10 \cdot 10 = 900$;

г) трёхзначных чисел, все цифры которых различны, существует $9 \cdot 9 \cdot 8$;

д) чётных трёхзначных чисел возможно $9 \cdot 10 \cdot 5$;

Урны и шарики. Есть урна (ящик), содержащая n пронумерованных объектов (шаров). Мы выбираем из этой урны k шаров; результатом выбора является набор из k шаров. Нас интересует, сколькими способами можно выбрать k шаров из n, или сколько различных результатов может получиться. На этот вопрос нельзя дать однозначный ответ, пока мы не определимся: а) с тем, как организован выбор (можно ли шары возвращать в урну), и б) с тем, что понимается под различными результатами выбора.

Рассмотрим следующие возможные способы выбора.

1. Выбор с возвращением: каждый вынутый шар возвращается в урну, каждый следующий шар выбирается из полной урны. В полученном наборе из k номеров шаров могут встречаться одни и те же номера.

2. Выбор без возвращения: вынутые шары в урну не возвращаются, и в полученном наборе не могут встречаться одни и те же номера.

Условимся, какие результаты выбора (наборы из k номеров шаров) мы будем считать различными. Есть ровно две возможности.

1. Выбор с учётом порядка: два набора номеров шаров считаются различными, если они отличаются составом или порядком номеров.

Так, при выборе трёх шаров из урны, содержащей 5 шаров, наборы $(1,\ 5,\ 2)$, $(2,\ 5,\ 1)$ и $(4,\ 4,\ 5)$ различны, если порядок учитывается.

2. Выбор без учёта порядка: два набора номеров шаров считаются различными, если они отличаются составом. Наборы, отличающиеся лишь порядком следования номеров, считаются одинаковыми.

Так, наборы $(1,\ 5,\ 2)$ и $(2,\ 5,\ 1)$ не различаются и образуют один и тот же результат выбора, если порядок не учитывается.

Подсчитаем, сколько возможно различных результатов для каждой из четырёх схем выбора (выбор с возвращением или без, и в каждом из этих случаев — с учётом порядка или без).

Упражнение. Перечислить все возможные результаты в каждой из четырёх схем при выборе двух шаров из четырёх.

Например, при выборе с возвращением и без учёта порядка: $(1,\ 1)$, $(1,\ 2)$, $(1,\ 3)$, $(1,\ 4)$, $(2,\ 2)$, $(2,\ 3)$, $(2,\ 4)$, $(3,\ 3)$, $(3,\ 4)$, $(4,\ 4)$.

Выбор без возвращения, с учётом порядка.

Теорема 2. *Общее количество различных наборов при выборе k элементов из n без возвращения и с учётом порядка равняется*

$$A_n^k = n(n-1)\cdot\ldots\cdot(n-k+1) = \frac{n!}{(n-k)!}$$

и называется числом размещений из n элементов по k элементов.

Доказательство. Первый шар можно выбрать n способами, его номер — любой из n возможных. При любом выборе первого шара есть $n-1$ способ выбрать второй шар. По теореме 1, число возможных пар

$$(\text{номер первого шара},\ \text{номер второго шара})$$

равно $n(n-1)$. Для каждой такой пары есть $n-2$ способа выбрать третий шар. По теореме 1, число возможных троек

$$((\text{номер первого шара},\ \text{номер второго шара}),\ \text{номер третьего шара})$$

равно произведению числа пар $n(n-1)$ и числа способов выбора третьего шара, т. е. равно $n(n-1)(n-2)$. Продолжая рассуждения, получим, что общее число возможных наборов из k шаров равно $n(n-1)\cdot\ldots\cdot(n-k+1)$. В этом произведении k сомножителей последний множитель $n-k+1$ есть число способов выбора k-го шара, когда уже выбраны предыдущие. □

Следствие 1. *Если в множестве n элементов, то существует ровно $n!$ перестановок этих элементов.*

Доказательство. Перестановка — результат выбора без возвращения и с учётом порядка n элементов из n. Поэтому общее число перестановок равно $A_n^n = n!$ □

Упражнение. Найти, сколько всего возможно различных результатов в следующих экспериментах:

а) из колоды в 36 карт без возвращения, с учётом порядка вынимают три карты;

б) Вася, Петя, Оля и Лена занимают какие-то четыре из десяти мест в классе;

в) из русского алфавита выбирают четыре разные буквы и составляют слово;

г) из различных цифр, не равных нулю, составляется трёхзначное число.

Выбор без возвращения и без учёта порядка.

Теорема 3. *Общее количество различных наборов при выборе k элементов из n без возвращения и без учёта порядка равняется*

$$C_n^k = \frac{A_n^k}{k!} = \frac{n!}{k!(n-k)!}$$

и называется числом сочетаний из n элементов по k элементов.

Доказательство. Согласно следствию 1, k различных номеров шаров можно упорядочить $k!$ способами. Поэтому из каждого набора, выбранного без возвращения и без учёта порядка, можно образовать $k!$ наборов, отличающихся друг от друга порядком следования номеров. Т. е. при выборе без возвращения и с учётом порядка возможно в $k!$ раз больше наборов, чем при выборе без учёта порядка. Поэтому число наборов при выборе без учёта порядка равно $A_n^k/k! = C_n^k$. □

Упражнение. Найти, сколько всего возможно различных результатов в следующих экспериментах:

а) из колоды в 36 карт без возвращения, без учёта порядка вынимают три карты;

б) из русского алфавита выбрасывают четыре буквы.

Выбор с возвращением и с учётом порядка.

Теорема 4. *Общее количество различных наборов при выборе k элементов из n с возвращением и с учётом порядка равняется n^k.*

Доказательство. Первый шар можно выбрать n способами. При каждом из этих способов второй шар можно выбрать также n способами, и так k раз. Общее число наборов равно $n \cdot n \cdot \ldots \cdot n = n^k$. □

Упражнение. Найти, сколько всего возможно различных результатов в следующих экспериментах:

а) из колоды в 36 карт тянут три раза карту с учётом порядка и с возвращением;

б) пятизначное число составляется из одних нечётных цифр;

в) обезьяна напечатала на машинке слово из десяти букв.

Выбор с возвращением и без учёта порядка. Рассмотрим урну с двумя пронумерованными шарами и перечислим результаты выбора двух шариков из этой урны при выборе с возвращением. Видим, что в схеме «без учёта порядка» получилось три различных результата, в отличие от четырёх результатов в схеме «с учётом порядка». Заметим также, что никаким делением на «число каких-нибудь перестановок», которое помогло избавиться от учёта порядка при выборе без возвращения, число 3 из числа 4 получить не удастся.

с учётом порядка	без учёта порядка
(1, 1)	(1, 1)
(2, 2)	(2, 2)
(1, 2)	} (1, 2)
(2, 1)	

Теорема 5. *Общее количество различных наборов при выборе k элементов из n с возвращением и без учёта порядка равняется*

$$C_{n+k-1}^{k} = C_{n+k-1}^{n-1}.$$

Упражнение. Проверить, что при $n = 2$ и $k = 2$ получается ровно 3.

Доказательство. Рассмотрим подробно, чем отличаются друг от друга два разных результата такой схемы выбора. Нам не важен порядок номеров, т. е. мы учитываем только, сколько раз в нашем наборе из k номеров шаров появился каждый номер. Поэтому результат выбора можно представить набором чисел $k_1, k_2, \ldots, k_n$, в котором k_i — число появлений шара номер i в наборе, и $k_1+\ldots+k_n=k$. Числа k_i принимают значения из множества $\mathbb{N}\cup\{0\}$. Два результата выбора в схеме выбора с возвращением и без учёта порядка различаются, если соответствующие им наборы $k_1, k_2, \ldots, k_n$ не совпадают (здесь порядок следования элементов k_i учитывается).

Представим себе другой эксперимент, имеющий точно такие же результаты, и посчитаем их количество. Есть n ящиков, в которых размещаются k шаров. Нас интересует только *число* шаров в каждом ящике. Результатом эксперимента снова является набор чисел $k_1, k_2, \ldots, k_n$, где k_i равно числу шаров в ящике с номером i, и $k_1+\ldots+k_n=k$. Числа k_i принимают натуральные значения или равны нулю.

А теперь изобразим результат такого размещения в виде схемы, в которой вертикальные линии обозначают перегородки между ящиками, а точки — находящиеся в ящиках шары:

$$| \bullet \bullet \bullet | \; | \bullet | \bullet \bullet | \bullet \bullet | \; | \bullet |$$

Мы видим результат размещения девяти шаров по семи ящикам. Первый ящик содержит три шара, второй и шестой ящики пусты, третий ящик содержит один шар, в четвёртом и пятом ящиках лежит по два шара. Переложим один шар из первого ящика во второй и изобразим таким же образом ещё два результата размещения:

$$| \bullet \bullet | \bullet | \bullet | \bullet \bullet | \bullet \bullet | \; | \bullet |$$

$$| \; | \; | \; | \; | \; | \; | \bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet |$$

Видим, что все размещения можно получить, меняя между собой шары и перегородки, или расставляя k шаров на $n-1+k$ местах. Число $n-1+k$ получается так: у n ящиков есть ровно $n+1$ перегородка, считая крайние, но из них перемещать можно лишь $n-1$ внутреннюю перегородку. Таким образом, имеется $n-1+k$ мест, которые можно занять шарами либо внутренними перегородками. Перебрав все возможные способы расставить k шаров на этих $n-1+k$ местах (заполняя оставшиеся места перегородками), переберем все нужные размещения.

Осталось заметить, что существует $C_{n-1+k}^{k}=C_{n+k-1}^{n-1}$ способов расставить k шаров на $n-1+k$ местах. Именно столько есть способов выбрать из $n-1+k$ номеров мест k номеров мест для шаров. $\square$

Упражнение.

а) Найти количество способов разложить натуральное число k в сумму n целых неотрицательных слагаемых, если важен порядок следования этих слагаемых.

б) Найти число различных производных порядка k функции n переменных.

в) Найти число возможных результатов подбрасывания двух игральных костей, если кости считаются неразличимыми. То же самое для трёх игральных костей.

§ 2. Элементарная теория вероятностей

Предмет теории вероятностей. Теория вероятностей изучает закономерности, возникающие в случайных экспериментах. Случайным называют эксперимент, результат которого нельзя предсказать заранее. Невозможность предсказать результат отличает *случайное* явление от *детерминированного*.

Не все случайные явления (эксперименты) можно изучать методами теории вероятностей, а лишь те, которые могут быть воспроизведены в одних и тех же условиях. Случайность и хаос — не одно и то же. Оказывается, что и в случайных экспериментах наблюдаются некоторые закономерности, например свойство «статистической устойчивости»: если A — некоторое событие, могущее произойти или не произойти в результате эксперимента, то доля $n(A)\,/\,n$ экспериментов, в которых данное событие произошло, имеет тенденцию стабилизироваться с ростом общего числа экспериментов n, приближаясь к некоторому числу $\mathsf{P}(A)$. Это число служит объективной характеристикой «степени возможности» событию A произойти.

Следует помнить, что мы занимаемся математикой и имеем дело не с реальностью, а лишь с её математической моделью. Мы и будем изучать только математические модели, а приложение их к реальности оставим на долю математической и практической статистики.

Пространство элементарных исходов.

Определение 1. *Пространством элементарных исходов* Ω («омега») называется множество, содержащее все возможные результаты данного случайного эксперимента, из которых в эксперименте происходит ровно один. Элементы этого множества называют *элементарными исходами* и обозначают буквой ω («омега»).

Определение 2. *Событиями* мы будем называть подмножества множества Ω. Говорят, что в результате эксперимента *произошло событие* $A \subseteq \Omega$, если в эксперименте произошел один из элементарных исходов, входящих в множество A.

Замечание 2. Вообще говоря, можно назвать событиями не обязательно любые подмножества множества Ω, а лишь элементы некоторого набора подмножеств. О смысле такого ограничения мы поговорим позднее.

Пример 1. Один раз подбрасывается кубик — игральная кость. Рассмотрим пространство элементарных исходов $\Omega = \{1, 2, 3, 4, 5, 6\} = \{$⚀, ⚁, ⚂, ⚃, ⚄, ⚅$\}$, элементарные исходы здесь соответствуют числу выпавших очков.

Примеры событий: $A = \{1, 2\} = \{$⚀, ⚁$\}$ — выпало одно или два очка; $B = \{1, 3, 5\} = \{$⚀, ⚂, ⚄$\}$ — выпало нечётное число очков.

Пример 2. Два раза подбрасывается игральная кость. Или, что то же самое, один раз подбрасываются две игральные кости. Будем считать пространством элементарных исходов множество пар чисел (i, j), где i (соответственно, j) есть число очков, выпавших при первом (втором) подбрасывании: $\Omega = \{(i, j),\ \text{где } 1 \leqslant i, j \leqslant 6\}$.

Примеры событий:
$A = \{(1,1), (1,2), (1,3), (1,4), (1,5), (1,6)\}$ — при первом подбрасывании выпало одно очко;
$B = \{(1,1), (2,1), (3,1), (4,1), (5,1), (6,1)\}$ — при втором подбрасывании выпало одно очко;
$C = \{(1,1), (2,2), (3,3), (4,4), (5,5), (6,6)\}$ — на костях выпало одинаковое число очков;
$D = \{(1,1), (1,3), (1,5), (3,1), (3,3), (3,5), (5,1), (5,3), (5,5)\}$ — на обеих костях выпало нечётное число очков.

Пример 3. На поверхность стола бросается монета. Результатом эксперимента можно считать координату центра монеты. Пространство элементарных исходов — множество точек стола. Если нам не безразличен угол поворота монеты, то можно добавить к множеству положений центра величину этого угла. В этом случае Ω есть множество пар $\{(x, \varphi)\}$, где $x \in \mathbb{R}^2$ — точка стола и $\varphi \in [0, 2\pi)$ — угол поворота. Число элементарных исходов такого эксперимента несчётно.

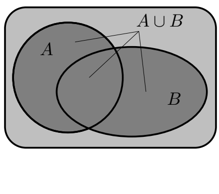
Пример 4. Монета подбрасывается до тех пор, пока не выпадет вверх гербом. Пространство элементарных исходов состоит из бесконечного, но счётного числа исходов: $\Omega = \{$ *г, рг, ррг, рррг, ррррг, рррррг,* $\ldots\}$, где *р* означает выпадение решки, а *г* — герба при одном подбрасывании.

Определение 3. 1. Достоверным называется событие, которое обязательно происходит в результате эксперимента, т. е. единственное событие, включающее все элементарные исходы — событие Ω.

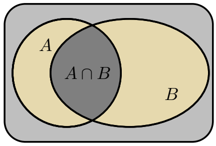
2. Невозможным называется событие, которое не может произойти в результате эксперимента, т. е. событие, не содержащее ни одного элементарного исхода («пустое множество» $\varnothing$). Заметим, что всегда $\varnothing \subset \Omega$.

Операции над событиями. В теории вероятностей существуют ровно те же операции над множествами, что и в теории множеств.

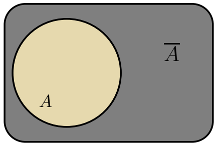
Определение 4. 1. Объединением $A \cup B$ событий A и B называется событие, состоящее в том, что произошло либо A, либо B, либо оба события одновременно. На языке теории множеств $A \cup B$ есть множество, содержащее как элементарные исходы из множества A, так и элементарные исходы из множества B.



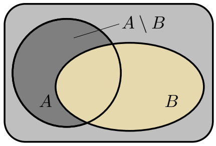
2. *Пересечением* $A \cap B$ событий A и B называется событие, состоящее в том, что произошли оба события A и B одновременно. На языке теории множеств $A \cap B$ есть множество, содержащее элементарные исходы, входящие в пересечение множеств A и B.



3. *Противоположным* (или дополнительным) к событию A называется событие $\overline{A} = \Omega \backslash A$, состоящее в том, что событие A в результате эксперимента не произошло. Т. е. множество $\overline{A}$ состоит из элементарных исходов, не входящих в A.



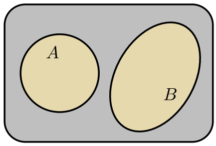
4. *Дополнением* $A \backslash B$ события B до A называется событие, состоящее в том, что произошло событие A, но не произошло B. Т. е. множество $A \backslash B$ содержит элементарные исходы, входящие в множество A, но не входящие в B.



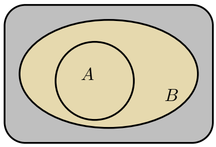
Определение 5.

1. События A и B называют *несовместными*, если $A \cap B = \varnothing$.

2. События $A_1, \ldots, A_n$ называются *попарно несовместными*, если для любых $i \neq j$, где $1 \leqslant i, j \leqslant n$, события A_i и A_j несовместны.



3. Говорят, что событие A *влечёт* событие B, и пишут $A \subseteq B$, если всегда, как только происходит событие A, происходит и событие B. На языке теории множеств это означает, что любой элементарный исход, входящий в множество A, одновременно входит и в множество B, т. е. A содержится в B.



Вероятность на дискретном пространстве элементарных исходов. Пространство элементарных исходов назовём *дискретным*, если оно конечно или счётно: $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n, \ldots\}$.

Так, эксперименты из примеров 1, 2 и 4 (но не 3) приводят к дискретным пространствам элементарных исходов.

Замечание 3. Множество счётно, если существует взаимно-однозначное соответствие между этим множеством и множеством всех натуральных чисел. Счётными множествами являются множество натуральных чисел, множество целых чисел (доказать), множество рациональных чисел (доказать), множество чётных чисел и т. д. Множество конечно, если оно состоит из конечного числа элементов.

Чтобы определить вероятность любого события на дискретном пространстве элементарных исходов, достаточно присвоить вероятность каждому элементарному исходу. Тогда вероятность любого события определяется как сумма вероятностей входящих в него элементарных исходов.

Определение 6. Поставим каждому элементарному исходу $\omega_i \in \Omega$ в соответствие число $p_i \in [0, 1]$ так, что
$$\sum_{\omega_i \in \Omega} p_i = 1.$$
Назовём число p_i вероятностью элементарного исхода ω_i. Вероятностью события A назовём число
$$\mathsf{P}(A) = \sum_{\omega_i \in A} p_i,$$
равное сумме вероятностей элементарных исходов, входящих в множество A. В случае $A = \varnothing$ положим $\mathsf{P}(A) = 0$.

Замечание 4. Позднее, познакомившись с аксиоматикой теории вероятностей, мы зададим вероятности событий непосредственно, а не через вероятности элементарных исходов. Тем более, что сложением вероятностей элементарных исходов можно получить лишь вероятность события, состоящего не более чем из счётного числа элементарных исходов (иначе само понятие суммирования не определено). Но на дискретном пространстве элементарных исходов определить вероятности событий так, как это сделано в определении 6, всегда возможно.

Перечислим очевидные в случае дискретного пространства свойства вероятности, которые мы скоро докажем сразу в общем случае.

1. $0 \leqslant \mathsf{P}(A) \leqslant 1$; $\mathsf{P}(\Omega) = 1$; $\mathsf{P}(\overline{A}) = 1 - \mathsf{P}(A)$;

2. Если A и B несовместны, то $\mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B)$;

3. В общем случае $\mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(A \cap B)$;

4. Если $A \subseteq B$, то $\mathsf{P}(A) \leqslant \mathsf{P}(B)$.

Упражнение. Доказать свойства 1 — 4, пользуясь определением 6.

Рассмотрим частный случай такой вероятности — так называемую «классическую вероятность».

Классическое определение вероятности. Предположим, что мы имеем дело с пространством элементарных исходов, состоящим из конечного числа N элементов: $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$. Предположим, что из каких-либо соображений мы можем считать элементарные исходы равновозможными. Тогда вероятность любого из них принимается равной $1 / N$. Эти соображения не имеют отношения к математической модели и основаны на какой-либо симметрии в эксперименте (симметричная монета, хорошо перемешанная колода карт, правильная кость).

Если событие $A = \{\omega_{i_1}, \ldots, \omega_{i_k}\}$ состоит из k элементарных исходов, то вероятность этого события равняется отношению k / N:
$$\mathsf{P}(A) = p_{i_1} + \ldots + p_{i_k} = k \cdot \frac{1}{N} = \frac{|A|}{|\Omega|},$$
где символом $|A|$ обозначено число элементов конечного множества A.

Определение 7. Говорят, что эксперимент удовлетворяет «классическому определению вероятности», если пространство элементарных исходов состоит из конечного числа $|\Omega| = N$ равновозможных исходов. В этом случае вероятность любого события A вычисляется по формуле

$$\mathsf{P}(A) = \frac{|A|}{|\Omega|},$$

называемой классическим определением вероятности.

Формулу $\mathsf{P}(A) = |A| \,/\, |\Omega|$ читают так: «вероятность события A равна отношению числа исходов, благоприятствующих событию A, к общему числу исходов». Полезно сравнить это определение с классической формулировкой Якоба Бернулли[1]: «Вероятность есть степень достоверности и отличается от неё как часть от целого» (Ars Conjectandi, 1713 г.)

Мы видим, что вычисление вероятности в классической схеме сводится к подсчёту общего числа «шансов» и числа шансов, благоприятствующих событию. Число шансов считают с помощью формул комбинаторики.

Рассмотрим урновые схемы из параграфа 1. Будем исходить из предположения о том, что появление любого шара равновозможно. Тогда три схемы: с возвращением и с учётом порядка, без возвращения и с учётом порядка, а также без возвращения и без учёта порядка, удовлетворяют классическому определению вероятности. Общее число элементарных исходов в этих схемах подсчитано в теоремах 4, 2, 3 и равно соответственно n^k, A_n^k, C_n^k. Четвёртая же схема — схема выбора с возвращением и без учёта порядка — имеет заведомо неравновозможные исходы.

Пример 5. Рассмотрим выбор двух шариков из двух или, что то же самое, дважды подбросим монету. Если учитывать порядок, то исходов получится четыре, и все они равновозможны, т. е. имеют вероятность по $1/4$:

(герб, герб), (решка, решка), (решка, герб), (герб, решка).

Если порядок не учитывать, то следует объявить два последних исхода одним и тем же результатом эксперимента, и получить не четыре, а три исхода:

(два герба), (две решки), (один герб и одна решка).

Первые два исхода имеют вероятности по $1/4$, а последний — вероятность $1/4 + 1/4 = 1/2$.

Упражнение. Посчитать число элементарных исходов в примере 2. Каким станет пространство элементарных исходов, если порядок костей не учитывать? Посчитать число элементарных исходов в таком пространстве (пользуясь теоремой 5 или прямым подсчётом). Убедиться, что их ровно $C_7^2 = 21$. Равновозможны ли эти исходы? Посчитать вероятность каждого.

[1] Jacob Bernoulli (27.12.1654 — 16.08.1705, Basel, Switzerland)

Гипергеометрическое распределение.

Пример 6. Из урны, в которой K белых и $N-K$ чёрных шаров, наудачу и без возвращения вынимают n шаров, $n \leqslant N$. Термин «наудачу» означает, что появление любого набора из n шаров равновозможно. Найти вероятность того, что будет выбрано k белых и $n-k$ чёрных шаров.

Решение. При $k > K$ или $n-k > N-K$ искомая вероятность равна нулю, так как соответствующее событие невозможно. Пусть $k \leqslant K$ и $n-k \leqslant N-K$.

Результатом эксперимента является набор из n шаров. Можно не учитывать или учитывать порядок следования шаров. Вероятность не должна зависеть от способа подсчёта.

Выбор без учёта порядка. Общее число элементарных исходов есть число n-элементных подмножеств множества, состоящего из N элементов: $|\Omega| = C_N^n$ по теореме 3.

Обозначим через A_k событие, вероятность которого требуется найти. Событию A_k благоприятствует появление любого набора, содержащего k белых шаров и $n-k$ чёрных. Число благоприятных исходов равно произведению (по теореме 1) числа способов выбрать k белых шаров из K и числа способов выбрать $n-k$ чёрных шаров из $N-K$, т. е. $|A_k| = C_K^k \cdot C_{N-K}^{n-k}$. Вероятность события A_k равна

$$\mathsf{P}(A) = \frac{C_K^k C_{N-K}^{n-k}}{C_N^n}. \tag{1}$$

Выбор с учётом порядка. Общее число элементарных исходов есть число способов разместить N элементов на n местах. По теореме 2

$$|\Omega| = A_N^n = N(N-1)\cdot\ldots\cdot(N-n+1).$$

При подсчёте числа благоприятных исходов нужно учесть число способов выбрать k белых и $n-k$ чёрных шаров и число способов расположить эти шары среди n. Можно, скажем, посчитать число способов выбрать k мест среди n, равное C_n^k, затем число способов разместить на этих k местах K белых шаров, равное A_K^k, и затем число способов разместить на оставшихся $n-k$ местах $N-K$ чёрных шаров, равное A_{N-K}^{n-k}. Перемножив (почему?) эти числа, получим

$$|A_k| = C_n^k \cdot A_K^k \cdot A_{N-K}^{n-k}, \qquad \mathsf{P}(A_k) = \frac{C_n^k A_K^k A_{N-K}^{n-k}}{A_N^n} = \frac{C_K^k C_{N-K}^{n-k}}{C_N^n}.$$

В рассмотренной задаче мы сопоставили каждому набору из k белых и $n-k$ чёрных шаров вероятность $\mathsf{P}(A_k)$ получить этот набор при выборе n шаров из урны, содержащей K белых и $N-K$ чёрных шаров.

Определение 8. Соответствие между числом k и вероятностью

$$\mathsf{P}(A_k) = \frac{C_K^k C_{N-K}^{n-k}}{C_N^n},$$

где k таково, что $0 \leqslant k \leqslant n$, $k \leqslant K$ и $n - k \leqslant N - K$, называется гипергеометрическим распределением.

Здесь мы в первый, но далеко не в последний раз встретились с термином «распределение» вероятностей. Это слово всегда обозначает некий способ разделить (распределить) общую единичную вероятность между какими-то точками или множествами на вещественной прямой.

В гипергеометрическом распределении единичная вероятность распределена между подходящими целыми числами k неравномерно. Каждому целому числу k сопоставлена своя вероятность $\mathsf{P}(A_k)$. На вещественной прямой можно единичную вероятность распределить по-разному. Этим одно распределение отличается от другого: тем, на каком множестве чисел «распределена» общая единичная вероятность, и тем, какие веса, или вероятности, присвоены отдельным точкам или частям этого множества.

Упражнение. Понять последний абзац.

Задание вероятностей на дискретном пространстве. Если пространство элементарных исходов счётно, но не конечно, нельзя всем элементарным исходам присвоить одинаковые вероятности (почему?). Приведём примеры того, какими могут быть вероятности на таком пространстве.

Пример 7. Пусть $\Omega = \mathbb{N}$. Зададим вероятность элементарного исхода $i \in \mathbb{N}$ так: $p_i = 1/2^i$. Проверим, что набор таких вероятностей удовлетворяет определению 6. По формуле суммы бесконечной геометрической прогрессии с первым членом $1/2$ и знаменателем $1/2 < 1$ имеем:

$$\sum_{i \in \mathbb{N}} p_i = \sum_{i=1}^{\infty} \frac{1}{2^i} = \frac{1/2}{1 - 1/2} = 1.$$

Пример 8. На том же самом множестве $\Omega = \mathbb{N}$ зададим вероятности так: $p_1 = \ldots = p_{100} = 0{,}01$, $p_i = 0$ для $i > 100$.

Пример 9. На том же $\Omega = \mathbb{N}$ положим $p_1 = 0{,}3$, $p_{1372} = 0{,}7$, остальные p_i равны нулю. Читатель легко найдёт вероятность события $A = \{1000, 1001, \ldots, 1500\} \subset \Omega$.

Пример 10. Пусть теперь $\Omega = \mathbb{N} \cup \{0\}$. Положим $p_i = \dfrac{7^i}{i!} e^{-7}$ для $i = 0, 1, 2, \ldots$ Проверим, равна ли единице сумма вероятностей всех элементарных исходов. Собрав разложенную в ряд Тейлора экспоненту, получим:

$$\sum_{\omega_i \in \Omega} p_i = e^{-7} \sum_{i=0}^{\infty} \frac{7^i}{i!} = e^{-7} e^7 = 1.$$

ГЛАВА 2

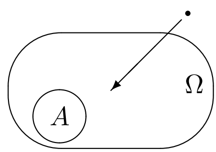
Геометрическая вероятность

Пусть будет поставлен следующий вопрос: если кто-нибудь стреляет наудачу, то какова вероятность, чтобы центр пули на пути своём прошел точно через центр того яблока, которое висит на этом дереве. Каждый должен признать, что многообразие всех возможных здесь случаев, отвечающих подобной $\langle\ldots\rangle$ вероятности, будет бесконечно, откуда следует, что степень этой вероятности имеет величину, которая равна $1/\infty$.

Б. Больцано, Парадоксы бесконечного

§ 1. Определения и примеры

Рассмотрим какую-нибудь область Ω в $\mathbb{R}^m$ (на прямой, на плоскости, в пространстве). Предположим, что «мера» Ω (длина, площадь, объем, соответственно) конечна. Пусть случайный эксперимент состоит в том, что мы наудачу бросаем в эту область точку. Термин «наудачу» означает, что вероятность попадания точки в любую часть $A\subset\Omega$ не зависит от формы или расположения A внутри Ω.



Определение 9. Эксперимент удовлетворяет условиям «геометрического определения вероятности», если его исходы можно изобразить точками некоторой области Ω в $\mathbb{R}^m$ так, что вероятность попадания точки в любую часть $A\subset\Omega$ не зависит от формы или расположения A внутри Ω, а зависит лишь от меры области A и пропорциональна этой мере:

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

где $\mu(A)$ обозначает меру области A (длину, площадь, объем и т.д.).

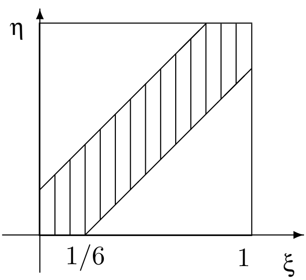
Если для точки, брошенной в область Ω, выполнены условия геометрического определения вероятности, то говорят, что точка равномерно распределена в области Ω.

Пример 11. Точка наудачу бросается на отрезок [0, 1]. Вероятность ей попасть в точку 0,5 равна нулю, так как равна нулю мера множества, состоящего из одной точки («длина точки»). Но попадание в точку 0,5 не является невозможным событием — это один из элементарных исходов эксперимента. Общее число элементарных исходов здесь бесконечно, но все они по-прежнему «равновозможны» — уже не в смысле классического определения вероятности, применить которое здесь нельзя из-за бесконечности числа исходов, а в смысле определения 9.

Задача о встрече.

Пример 12. Два лица X и Y условились встретиться в определённом месте между двумя и тремя часами дня. Пришедший первым ждет другого в течение 10 минут, после чего уходит. Чему равна вероятность встречи этих лиц, если каждый из них может прийти в любое время в течение указанного часа независимо от другого?

Решение. Будем считать интервал с 14 до 15 часов отрезком $[0,\ 1]$ длиной в 1 час. Пусть ξ («кси») и η («эта») — моменты прихода X и Y — точки отрезка $[0,\ 1]$. Пространством элементарных исходов будет квадрат со стороной 1: $\Omega = \{(\xi,\eta) \mid 0 \leqslant \xi \leqslant 1,\ 0 \leqslant \eta \leqslant 1\} = [0,\ 1]\times[0,\ 1]$. Можно считать, что эксперимент сводится к бросанию точки наудачу в квадрат. При этом благоприятными исходами являются точки множества A:



$$A = \{(\xi,\eta) \mid |\xi - \eta| \leqslant 1/6\}$$

(10 минут $=$ 1/6 часа). Попадание в множество A наудачу брошенной в квадрат точки означает, что X и Y встретятся. Тогда вероятность встречи равна

$$\mathsf{P}(A) = \frac{\mu(A)}{\mu(\Omega)} = \frac{1-(5/6)^2}{1} = \frac{11}{36}.$$

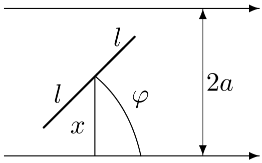
Задача Бюффона[2].

Пример 13. На плоскости начерчены параллельные прямые, находящиеся друг от друга на расстоянии $2a$. На плоскость наудачу брошена игла длины $2l < 2a$. Какова вероятность того, что игла пересечёт какую-нибудь прямую?

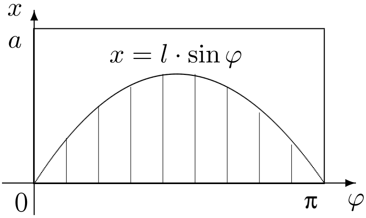
Решение. Поймем, что означает здесь «наудачу брошена игла». Возможные положения иглы (отрезка) на плоскости полностью определяются положением середины иглы и углом поворота иглы относительно какого-либо направления. Причём две эти переменные (поло-

[2] Georges Louis Leclerc Comte de Buffon (7.09.1707 — 16.04.1788, France)

жение центра и угол поворота) меняются независимо друг от друга. Обозначим через $x \in [0, a]$ расстояние от середины иглы до ближайшей прямой, а через $\varphi \in [0, \pi]$ — угол между каким-то направлением прямых и иглой. Множество возможных положений иглы целиком определяется выбором наудачу точки из прямоугольника $\Omega = [0, \pi] \times [0, a]$.



Игла пересекает ближайшую прямую, если координаты выбранной наудачу точки удовлетворяют неравенству: $x \leqslant l \sin\varphi$. Площадь области $A \subset \Omega$, точки которой удовлетворяют такому неравенству, равна



$$\mu(A) = \int_0^\pi l \cdot \sin\varphi \, d\varphi = -l \cdot \cos\varphi \Big|_0^\pi = 2l.$$

Поделим на $\mu(\Omega) = a\pi$ и получим, что искомая вероятность равна $\mathsf{P}(A) = 2l \,/\, a\pi$.

§ 2. Существование неизмеримых множеств

Заканчивая обсуждение понятия геометрической вероятности, сделаем очень важное для дальнейшего замечание.

Замечание 5. Если даже эксперимент удовлетворяет геометрическому определению вероятности, далеко не для всех множеств $A \subset \Omega$ вероятность может быть вычислена как отношение меры A к мере Ω. Причиной этого является существование так называемых «неизмеримых» множеств, т. е. множеств, мера которых не существует. Не путать с точкой — множеством нулевой меры!

Пример 14 (множество Витали[3]). В этом примере мы построим множество на отрезке, «длина» которого не существует. Нам понадобятся лишь следующие очевидные свойства «длины» множества: длина множества остается неизменной при сдвиге всех точек этого множества; длина множества, составленного из счётного объединения попарно непересекающихся множеств, равняется сумме длин этих множеств.

Рассмотрим окружность единичного радиуса (то же, что отрезок $[0, 2\pi]$). Возьмём любое иррациональное число α. Поскольку оно иррационально, число $n\alpha$ не является целым ни при каком целом $n \neq 0$ (т. е. число $2\pi n\alpha$ равно $2\pi k$ лишь при $n = k = 0$).

Поэтому если взять произвольную точку $x \in [0, 2\pi]$ на окружности и перечислить все точки, которые получаются поворотом точки x на угол $2\pi n\alpha$, $n = \pm1, \pm2, \ldots$, то мы ни разу не вернёмся в точку x. Точек, получившихся из точки x такими поворотами, счётное число. Объединим их в один класс

[3] Giuseppe Vitali (26.08.1875 — 29.02.1932, Italy)

точек. С любой другой точкой окружности можно тоже связать класс точек, получающихся из неё поворотами на $2\pi n\alpha$ при целых n. Таким образом, вся окружность разбивается на классы точек. В каждом классе счётное число точек, и все точки в одном классе получаются друг из друга такими поворотами. Разные классы не пересекаются. Заметим, что таких классов несчётное число, т. к. объединением счётного числа счётных множеств нельзя получить несчётное число точек окружности.

Искомое множество A_0 определим так: возьмём из каждого такого класса ровно по одной точке. Пусть множество A_n получается поворотом всех точек множества A_0 на угол $2\pi n\alpha$, $n = \pm1, \pm2, \ldots$.

Так как все точки одного класса можно получить, поворачивая любую из них на угол $2\pi n\alpha$, $n = \pm1, \pm2, \ldots$, а в множестве A_0 собрано по одной точке из каждого класса, то поворачивая это множество, получим все точки окружности.

Очевидно, что $\bigcup\limits_{n=-\infty}^{\infty} A_n = [0, 2\pi]$. Предположим, что «длина» $l(A_0)$ множества A_0 существует. Тогда все множества A_n имеют ту же длину, так как получены из A_0 поворотом. И так как все эти множества не пересекаются, то «длина» их объединения равна сумме их длин:

$$2\pi = l\left(\bigcup_{n=-\infty}^{\infty} A_n\right) = \sum_{n=-\infty}^{\infty} l(A_n) = \sum_{n=-\infty}^{\infty} l(A_0) = \begin{cases} \infty, & \text{если } l(A_0) > 0, \\ 0, & \text{если } l(A_0) = 0. \end{cases}$$

Полученное противоречие означает, что длина множества A_0 просто не существует.

Итак, мы построили множество на отрезке, длина которого не существует (неизмеримое множество). Пользуясь геометрическим определением вероятности, мы не можем определить вероятность попадания точки в такое неизмеримое множество. А если не для всех подмножеств Ω мы можем определить вероятности, следует сузить класс множеств, называемых «событиями», оставив в этом классе только те множества, вероятность которых определена.

В следующей главе мы займёмся, следуя Колмогорову[4], построением аксиоматики теории вероятностей: познакомимся с понятиями σ-алгебры (или поля) событий, вероятностной меры, вероятностного пространства, а также докажем сформулированные в параграфе 2 главы 1 свойства вероятности.

[4] Андрей Николаевич Колмогоров (25.04.1903 — 20.10.1987)

ГЛАВА 3

Аксиоматика теории вероятностей

Математик должен знать меру, норму и предел
(фольклор ММФ НГУ)

§ 1. Алгебра и сигма-алгебра событий

Алгебра событий. Пусть Ω — пространство элементарных исходов некоторого случайного эксперимента (т. е. непустое множество произвольной природы). Мы собираемся определить набор подмножеств Ω, которые будут называться событиями, и затем задать вероятность как функцию, определённую т о л ь к о на множестве событий.

Итак, событиями мы будем называть не любые подмножества Ω, а лишь элементы некоторого выделенного набора подмножеств Ω. При этом необходимо позаботиться, чтобы этот набор подмножеств был з а м к н у т относительно обычных операций над событиями, т. е. чтобы объединение, пересечение, дополнение событий снова давало событие. Сначала введём понятие алгебры событий.

О п р е д е л е н и е 10. Множество $\mathcal{A}$, элементами которого являются подмножества множества Ω (не обязательно все) называется а л г е б р о й (алгеброй событий), если оно удовлетворяет следующим условиям:

(A1) $\Omega \in \mathcal{A}$ (алгебра событий содержит достоверное событие);

(A2) если $A \in \mathcal{A}$, то $\overline{A} \in \mathcal{A}$ (вместе с любым событием алгебра содержит противоположное событие);

(A3) если $A \in \mathcal{A}$ и $B \in \mathcal{A}$, то $A \cup B \in \mathcal{A}$ (вместе с любыми двумя событиями алгебра содержит их объединение).

Из (A1) и (A2) следует, что пустое множество $\varnothing = \overline{\Omega}$ также содержится в $\mathcal{A}$. Из (A3) следует, что вместе с любым к о н е ч н ы м набором событий алгебра содержит их объединение: для любого $n \geqslant 2$, для любых $A_1, \ldots, A_n \in \mathcal{A}$ выполнено $A_1 \cup \ldots \cup A_n \in \mathcal{A}$. Вместо замкнутости относительно объединения можно требовать замкнутость относительно пересечения.

С в о й с т в о 1. *В определении* 10 можно заменить (A3) на (A4):

(A4) если $A \in \mathcal{A}$ и $B \in \mathcal{A}$, то $A \cap B \in \mathcal{A}$.

Доказательство. Докажем, что при выполнении (A1) и (A2) из (A3) следует (A4). Если $A, B \in \mathcal{A}$, то $\overline{A} \in \mathcal{A}$, $\overline{B} \in \mathcal{A}$ по свойству (A2). Тогда из (A3) следует, что $\overline{A} \cup \overline{B} \in \mathcal{A}$, и, по (A2), дополнение $\overline{\overline{A} \cup \overline{B}}$ к этому множеству также принадлежит $\mathcal{A}$. В силу формул двойственности, дополнение к объединению как раз и есть пересечение дополнений:

$$A \cap B = \overline{\overline{A} \cup \overline{B}} \in \mathcal{A}.$$

Аналогично доказывается, что при выполнении (A1) и (A2) из (A4) следует (A3), т. е. эти два свойства в определении взаимозаменяемы. □

Пример 15. Пусть $\Omega = \{\spadesuit, \clubsuit, \diamondsuit, \heartsuit\}$ — пространство элементарных исходов. Следующие наборы подмножеств Ω являются алгебрами (проверьте это по определению):

1. $\mathcal{A} = \{\Omega, \varnothing\} = \{\{\spadesuit, \clubsuit, \diamondsuit, \heartsuit\}, \varnothing\}$ — тривиальная алгебра.
2. $\mathcal{A} = \{\Omega, \varnothing, \{\diamondsuit\}, \Omega \setminus \{\diamondsuit\}\} = \{\{\spadesuit, \clubsuit, \diamondsuit, \heartsuit\}, \varnothing, \{\diamondsuit\}, \{\spadesuit, \clubsuit, \heartsuit\}\}$.
3. $\mathcal{A} = \{\Omega, \varnothing, A, \overline{A}\} = \big\{\{\spadesuit, \clubsuit, \diamondsuit, \heartsuit\}, \varnothing, A, \overline{A}\big\}$, где A — произвольное подмножество Ω (в предыдущем примере $A = \{\diamondsuit\}$).
4. $\mathcal{A} = 2^{\Omega}$ — множество всех подмножеств Ω.

Упражнение. Доказать, что если Ω состоит из n элементов, то в множестве всех его подмножеств ровно 2^n элементов.

Сигма-алгебра событий. В теории вероятностей часто возникает необходимость объединять счётные наборы событий и считать событием результат такого объединения. При этом свойства (A3) алгебры оказывается недостаточно: из него не вытекает, что объединение счётной последовательности множеств из алгебры снова принадлежит алгебре. Поэтому разумно наложить более суровые ограничения на класс событий.

Определение 11. Множество $\mathcal{F}$, элементами которого являются подмножества множества Ω (не обязательно все) называется σ-алгеброй (σ-алгеброй событий), если выполнены следующие условия:

(S1) $\Omega \in \mathcal{F}$ (σ-алгебра событий содержит достоверное событие);

(S2) если $A \in \mathcal{F}$, то $\overline{A} \in \mathcal{F}$ (вместе с любым событием σ-алгебра содержит противоположное событие);

(S3) если $A_1, A_2, \ldots \in \mathcal{F}$, то $A_1 \cup A_2 \cup \ldots \in \mathcal{F}$ (вместе с любым счётным набором событий σ-алгебра содержит их объединение).

Упражнение.

а) Доказать, что вместо (S1) достаточно предположить непустоту множества $\mathcal{F}$.

б) Вывести из (S1) и (S2), что $\varnothing \in \mathcal{F}$.

Этого набора аксиом достаточно для замкнутости множества $\mathcal{F}$ относительно счётного числа любых других операций над событиями. В частности, аналогично свойству 1 проверяется следующее утверждение.

Свойство 2. *В определении* 11 можно заменить (S3) на (S4):

(S4) если $A_1, A_2, \ldots \in \mathcal{F}$, то $A_1 \cap A_2 \cap \ldots \in \mathcal{F}$.

Как показывает следующее свойство, всякая σ-алгебра есть алгебра.

Свойство 3. *Если $\mathcal{F}$ — σ-алгебра, то она удовлетворяет свойству* (A3), т. е. для любых $A \in \mathcal{F}$ и $B \in \mathcal{F}$ выполняется $A \cup B \in \mathcal{F}$.

Доказательство. Превратим пару A, B в счётную последовательность событий так: $A, B, B, B, B, \ldots$, т. е. положим $A_1 = A$, $A_i = B$ при всех $i \geqslant 2$. Объединение $A \cup B$ совпадает с объединением всех множеств A_i из этой бесконечной последовательности. А так как $\mathcal{F}$ — σ-алгебра, то

$$A \cup B = \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}. \qquad \square$$

Упражнение. Докажите, что для любых $A, B \in \mathcal{F}$ выполнено $A \setminus B \in \mathcal{F}$.

Итак, всякая σ-алгебра автоматически является алгеброй, но не наоборот. Приведём пример алгебры, не являющейся σ-алгеброй.

Пример 16. Пусть $\Omega = \mathbb{R}$, и пусть $\mathcal{A}$ — множество, содержащее любые конечные подмножества $\mathbb{R}$ (т. е. состоящие из конечного числа точек, в том числе пустое) и их дополнения. В частности, множество $\{0, 2, \pi\}$ принадлежит $\mathcal{A}$, множество $(-\infty, -7{,}2) \cup (-7{,}2, 5) \cup (5, \infty)$ принадлежит $\mathcal{A}$.

Легко проверить, что множество $\mathcal{A}$ является алгеброй. Действительно, пустое множество и само $\Omega = \mathbb{R}$ там содержатся, дополнение к любому конечному подмножеству множества вещественных чисел содержится в $\mathcal{A}$ по определению, дополнение к множеству вида $\mathbb{R} \setminus A$ для конечных A совпадает с A и также принадлежит $\mathcal{A}$ по определению. Свойство (A3) проверяется непосредственно: объединение любых конечных множеств снова конечно и поэтому принадлежит $\mathcal{A}$. Объединение конечного множества с множеством вида $\mathbb{R} \setminus A$, где A конечно, есть снова множество вида $\mathbb{R} \setminus B$, где B конечно (или пусто). Объединение двух множеств $\mathbb{R} \setminus A$ и $\mathbb{R} \setminus B$, являющихся дополнениями до $\mathbb{R}$ конечных множеств A и B, есть снова множество такого же вида.

Однако алгебра $\mathcal{A}$ не содержит ни одного счётного множества точек. Действительно, объединяя конечные множества в конечном числе, мы можем получить только конечное множество. Например, натуральный ряд $\mathbb{N}$ не принадлежит $\mathcal{A}$. Поэтому $\mathcal{A}$ не является σ-алгеброй: для бесконечной, но счётной последовательности одноточечных множеств $A_i = \{i\}$ из $\mathcal{A}$ их объединение $\mathbb{N} = A_1 \cup A_2 \cup \ldots$ не принадлежит $\mathcal{A}$.

Все алгебры из примера 15 являются σ-алгебрами, поскольку содержат лишь конечное число элементов. Вообще, на конечном множестве Ω понятия алгебры и σ-алгебры совпадают. Множество всех подмножеств Ω является σ-алгеброй для любого Ω.

Борелевская[5] σ-алгебра в $\mathbb{R}$. Приведём пример σ-алгебры, которая нам будет необходима в дальнейшем,— σ-алгебры борелевских множеств на вещественной прямой.

Борелевской сигма-алгеброй в $\mathbb{R}$ называется самая маленькая среди всех возможных σ-алгебр, содержащих любые интервалы на прямой. Разумеется, σ-алгебры, содержащие все интервалы, существуют. Например, множество всех подмножеств $\mathbb{R}$ — это σ-алгебра, и она содержит все интервалы. Что же такое «самая маленькая σ-алгебра» из нескольких данных? Обратимся к примерам.

Пример 17. Пусть $\Omega = \mathbb{R}$ — вещественная прямая. Рассмотрим некоторые наборы множеств, не являющиеся σ-алгебрами, и увидим, как их можно дополнить до σ-алгебр.

1. Множество $\mathfrak{A} = \{\mathbb{R}, \varnothing, [0, 1], \{0\}\}$ не является σ-алгеброй, так как, например, $\overline{[0, 1]} = \mathbb{R} \setminus [0, 1] = (-\infty, 0) \cup (1, \infty) \not\in \mathfrak{A}$. Самый маленький набор множеств, содержащий $\mathfrak{A}$ и являющийся σ-алгеброй (минимальная σ-алгебра), получится, если включить в него всевозможные объединения, пересечения и дополнения множеств из $\mathfrak{A}$: $\mathcal{F} = \{\mathbb{R}, \varnothing, [0, 1], \{0\}, (-\infty, 0) \cup (1, \infty), (0, 1], (-\infty, 0] \cup (1, \infty), (-\infty, 0) \cup (0, \infty)\}$. Более точно:

Определение 12. Минимальной σ-алгеброй, содержащей набор множеств $\mathfrak{A}$, называется пересечение всех σ-алгебр, содержащих $\mathfrak{A}$.

Ещё раз напомним, что пересекать в определении 12 есть что: хотя бы одна σ-алгебра, содержащая данный набор множеств, всегда найдётся — это σ-алгебра всех подмножеств Ω (в данном случае $\Omega = \mathbb{R}$).

Упражнение. Доказать, что пересечение двух σ-алгебр, содержащих набор множеств $\mathfrak{A}$, снова является σ-алгеброй (невероятно!), содержащей $\mathfrak{A}$.

Упражнение. Найти минимальную σ-алгебру, содержащую следующий набор подмножеств $\mathbb{R}$: $\mathfrak{A} = \{\mathbb{R}, \varnothing, [0, 1], \{3\}\}$.

2. Пусть множество $\mathfrak{A}$ подмножеств вещественной прямой $\mathbb{R}$ состоит из всевозможных открытых интервалов (a, b), где $a < b$:

$$\mathfrak{A} = \{(a, b) \mid -\infty < a < b < \infty\}.$$

Упражнение. Проверить, что множество $\mathfrak{A}$ всех интервалов ни в коем случае не является ни алгеброй, ни σ-алгеброй! Указание: привести примеры двадцати множеств из $\mathfrak{A}$, дополнения к которым не принадлежат $\mathfrak{A}$; привести примеры пяти множеств из $\mathfrak{A}$, любые объединения которых не принадлежат $\mathfrak{A}$.

Определение 13. Минимальная σ-алгебра, содержащая множество $\mathfrak{A}$ всех интервалов на вещественной прямой, называется борелевской σ-алгеброй в $\mathbb{R}$ и обозначается $\mathfrak{B}(\mathbb{R})$.

[5] Félix Edouard Justin Emile Borel (7.01.1871 — 3.02.1956, France)

Перечислим некоторые множества на прямой, содержащиеся в $\mathfrak{B}(\mathbb{R})$ по определению. Таковы все привычные нам множества. Чтобы получить множество, не содержащееся в $\mathfrak{B}(\mathbb{R})$, требуются специальные построения. Итак, мы знаем, что все интервалы на прямой принадлежат $\mathfrak{B}(\mathbb{R})$, и $\mathfrak{B}(\mathbb{R})$ — σ-алгебра. Отсюда сразу следует, что $\mathfrak{B}(\mathbb{R})$ содержит любое множество, которое можно получить из интервалов с помощью счётного числа операций объединения или пересечения, а также взятием дополнения.

В частности, $\mathbb{R}$ принадлежит $\mathfrak{B}(\mathbb{R})$. Это сразу следует из свойства (S1) σ-алгебры, но может быть доказано и исходя из свойств (S2), (S3).

$\triangleright$ Интервал $(-n,\, n)$ принадлежит $\mathfrak{A}$, а значит, принадлежит и $\mathfrak{B}(\mathbb{R})$ при любом $n \in \mathbb{N}$, т. е. $(-n,\, n) \in \mathfrak{B}(\mathbb{R})$. Но $\mathfrak{B}(\mathbb{R})$ — σ-алгебра, и содержит счётное объединение любых своих элементов, поэтому

$$\mathbb{R} = \bigcup_{n=1}^{\infty} (-n, n) \in \mathfrak{B}(\mathbb{R}). \qquad \square$$

Далее, любой интервал вида $(a,\, b]$ (или $[a,\, b)$, или $[a,\, b]$), где $a < b$, принадлежит $\mathfrak{B}(\mathbb{R})$.

$\triangleright$ Интервал $(a,\, b + 1/n)$ принадлежит $\mathfrak{B}(\mathbb{R})$ при любом $n \in \mathbb{N}$. Тогда счётное пересечение этих интервалов

$$(a,\, b] = \bigcap_{n=1}^{\infty} \left(a,\, b + \frac{1}{n}\right)$$

по свойству (S4) также принадлежит $\mathfrak{B}(\mathbb{R})$. $\square$

Упражнение. Докажите, что $(a,\, b] = \bigcap\limits_{n=1}^{\infty} (a,\, b + 1/n)$ по определению пересечения множеств: $x \in A \cap B$ тогда и только тогда, когда $x \in A$ и $x \in B$.

Любое одноточечное подмножество $\{b\} \subset \mathbb{R}$ принадлежит $\mathfrak{B}(\mathbb{R})$.

$\triangleright$ Действительно, $\{b\} = (a,\, b] \setminus (a,\, b)$, а разность $A \setminus B = A \cap \overline{B}$ двух множеств из σ-алгебры снова принадлежит σ-алгебре. $\square$

Упражнение. Докажите, что множества вида $(a_1,\, b_1) \cup (a_2,\, b_2)$ принадлежат $\mathfrak{B}(\mathbb{R})$, что множество натуральных чисел $\mathbb{N}$ принадлежит $\mathfrak{B}(\mathbb{R})$, множество рациональных чисел $\mathbb{Q}$ принадлежит $\mathfrak{B}(\mathbb{R})$.

3. Борелевская σ-алгебра в $\mathbb{R}^n$ строится совершенно так же, как в $\mathbb{R}$. Это должна быть минимальная σ-алгебра, содержащая все множества вида $(a_1,\, b_1) \times \ldots \times (a_n,\, b_n)$ — уже не интервалы, как в $\mathbb{R}$, а прямоугольники в $\mathbb{R}^2$, параллелепипеды в $\mathbb{R}^3$ и т. д. Вместе с ними $\mathfrak{B}(\mathbb{R}^n)$ содержит любые множества, являющиеся «предельными» для объединений измельчающихся прямоугольников. Например, круг в $\mathbb{R}^2$ является борелевским множеством — можно изнутри или снаружи приблизить его объединениями прямоугольников.

Итак, мы определили специальный класс $\mathcal{F}$ подмножеств пространства элементарных исходов Ω, названный σ-алгеброй событий, причём применение счётного числа любых операций (объединений, пересечений, дополнений) к множествам из $\mathcal{F}$ снова дает множество из $\mathcal{F}$, т. е. не выводит за рамки этого класса. Событиями будем называть только множества $A \in \mathcal{F}$.

Определим теперь понятие вероятности как функции, определённой на множестве событий (функции, которая каждому событию ставит в соответствие число — вероятность этого события).

А чтобы читателю сразу стало понятно, о чём пойдёт речь, добавим: вероятность мы определим как неотрицательную нормированную меру, заданную на σ-алгебре $\mathcal{F}$ подмножеств Ω. Следующий параграф познакомит нас с понятиями меры и вероятностной меры.

§ 2. Мера и вероятностная мера

Мера как неотрицательная σ-аддитивная функция множеств.

Определение 14. Пусть Ω — некоторое множество и $\mathcal{F}$ — σ-алгебра его подмножеств. Функция $\mu : \mathcal{F} \to \mathbb{R} \cup \{+\infty\}$ называется мерой на $(\Omega, \mathcal{F})$, если она удовлетворяет условиям:

(μ1) для любого множества $A \in \mathcal{F}$ его мера неотрицательна: $\mu(A) \geqslant 0$;

(μ2) для любого счётного набора попарно непересекающихся множеств $A_1, A_2, A_3, \ldots \in \mathcal{F}$ (т. е. такого, что $A_i \cap A_j = \varnothing$ при всех $i \neq j$) мера их объединения равна сумме их мер:

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$$

(«счётная аддитивность» или «σ-аддитивность» меры).

Упражнение. Зачем в свойстве (μ2) требуется, чтобы события не пересекались? Может ли какая-нибудь функция $\mu : \mathcal{F} \to \mathbb{R}$ удовлетворять свойству $\mu(A \cup B) = \mu(A) + \mu(B)$ при любых событиях A и B?

Упражнение. Указать область определения и область значений функции μ. Для каких $A \subset \Omega$ определено значение $\mu(A)$?

Пример 18. Пусть $\Omega = \{a, b, c\}$, $\mathcal{F} = 2^{\Omega}$ — множество всех подмножеств Ω. Зададим меру μ на $\mathcal{F}$ так: $\mu\{a\} = 3$, $\mu\{b\} = 17$, $\mu\{c\} = 1$, $\mu\{a, b\} = 20$, $\mu\{a, c\} = 4$, $\mu\{b, c\} = 18$, $\mu\{a, b, c\} = 21$, $\mu(\varnothing) = 0$. Для краткости записи мы вместо $\mu(\{a\})$ писали всюду $\mu\{a\}$.

Пример 19. Пусть $\Omega = \mathbb{N}$, $\mathcal{F} = 2^{\mathbb{N}}$ — множество всех подмножеств натурального ряда. Зададим меру μ на $\mathcal{F}$ так: $\mu(A) = |A|$ — число элементов в множестве A ($\mu(A) = \infty$, если множество A не является конечным).

[6] Henri Léon Lebesgue (28.06.1875 — 26.07.1941, France)

Пример 20 (мера Лебега[6]). Когда мы говорили о геометрической вероятности, мы использовали термин «мера области A в $\mathbb{R}^m$», имея в виду «длину» на прямой, «площадь» на плоскости, «объем» в трёхмерном пространстве. Являются ли все эти «длины-площади-объемы» настоящими мерами в смысле определения 14? Мы решим этот вопрос для прямой, оставляя плоскость и пространство большей размерности читателю.

Замечание 6. Если вам уже расхотелось читать дальше, сообщаем: мерой Лебега в задачниках и учебниках называют как раз «длину-площадь-объем», так что всё в порядке, дальнейшее до п. 2 можно смело пропустить.

Рассмотрим вещественную прямую с σ-алгеброй борелевских множеств. Эта σ-алгебра, по определению, есть наименьшая σ-алгебра, содержащая любые интервалы. Для каждого интервала $(a, b) \subset \mathbb{R}$ число $b - a$ назовём длиной интервала (a, b).

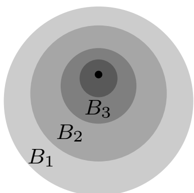
Мы не станем доказывать следующее утверждение:

Лемма 1. *Существует единственная мера λ на $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$, значение которой на любом интервале равно его длине: $\lambda(a, b) = b - a$. Эта мера называется мерой Лебега.*

Замечание 7. Это утверждение является следствием теоремы Каратеодори[7] о продолжении меры с алгебры на σ-алгебру, применительно к $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$.

Нам пригодится свойство, которым обладает любая мера. Это свойство непрерывности меры иногда называют аксиомой непрерывности, имея в виду, что ею можно заменить $(\mu 2)$ в определении 14.

Лемма 2 (свойство непрерывности меры). *Пусть дана убывающая последовательность $B_1 \supseteq B_2 \supseteq B_3 \supseteq \ldots$ множеств из $\mathcal{F}$ такая, что $\mu(B_1) < \infty$ и $B = \bigcap\limits_{n=1}^{\infty} B_n$. Тогда $\mu(B) = \lim\limits_{n\to\infty} \mu(B_n)$.*



Доказательство. Обозначим через C_n кольца: $C_n = B_n \setminus B_{n+1}$. Множества $B, C_1, C_2, C_3, \ldots$ попарно не пересекаются. Тогда из представлений

$$B_1 = B \cup \left(\bigcup_{i=1}^{\infty} C_i \right), \qquad B_n = B \cup \left(\bigcup_{i=n}^{\infty} C_i \right)$$

вытекают, в силу аксиомы $(\mu 2)$, соответствующие равенства и для мер:

$$\mu(B_1) = \mu(B) + \sum_{i=1}^{\infty} \mu(C_i), \qquad \mu(B_n) = \mu(B) + \sum_{i=n}^{\infty} \mu(C_i).$$

Первая сумма $\sum\limits_{i=1}^{\infty} \mu(C_i)$ в силу условия $\mu(B_1) < \infty$ есть сумма абсолютно сходящегося ряда (составленного из неотрицательных слагаемых). Из схо-

[7] Constantin Carathéodory (13.09.1873 — 2.02.1950, Germany)

димости этого ряда следует, что «хвост» ряда, равный $\sum\limits_{i=n}^{\infty}\mu(C_i)$, стремится к нулю при $n\to\infty$. Поэтому

$$\mu(B_n)=\mu(B)+\sum_{i=n}^{\infty}\mu(C_i)\underset{n\to\infty}{\longrightarrow}\mu(B)+0=\mu(B).\qquad\square$$

В полезности этого свойства легко убедиться упражнениями.

Упражнение. Используя аксиому непрерывности меры для убывающей последовательности множеств $B_n=(x-1/n,\,x+1/n)$, доказать, что мера Лебега одноточечного подмножества $\{x\}$ вещественной прямой равна нулю: $\lambda\{x\}=0$. Используя этот факт, доказать, что $\lambda(\mathbb{N})=0$, $\lambda(\mathbb{Z})=0$, $\lambda(\mathbb{Q})=0$, $\lambda(a,\,b)=\lambda[a,\,b]$.

Замечание 8. В отсутствие предположения $\mu(B_1)<\infty$ (или $\mu(B_n)<\infty$ для некоторого $n\geqslant 1$), заставляющего меры вложенных множеств быть конечными, свойство $\mu(B)=\lim\limits_{n\to\infty}\mu(B_n)$ может не выполняться.

Например, зададим меру на $\mathfrak{B}(\mathbb{R})$ так: $\mu(B)=0$, если B не более чем счётно, иначе $\mu(B)=\infty$. Тогда для множеств $B_n=(x-1/n,\,x+1/n)$ имеем:

$$B=\bigcap_{n=1}^{\infty}B_n=\{x\},\quad \mu(B_n)=\infty\not\to\mu(B)=0.$$

И вот наконец мы в состоянии определить понятие вероятности.

Вероятность как нормированная мера.

Определение 15. Пусть Ω — непустое множество и $\mathcal{F}$ — σ-алгебра его подмножеств. Мера $\mu:\mathcal{F}\to\mathbb{R}$ называется нормированной, если $\mu(\Omega)=1$. Другое название нормированной меры — вероятность или вероятностная мера.

То же самое ещё раз и подробно:

Определение 16. Пусть Ω — пространство элементарных исходов, $\mathcal{F}$ — σ-алгебра его подмножеств (событий). Вероятностью или вероятностной мерой на $(\Omega,\mathcal{F})$ называется функция $\mathsf{P}:\mathcal{F}\to\mathbb{R}$, обладающая свойствами:

(P1) для любого события $A\in\mathcal{F}$ выполняется неравенство $\mathsf{P}(A)\geqslant 0$;

(P2) для любого счётного набора попарно несовместных событий $A_1,\,A_2,\,A_3,\,\ldots\in\mathcal{F}$ имеет место равенство

$$\mathsf{P}\Bigl(\bigcup_{i=1}^{\infty}A_i\Bigr)=\sum_{i=1}^{\infty}\mathsf{P}(A_i);$$

(P3) вероятность достоверного события равна единице: $\mathsf{P}(\Omega)=1$.

Свойства (P1) — (P3) называют аксиомами вероятности.

Определение 17. Тройка $\langle\Omega,\mathcal{F},\mathsf{P}\rangle$, в которой Ω — пространство элементарных исходов, $\mathcal{F}$ — σ-алгебра его подмножеств и P — вероятностная мера на $\mathcal{F}$, называется вероятностным пространством.

Докажем свойства вероятности, вытекающие из аксиом. Ниже мы не будем всякий раз оговаривать, что имеем дело только с событиями.

Свойство 0. $\mathsf{P}(\varnothing) = 0$.

$\triangleright$ События $A_i = \varnothing$, где $i \geqslant 1$, попарно несовместны, и их объединение есть также пустое множество. По аксиоме (P2),

$$\mathsf{P}(\varnothing) = \sum_{i=1}^{\infty} \mathsf{P}(A_i) = \sum_{i=1}^{\infty} \mathsf{P}(\varnothing).$$

Это возможно только в случае $\mathsf{P}(\varnothing) = 0$. $\square$

Аксиома счётной аддитивности вероятности (P2) тем более верна для конечного набора попарно несовместных событий.

Свойство 1. Для любого конечного набора попарно несовместных событий $A_1, \ldots, A_n \in \mathcal{F}$ имеет место равенство

$$\mathsf{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathsf{P}(A_i).$$

$\triangleright$ Положим $A_i = \varnothing$ при любом $i > n$. Вероятности этих событий, по свойству 0, равны нулю. События $A_1, \ldots, A_n, \varnothing, \varnothing, \varnothing, \ldots$ попарно несовместны, и по аксиоме (P2),

$$\mathsf{P}\left(\bigcup_{i=1}^{n} A_i\right) = \mathsf{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathsf{P}(A_i) = \sum_{i=1}^{n} \mathsf{P}(A_i). \qquad \square$$

Сразу несколько следствий можно получить из этого свойства.

Свойство 2. Для любого события A выполнено: $\mathsf{P}(\overline{A}) = 1 - \mathsf{P}(A)$.

$\triangleright$ Поскольку $A \cup \overline{A} = \Omega$, и события A и $\overline{A}$ несовместны, из аксиомы (P3) и свойства 1 получим $\mathsf{P}(A) + \mathsf{P}(\overline{A}) = \mathsf{P}(\Omega) = 1$. $\square$

Свойство 3. Если $A \subseteq B$, то $\mathsf{P}(B \setminus A) = \mathsf{P}(B) - \mathsf{P}(A)$.

$\triangleright$ Представим B в виде объединения двух несовместных событий: $B = A \cup (B \setminus A)$. По свойству 1, $\mathsf{P}(B) = \mathsf{P}(A) + \mathsf{P}(B \setminus A)$. $\square$

Сразу же заметим, что по аксиоме (P1) выражение в правой части равенства $\mathsf{P}(B) = \mathsf{P}(A) + \mathsf{P}(B \setminus A)$ больше либо равно $\mathsf{P}(A)$, что доказывает следующее свойство монотонности вероятности.

Свойство 4. Если $A \subseteq B$, то $\mathsf{P}(A) \leqslant \mathsf{P}(B)$.

Свойство 5. Для любого события A выполнено: $0 \leqslant \mathsf{P}(A) \leqslant 1$.

$\triangleright$ $\mathsf{P}(A) \geqslant 0$ по (P1). А так как $A \subseteq \Omega$, то $\mathsf{P}(A) \leqslant \mathsf{P}(\Omega) = 1$. $\square$

Свойство 6. Всегда $\mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(A \cap B)$.

$\triangleright$ Имеем $A \cap B \subseteq B$, поэтому $\mathsf{P}(B \setminus (A \cap B)) = \mathsf{P}(B) - \mathsf{P}(A \cap B)$ по свойству 3. Но $A \cup B = A \cup (B \setminus (A \cap B))$, причём A и $B \setminus (A \cap B)$ несовместны. Снова пользуясь свойством 1, получим:

$$\mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B \setminus (A \cap B)) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(A \cap B). \qquad \square$$

Из этого свойства и аксиомы (P1) следуют два полезных свойства. Свойство 8 читатель докажет с помощью свойства 7.

Свойство 7. Всегда $\mathsf{P}(A \cup B) \leqslant \mathsf{P}(A) + \mathsf{P}(B)$.

Свойство 8. Совершенно всегда $\mathsf{P}(A_1 \cup \ldots \cup A_n) \leqslant \sum\limits_{i=1}^{n} \mathsf{P}(A_i)$.

Следующее свойство называют формулой *включения и исключения*. Она оказывается весьма полезной в случае, когда для вычисления вероятности некоторого события A нельзя разбить это событие на удобные попарно несовместные события, но удаётся разбить событие A на простые составляющие, которые, однако, совместны.

Свойство 9. Для любого конечного набора событий $A_1, \ldots, A_n$ имеет место равенство:

$$\mathsf{P}(A_1 \cup \ldots \cup A_n) = \sum_{i=1}^{n} \mathsf{P}(A_i) - \sum_{i<j} \mathsf{P}(A_i A_j) + \\ + \sum_{i<j<m} \mathsf{P}(A_i A_j A_m) - \ldots + (-1)^{n-1} \mathsf{P}(A_1 A_2 \ldots A_n). \quad (2)$$

▷ Воспользуемся методом математической индукции. Базис индукции при $n = 2$ — свойство 6. Пусть свойство 9 верно при $n = k - 1$. Докажем, что тогда оно верно при $n = k$. По свойству 6,

$$\mathsf{P}\left(\bigcup_{i=1}^{k} A_i\right) = \mathsf{P}\left(\bigcup_{i=1}^{k-1} A_i\right) + \mathsf{P}\left(A_k\right) - \mathsf{P}\left(A_k \cap \bigcup_{i=1}^{k-1} A_i\right). \quad (3)$$

По предположению индукции, первое слагаемое в правой части (3) равно

$$\mathsf{P}\left(\bigcup_{i=1}^{k-1} A_i\right) = \sum_{i=1}^{k-1} \mathsf{P}(A_i) - \sum_{1\leqslant i<j\leqslant k-1} \mathsf{P}(A_i A_j) + \\ + \sum_{1\leqslant i<j<m\leqslant k-1} \mathsf{P}(A_i A_j A_m) - \ldots + (-1)^{k-2} \mathsf{P}(A_1 A_2 \ldots A_{k-1}). \quad (4)$$

Вычитаемое в правой части (3) равно

$$\mathsf{P}\left(A_k \cap \bigcup_{i=1}^{k-1} A_i\right) = \mathsf{P}\left(\bigcup_{i=1}^{k-1} A_i A_k\right) = \sum_{i=1}^{k-1} \mathsf{P}(A_i A_k) - \sum_{1\leqslant i<j\leqslant k-1} \mathsf{P}(A_i A_j A_k) + \\ + \sum_{1\leqslant i<j<m\leqslant k-1} \mathsf{P}(A_i A_j A_m A_k) - \ldots + (-1)^{k-2} \mathsf{P}(A_1 A_2 \ldots A_{k-1} A_k). \quad (5)$$

Упражнение. Подставить (4), (5) в (3) и довести до конца шаг индукции. □

Приведём пример задачи, в которой использование свойства 9 — самый простой путь решения. Это известная «задача о рассеянной секретарше».

Пример 21. Есть n писем и n подписанных конвертов. Письма раскладываются в конверты наудачу по одному. Найти вероятность того, что хотя бы одно письмо попадет в предназначенный ему конверт, и предел этой вероятности при $n \to \infty$.

Решение. Пусть событие A_i, $i = 1, \ldots, n$, означает, что i-е письмо попало в свой конверт. Тогда

$$A = \{\textit{хотя бы одно письмо попало в свой конверт}\} = A_1 \cup \ldots \cup A_n.$$

Так как события $A_1, \ldots, A_n$ совместны, придётся использовать формулу (2). По классическому определению вероятности вычислим вероятности всех событий A_i и их пересечений. Элементарными исходами будут всевозможные перестановки (размещения) n писем по n конвертам. Их общее число есть $|\Omega| = n!$, и событию A_i благоприятны $(n-1)!$ из них, а именно любые перестановки всех писем, кроме i-го, лежащего в своём конверте. Поэтому $\mathsf{P}(A_i) = (n-1)!/n! = 1/n$ для всех i. Совершенно так же получим, что при любых $i \neq j$

$$\mathsf{P}(A_iA_j) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}.$$

Вероятность пересечения любых трёх событий равна

$$\mathsf{P}(A_iA_jA_m) = \frac{(n-3)!}{n!} = \frac{1}{n(n-1)(n-2)}.$$

Аналогично посчитаем вероятности пересечений любого другого числа событий, в том числе $\mathsf{P}(A_1 \ldots A_n) = 1/n!$

Вычислим количество слагаемых в каждой сумме в формуле (2). Например, в сумме по $1 \leqslant i < j < m \leqslant n$ ровно C_n^3 слагаемых — ровно столько трёхэлементных множеств можно образовать из n элементов, и каждое такое множество $\{i, j, m\}$ встречается в индексах данной суммы единожды. Подставляя все вероятности в формулу (2), получим:

$$\mathsf{P}(A) = n\,\frac{1}{n} - C_n^2\,\frac{1}{n(n-1)} + C_n^3\,\frac{1}{n(n-1)(n-2)} - \ldots + (-1)^{n-1}\,\frac{1}{n!} =$$

$$= 1 - \frac{1}{2!} + \frac{1}{3!} - \ldots + (-1)^{n-1}\,\frac{1}{n!} \longrightarrow 1 - e^{-1} \text{ при } n \to \infty$$

Упражнение. Выписать разложение e^{-1} в ряд Тейлора и убедиться в том, что $\mathsf{P}(A) \longrightarrow 1 - e^{-1}$ при $n \to \infty$.

ГЛАВА 4

Условная вероятность, независимость

Здесь является вопрос, возбуждённый несколькими философами, относительно влияния прошлого на вероятность будущего.

П. Лаплас, Опыт философии теории вероятностей

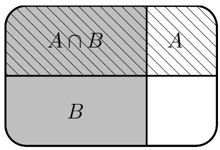
§ 1. Условная вероятность

Пример 22. Игральная кость подбрасывается один раз. Известно, что выпало более трёх очков. Какова при этом вероятность того, что выпало чётное число очков?

Зная, что выпало более трёх очков, мы можем сузить множество всех возможных элементарных исходов до трёх одинаково вероятных исходов: $\Omega = \{4, 5, 6\}$, из которых событию $A = \{$выпало чётное число очков$\}$ благоприятствуют ровно два: $A = \{4, 6\}$. Поэтому $\mathsf{P}(A) = 2/3$.

Посмотрим на вопрос с точки зрения первоначального эксперимента. Пространство элементарных исходов при одном подбрасывании кубика состоит из шести точек: $\Omega = \{1, 2, 3, 4, 5, 6\}$. Слова «известно, что выпало более трёх очков» означают, что в эксперименте произошло событие $B = \{4, 5, 6\}$. Слова «какова при этом вероятность того, что выпало чётное число очков?» означают, что нас интересует, в какой доле случаев при осуществлении B происходит и A. Вероятность события A, вычисленную в предположении, что нечто о результате эксперимента уже известно (событие B произошло), мы будем обозначать через $\mathsf{P}(A \,|\, B)$.

Мы хотим найти, какую часть составляют исходы, благоприятствующие A внутри B (т. е. одновременно A и B), среди исходов, благоприятствующих B.



$$\mathsf{P}(A \,|\, B) = \frac{2}{3} = \frac{2/6}{3/6} = \frac{\mathsf{P}(A \cap B)}{\mathsf{P}(B)}.$$

Мы пришли к выражению, которое можно считать определением условной вероятности.

Определение 18. Условной вероятностью события A при условии, что произошло событие B, называется число

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

Условная вероятность определена только в случае, когда $\mathsf{P}(B) > 0$.

Это определение бывает полезно использовать не для вычисления условной вероятности, а для последовательного вычисления вероятности нескольким событиям случиться одновременно, если известны соответствующие условные вероятности. А именно, справедливы следующие «теоремы умножения вероятностей».

Теорема 6. *Если* $\mathsf{P}(B) > 0$ *и* $\mathsf{P}(A) > 0$*, то*

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

Теорема 7. *Для любых событий* $A_1, \ldots, A_n$ *верно равенство:*

$$\mathsf{P}(A_1 \ldots A_n) = \mathsf{P}(A_1)\,\mathsf{P}(A_2 \mid A_1)\,\mathsf{P}(A_3 \mid A_1 A_2) \cdot \ldots \cdot \mathsf{P}(A_n \mid A_1 \ldots A_{n-1}),$$

если все участвующие в нём условные вероятности определены.

Упражнение. Доказать теорему 7 методом математической индукции.

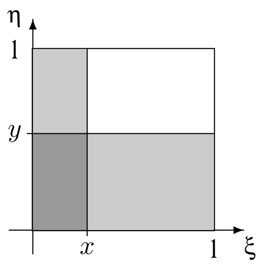
Упражнение. Доказать, что все условные вероятности в теореме 7 определены тогда и только тогда, когда $\mathsf{P}(A_1 \ldots A_{n-1}) > 0$.

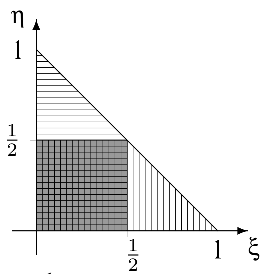
§ 2. Независимость

Определение 19. События A и B называются независимыми, если $\mathsf{P}(A \cap B) = \mathsf{P}(A)\mathsf{P}(B)$.

Пример 23. 1. Точка с координатами ξ, η бросается наудачу в единичный квадрат со сторонами, параллельными осям координат. Доказать, что для любых $x, y \in \mathbb{R}$ события $A = \{\xi < x\}$ и $B = \{\eta < y\}$ независимы.

2. Точка с координатами ξ, η бросается наудачу в треугольник с вершинами $(1, 0)$, $(0, 0)$ и $(0, 1)$. Доказать, что события $A = \{\xi < 1/2\}$ и $B = \{\eta < 1/2\}$ зависимы.





1. Рассмотрим $x, y \in [0, 1]$ (разобрать остальные случаи). Тогда $\mathsf{P}(A) = x$, $\mathsf{P}(B) = y$, $\mathsf{P}(A \cap B) = xy$, т. е. события $A = \{\xi < x\}$ и $B = \{\eta < y\}$ независимы.

2. Вычислив соответствующие площади в треугольнике, получим: $\mathsf{P}(A) = 3/4$, $\mathsf{P}(B) = 3/4$, $\mathsf{P}(A \cap B) = 1/2 \neq (3/4)^2$, т. е. события $A = \{\xi < 1/2\}$ и $B = \{\eta < 1/2\}$ зависимы.

Естественно считать события A и B независимыми, когда условная вероятность A при условии, что B произошло, остаётся такой же, как и безусловная. Убедимся, что этим свойством обладают события, независимые согласно определению 19.

Свойство 4. *Пусть* $\mathsf{P}(B) > 0$. *Тогда события A и B независимы тогда и только тогда, когда* $\mathsf{P}(A \mid B) = \mathsf{P}(A)$.

Если $\mathsf{P}(A) > 0$, *то события A и B независимы тогда и только тогда, когда* $\mathsf{P}(B \mid A) = \mathsf{P}(B)$.

Упражнение. Доказать, пользуясь определением условной вероятности.

Свойство 5. *Пусть события A и B несовместны. Тогда независимыми они будут только в том случае, если* $\mathsf{P}(A) = 0$ *или* $\mathsf{P}(B) = 0$.

Это свойство (а вы его доказали?) означает, что в невырожденном случае (когда вероятности событий положительны) несовместные события не могут быть независимыми. Зависимость между ними — просто причинно-следственная: если $A \cap B = \varnothing$, то $A \subseteq \overline{B}$, т. е. при выполнении A событие B не происходит. Это свойство можно сформулировать иначе: *в невырожденном случае независимые события просто обязаны пересекаться, т. е. быть совместными.*

Упражнение. Доказать с помощью свойства монотонности вероятности, что событие A, вероятность которого равна нулю или единице, не зависит ни от какого события B, в том числе и от самого себя.

Свойство 6. *Если события A и B независимы, то независимы и события A и $\overline{B}$, $\overline{A}$ и B, $\overline{A}$ и $\overline{B}$.*

Доказательство. Так как $A = (A \cap B) \cup (A \cap \overline{B})$, и события $A \cap B$ и $A \cap \overline{B}$ несовместны, то $\mathsf{P}(A) = \mathsf{P}(A \cap B) + \mathsf{P}(A \cap \overline{B})$. Поэтому $\mathsf{P}(A \cap \overline{B}) = \mathsf{P}(A) - \mathsf{P}(A \cap B) = \mathsf{P}(A) - \mathsf{P}(A)\mathsf{P}(B) = \mathsf{P}(A)(1 - \mathsf{P}(B)) = \mathsf{P}(A)\mathsf{P}(\overline{B})$.

Вывести отсюда остальные утверждения. □

Если у нас не два, а большее число событий, выполнение только одного равенства $\mathsf{P}(A_1 \cap \ldots \cap A_n) = \mathsf{P}(A_1) \cdot \ldots \cdot \mathsf{P}(A_n)$ вовсе не означает независимости этих событий. Например, при таком равенстве события A_1 и A_2 вполне могут оказаться зависимыми.

Пример 24. Пусть $0 < \mathsf{P}(A) < 1$. События $A_1 = A$, $A_2 = A$, $A_3 = \emptyset$ обладают свойством

$$0 = \mathsf{P}(A_1 \cap A_2 \cap A_3) = \mathsf{P}(A_1) \cdot \mathsf{P}(A_2) \cdot \mathsf{P}(A_3) = 0,$$

что не мешает событиям A_1 и A_2 быть зависимыми:

$$\mathsf{P}(A_1 \cap A_2) = \mathsf{P}(A) \neq \mathsf{P}(A_1) \cdot \mathsf{P}(A_2) = \big(\mathsf{P}(A)\big)^2.$$

Хотелось бы независимостью нескольких событий считать такое свойство, при котором любые комбинации этих событий будут независимы между собой: например, независимы $A_1 \cap A_2$ и $A_3 \cup A_4 \cup A_5$.

Определение 20. События $A_1, \ldots, A_n$ называются независимыми в совокупности, если для любого $1 \leqslant k \leqslant n$ и любого набора различных меж собой индексов $1 \leqslant i_1, \ldots, i_k \leqslant n$ имеет место равенство:

$$\mathsf{P}(A_{i_1} \cap \ldots \cap A_{i_k}) = \mathsf{P}(A_{i_1}) \cdot \ldots \cdot \mathsf{P}(A_{i_k}). \tag{6}$$

Замечание 9. Если события $A_1, \ldots, A_n$ независимы в совокупности, то они попарно независимы, т. е. любые два события A_i, A_j независимы. Достаточно в равенстве (6) взять $k = 2$. Обратное, как показывает следующий пример, неверно: из попарной независимости не вытекает независимость в совокупности.

Пример 25 (пример Бернштейна[8]). Рассмотрим правильный тетраэдр, три грани которого окрашены соответственно в красный, синий, зелёный цвета, а четвёртая грань содержит все три цвета. Событие A (соответственно B, C) означает, что выпала грань, содержащая красный (соответственно синий, зелёный) цвета.

Вероятность каждого из этих событий равна $1/2$, так как каждый цвет есть на двух гранях из четырёх. Вероятность пересечения любых двух из них равна $1/4$, так как только одна грань из четырёх содержит два цвета. А так как $1/4 = 1/2 \cdot 1/2$, то все события попарно независимы.

Но вероятность пересечения всех трёх тоже равна $1/4$, а не $1/8$, т. е. события не являются независимыми в совокупности.

Заметьте, что равенство (6) выполнено при $k = 2$, но не при $k = 3$.

§ 3. Формула полной вероятности

Пример 26. Есть три завода, производящих одну и ту же продукцию. При этом первый завод производит 25%, второй завод — 35% и третий — 40% всей производимой продукции. Брак составляет 5% от продукции первого завода, 3% от продукции второго и 4% от продукции третьего завода. Вся продукция смешивается и поступает в продажу. Найти а) вероятность купить бракованное изделие; б) условную вероятность того, что купленное изделие изготовлено первым заводом, если это изделие бракованное.

Первая вероятность равна доле бракованных изделий в объеме всей продукции, т. е. $0{,}05{\cdot}0{,}25 + 0{,}03{\cdot}0{,}35 + 0{,}04{\cdot}0{,}4$. Вторая вероятность равна доле брака первого завода среди всего брака, т. е.

$$\frac{0{,}05{\cdot}0{,}25}{0{,}05{\cdot}0{,}25 + 0{,}03{\cdot}0{,}35 + 0{,}04{\cdot}0{,}4}.$$

[8] Сергей Натанович Бернштейн (5.03.1880 — 26.10.1968)

Заметим, что найти эти вероятности можно безо всякого знания теории вероятностей вообще. Достаточно элементарного здравого смысла.

Определение 21. Конечный или счётный набор попарно несовместных событий $H_1, H_2, \ldots$ таких, что $\mathsf{P}(H_i) > 0$ для всех i и $H_1 \cup H_2 \cup \ldots = \Omega$, называется полной группой событий или разбиением пространства Ω.

События $H_1, H_2, \ldots$, образующие полную группу событий, часто называют гипотезами. При подходящем выборе гипотез для произвольного события A могут быть сравнительно просто вычислены $\mathsf{P}(A \mid H_i)$ (вероятность событию A произойти при выполнении «гипотезы» H_i) и собственно $\mathsf{P}(H_i)$ (вероятность выполнения «гипотезы» H_i). Как, используя эти данные, посчитать вероятность события A?

Теорема 8 (формула полной вероятности). *Пусть дана полная группа событий H_1, H_2, … Тогда вероятность любого события A может быть вычислена по формуле:*

$$\mathsf{P}(A) = \sum_{i=1}^{\infty} \mathsf{P}(H_i)\, \mathsf{P}(A \mid H_i).$$

Доказательство. Заметим, что

$$A = A \cap \Omega = A \cap \left(\bigcup_{i=1}^{\infty} H_i \right) = \bigcup_{i=1}^{\infty} (A \cap H_i),$$

и события $(A \cap H_1)$, $(A \cap H_2)$, … попарно несовместны. Поэтому

$$\mathsf{P}(A) = \mathsf{P}\left(\bigcup_{i=1}^{\infty} (A \cap H_i) \right) = \sum_{i=1}^{\infty} \mathsf{P}(A \cap H_i) = \sum_{i=1}^{\infty} \mathsf{P}(H_i)\, \mathsf{P}(A \mid H_i).$$

Во втором равенстве мы использовали σ-аддитивность вероятностной меры (а что это?), а в третьем — теорему 6 умножения вероятностей. □

§ 4. Формула Байеса[9]

Теорема 9 (формула Байеса). *Пусть $H_1, H_2, \ldots$ — полная группа событий, и A — некоторое событие, вероятность которого положительна. Тогда условная вероятность того, что имело место событие H_k, если в результате эксперимента наблюдалось событие A, может быть вычислена по формуле:*

$$\mathsf{P}(H_k \mid A) = \frac{\mathsf{P}(H_k)\, \mathsf{P}(A \mid H_k)}{\sum\limits_{i=1}^{\infty} \mathsf{P}(H_i)\, \mathsf{P}(A \mid H_i)}.$$

[9] Thomas Bayes (1702 — 17.04.1761, England)

Доказательство. По определению условной вероятности,

$$\mathsf{P}(H_k \,|\, A) = \frac{\mathsf{P}(H_k \cap A)}{\mathsf{P}(A)} = \frac{\mathsf{P}(H_k)\,\mathsf{P}(A \,|\, H_k)}{\sum\limits_{i=1}^{\infty} \mathsf{P}(H_i)\,\mathsf{P}(A \,|\, H_i)}. \qquad \square$$

Пример 27. Вернёмся к примеру 26. Изделие выбирается наудачу из всей произведённой продукции. Рассмотрим три гипотезы: $H_i = \{$изделие изготовлено i-м заводом$\}$, $i = 1, 2, 3$. Вероятности этих событий даны: $\mathsf{P}(H_1) = 0{,}25$, $\mathsf{P}(H_2) = 0{,}35$, $\mathsf{P}(H_3) = 0{,}4$.

Пусть $A = \{$изделие оказалось бракованным$\}$. Даны также условные вероятности $P(A \,|\, H_1) = 0{,}05$, $P(A \,|\, H_2) = 0{,}03$, $P(A \,|\, H_3) = 0{,}04$.

Убедитесь, что полученные нами в примере 26 вероятности совпадают с вероятностями, вычисленными по формуле полной вероятности и по формуле Байеса.

Пример 28. Два стрелка подбрасывают монетку и выбирают, кто из них стреляет по мишени (одной пулей). Первый стрелок попадает по мишени с вероятностью 1, второй стрелок — с вероятностью 0,00001. Можно сделать два предположения об эксперименте: $H_1 = \{$стреляет 1-й стрелок$\}$ и $H_2 = \{$стреляет 2-й стрелок$\}$. Априорные (a'priori — «до опыта») вероятности этих гипотез одинаковы: $\mathsf{P}(H_1) = \mathsf{P}(H_2) = 1/2$.

Рассмотрим событие $A = \{$пуля попала в мишень$\}$. Известно, что

$$\mathsf{P}(A \,|\, H_1) = 1, \quad \mathsf{P}(A \,|\, H_2) = 0{,}00001.$$

Поэтому вероятность пуле попасть в мишень

$$\mathsf{P}(A) = 1/2 \cdot 1 + 1/2 \cdot 0{,}00001.$$

Предположим, что событие A произошло. Какова теперь апостериорная (a'posteriori — «после опыта») вероятность каждой из гипотез H_i? Очевидно, что первая из этих гипотез много вероятнее второй (а именно, в 100 000 раз). Действительно,

$$\mathsf{P}(H_1 \,|\, A) = \frac{1/2 \cdot 1}{1/2 \cdot 1 + 1/2 \cdot 0{,}00001} = \frac{1}{1 + 0{,}00001} \approx 0{,}99999,$$

$$\mathsf{P}(H_2 \,|\, A) = \frac{1/2 \cdot 0{,}00001}{1/2 \cdot 1 + 1/2 \cdot 0{,}00001} = \frac{0{,}00001}{1 + 0{,}00001} \approx 0{,}00001.$$

ГЛАВА 5
Схема Бернулли

На дне глубокого сосуда лежат спокойно n шаров.
Поочередно их оттуда таскают двое дураков.
Сия работа им приятна, они таскают t минут,
И, вынув шар, его обратно тотчас немедленно кладут.
Ввиду занятия такого, сколь вероятность велика,
Что первый был глупей второго, когда шаров он вынул k?

Фольклор, СПбГУ

§ 1. Распределение числа успехов в n испытаниях

Определение 22. Схемой Бернулли называется последовательность независимых в совокупности испытаний, в каждом из которых возможны лишь два исхода — «успех» и «неудача», при этом успех в одном испытании происходит с вероятностью $p \in (0, 1)$, а неудача — с вероятностью $q = 1 - p$.

Под независимостью в совокупности испытаний понимается независимость в совокупности любых событий, относящихся к разным испытаниям. В испытаниях схемы Бернулли, когда с одним испытанием можно связать только два взаимоисключающих события, независимость в совокупности испытаний означает, что при любом n независимы в совокупности события $A_1 = \{$успех в первом испытании$\}$, ..., $A_n = \{$успех в n-ом испытании$\}$. Эти события принадлежат одному и тому же пространству элементарных исходов, полученному декартовым произведением бесконечного числа двухэлементных множеств $\{у, н\}$:

$$\Omega = \{(a_1, a_2, \ldots) \mid a_i \in \{у, н\}\}.$$

Здесь буквами «$у$» и «$н$» обозначены успешный и неудачный результаты испытаний соответственно.

Обозначим через ν_n число успехов, случившихся в n испытаниях схемы Бернулли. Эта величина может принимать целые значения от нуля до n в зависимости от результата n испытаний. Например, если все n испытаний завершились неудачей, то величина ν_n равна нулю.

Теорема 10 (формула Бернулли). *Для любого* $k = 0, 1, \ldots, n$ *имеет место равенство:*

$$\mathsf{P}(\nu_n = k) = C_n^k\, p^k\, (1-p)^{n-k} = C_n^k\, p^k\, q^{n-k}.$$

Доказательство. Событие $A = \{\nu_n = k\}$ означает, что в n испытаниях схемы Бернулли произошло ровно k успехов. Рассмотрим один из благоприятствующих событию A элементарных исходов:

$$(\underbrace{у,\ у,\ \ldots,\ у}_{k},\ \underbrace{н, н,\ \ldots,\ н}_{n-k}),$$

когда первые k испытаний завершились успехом, остальные неудачей. Поскольку испытания независимы, вероятность такого элементарного исхода равна $p^k(1-p)^{n-k}$. Другие благоприятствующие событию A элементарные исходы отличаются лишь расположением k успехов на n местах. Есть ровно C_n^k способов расположить k успехов на n местах. Поэтому событие A состоит из C_n^k элементарных исходов, вероятность каждого из которых также равна $p^k(1-p)^{n-k}$. □

Определение 23. Набор чисел $\left\{C_n^k p^k (1-p)^{n-k},\ k = 0, 1, \ldots, n\right\}$ называется биномиальным распределением вероятностей.

§ 2. Номер первого успешного испытания

Рассмотрим схему Бернулли с вероятностью успеха p в одном испытании. Введём величину τ со значениями $1, 2, 3, \ldots$, равную номеру первого успешного испытания.

Теорема 11. *Вероятность того, что первый успех произойдёт в испытании с номером* $k \in \mathbb{N} = \{1, 2, 3, \ldots\}$, *равна* $\mathsf{P}(\tau = k) = pq^{k-1}$.

Доказательство. Вероятность первым $k-1$ испытаниям завершиться неудачей, а последнему — успехом, равна

$$\mathsf{P}(\tau = k) = \mathsf{P}(н,\ \ldots,\ н,\ у) = pq^{k-1}. \qquad \square$$

Определение 24. Набор чисел $\{pq^{k-1},\ k = 1, 2, 3, \ldots\}$ называется геометрическим распределением вероятностей.

Геометрическое распределение вероятностей обладает интересным свойством, которое можно назвать свойством «нестарения».

Теорема 12. *Пусть* $\mathsf{P}(\tau = k) = pq^{k-1}$ *для любого* $k \in \mathbb{N}$. *Тогда для любых неотрицательных целых* n *и* k *имеет место равенство:*

$$\mathsf{P}(\tau > n + k \mid \tau > n) = \mathsf{P}(\tau > k).$$

Если, например, считать величину τ временем безотказной работы (измеряемым целым числом часов) некоторого устройства, то данному равенству можно придать следующее звучание: *вероятность работающему* *устройству проработать ещё сколько-то часов не зависит от того момента, когда мы начали отсчёт времени, или от того, сколько уже работает устройство*. Общепринятое название этого свойства — свойство «отсутствия последействия».

Доказательство. По определению условной вероятности,

$$\mathsf{P}(\tau > n+k \,|\, \tau > n) = \frac{\mathsf{P}(\tau > n+k,\, \tau > n)}{\mathsf{P}(\tau > n)} = \frac{\mathsf{P}(\tau > n+k)}{\mathsf{P}(\tau > n)}. \qquad (7)$$

Последнее равенство следует из того, что событие $\{\tau > n+k\}$ влечёт событие $\{\tau > n\}$, поэтому пересечение этих событий есть $\{\tau > n+k\}$. Найдём для целого $m \geqslant 0$ вероятность $\mathsf{P}(\tau > m)$:

$$\mathsf{P}(\tau > m) = \sum_{i=m+1}^{\infty} \mathsf{P}(\tau = i) = \sum_{i=m+1}^{\infty} p\, q^{i-1} = \frac{p\, q^m}{1-q} = q^m.$$

Можно получить $\mathsf{P}(\tau > m)$ и ещё проще: событие $\{\tau > m\}$ означает в точности, что в схеме Бернулли первые m испытаний завершились «неудачами», т. е. его вероятность равна q^m. Возвращаясь к (7), получим

$$\mathsf{P}(\tau > n+k \,|\, \tau > n) = \frac{\mathsf{P}(\tau > n+k)}{\mathsf{P}(\tau > n)} = \frac{q^{n+k}}{q^n} = q^k = \mathsf{P}(\tau > k). \qquad \square$$

§ 3. Независимые испытания с несколькими исходами

Рассмотрим следующий пример, когда из двух очень похожих вопросов на один можно ответить, пользуясь формулой Бернулли, а для другого этой формулы оказывается недостаточно.

Пример 29. Игральная кость подбрасывается пятнадцать раз. Найти вероятности следующих событий: а) выпадет ровно десять троек; б) выпадет ровно десять троек и три единицы.

Решение. а) Есть пятнадцать испытаний схемы Бернулли с вероятностью успеха $1/6$ (здесь успех — выпадение тройки). Вероятность десяти успехов в пятнадцати испытаниях равна

$$\mathsf{P}(\nu_{15} = 10) = C_{15}^{10}\, (1/6)^{10}\, (1-1/6)^{15-10}.$$

б) Здесь каждое испытание имеет три, а не два исхода: выпадение тройки, выпадение единицы, выпадение остальных граней. Воспользоваться формулой для числа успехов в схеме Бернулли не удаётся — перед нами уже не схема Бернулли.

Осталось изобрести формулу для подсчёта вероятности каждому исходу в нескольких независимых испытаниях выпасть нужное число раз, если в одном испытании возможно не два, а более исходов.

Пусть в одном испытании возможны m исходов: $1, 2, \ldots, m$, и исход i в одном испытании случается с вероятностью p_i, где $p_1 + \ldots + p_m = 1$.

Обозначим через $P(n_1, \ldots, n_m)$ искомую вероятность того, что в $n = n_1 + \ldots + n_m$ независимых испытаниях исход 1 появился n_1 раз, исход 2 — n_2 раз, и т. д., исход m — n_m раз.

Теорема 13. *Для любого n и любых целых $n_1 \geqslant 0, \ldots, n_m \geqslant 0$ таких, что $n_1 + \ldots + n_m = n$, верна формула:*

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

Доказательство. Рассмотрим один элементарный исход, благоприятствующий выпадению n_1 единиц, n_2 двоек, …, n_m раз m-ок:

$$(\underbrace{1, \ldots, 1}_{n_1}, \underbrace{2, \ldots, 2}_{n_2}, \ldots, \underbrace{m, \ldots, m}_{n_m}).$$

Это результат n экспериментов, когда все нужные исходы появились в некотором заранее заданом порядке. Вероятность такого результата n независимых испытаний равна $p_1^{n_1} \cdot \ldots \cdot p_m^{n_m}$.

Остальные благоприятные исходы отличаются лишь расположением чисел $1, 2, \ldots, m$ на n местах. Число таких исходов равно числу способов расставить на n местах n_1 единиц, n_2 двоек, …, n_m чисел m, т. е.

$$C_n^{n_1} \cdot C_{n-n_1}^{n_2} \cdot C_{n-n_1-n_2}^{n_3} \cdot \ldots \cdot C_{n-n_1-\ldots-n_{m-1}}^{n_m} = \frac{n!}{n_1! \ldots n_m!}. \quad \square$$

Теперь мы можем вернуться к примеру 29(б) и выписать ответ: так как вероятности выпадения тройки и единицы равны по $1/6$, а вероятность третьего исхода (выпали любые другие грани) равна $4/6$, то вероятность получить десять троек, три единицы и ещё два других очка равна

$$P(10, 3, 2) = \frac{15!}{10!\, 3!\, 2!}\, \frac{1}{6^{10}}\, \frac{1}{6^3}\, \left(\frac{4}{6}\right)^2.$$

§ 4. Приближение гипергеометрического распределения биномиальным

Рассмотрим урну, содержащую N шаров, из которых K шаров — белые, а оставшиеся $N - K$ шаров — чёрные. Из урны наудачу (без возвращения) выбираются n шаров. Вероятность $P_{N,K}(n, k)$ того, что будет выбрано ровно k белых и $n - k$ чёрных шаров, находится по формуле (см. определение 8 гипергеометрического распределения вероятностей):

$$P_{N,K}(n, k) = \frac{C_K^k\, C_{N-K}^{n-k}}{C_N^n}.$$

Если число шаров в урне очень велико, то извлечение одного, двух, трёх шаров почти не меняет пропорцию белых и чёрных шаров в урне, так что вероятности $P_{N,K}(n,\, k)$ не очень отличаются от вероятностей в процедуре выбора с возвращением:

$$\mathsf{P}(\nu_n = k) = C_n^k \left(\frac{K}{N}\right)^k \left(1 - \frac{K}{N}\right)^{n-k}.$$

Сформулируем и докажем нашу первую предельную теорему.

Теорема 14. *Если $N \to \infty$ и $K \to \infty$ так, что $K/N \to p \in (0,1)$, то для любых фиксированных n и $0 \leqslant k \leqslant n$*

$$P_{N,K}(n,\, k) = \frac{C_K^k C_{N-K}^{n-k}}{C_N^n} \longrightarrow C_n^k p^k (1-p)^{n-k}.$$

Доказательство. Перепишем $P_{N,K}(n,\, k)$ следующим образом:

$$P_{N,K}(n,\, k) = \frac{K!}{k!(K-k)!}\,\frac{(N-K)!}{(n-k)!(N-K-(n-k))!}\,\frac{n!(N-n)!}{N!} =$$

$$= C_n^k \frac{\overbrace{K(K-1)\dots(K-k+1)}^{k}\,\overbrace{(N-K)(N-K-1)\dots(N-K-(n-k)+1)}^{n-k}}{\underbrace{N(N-1)\dots(N-n+1)}_{n=k+(n-k)}}.$$

И в числителе, и в знаменателе дроби — произведение фиксированного числа n сомножителей, поэтому и дробь есть произведение n сомножителей. Каждый из первых k сомножителей имеет вид $(K-a)\,/\,(N-b)$ при некоторых фиксированных a и b и стремится к p при $K/N \to p$. Каждый из оставшихся $n-k$ сомножителей имеет вид $(N-K-a)\,/\,(N-b)$ и стремится к $1-p$ при $K/N \to p$. Окончательно имеем

$$P_{N,K}(n,k) \to C_n^k p^k (1-p)^{n-k} \text{ при } \frac{K}{N} \to p. \qquad \square$$

§ 5. Теорема Пуассона[10] для схемы Бернулли

Предположим, нам нужна вероятность получить не менее семи успехов в тысяче испытаний схемы Бернулли с вероятностью успеха 0,003. Вероятность этого события равна любому из следующих выражений, вычислить которые довольно сложно:

$$\sum_{k=7}^{1\,000} C_{1\,000}^k\,(0{,}003)^k\,(0{,}997)^{1\,000-k} = 1 - \sum_{k=0}^{6} C_{1\,000}^k\,(0{,}003)^k\,(0{,}997)^{1\,000-k}.$$

Сформулируем теорему о приближённом вычислении вероятности какого-либо числа успехов в большом числе испытаний схемы Бернулли

[10] Siméon Denis Poisson (21.06.1781 — 25.04.1840, France)

с маленькой вероятностью успеха. Термин «большое число» должен означать $n \to \infty$. Если при этом $p = p_n \not\to 0$, то, очевидно, вероятность получить любое конечное число успехов при растущем числе испытаний стремится к нулю. Необходимо чтобы вероятность успеха $p = p_n$ стремилась к нулю одновременно с ростом числа испытаний. Но от испытания к испытанию вероятность успеха меняться не может (см. определение схемы Бернулли). Поэтому нам придётся рассмотреть так называемую «схему серий»: если испытание одно, то вероятность успеха в нём равна p_1, если испытаний два, то вероятность успеха в каждом — p_2, и т. д. Если испытаний n, то в каждом из них вероятность успеха равна p_n. Вероятность успеха меняется не внутри одной серии испытаний, а от серии к серии, когда меняется общее число испытаний.

Теорема 15 (теорема Пуассона). *Пусть $n \to \infty$ и $p_n \to 0$ так, что $np_n \to \lambda > 0$. Тогда для любого $k \geqslant 0$ вероятность получить k успехов в n испытаниях схемы Бернулли с вероятностью успеха p_n стремится к величине $\lambda^k\, e^{-\lambda} / k!$:*

$$C_n^k\, p_n^k\, (1-p_n)^{n-k} \longrightarrow \frac{\lambda^k}{k!}\, e^{-\lambda}.$$

Доказательство. Положим $\lambda_n = n \cdot p_n \to \lambda > 0$. Тогда $p_n = \lambda_n / n$ и

$$C_n^k p_n^k (1-p_n)^{n-k} = C_n^k \frac{\lambda_n^k}{n^k} \left(1 - \frac{\lambda_n}{n}\right)^{n-k} =$$

$$= \underbrace{\frac{n(n-1)\ldots(n-k+1)}{n^k}}_{\substack{\downarrow \\ 1}} \frac{\lambda_n^k}{k!} \underbrace{\left(1 - \frac{\lambda_n}{n}\right)^n}_{\substack{\downarrow \\ e^{-\lambda}}} \underbrace{\left(1 - \frac{\lambda_n}{n}\right)^{-k}}_{\substack{\downarrow \\ 1}} \longrightarrow \frac{\lambda^k}{k!}\, e^{-\lambda}. \qquad (8)$$

В соотношении (8) мы воспользовались тем, что $\lambda_n^k \to \lambda^k$ и замечательным пределом $(1 - \lambda_n/n)^n \to e^{-\lambda}$. Докажем последнее свойство:

$$\ln\left(1 - \frac{\lambda_n}{n}\right)^n = n \ln\left(1 - \frac{\lambda_n}{n}\right) = n\left(-\frac{\lambda_n}{n} + O\left(\frac{\lambda_n^2}{n^2}\right)\right) \to -\lambda. \qquad \square$$

Определение 25. Набор чисел $\left\{\dfrac{\lambda^k}{k!}\, e^{-\lambda},\ k = 0, 1, 2, \ldots\right\}$ называется распределением Пуассона с параметром $\lambda > 0$.

По теореме 15 можно приближённо посчитать вероятность получить не менее семи успехов в тысяче испытаний схемы Бернулли с вероятностью успеха 0,003, с вычисления которой мы начали. Поскольку $n = 1\,000$ «велико», а $p_n = 0{,}003$ «мало», то, взяв $\lambda = np_n = 3$, можно записать приближённое равенство

$$1 - \sum_{k=0}^{6} C_{1\,000}^{k}\,(0{,}003)^k\,(0{,}997)^{1\,000-k} \approx 1 - \sum_{k=0}^{6} \frac{3^k}{k!}\,e^{-3} = \sum_{k=7}^{\infty} \frac{3^k}{k!}\,e^{-3} = \\ = \text{ табличное значение } \overline{\Pi}_3(7) \approx 0{,}034. \qquad (9)$$

Осталось решить, а достаточно ли $n = 10^3$ велико, а $p_n = 0{,}003$ мало, чтобы заменить точную вероятность $\mathsf{P}(\nu_n = k)$ на приближённое значение $\lambda^k e^{-\lambda} / k!$ Для этого нужно уметь оценивать разницу между этими двумя вероятностями.

Следующую очень полезную теорему мы, исключительно из экономии времени, доказывать не станем.

Теорема 16 (уточнённая теорема Пуассона). *Пусть A — произвольное множество целых неотрицательных чисел, ν_n — число успехов в n испытаниях схемы Бернулли с вероятностью успеха p, и пусть $\lambda = np$. Справедливо неравенство:*

$$\left| \mathsf{P}(\nu_n \in A) - \sum_{k\in A} \frac{\lambda^k}{k!}\,e^{-\lambda} \right| = \left| \sum_{k\in A} C_n^k p^k (1-p)^{n-k} - \sum_{k\in A} \frac{\lambda^k}{k!}\,e^{-\lambda} \right| \leqslant np^2.$$

Таким образом, теорема 16 предоставляет нам возможность самим решать, достаточно ли n велико, а p мало, руководствуясь полученной величиной погрешности. Какова же погрешность в формуле (9)? Взяв $A = \{0, 1, \ldots, 6\}$, имеем:

$$\left| \mathsf{P}(\nu_{1000} \geqslant 7) - 0{,}034 \right| = \left| \Bigl(1 - \mathsf{P}(\nu_{1000} \leqslant 6)\Bigr) - \Bigl(1 - \sum_{k=0}^{6} \frac{3^k}{k!}\,e^{-3}\Bigr) \right| = \\ = \left| \mathsf{P}(\nu_{1000} \leqslant 6) - \sum_{k=0}^{6} \frac{3^k}{k!}\,e^{-3} \right| \leqslant np^2 = 0{,}009.$$

Можно утверждать, что искомая вероятность заключена в границах $(0{,}034 - 0{,}009,\ 0{,}034 + 0{,}009) = (0{,}025,\ 0{,}043)$.

На самом деле можно уточнить оценку в теореме 16. Например, можно доказать, что погрешность даже меньше, чем $\min(p,\ np^2)$. В нашем примере это втрое уменьшает оценку для погрешности — 0,003 вместо 0,009, уточняя границы для истинной вероятности: $(0{,}031,\ 0{,}037)$.

ГЛАВА 6

Случайные величины и их распределения

Абстрагировать — это, по-видимому, значит переходить к сути дела. Это значит освобождаться от случайных черт и сосредотачивать внимание на особо важных свойствах.

М. Кац, Статистическая независимость в теории вероятностей, анализе и теории чисел

§ 1. Случайные величины

Мы уже видели, что для многих экспериментов нет никаких различий в подсчёте в е р о я т н о с т е й событий, тогда как элементарные исходы в этих экспериментах очень различаются. Но нас и должны интересовать именно вероятности событий, а не структура пространства элементарных исходов. Поэтому пора во всех таких «похожих» экспериментах вместо самых разных элементарных исходов использовать, например, числа. Иначе говоря, каждому элементарному исходу поставить в соответствие некоторое вещественное число, и работать только с числами.

Пусть задано вероятностное пространство $\langle\Omega, \mathcal{F}, \mathsf{P}\rangle$.

Определение 26. Функция $\xi : \Omega \to \mathbb{R}$ называется с л у ч а й н о й в е л и ч и н о й, если для любого борелевского множества $B \in \mathfrak{B}(\mathbb{R})$ множество $\xi^{-1}(B)$ является событием, т. е. принадлежит σ-алгебре $\mathcal{F}$.

Множество $\xi^{-1}(B) = \{\omega \mid \xi(\omega) \in B\}$, состоящее из тех элементарных исходов ω, для которых $\xi(\omega)$ принадлежит B, называется полным прообразом множества B.

Замечание 10. Вообще, пусть функция f действует из множества X в множество Y, и заданы σ-алгебры $\mathcal{F}$ и $\mathcal{G}$ подмножеств X и Y соответственно. Функция f называется измеримой, если для любого множества $B \in \mathcal{G}$ его полный прообраз $f^{-1}(B)$ принадлежит $\mathcal{F}$.

Замечание 11. Читатель, не желающий забивать себе голову абстракциями, связанными с σ-алгебрами событий и с измеримостью, может смело считать, что любое множество элементарных исходов есть событие, и, следовательно, случайная величина есть произвольная функция из Ω в $\mathbb{R}$. Неприятностей на практике это не влечёт, так что всё дальнейшее в этом параграфе можно пропустить.

Теперь, избавившись от нелюбопытных читателей, попробуем понять, зачем случайной величине нужна измеримость.

Если задана случайная величина ξ, нам может потребоваться вычислить вероятности вида $\mathsf{P}(\xi = 5) = \mathsf{P}\{\omega \mid \xi(\omega) = 5\}$, $\mathsf{P}(\xi \in [-3, 7])$, $\mathsf{P}(\xi \geqslant 3{,}2)$, $\mathsf{P}(\xi < 0)$ (и вообще самые разные вероятности попадания в борелевские множества на прямой). Это возможно лишь если множества, стоящие под знаком вероятности, являются событиями — ведь вероятность есть функция, определённая только на σ-алгебре событий. Требование измеримости равносильно тому, что для любого борелевского множества B определена вероятность $\mathsf{P}(\xi \in B)$.

Можно потребовать в определении 26 чего-нибудь другого. Например, чтобы событием было попадание в любой интервал: $\{\omega \mid \xi(\omega) \in (a, b)\} \in \mathcal{F}$, или в любой полуинтервал: $\{\omega \mid \xi(\omega) < x\} \in \mathcal{F}$.

Убедимся, например, что эквивалентны определения 26 и 27:

Определение 27. Функция $\xi : \Omega \to \mathbb{R}$ называется случайной величиной, если для любых вещественных $a < b$ множество $\{\omega \mid \xi(\omega) \in (a, b)\}$ принадлежит σ-алгебре $\mathcal{F}$.

Доказательство эквивалентности определений 26, 27. Если ξ — случайная величина в смысле определения 26, то она будет случайной величиной и в смысле определения 27, поскольку любой интервал (a, b) является борелевским множеством.

Докажем, что верно и обратное. Пусть для любого интервала (a, b) выполнено $\xi^{-1}((a, b)) \in \mathcal{F}$. Мы должны доказать, что то же самое верно и для любых борелевских множеств. Соберём в множестве $\mathcal{A} = \{B \subseteq \mathbb{R} \mid \xi^{-1}(B) \in \mathcal{F}\}$ все подмножества вещественной прямой, прообразы которых являются событиями. Множество $\mathcal{A}$ уже содержит все интервалы (a, b). Покажем теперь, что множество $\mathcal{A}$ является σ-алгеброй.

По определению, $B \in \mathcal{A}$ тогда и только тогда, когда множество $\xi^{-1}(B) = \{\omega \mid \xi(\omega) \in B\}$ принадлежит $\mathcal{F}$.

1. Убедимся, что $\mathbb{R} \in \mathcal{A}$. Но $\xi^{-1}(\mathbb{R}) = \Omega \in \mathcal{F}$ и, следовательно, $\mathbb{R} \in \mathcal{A}$.

2. Убедимся, что $\overline{B} \in \mathcal{A}$ для любого $B \in \mathcal{A}$. Пусть $\xi^{-1}(B) \in \mathcal{F}$. Тогда $\xi^{-1}(\overline{B}) = \{\omega \mid \xi(\omega) \notin B\} = \Omega \setminus \xi^{-1}(B) \in \mathcal{F}$, так как $\mathcal{F}$ — σ-алгебра.

3. Убедимся, что $B_1 \cup B_2 \cup \ldots \in \mathcal{A}$ для любых $B_1, B_2, \ldots \in \mathcal{A}$. Пусть $\xi^{-1}(B_i) \in \mathcal{F}$ для всех $i \geqslant 1$. Но $\mathcal{F}$ — σ-алгебра, поэтому

$$\xi^{-1}\Bigl(\bigcup_{i=1}^{\infty} B_i\Bigr) = \Bigl\{\omega \Bigm| \xi(\omega) \in \bigcup_{i=1}^{\infty} B_i\Bigr\} = \bigcup_{i=1}^{\infty} \xi^{-1}(B_i) \in \mathcal{F}.$$

Мы доказали, что $\mathcal{A}$ — σ-алгебра и содержит все интервалы на прямой. Но $\mathfrak{B}(\mathbb{R})$ — наименьшая из σ-алгебр, содержащих все интервалы на прямой. Следовательно, $\mathcal{A}$ содержит все множества из $\mathfrak{B}(\mathbb{R})$: $\mathfrak{B}(\mathbb{R}) \subseteq \mathcal{A}$. □

Приведём примеры измеримых и неизмеримых функций.

Пример 30. Подбрасываем кубик. Пусть $\Omega = \{1, 2, 3, 4, 5, 6\}$, и две функции из Ω в $\mathbb{R}$ заданы так: $\xi(\omega) = \omega$, $\eta(\omega) = \omega^2$. Пока не задана σ-алгебра $\mathcal{F}$, нельзя говорить об измеримости. Функция, измеримая относительно какой-то σ-алгебры $\mathcal{F}$, может не быть таковой для другой $\mathcal{F}$.

1. Если $\mathcal{F}$ есть множество всех подмножеств Ω, то ξ и η являются случайными величинами, поскольку любое множество элементарных исходов принадлежит $\mathcal{F}$, в том числе и $\{\omega \mid \xi(\omega) \in B\}$ или $\{\omega \mid \eta(\omega) \in B\}$. Можно записать соответствие между значениями случайных величин ξ и η и вероятностями принимать эти значения в виде «таблицы распределения вероятностей» или, коротко, «таблицы распределения»:

ξ	1	2	3	4	5	6
P	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

η	1	4	9	16	25	36
P	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Здесь $\mathsf{P}(\xi = 1) = \ldots = \mathsf{P}(\xi = 6) = \mathsf{P}(\eta = 1) = \ldots = \mathsf{P}(\eta = 36) = \frac{1}{6}$.

2. Пусть σ-алгебра событий $\mathcal{F}$ состоит из четырёх множеств:

$$\mathcal{F} = \big\{\Omega,\ \varnothing,\ \{1, 3, 5\},\ \{2, 4, 6\}\big\},$$

т. е. событием является, кроме достоверного и невозможного событий, выпадение чётного или нечётного числа очков. Убедимся, что при такой сравнительно бедной σ-алгебре ни ξ, ни η не являются случайными величинами, поскольку они неизмеримы. Возьмём, скажем, $B = \{4\}$. Видим, что $\{\omega \mid \xi(\omega) = 4\} = \{4\} \notin \mathcal{F}$ и $\{\omega \mid \eta(\omega) = 4\} = \{2\} \notin \mathcal{F}$.

Упражнение.

а) Какие функции измеримы относительно $\mathcal{F} = \big\{\Omega,\ \varnothing,\ \{1, 3, 5\},\ \{2, 4, 6\}\big\}$?

б) Доказать, что ξ и η не являются случайными величинами, если $\mathcal{F} = \{\Omega,\ \varnothing\}$.

в) Доказать, что относительно тривиальной σ-алгебры измеримы только функции вида $\xi(\omega) = c$ (постоянные).

Пример 31. Пусть $\Omega = [0, 2\pi]$, $\mathcal{F} = \mathfrak{B}(\mathbb{R}) \cap [0, 2\pi]$ — сигма-алгебра борелевских подмножеств отрезка $[0, 2\pi]$, $\mathsf{P}(\cdot) = \lambda(\cdot)$ — мера Лебега на $\mathcal{F}$ и A — неизмеримое множество Витали, построенное нами в примере 14. Функция

$$\xi(\omega) = I_A(\omega) = \begin{cases} 1, & \text{если } \omega \in A, \\ 0, & \text{если } \omega \notin A \end{cases}$$

не является случайной величиной, поскольку, например, прообраз единицы $\xi^{-1}(\{1\}) = \{\omega \mid \xi(\omega) = 1\} = A$ не принадлежит $\mathcal{F}$. И вероятность для ξ попасть в единицу $\mathsf{P}(\xi = 1) = \lambda(A)$ просто не существует.

Познакомимся с важным понятием — «распределение» случайной величины и опишем различные типы распределений случайных величин.

§ 2. Распределения случайных величин

Определение 28. Распределением случайной величины ξ называется вероятностная мера $\mu(B) = \mathsf{P}(\xi \in B)$ на множестве борелевских подмножеств $\mathbb{R}$.

Можно представлять себе распределение случайной величины ξ как соответствие между множествами $B \in \mathfrak{B}(\mathbb{R})$ и вероятностями $\mathsf{P}(\xi \in B)$.

Распределения случайных величин суть основные объекты изучения в теории вероятностей. Мы не будем, как правило, интересоваться тем, из какого множества действует функция и каким именно элементарным исходам сопоставляет свои возможные значения. Нас будет интересовать лишь, с какой вероятностью эти значения принимаются. Приведём несколько примеров совершенно разных случайных величин, имеющих одно и то же распределение (одинаково распределённых).

Пример 32.

1. Один раз бросается правильная монета. Пространство Ω состоит из двух элементарных исходов — *герб* и *решка*. В качестве σ-алгебры рассмотрим множество всех подмножеств Ω. Вероятность зададим как в классической схеме. Построим две случайные величины ξ и η так: положим

$\xi(\omega) = 1$, если $\omega =$ *герб*, и $\xi(\omega) = 0$, если $\omega =$ *решка*;

$\eta(\omega) = 0$, если $\omega =$ *герб*, и $\eta(\omega) = 1$, если $\omega =$ *решка*.

Очевидно, что для любого множества $B \subseteq \mathbb{R}$ вероятности принадлежать B для ξ и η одинаковы. Тем не менее ни для одного элементарного исхода ω значения $\xi(\omega)$ и $\eta(\omega)$ не совпадают. Т. е. ξ и η одинаково распределены, но не одинаковы как функции.

2. Точка наудачу бросается на отрезок $[0, 1]$. В этом случае Ω есть отрезок $[0, 1]$ с σ-алгеброй борелевских подмножеств $[0, 1]$ и мерой Лебега в качестве вероятности. Предлагаю читателю убедиться, что две совершенно разные функции: $\xi(\omega) = \omega$ и $\eta(\omega) = 1 - \omega$ (расстояния до упавшей точки от левого и от правого концов отрезка соответственно) обладают одинаковыми вероятностями принимать значения внутри любых борелевских множеств B. Вероятности эти равны мере Лебега пересечения множеств B и $[0, 1]$. Таким образом, эти случайные величины снова одинаково распределены, но не одинаковы: их значения совпадают лишь при одном элементарном исходе $\omega = 0{,}5$ (нарисовать графики функций $\xi(\omega)$ и $\eta(\omega)$).

3. На том же отрезке $[0, 1]$ построим две функции: $\xi(\omega) = 0$ при всех ω; $\eta(\omega) = 0$ при всех ω, кроме $\omega = 0{,}5$, а в точке $\omega = 0{,}5$ положим $\eta(\omega) = -17$.

Поскольку мера Лебега точки (она же — вероятность) равна нулю, распределения величин ξ и η одинаковы. Теперь $\xi(\omega)$ и $\eta(\omega)$ снова не совпадают как функции, но отличаются их значения лишь на множестве нулевой вероятности — только в точке $\omega = 0{,}5$. В этом случае говорят, что ξ и η совпадают «почти наверное»: $\mathsf{P}(\xi = \eta) = 1$.

Опишем различные типы распределений случайных величин. Вся вероятностная масса может быть сосредоточена в нескольких точках прямой, может быть «размазана» по некоторому интервалу или по всей прямой. В зависимости от типа множества, на котором сосредоточена вся единичная вероятностная масса, распределения делят на дискретные, абсолютно непрерывные, сингулярные и их смеси.

Определение 29. Случайная величина ξ имеет дискретное распределение, если существует конечный или счётный набор чисел $\{a_1, a_2, \ldots\}$ такой, что

$$\sum_{i=1}^{\infty} \mathsf{P}(\xi = a_i) = 1.$$

Итак, случайная величина ξ имеет дискретное распределение, если она принимает не более чем счётное число значений. Значения эти иначе называют атомами: ξ имеет атом в точке x, если $\mathsf{P}(\xi = x) > 0$.

Если случайная величина ξ имеет дискретное распределение, то для любого $B \subseteq \mathbb{R}$

$$\mathsf{P}(\xi \in B) = \sum_{a_i \in B} \mathsf{P}(\xi = a_i).$$

Дискретное распределение удобно задавать следующей таблицей, в которой $p_i = \mathsf{P}(\xi = a_i)$:

ξ	a_1	a_2	a_3	$\ldots$
P	p_1	p_2	p_3	$\ldots$

Определение 30. Случайная величина ξ имеет абсолютно непрерывное распределение, если существует неотрицательная функция $f_\xi(x)$ такая, что для любого борелевского множества B имеет место равенство:

$$\mathsf{P}(\xi \in B) = \int\limits_B f_\xi(x)\, dx.$$

Функцию $f_\xi(x)$ называют плотностью распределения величины ξ.

Замечание 12. Интеграл выше есть интеграл Лебега, а не Римана. Вполне достаточно, если читатель, не знакомый с интегралом Лебега, будет представлять его себе просто как площадь под графиком подынтегральной функции над множеством B. При этом площадь над множеством B, имеющим нулевую меру Лебега, равна нулю. Заметим, что любая функция, отличающаяся от функции $f_\xi(x)$ лишь в конечном или счётном числе точек (или на множестве нулевой меры Лебега), будет являться плотностью того же распределения, так как интеграл не изменится от изменения подынтегральной функции на множестве меры нуль.

Теорема 17. *Плотность распределения обладает свойствами*:

(f1) $f_\xi(x) \geqslant 0$ *для любого* x; (f2) $\int\limits_{-\infty}^{\infty} f_\xi(t)\,dt = 1.$

Доказательство. (f1) выполнено по определению плотности, (f2) также следует из определения:

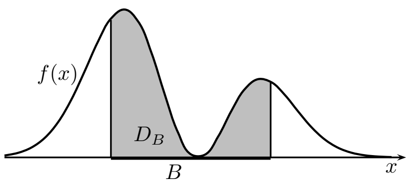
$$\mathsf{P}(\xi \in \mathbb{R}) = 1 = \int\limits_{\mathbb{R}} f_\xi(x)\,dx. \qquad \square$$

Эти два свойства полностью характеризуют класс плотностей:

Теорема 18. *Если функция* f *обладает свойствами* (f1) *и* (f2), *то существует вероятностное пространство и случайная величина* ξ *на нём, для которой* f *является плотностью распределения.*

Доказательство. Пусть Ω есть область, заключенная между осью абсцисс и графиком функции f. Площадь области Ω равна единице по свойству (f2). Пусть $\mathfrak{F}$ — множество борелевских подмножеств Ω, а P — мера Лебега (площадь) на множествах из $\mathfrak{F}$. И пусть случайная величина ξ есть абсцисса точки, наудачу брошенной в эту область. Тогда для любого $B \in \mathfrak{B}(\mathbb{R})$ выполнено:

$$\mathsf{P}(\xi \in B) = \mathsf{P}(\text{точка попала в } D_B) = \frac{\text{площадь } D_B}{\text{площадь } \Omega} = \int\limits_B f(x)\,dx. \tag{10}$$



Здесь область D_B есть криволинейная трапеция под графиком плотности, с основанием B. По определению, равенство (10) означает, что функция f является плотностью распределения случайной величины ξ. $\square$

Свойство 7. *Если случайная величина* ξ *имеет абсолютно непрерывное распределение, то* $\mathsf{P}(\xi = x) = 0$ *для любого* $x \in \mathbb{R}$.

Доказательство сразу следует из определения 30 и замечания 12, так как интеграл по области интегрирования, состоящей из одной точки, равен нулю. $\square$

Можно выделить ещё один особый класс распределений, сосредоточенных, в отличие от абсолютно непрерывных распределений, на множестве нулевой меры Лебега, но не имеющих, в отличие от дискретных, атома ни в одной точке этого множества.

Определение 31. Говорят, что случайная величина ξ имеет *сингулярное* распределение, если существует борелевское множество B с нулевой лебеговой мерой $\lambda(B) = 0$ такое, что $\mathsf{P}(\xi \in B) = 1$, но при этом $\mathsf{P}(\xi = x) = 0$ для любой точки $x \in B$.

Можно отметить следующее свойство сингулярных распределений. Множество B, на котором сосредоточено всё распределение, не может состоять из конечного или счётного числа точек. Действительно, если B конечно или счётно, то $\mathsf{P}(\xi \in B) = \sum \mathsf{P}(\xi = x_i)$, где суммирование ведётся по всем $x_i \in B$. Последняя сумма равна нулю как сумма счётного числа нулей, что противоречит предположению $\mathsf{P}(\xi \in B) = 1$.

Таким образом, любое сингулярное распределение сосредоточено на несчётном множестве с нулевой мерой Лебега. Примером такого множества может служить канторовское совершенное множество, а примером такого распределения — лестница Кантора (выяснить, что это такое!).

Наконец, распределение может быть выпуклой линейной комбинацией дискретного, абсолютно непрерывного и сингулярного распределений.

Определение 32. Говорят, что случайная величина ξ имеет *смешанное* распределение, если найдутся такие случайные величины ξ_1, ξ_2 и ξ_3 — с дискретным, абсолютно непрерывным и сингулярным распределениями соответственно (или такие три распределения), и числа $p_1, p_2, p_3 \in [0, 1)$, $p_1 + p_2 + p_3 = 1$, что для любого $B \in \mathfrak{B}(\mathbb{R})$ имеет место равенство:

$$\mathsf{P}(\xi \in B) = p_1\mathsf{P}(\xi_1 \in B) + p_2\mathsf{P}(\xi_2 \in B) + p_3\mathsf{P}(\xi_3 \in B).$$

По заданным на одном вероятностном пространстве случайным величинам ξ_1, ξ_2, ξ_3 и числам $p_1+p_2+p_3 = 1$ можно построить случайную величину со смешанным распределением так: пусть φ — случайная величина с дискретным распределением на том же вероятностном пространстве такая, что $\mathsf{P}(\varphi = k) = p_k$ для $k = 1, 2, 3$, и при любом k и любом $B \in \mathfrak{B}(\mathbb{R})$ события $\{\varphi = k\}$ и $\{\xi_k \in B\}$ независимы.

Построим случайную величину ξ так: $\xi(\omega) = \xi_k(\omega)$, если $\varphi(\omega) = k$, где $k = 1, 2, 3$. Её распределение найдём по формуле полной вероятности:

$$\mathsf{P}(\xi \in B) = \mathsf{P}(\xi_1 \in B, \varphi = 1) + \mathsf{P}(\xi_2 \in B, \varphi = 2) + \mathsf{P}(\xi_3 \in B, \varphi = 3).$$

В силу независимости событий под знаком каждой из вероятностей,

$$\mathsf{P}(\xi \in B) = p_1\mathsf{P}(\xi_1 \in B) + p_2\mathsf{P}(\xi_2 \in B) + p_3\mathsf{P}(\xi_3 \in B).$$

Никаких других видов распределений, кроме перечисленных выше, не существует (доказано Лебегом).

§ 3. Функция распределения

Описание распределения набором вероятностей $\mathsf{P}(\xi \in B)$ не очень удобно: слишком много существует борелевских множеств. Мы описали дискретные распределения таблицей распределения, абсолютно непрерывные — плотностью распределения. Попробуем поискать какой-нибудь универсальный способ описать любое возможное распределение.

Можно поставить вопрос иначе: распределение есть набор вероятностей попадания в любые борелевские множества на прямой. Нельзя ли обойтись знанием вероятностей попадания в какой-нибудь меньший набор множеств на прямой? Борелевская σ-алгебра $\mathfrak{B}(\mathbb{R})$ порождается интервалами (равно как и лучами $(-\infty, x)$), поэтому можно ограничиться только вероятностями попадания в такие лучи для всех $x \in \mathbb{R}$. А уже с их помощью можно будет определить и вероятность попасть в любое борелевское множество.

Замечание 13. Можно с таким же успехом ограничиться набором вероятностей попадания в интервалы $(-\infty, x]$, или в (x, ∞), или в $[x, \infty)$.

Определение 33. *Функцией распределения* случайной величины ξ называется функция $F_\xi : \mathbb{R} \to [0, 1]$, при каждом $x \in \mathbb{R}$ равная вероятности случайной величине ξ принимать значения, меньшие x:

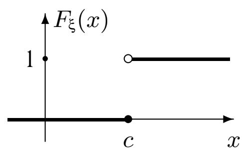
$$F_\xi(x) = \mathsf{P}(\xi < x) = \mathsf{P}\{\omega : \xi(\omega) < x\}.$$

Перечислим основные дискретные и абсолютно непрерывные распределения и найдём их функции распределения.

§ 4. Примеры дискретных распределений

Вырожденное распределение. Говорят, что случайная величина ξ имеет вырожденное распределение в точке $c \in \mathbb{R}$, и пишут: $\xi \in \mathrm{I}_c$, если ξ принимает единственное значение c с вероятностью 1, т. е. $\mathsf{P}(\xi = c) = 1$. Функция распределения ξ имеет вид:

$$F_\xi(x) = \mathsf{P}(\xi < x) = \mathsf{P}(c < x) = \begin{cases} 0, & x \leqslant c; \\ 1, & x > c. \end{cases}$$



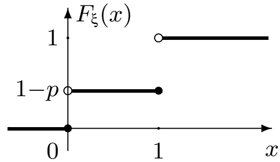
Распределение Бернулли. Говорят, что случайная величина ξ имеет распределение Бернулли с параметром p, и пишут: $\xi \in \mathrm{B}_p$, если ξ принимает значения 1 и 0 с вероятностями p и $1-p$ соответственно. Случайная величина ξ с таким распределением равна *числу успехов* в одном испытании схемы Бернулли с вероятностью успеха p: ни одного успеха или один успех. Таблица распределения ξ имеет вид:

ξ	0	1
P	$1-p$	p

.

Функция распределения случайной величины ξ такова:

$$F_\xi(x) = \mathsf{P}(\xi < x) = \begin{cases} 0, & x \leqslant 0; \\ 1-p, & 0 < x \leqslant 1 \\ 1, & x > 1. \end{cases}$$



Биномиальное распределение. Говорят, что случайная величина ξ имеет биномиальное распределение с параметрами $n \in \mathbb{N}$ и $p \in (0, 1)$, и пишут: $\xi \in \mathrm{B}_{n,p}$, если ξ принимает значения $k = 0, 1, \ldots, n$ с вероятностями $\mathsf{P}(\xi = k) = C_n^k p^k (1-p)^{n-k}$. Случайная величина с таким распределением имеет смысл *числа успехов в n испытаниях* схемы Бернулли с вероятностью успеха p. Таблица распределения ξ имеет вид:

ξ	0	1	$\ldots$	k	$\ldots$	n
P	$(1-p)^n$	$np(1-p)^{n-1}$	$\ldots$	$C_n^k p^k (1-p)^{n-k}$	$\ldots$	p^n

.

Распределение Бернулли совпадает с распределением $\mathrm{B}_{1,\,p}$.

Геометрическое распределение. Говорят, что случайная величина τ имеет геометрическое распределение с параметром $p \in (0, 1)$, и пишут $\tau \in \mathrm{G}_p$, если τ принимает значения $k = 1, 2, 3, \ldots$ с вероятностями $\mathsf{P}(\tau = k) = p(1-p)^{k-1}$. Случайная величина с таким распределением имеет смысл *номера первого успешного испытания* в схеме Бернулли с вероятностью успеха p. Таблица распределения τ имеет вид:

τ	1	2	$\ldots$	k	$\ldots$
P	p	$p(1-p)$	$\ldots$	$p(1-p)^{k-1}$	$\ldots$

.

Распределение Пуассона. Говорят, что случайная величина ξ имеет распределение Пуассона с параметром λ, где $\lambda > 0$, и пишут: $\xi \in \Pi_\lambda$, если ξ принимает значения $k = 0, 1, 2, \ldots$ с вероятностями $\mathsf{P}(\xi = k) = \dfrac{\lambda^k}{k!} e^{-\lambda}$. Таблицу распределения ξ читатель может нарисовать самостоятельно.

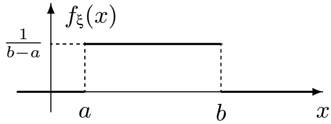
Гипергеометрическое распределение. Говорят, что случайная величина ξ имеет гипергеометрическое распределение с параметрами n, N и K, где $K \leqslant N$, $n \leqslant N$, если ξ принимает целые значения k такие, что $0 \leqslant k \leqslant K, 0 \leqslant n-k \leqslant N-K$, с вероятностями $\mathsf{P}(\xi = k) = C_K^k C_{N-K}^{n-k} \,/\, C_N^n$. Случайная величина с таким распределением имеет смысл числа белых шаров среди n шаров, выбранных наудачу и без возвращения из урны, содержащей K белых шаров и $N-K$ не белых.

Упражнение. Построить графики функций распределения для распределения Пуассона, биномиального и геометрического распределения.

§ 5. Примеры абсолютно непрерывных распределений

Равномерное распределение. Говорят, что ξ имеет равномерное распределение на отрезке $[a, b]$, и пишут: $\xi \in \mathrm{U}_{a,b}$, если плотность распределения ξ постоянна на отрезке $[a, b]$ и равна нулю вне него:

$$f_\xi(x) = \begin{cases} \dfrac{1}{b-a}, & \text{если } x \in [a, b], \\ 0, & \text{если } x \notin [a, b]. \end{cases}$$



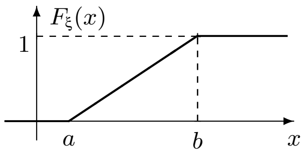
Очевидно, что площадь под графиком этой функции равна единице и $f_\xi(x) \geqslant 0$. Поэтому $f_\xi(x)$ является плотностью распределения.

Случайная величина $\xi \in \mathrm{U}_{a,b}$ имеет смысл координаты точки, выбранной наудачу на отрезке $[a, b]$. Вычислим по определению 30 функцию распределения случайной величины ξ:

$$F_\xi(x) = \mathsf{P}(\xi < x) = \int\limits_{-\infty}^{x} f_\xi(t)\,dt = \begin{cases} \int\limits_{-\infty}^{x} 0\,dt, & x < a, \\ \int\limits_{-\infty}^{0} 0\,dt + \int\limits_{a}^{x} \dfrac{1}{b-a}\,dt, & a \leqslant x \leqslant b, \\ \int\limits_{-\infty}^{a} 0\,dt + \int\limits_{a}^{b} \dfrac{1}{b-a}\,dt + \int\limits_{b}^{x} 0\,dt, & x > b. \end{cases}$$

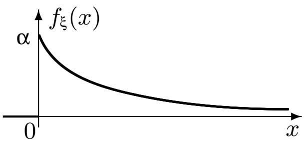
Получим следующую непрерывную функцию распределения:

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



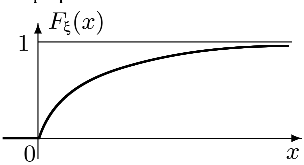
Показательное распределение. Говорят, что ξ имеет показательное (экспоненциальное) распределение с параметром $\alpha > 0$, и пишут: $\xi \in \mathrm{E}_\alpha$, если ξ имеет следующую плотность распределения:

$$f_\xi(x) = \begin{cases} 0, & \text{если } x < 0, \\ \alpha e^{-\alpha x}, & \text{если } x \geqslant 0. \end{cases}$$



Функция распределения случайной величины ξ непрерывна:

$$F_\xi(x) = \mathsf{P}(\xi < x) = \begin{cases} 0, & \text{если } x < 0, \\ 1 - e^{-\alpha x}, & \text{если } x \geqslant 0. \end{cases}$$



Показательное распределение является единственным абсолютно непрерывным распределением, для которого выполнено свойство «нестарения» (и в этом смысле оно является непрерывным аналогом дискретного геометрического распределения).

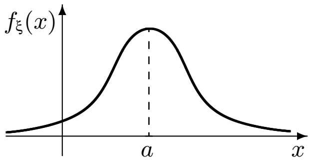
Теорема 19. *Пусть* $\xi \in \mathrm{E}_\alpha$. *Тогда для любых* $x, y > 0$

$$\mathsf{P}(\xi > x + y \mid \xi > x) = \mathsf{P}(\xi > y). \qquad (11)$$

Упражнение. Доказать теорему 19. Доказать далее, что если неотрицательная случайная величина ξ с абсолютно непрерывным распределением обладает свойством (11) при любых $x, y > 0$, то она имеет показательное распределение с некоторым параметром α.

Нормальное распределение. Говорят, что ξ имеет нормальное (гауссовское[11]) распределение с параметрами a и σ^2, где $a \in \mathbb{R}$, $\sigma > 0$, и пишут: $\xi \in \mathrm{N}_{a,\sigma^2}$, если ξ имеет следующую плотность распределения:

$$f_\xi(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-a)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$



Убедимся, что $f_\xi(x)$ является плотностью распределения. Так как $f_\xi(x) > 0$ для всех $x \in \mathbb{R}$, то свойство (f1) выполнено. Проверим (f2):

$$\int\limits_{-\infty}^{\infty} f_\xi(x)\,dx = \int\limits_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-a)^2}{2\sigma^2}}\,dx = \left[\begin{array}{c}\text{замена переменных}\\ t = \dfrac{x-a}{\sigma},\ dx = \sigma\,dt\end{array}\right] =$$

$$= \int\limits_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}}\, e^{-t^2/2}\sigma\,dt = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} e^{-t^2/2}\,dt = \frac{I}{\sqrt{2\pi}} = 1,$$

где через I обозначен табличный интеграл (интеграл Пуассона[12])

$$I = \int\limits_{-\infty}^{\infty} e^{-x^2/2}\,dx = \sqrt{2\pi}.$$

---

[11] Johann Carl Friedrich Gauss (30.04.1777 — 23.02.1855, Germany)

[12] Этот интеграл вычисляется так:

$$I^2 = \int\limits_{-\infty}^{\infty} e^{-x^2/2}\,dx \int\limits_{-\infty}^{\infty} e^{-y^2/2}\,dy = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} e^{-(x^2+y^2)/2}\,dx\,dy.$$

Далее полярная замена переменных: $x = r\cos\varphi$, $y = r\sin\varphi$, $dx\,dy = r\,dr\,d\varphi$, $x^2 + y^2 = r^2$:

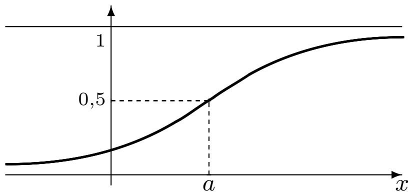
$$I^2 = \int\limits_0^{2\pi}\int\limits_0^{\infty} re^{-r^2/2}\,dr\,d\varphi = \int\limits_0^{2\pi}\int\limits_0^{\infty} e^{-r^2/2}\,d(r^2/2)\,d\varphi = 2\pi, \quad I = \sqrt{2\pi}.$$

Нормальное распределение $\mathrm{N}_{0,1}$ с параметрами $a = 0$ и $\sigma^2 = 1$ называется с т а н д а р т н ы м н о р м а л ь н ы м распределением. Плотность стандартного нормального распределения равна $f_\xi(x) = \dfrac{1}{\sqrt{2\pi}}\, e^{-x^2/2}$.

Ввиду особой роли нормального распределения в теории вероятностей (мы ещё узнаем о ней) существует даже специальное обозначение $\Phi_{a,\sigma^2}(x)$ для функции распределения нормального закона N_{a,σ^2}. Из курса математического анализа читателю известно, что первообразная функции e^{-x^2} не может быть выражена через элементарные функции. Поэтому функцию $\Phi_{a,\sigma^2}(x)$ можно записать лишь в виде интеграла:

$$\Phi_{a,\sigma^2}(x) = \int\limits_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(t-a)^2}{2\sigma^2}}\, dt,$$

$$\Phi_{0,1}(x) = \int\limits_{-\infty}^{x} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{t^2}{2}}\, dt.$$



Функция $\Phi_{0,1}(x)$ табулирована, т. е. её значения при различных вещественных x вычислены. Их можно найти в соответствующих таблицах.

Гамма-распределение. Говорят, что ξ имеет гамма-распределение с параметрами $\alpha > 0$, $\lambda > 0$, и пишут: $\xi \in \Gamma_{\alpha,\lambda}$, если ξ имеет следующую плотность распределения:

$$f_\xi(x) = \begin{cases} 0, & \text{если } x \leqslant 0, \\ c \cdot x^{\lambda-1} e^{-\alpha x}, & \text{если } x > 0, \end{cases}$$

где постоянная c вычисляется из свойства (f2) плотности так:

$$1 = \int\limits_{-\infty}^{\infty} f_\xi(x)\, dx = c \int\limits_{0}^{\infty} x^{\lambda-1} e^{-\alpha x}\, dx = \frac{c}{\alpha^\lambda} \int\limits_{0}^{\infty} (\alpha x)^{\lambda-1} e^{-\alpha x}\, d(\alpha x) = \frac{c}{\alpha^\lambda}\, \Gamma(\lambda),$$

откуда $c = \alpha^\lambda / \Gamma(\lambda)$. Здесь через $\Gamma(\lambda)$ обозначен интеграл

$$\Gamma(\lambda) = \int\limits_{0}^{\infty} x^{\lambda-1} e^{-x}\, dx = (\lambda - 1)\Gamma(\lambda - 1),$$

называемый гамма-функцией Эйлера[13]; $\Gamma(k) = (k-1)!$ при целых положительных k, $\Gamma(1) = 1$. Замена в интеграле Пуассона даст $\Gamma(1/2) = \sqrt{\pi}$.

Полезно отметить, что показательное распределение есть частный случай гамма-распределения: $\mathrm{E}_\alpha = \Gamma_{\alpha,1}$.

---

[13]Leonhard Euler (15.04.1707 — 18.09.1783, Switzerland, Россия)

Упражнение. Нарисовать график плотности распределения $\Gamma_{\alpha,\lambda}$ при $\lambda < 1$, при $\lambda = 1$ и при $\lambda > 1$, отметить на этом графике точки экстремума, точки перегиба и иные особенности графика.

Функцию распределения гамма-распределения можно записать, вообще говоря, только в виде интеграла:

$$F_\xi(x) = \frac{\alpha^\lambda}{\Gamma(\lambda)} \int\limits_0^x t^{\lambda-1} e^{-\alpha t}\, dt.$$

Но при целых значениях параметра λ интегрированием по частям этот интеграл можно превратить в сумму:

$$F_\xi(x) = 1 - \sum_{k=0}^{\lambda-1} \frac{(\alpha x)^k}{k!}\, e^{-\alpha x} = \sum_{k=\lambda}^{\infty} \frac{(\alpha x)^k}{k!}\, e^{-\alpha x}. \tag{12}$$

Упражнение. Доказать первое из равенств (12) при целых значениях λ. Доказать следующее забавное равенство: $F_\xi(x) = \mathsf{P}(\eta \geqslant \lambda)$, где $\eta \in \Pi_{\alpha x}$.

Распределение Коши. Говорят, что ξ имеет распределение Коши[14] с параметрами $a \in \mathbb{R}$, $\sigma > 0$, и пишут: $\xi \in \mathrm{C}_{a,\sigma}$, если ξ имеет следующую плотность распределения:

$$f_\xi(x) = \frac{1}{\pi}\, \frac{\sigma}{\sigma^2 + (x-a)^2} \quad \text{для любого } x \in \mathbb{R}.$$

Плотность распределения Коши симметрична относительно прямой $x = a$ и похожа на плотность нормального распределения, но имеет более толстые «хвосты» на $\pm\infty$. Функция распределения случайной величины ξ с распределением Коши равна $F_\xi(x) = \dfrac{1}{2} + \dfrac{1}{\pi} \operatorname{arctg}\left(\dfrac{x-a}{\sigma}\right)$ при всех x.

Распределение Парето. Говорят, что ξ имеет распределение Парето[15] с параметром $\alpha > 0$, если ξ имеет следующие плотность и функцию распределения:

$$f_\xi(x) = \begin{cases} \dfrac{\alpha}{x^{\alpha+1}}, & \text{если } x \geqslant 1,\\ 0, & \text{если } x < 1; \end{cases} \qquad F_\xi(x) = \begin{cases} 1 - \dfrac{1}{x^\alpha}, & \text{если } x \geqslant 1,\\ 0, & \text{если } x < 1. \end{cases}$$

Часто рассматривают более широкий класс распределений Парето, сосредоточенных не на $[1, \infty)$, а на $[c, \infty)$ при $c > 0$.

С другими абсолютно непрерывными распределениями (хи-квадрат Пирсона, распределениями Стьюдента, Фишера, Колмогорова, Лапласа) мы познакомимся при изучении математической статистики. С распределе-

[14] Augustin Louis Cauchy (21.08.1789 — 23.05.1857, France)
[15] Vilfredo Pareto (15.07.1848 — 20.08.1923, France, Italy, Switzerland)

ниями Вейбулла, логарифмически нормальным и некоторыми другими читатель познакомится в дальнейших курсах.

§ 6. Свойства функций распределения

Общие свойства функций распределения. Функцией распределения случайной величины ξ мы назвали функцию $F_\xi(x) = \mathsf{P}(\xi < x)$. Основные свойства этой функции заключены в теореме:

Теорема 20. *Любая функция распределения обладает следующими свойствами:*

(F1) *она не убывает: если* $x_1 < x_2$, *то* $F_\xi(x_1) \leqslant F_\xi(x_2)$;

(F2) *существуют пределы* $\lim\limits_{x\to-\infty} F_\xi(x) = 0$ *и* $\lim\limits_{x\to+\infty} F_\xi(x) = 1$;

(F3) *она в любой точке непрерывна слева:*

$$F_\xi(x_0 - 0) = \lim_{x\to x_0-0} F_\xi(x) = F_\xi(x_0).$$

Доказательство свойства (F1). Для любых чисел $x_1 < x_2$ событие $\{\xi < x_1\}$ влечёт событие $\{\xi < x_2\}$, т. е. $\{\xi < x_1\} \subseteq \{\xi < x_2\}$. Но вероятность — монотонная функция событий, поэтому

$$F_\xi(x_1) = \mathsf{P}\{\xi < x_1\} \leqslant \mathsf{P}\{\xi < x_2\} = F_\xi(x_2). \qquad \square$$

Для доказательства остальных свойств нам понадобится свойство непрерывности вероятностной меры (глава 3, 2, стр. 28, лемма 2).

Доказательство свойства (F2). Заметим сначала, что существование пределов в свойствах (F2), (F3) вытекает из монотонности и ограниченности функции $F_\xi(x)$. Остается лишь доказать равенства $\lim\limits_{x\to-\infty} F_\xi(x) = 0$, $\lim\limits_{x\to+\infty} F_\xi(x) = 1$ и $\lim\limits_{x\to x_0-0} F_\xi(x) = F_\xi(x_0)$. Для этого в каждом случае достаточно найти предел по какой-нибудь подпоследовательности $\{x_n\}$, так как существование предела влечёт совпадение всех частичных пределов.

Докажем, что $F_\xi(-n) \to 0$ при $n \to \infty$. Рассмотрим вложенную убывающую последовательность событий $B_n = \{\xi < -n\}$:

$$B_{n+1} = \bigl\{\xi < -(n+1)\bigr\} \subseteq B_n = \bigl\{\xi < -n\bigr\} \text{ для любых } n \geqslant 1.$$

Пересечение B всех этих событий состоит из тех и только тех ω, для которых $\xi(\omega)$ меньше любого вещественного числа. Но для любого элементарного исхода ω значение $\xi(\omega)$ вещественно, и не может быть меньше всех вещественных чисел. Иначе говоря, пересечение событий B_n не содержит элементарных исходов, т. е. $B = \bigcap B_n = \varnothing$. По свойству непрерывности меры, $F_\xi(-n) = \mathsf{P}(B_n) \to \mathsf{P}(B) = 0$ при $n \to \infty$.

Точно так же докажем остальные свойства.

Покажем, что $F_\xi(n) \to 1$ при $n \to \infty$, т. е. $1-F_\xi(n) = \mathsf{P}(\xi \geqslant n) \to 0$. Обозначим через B_n событие $B_n = \{\xi \geqslant n\}$. События B_n вложены:

$$B_{n+1} = \big\{\xi \geqslant (n+1)\big\} \subseteq B_n = \big\{\xi \geqslant n\big\} \text{ для любых } n \geqslant 1,$$

а пересечение B этих событий снова пусто — оно означает, что ξ больше любого вещественного числа. По свойству непрерывности меры, $1-F_\xi(n) = \mathsf{P}(B_n) \to \mathsf{P}(B) = 0$ при $n \to \infty$. □

Доказательство свойства (F3). Достаточно доказать, что $F_\xi(x_0 - 1/n) \to F_\xi(x_0)$ при $n \to \infty$. Иначе говоря, доказать сходимость к нулю следующей разности:

$$F_\xi(x_0) - F_\xi\Big(x_0 - \frac{1}{n}\Big) = \mathsf{P}(\xi < x_0) - \mathsf{P}\Big(\xi < x_0 - \frac{1}{n}\Big) = \mathsf{P}\Big(x_0 - \frac{1}{n} \leqslant \xi < x_0\Big).$$

Упражнение. Обозначьте событие $\{x_0 - 1/n \leqslant \xi < x_0\}$ через B_n, и попробуйте снова воспользоваться свойством непрерывности меры. □

Следующая теорема говорит о том, что три доказанных свойства полностью описывают класс функций распределения. То, что любая функция распределения ими обладает, мы с вами доказали, а теорема утверждает, что любая функция с такими свойствами есть функция распределения.

Теорема 21. *Если функция $F : \mathbb{R} \to [0,1]$ удовлетворяет свойствам* (F1)–(F3)*, то F есть функция распределения некоторой случайной величины ξ, т. е. найдётся вероятностное пространство $\langle \Omega, \mathcal{F}, \mathsf{P} \rangle$ и случайная величина ξ на нём такая, что $F(x) \equiv F_\xi(x)$.*

Эту теорему мы доказывать не станем. Хотя её можно попробовать доказать конструктивно — предъявив то вероятностное пространство (проще всего отрезок $\Omega = [0,1]$ с σ-алгеброй борелевских множеств и мерой Лебега) и ту случайную величину, о существовании которых идёт речь.

Упражнение. Непременно попробуйте сделать это! Например, можно попробовать, не подойдёт ли $\xi(\omega) = \sup\{x : F(x) < \omega\}$.

Помимо отмеченных в теореме 20, функции распределения обладают следующими свойствами:

(F4) В любой точке x_0 разница $F_\xi(x_0 + 0) - F_\xi(x_0)$ равна $\mathsf{P}(\xi = x_0)$:

$$F_\xi(x_0 + 0) - F_\xi(x_0) = \lim_{x \to x_0 + 0} F_\xi(x) - F_\xi(x_0) = \mathsf{P}(\xi = x_0),$$

или, иначе говоря, $F_\xi(x_0 + 0) = F_\xi(x_0) + \mathsf{P}(\xi = x_0) = \mathsf{P}(\xi \leqslant x_0)$.

Упражнение. Докажите сами (так же, как мы доказывали (F2) и (F3)).

Заметим, что разница $F_\xi(x_0+0)-F_\xi(x_0)$ между пределом при стремлении к x_0 справа и значением в точке x_0 есть величина скачка функции распределения. Эта величина равна нулю, если функция распределения непрерывна (справа) в точке x_0. Слева функция распределения непрерывна всегда.

Замечание 14. Очень часто функцией распределения называют $\mathsf{P}(\xi \leqslant x)$. Эта функция отличается от определённой выше лишь тем, что она непрерывна справа, а не слева. Соответственно, вероятность $\mathsf{P}(\xi = x_0)$ для неё равна величине скачка слева, а не справа.

(F5) Для любой случайной величины ξ имеет место равенство:

$$\mathsf{P}(a \leqslant \xi < b) = F_\xi(b) - F_\xi(a). \tag{13}$$

Если же функция распределения $F_\xi(x)$ непрерывна в точках a и b, то

$$\mathsf{P}(a \leqslant \xi < b) = \mathsf{P}(a < \xi < b) = \mathsf{P}(a \leqslant \xi \leqslant b) = \mathsf{P}(a < \xi \leqslant b) = F_\xi(b) - F_\xi(a).$$

Доказательство. Докажем только равенство (13). Все остальные равенства следуют из него и свойства (F4).

Заметим, что $\{\xi < a\} \cup \{a \leqslant \xi < b\} = \{\xi < b\}$, и первые два события несовместны. Поэтому $\mathsf{P}\{\xi < a\} + \mathsf{P}\{a \leqslant \xi < b\} = \mathsf{P}\{\xi < b\}$, или $F_\xi(a) + \mathsf{P}\{a \leqslant \xi < b\} = F_\xi(b)$, что и требовалось доказать. □

Функция распределения дискретного распределения. Мы видели, как выглядят функции распределения некоторых дискретных распределений. Согласно определению дискретного распределения, его функция распределения может быть найдена по таблице распределения так:

$$F_\xi(x) = \mathsf{P}(\xi < x) = \sum_{k\,:\,a_k < x} \mathsf{P}(\xi = a_k).$$

Из свойств (F4) и (F5) получаем следующее свойство.

Свойство 8. *Случайная величина ξ имеет дискретное распределение тогда и только тогда, когда функция распределения $F_\xi(x)$ — ступенчатая функция. При этом значения ξ суть точки a_i скачков $F_\xi(x)$, и $p_i = \mathsf{P}(\xi = a_i) = F_\xi(a_i + 0) - F_\xi(a_i)$ — величины скачков.*

Упражнение. Доказать, что любая функция распределения имеет не более чем счётное число точек разрыва (или «скачков»). Сколько скачков величиной более $1/2$ может иметь функция распределения? Не больше одного или не больше двух? А скачков величиной более $1/3$? Более $1/4$?

Свойства абсолютно непрерывного распределения. Пусть случайная величина ξ имеет абсолюютно непрерывное распределение с плотностью $f_\xi(t)$. Тогда функция распределения в любой точке x может быть найдена по плотности распределения так:

$$F_\xi(x) = \mathsf{P}(\xi < x) = \mathsf{P}(\xi \in (-\infty,\, x)) = \int\limits_{-\infty}^{x} f_\xi(t)\,dt. \tag{14}$$

Поскольку функция распределения однозначно определяет распределение случайной величины, можно считать возможность представить функцию распределения интегралом (14) от неотрицательной функции определением абсолютно непрерывного распределения.

(f3) Если случайная величина ξ имеет абсолютно непрерывное распределение, то её функция распределения всюду непрерывна.

Доказательство. Этот факт следует из свойства 7 и (F4). Заметим, что (f3) есть также следствие представления (14) и непрерывности интеграла как функции верхнего предела. □

(f4) Если случайная величина ξ имеет абсолютно непрерывное распределение, то её функция распределения дифференцируема почти всюду, и

$$f_\xi(x) = F'_\xi(x) = \frac{d}{dx}\, F_\xi(x) \text{ для почти всех } x.$$

Замечание 15. Термин для «почти всех» означает «для всех, кроме (возможно) x из некоторого множества нулевой меры Лебега».

Заметим, что любая функция распределения дифференцируема почти всюду. Например, функции распределения равномерного распределения и распределения Бернулли дифференцируемы всюду, кроме двух точек. Но у равномерного распределения плотность существует, а у распределения Бернулли — нет. Поэтому возможность дифференцировать функцию распределения никакого отношения к существованию плотности не имеет. Даже если мы дополнительно потребуем непрерывности функции распределения, этого не будет достаточно для абсолютной непрерывности распределения. Например, далее мы увидим, что функция распределения сингулярного распределения непрерывна и дифференцируема почти всюду, однако плотности у этого распределения нет, так как производная функции распределения почти всюду равна нулю.

Опираясь на свойства (f4) и (14), можно сформулировать такой критерий абсолютной непрерывности распределения: распределение с функцией распределения $F_\xi(x)$ абсолютно непрерывно, если при всех x

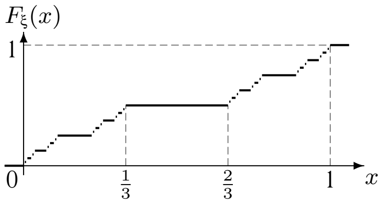
$$F_\xi(x) = \int\limits_{-\infty}^{x} F'_\xi(t)\, dt.$$

Из определения абсолютно непрерывного распределения и свойства 7 сразу следует свойство:

(f5) Если случайная величина ξ имеет абсолютно непрерывное распределение, то для любых $a < b$ имеют место равенства:

$$\mathsf{P}(a < \xi < b) = \mathsf{P}(a \leqslant \xi < b) = \mathsf{P}(a < \xi \leqslant b) = \mathsf{P}(a \leqslant \xi \leqslant b) = \int\limits_a^b f_\xi(t)\, dt.$$

Функция распределения сингулярного распределения. Для полноты картины посмотрим, какие свойства имеет функция распределения сингулярного распределения. Согласно определению 31, случайная величина ξ с сингулярным распределением принимает с единичной вероятностью лишь значения из некоторого борелевского множества B с нулевой лебеговой мерой. Поэтому $\mathsf{P}(\xi \in \mathbb{R} \setminus B) = 0$. Но согласно равенству (13), если $\mathsf{P}(\xi \in [a, b)) = 0$, то $F_\xi(b) = F_\xi(a)$, т. е. расти функция распределения может лишь в точках множества B. На всём остальном множестве $\mathbb{R} \setminus B$ функция распределения имеет нулевую производную (в точках, где эта производная существует, т. е. почти всюду). Тем не менее, $F_\xi(x)$ всюду непрерывна, поскольку $\mathsf{P}(\xi = x) = 0$ для любой точки $x \in \mathbb{R}$. Примером такой функции распределения служит лестница Кантора:



Функция распределения смешанного распределения. Функция распределения смешанного распределения есть линейная комбинация функций распределения дискретного, абсолютно непрерывного и сингулярного распределений.Если смешивать только дискретное и абсолютно непрерывное распределения, то функция распределения будет иметь разрывы в точках значений дискретного распределения и участки непрерывного роста, приращение функции на которых восстанавливается по её производной.

§ 7. Свойства нормального распределения

Установим связь между функциями $\Phi_{a,\sigma^2}(x)$ и $\Phi_{0,1}(x)$.

Свойство 9. *Для любого* $x \in \mathbb{R}$ *справедливо соотношение:*

$$\Phi_{a,\sigma^2}(x) = \Phi_{0,1}\left(\frac{x-a}{\sigma}\right).$$

Доказательство.

$$\Phi_{a,\sigma^2}(x) = \int\limits_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}}\, e^{-(t-a)^2/\,2\sigma^2}\, dt = \int\limits_{-\infty}^{\frac{x-a}{\sigma}} \frac{1}{\sqrt{2\pi}}\, e^{-y^2/\,2} dy = \Phi_{0,1}\left(\frac{x-a}{\sigma}\right).$$

Мы сделали замену переменных $y = (t-a)\,/\,\sigma$, $dt = \sigma\, dy$, верхняя граница интегрирования $t = x$ при такой замене перешла в $y = (x-a)\,/\,\sigma$. □

То же самое для случайных величин можно сформулировать так:

Следствие 2. *Если* $\xi \in \mathrm{N}_{a,\sigma^2}$, *то* $\eta = \dfrac{\xi - a}{\sigma} \in \mathrm{N}_{0,1}$.

Доказательство. Убедимся, что случайная величина η имеет функцию распределения $\Phi_{0,1}(x)$:

$$\begin{aligned} F_\eta(x) &= \mathsf{P}(\eta < x) = \mathsf{P}\left(\frac{\xi - a}{\sigma} < x\right) = \mathsf{P}(\xi < \sigma x + a) = \\ &= \Phi_{a,\sigma^2}(\sigma x + a) = \Phi_{0,1}\left(\frac{\sigma x + a - a}{\sigma}\right) = \Phi_{0,1}(x). \quad \square \end{aligned}$$

Следствие 3. *Если* $\xi \in \mathrm{N}_{a,\sigma^2}$, *то*

$$\mathsf{P}(x_1 < \xi < x_2) = \Phi_{a,\sigma^2}(x_2) - \Phi_{a,\sigma^2}(x_1) = \Phi_{0,1}\left(\frac{x_2 - a}{\sigma}\right) - \Phi_{0,1}\left(\frac{x_1 - a}{\sigma}\right).$$

Видим, что вычисление любых вероятностей для нормально распределённой случайной величины сводится к вычислению функции распределения $\Phi_{0,1}(x)$. Она обладает следующими свойствами (нарисуйте их на графике плотности стандартного нормального распределения):

Свойство 10. $\Phi_{0,1}(0) = 0{,}5$, $\quad \Phi_{0,1}(-x) = 1 - \Phi_{0,1}(x)$.

Свойство 11. *Если* $\xi \in \mathrm{N}_{0,1}$, *то для любого* $x > 0$

$$\mathsf{P}(|\xi| < x) = 1 - 2\Phi_{0,1}(-x) = 2\Phi_{0,1}(x) - 1.$$

Доказательство. При $x > 0$ имеем:

$$\begin{aligned} \mathsf{P}(|\xi| < x) &= \mathsf{P}(-x < \xi < x) = \Phi_{0,1}(x) - \Phi_{0,1}(-x) = \\ &= 1 - 2\Phi_{0,1}(-x) = 2\Phi_{0,1}(x) - 1. \qquad \square \end{aligned}$$

Свойство 12 (правило трех сигм). *Если* $\xi \in \mathrm{N}_{a,\sigma^2}$, *то*

$$\mathsf{P}(|\xi - a| \geqslant 3\sigma) = 0{,}0027 \quad \textit{(совсем мало)}.$$

Доказательство. Перейдём к противоположному событию:

$$\mathsf{P}\big(|\xi - a| \geqslant 3\sigma\big) = 1 - \mathsf{P}\big(|\xi - a| < 3\sigma\big) = 1 - \mathsf{P}\left(\left|\frac{\xi - a}{\sigma}\right| < 3\right).$$

Но величина $\eta = (\xi - a)/\sigma$ имеет стандартное нормальное распределение, и можно использовать свойство 11: $1 - \mathsf{P}(|\eta| < 3) = 1 - (1 - 2\Phi_{0,1}(-3)) = 2\Phi_{0,1}(-3) = 2 \cdot 0{,}00135 = 0{,}0027$ (найти в таблице!). $\square$

Большого смысла в запоминании числа 0,0027 нет, но полезно помнить, что почти вся масса нормального распределения сосредоточена в границах от $a - 3\sigma$ до $a + 3\sigma$.

ГЛАВА 7
Преобразования случайных величин

...По данному соотношению между флюентами определить соотношение между флюксиями.

И. Ньютон, Метод флюксий и бесконечных рядов...

Пусть на вероятностном пространстве $\langle \Omega, \mathcal{F}, \mathsf{P} \rangle$ задана случайная величина ξ. Если функция $g : \mathbb{R} \to \mathbb{R}$ такова, что $g(\xi)$ — случайная величина, то нужно уметь находить распределение $g(\xi)$ по распределению ξ. Эта проблема возникает, например, при моделировании случайных величин с заданным распределением. Датчик случайных чисел может генерировать лишь значения случайных величин с равномерным распределением. А если нам необходимы значения показательно распределённой случайной величины, нужно знать, какое преобразование применить, чтобы из равномерного распределения получить показательное.

§ 1. Измеримость функций от случайных величин

Вопрос об измеримости $g(\xi)$ решает следующая теорема.

Теорема 22. *Пусть ξ — случайная величина, а $g : \mathbb{R} \to \mathbb{R}$ — борелевская (измеримая по Борелю) функция, т. е. такая, что для всякого борелевского множества B его прообраз $g^{-1}(B)$ есть снова борелевское множество. Тогда $g(\xi)$ — случайная величина.*

Доказательство. Проверим, что прообраз любого борелевского множества при отображении $g(\xi) : \Omega \to \mathbb{R}$ является событием. Возьмём произвольное $B \in \mathfrak{B}(\mathbb{R})$ и положим $B_1 = g^{-1}(B)$. Множество B_1 борелевское, так как функция g измерима по Борелю. Найдём $(g(\xi))^{-1}(B)$:

$$\{\omega \mid g(\xi(\omega)) \in B\} = \{\omega \mid \xi(\omega) \in g^{-1}(B) = B_1\} = \xi^{-1}(B_1) \in \mathcal{F},$$

поскольку $B_1 \in \mathfrak{B}(\mathbb{R})$ и ξ — случайная величина. □

Борелевскими являются все привычные нам функции. Функцией, неизмеримой по Борелю, будет, например, индикаторная функция неизмеримого множества Витали. Вообще говоря, неизмеримые функции суть объекты экзотические, в обычной жизни не встречающиеся.

§ 2. Распределения функций от случайных величин

Линейные и монотонные преобразования. Если ξ имеет дискретное распределение, то для любой борелевской функции g величина $g(\xi)$ также имеет дискретное распределение, и таблица её распределения находится просто по определению. Поэтому мы будем рассматривать преобразования случайных величин с абсолютно непрерывными распределениями.

Пусть случайная величина ξ имеет функцию распределения $F_\xi(x)$ и плотность распределения $f_\xi(x)$. Построим с помощью борелевской функции $g : \mathbb{R} \to \mathbb{R}$ случайную величину $\eta = g(\xi)$. Требуется найти функцию распределения и, если существует, плотность распределения величины η.

Замечание 16. Плотность распределения случайной величины $\eta = g(\xi)$ существует далеко не при любых функциях g. Так, если функция g кусочно-постоянна, то η имеет дискретное распределение, и плотность её распределения не существует.

Упражнение. Привести пример плотности распределения случайной величины ξ и непрерывной функции g таких, что $\eta = g(\xi)$ имеет:

а) дискретное распределение; б) невырожденное дискретное распределение.

Плотность распределения величины $\eta = g(\xi)$ заведомо существует, если, например, функция g (строго) монотонна. В общем случае мы не можем просто продифференцировать функцию распределения, поскольку не знаем, существует ли плотность. Следует доказать, что распределение абсолютно непрерывно. Но доказывая это, мы попутно найдём и плотность распределения. Действительно, у нас есть следующий путь доказательства абсолютной непрерывности распределения. Если, согласно равенству (14), можно для любого x представить функцию распределения величины η в виде

$$F_\eta(x) = \int\limits_{-\infty}^{x} h(y)\, dy, \qquad \text{где } h(y) \geqslant 0,$$

то плотность распределения величины η существует и равна подынтегральной функции: $f_\eta(x) = h(x)$. Другой путь — продифференцировать функцию распределения и убедиться, что производная является плотностью распределения, т. е. обладает свойствами (f1) и (f2).

Теорема 23. *Пусть ξ имеет функцию распределения $F_\xi(x)$ и плотность распределения $f_\xi(x)$, и постоянная a отлична от нуля. Тогда случайная величина $\eta = a\xi + b$ имеет плотность распределения*

$$f_\eta(x) = \frac{1}{|a|}\, f_\xi\Big(\frac{x-b}{a}\Big).$$

Доказательство. Пусть сначала $a > 0$.

$$F_\eta(x) = \mathsf{P}(a\xi + b < x) = \mathsf{P}\Big(\xi < \frac{x-b}{a}\Big) = F_\xi\Big(\frac{x-b}{a}\Big) = \int\limits_{-\infty}^{(x-b)/a} f_\xi(t)\, dt.$$

Сделаем замену переменной в последнем интеграле. Переменную t заменим на новую переменную y так: $t = (y-b)\,/\,a$. Тогда $dt = dy\,/\,a$, верхняя граница области интегрирования $t = (x-b)\,/\,a$ перейдёт в $y = x$, нижняя $t = -\infty$ перейдёт в $y = -\infty$. Получим

$$F_\eta(x) = \int\limits_{-\infty}^{x} \frac{1}{a}\, f_\xi\Bigl(\frac{y-b}{a}\Bigr)\, dy.$$

Функция под интегралом — плотность распределения $f_\eta(y)$ случайной величины $\eta = a\xi + b$ при $a > 0$.

Пусть теперь $a < 0$.

$$F_\eta(x) = \mathsf{P}(a\xi + b < x) = \mathsf{P}\Bigl(\xi > \frac{x-b}{a}\Bigr) = \int\limits_{(x-b)/a}^{+\infty} f_\xi(t)\, dt.$$

Сделаем ту же замену переменной $t = (y-b)\,/\,a$, $y = at + b$. Но теперь граница интегрирования $t = +\infty$ перейдёт в $y = -\infty$, поскольку $a < 0$. Получим

$$F_\eta(x) = \int\limits_{x}^{-\infty} \frac{1}{a}\, f_\xi\Bigl(\frac{y-b}{a}\Bigr)\, dy = \int\limits_{-\infty}^{x} \frac{1}{|a|}\, f_\xi\Bigl(\frac{y-b}{a}\Bigr)\, dy.$$

Функция под интегралом и есть плотность распределения $f_\eta(y)$ случайной величины $\eta = a\xi + b$ при $a < 0$. □

Для произвольной монотонной функции g справедливо утверждение:

Теорема 24. *Пусть ξ имеет плотность распределения $f_\xi(x)$, и функция $g : \mathbb{R} \to \mathbb{R}$ монотонна. Тогда случайная величина $\eta = g(\xi)$ имеет плотность распределения*

$$f_\eta(x) = \bigl(g^{-1}(x)\bigr)'\; f_\xi\left(g^{-1}(x)\right).$$

Здесь g^{-1} — функция, обратная к g, и $\bigl(g^{-1}(x)\bigr)'$ — её производная.

Упражнение. Доказать теорему 24.

Из теоремы 23 следуют уже знакомые нам утверждения:

Следствие 4. *Если $\xi \in \mathrm{N}_{0,1}$, то $\eta = \sigma\xi + a \in \mathrm{N}_{a,\sigma^2}$.*

Доказательство. Действительно,

$$f_\eta(x) = \frac{1}{\sigma}\, f_\xi\Bigl(\frac{x-a}{\sigma}\Bigr) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-a)^2}{2\sigma^2}}. \qquad \square$$

Следствие 5. *Если $\xi \in \mathrm{N}_{a,\sigma^2}$, то $(\xi - a)\,/\,\sigma \in \mathrm{N}_{0,1}$.*

Следствие 6. *Если $\xi \in \mathrm{U}_{0,\,1}$, то $a\xi + b \in \mathrm{U}_{b,\,a+b}$ при $a > 0$.*

Следствие 7. *Если $\xi \in \mathrm{E}_\alpha$, то $\alpha\xi \in \mathrm{E}_1$.*

Квантильное преобразование.

Теорема 25. *Пусть функция распределения* $F(x) = F_\xi(x)$ *непрерывна. Тогда случайная величина* $\eta = F(\xi)$ *имеет равномерное на отрезке* $[0,\,1]$ *распределение.*

Доказательство. Найдём функцию распределения случайной величины η. Заметим, что всегда $0 \leqslant \eta \leqslant 1$. Предположим сначала, что функция F всюду возрастает. Тогда она обратима, и поэтому

$$F_\eta(x) = \mathsf{P}(F(\xi) < x) = \begin{cases} 0, & \text{если } x \leqslant 0, \\ \mathsf{P}(\xi < F^{-1}(x)), & \text{если } x \in (0,\,1), \\ 1, & \text{если } x \geqslant 1. \end{cases} \tag{15}$$

Но $\mathsf{P}(\xi < F^{-1}(x)) = F\left(F^{-1}(x)\right) = x$, т. е. $\eta \in \mathrm{U}_{0,1}$.

Если функция F не является всюду возрастающей, то у неё есть участки постоянства. В этом случае просто обозначим через $F^{-1}(x)$ самую левую точку из замкнутого множества $\{t \,|\, F(t) = x\}$ прообразов точки $x \in (0,\,1)$. При таком понимании «обратной» функции равенства (15) остаются справедливыми вместе с равенством $\mathsf{P}(\xi < F^{-1}(x)) = F\left(F^{-1}(x)\right) = x$ для любого $x \in (0,\,1)$. □

Теорему 25 можно использовать для построения случайных величин с заданным распределением по равномерно распределённой случайной величине (например, по результату датчика случайных чисел). Следующее утверждение верно не только для непрерывных, но для любых функций распределения F. Обозначим через $F^{-1}(x)$ точную нижнюю грань множества тех t, для которых $F(t) \geqslant x$:

$$F^{-1}(x) = \inf\{t \,|\, F(t) \geqslant x\}.$$

Для непрерывной функции F это определение «обратной функции» совпадает с введённым в доказательстве теоремы 25.

Теорема 26. *Пусть* $\eta \in \mathrm{U}_{0,1}$, *а* F *— произвольная функция распределения. Тогда случайная величина* $\xi = F^{-1}(\eta)$ *(«квантильное преобразование» над* η*) имеет функцию распределения* F.

Следствие 8. *Пусть* $\eta \in \mathrm{U}_{0,\,1}$. *Верны соотношения:*

$$-\frac{1}{\alpha}\ln(1-\eta) \in \mathrm{E}_\alpha, \qquad a + \sigma\,\mathrm{tg}(\pi\eta - \pi/2) \in C_{a,\sigma}, \qquad \Phi^{-1}_{0,1}(\eta) \in \mathrm{N}_{0,1}.$$

Упражнение. Доказать теорему 26 и следствие 8, а также продолжить список соотношений. Как получить случайную величину с распределением Парето? А с нормальным распределением? (Указание: так её никто не получает).

ГЛАВА 8
Многомерные распределения

Не следует множить сущности сверх необходимости.
Принцип «бритва У. Оккама»

§ 1. Совместное распределение

Пусть случайные величины $\xi_1, \ldots, \xi_n$ заданы на одном вероятностном пространстве $\langle \Omega, \mathcal{F}, \mathsf{P} \rangle$.

Определение 34. Функция

$$F_{\xi_1, \ldots, \xi_n}(x_1, \ldots, x_n) = \mathsf{P}(\xi_1 < x_1, \ldots, \xi_n < x_n)$$

называется функцией распределения вектора $(\xi_1, \ldots, \xi_n)$ или функцией совместного распределения случайных величин $\xi_1, \ldots, \xi_n$.

Перечислим свойства функции совместного распределения. Для простоты обозначений ограничимся вектором (ξ_1, ξ_2) из двух величин.

(F0) Для любых x_1, x_2 верно неравенство: $0 \leqslant F_{\xi_1, \xi_2}(x_1, x_2) \leqslant 1$.

(F1) $F_{\xi_1, \xi_2}(x_1, x_2)$ не убывает по каждой координате вектора (x_1, x_2).

(F2) Для любого $i = 1, 2$ существует $\lim\limits_{x_i \to -\infty} F_{\xi_1, \xi_2}(x_1, x_2) = 0$. Существует двойной предел $\lim\limits_{x_1 \to +\infty} \lim\limits_{x_2 \to +\infty} F_{\xi_1, \xi_2}(x_1, x_2) = 1$.

(F3) Функция $F_{\xi_1, \xi_2}(x_1, x_2)$ по каждой координате вектора (x_1, x_2) непрерывна слева.

(F4) Чтобы по функции совместного распределения восстановить функции распределения ξ_1 и ξ_2 в отдельности, следует устремить мешающую переменную к $+\infty$:

$$\lim_{x_1 \to +\infty} F_{\xi_1, \xi_2}(x_1, x_2) = F_{\xi_2}(x_2), \quad \lim_{x_2 \to +\infty} F_{\xi_1, \xi_2}(x_1, x_2) = F_{\xi_1}(x_1). \quad (16)$$

Доказательство всех этих свойств совершенно аналогично одномерному случаю. Но теперь свойств (F0)—(F3) не хватает для описания класса функций совместного распределения. Т. е. выполнение этих свойств для некоторой функции $F : \mathbb{R}^2 \to \mathbb{R}$ не гарантирует, что эта функция является функцией распределения некоторого случайного вектора.

Упражнение. Доказать, что функция

$$F(x_1, x_2) = \begin{cases} 0, & \text{если } x_1 \leqslant 0 \text{ или } x_2 \leqslant 0 \text{ или } x_1 + x_2 \leqslant 1; \\ 1, & \text{если одновременно } x_1 > 0,\ x_2 > 0,\ x_1 + x_2 > 1. \end{cases}$$

удовлетворяет всем свойствам (F0)—(F3), но не является функцией распределения никакого вектора (ξ_1, ξ_2) хотя бы потому, что, найдись такой вектор, найдётся и прямоугольник $[a_1, b_1) \times [a_2, b_2)$, «вероятность» попасть в который (вычисленная с помощью этой «функции распределения») отрицательна:

$$\mathsf{P}(a_1 \leqslant \xi_1 < b_1,\ a_2 \leqslant \xi_2 < b_2) < 0.$$

Легко доказать (убедиться, что легко), что для любых $a_1 < b_1$, $a_2 < b_2$ справедливо равенство: $\mathsf{P}(a_1 \leqslant \xi_1 < b_1,\ a_2 \leqslant \xi_2 < b_2) = F_{\xi_1, \xi_2}(b_1, b_2) + F_{\xi_1, \xi_2}(a_1, a_2) - F_{\xi_1, \xi_2}(a_1, b_2) - F_{\xi_1, \xi_2}(b_1, a_2)$.

Дополнительно к свойствам (F0)—(F3) от функции F требуют неотрицательности этого выражения (при любых $a_1 < b_1$, $a_2 < b_2$).

§ 2. Типы многомерных распределений

Ограничимся рассмотрением двух типичных случаев: когда *совместное* распределение координат случайного вектора (ξ_1, ξ_2) либо дискретно, либо абсолютно непрерывно. Заметим, что сингулярные совместные распределения тоже не являются редкостью, в отличие от одномерного случая: стоит бросить точку наудачу на отрезок на плоскости, и мы получим сингулярное совместное распределение (доказать).

Дискретное совместное распределение.

Определение 35. Говорят, что случайные величины ξ_1, ξ_2 имеют дискретное совместное распределение, если существует конечный или счётный набор $\{a_i, b_j\}$ такой, что

$$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \mathsf{P}(\xi_1 = a_i,\ \xi_2 = b_j) = 1.$$

Таблицу, на пересечении i-й строки и j-го столбца которой стоит вероятность $\mathsf{P}(\xi_1 = a_i,\ \xi_2 = b_j)$, называют таблицей совместного распределения случайных величин ξ_1 и ξ_2.

Таблицы распределения каждой из случайных величин ξ_1, ξ_2 в отдельности (таблицы частных, или *маргинальных* распределений) восстанавливаются по таблице совместного распределения с помощью формул:

$$\mathsf{P}(\xi_1 = a_i) = \sum_{j=1}^{\infty} \mathsf{P}(\xi_1 = a_i, \xi_2 = b_j), \quad \mathsf{P}(\xi_2 = b_j) = \sum_{i=1}^{\infty} \mathsf{P}(\xi_1 = a_i, \xi_2 = b_j).$$

Так, первое равенство следует из того, что набор $\{\xi_2 = b_1\}$, $\{\xi_2 = b_2\}, \ldots$ есть полная группа событий, и поэтому событие $\{\xi_1 = a_i\}$ раскладывается в объединение попарно несовместных событий:

$$\{\xi_1 = a_i\} = \bigcup_{j=1}^{\infty} \{\xi_1 = a_i,\, \xi_2 = b_j\}.$$

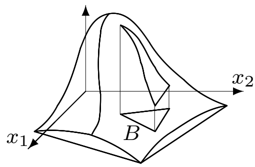
Абсолютно непрерывное совместное распределение.

Определение 36. Говорят, что случайные величины ξ_1, ξ_2 имеют абсолютно непрерывное совместное распределение, если существует неотрицательная функция $f_{\xi_1,\xi_2}(x_1,\, x_2)$ такая, что для любого множества $B \in \mathfrak{B}(\mathbb{R}^2)$ имеет место равенство

$$\mathsf{P}((\xi_1,\, \xi_2) \in B) = \iint\limits_B f_{\xi_1,\xi_2}(s_1,\, s_2)\, ds_1\, ds_2.$$

Если такая функция $f_{\xi_1,\xi_2}(x_1,\, x_2)$ существует, она называется плотностью совместного распределения случайных величин $\xi_1,\, \xi_2$.

Достаточно, если двойной интеграл по множеству B читатель будет понимать как объём области под графиком функции $f_{\xi_1,\xi_2}(x_1,\, x_2)$ над множеством B в плоскости переменных $(x_1,\, x_2)$, как показано на рисунке справа.



Если случайные величины $\xi_1,\, \xi_2$ имеют абсолютно непрерывное совместное распределение, то для любых $x_1,\, x_2$ имеет место равенство:

$$F_{\xi_1,\xi_2}(x_1,\, x_2) = \mathsf{P}(\xi_1 < x_1,\, \xi_2 < x_2) = \int\limits_{-\infty}^{x_1} \left(\int\limits_{-\infty}^{x_2} f_{\xi_1,\xi_2}(s_1,\, s_2)\, ds_2 \right) ds_1.$$

Плотность совместного распределения обладает такими же свойствами, как и плотность распределения одной случайной величины:

(f1) Неотрицательность: $f_{\xi_1,\xi_2}(x_1,\, x_2) \geqslant 0$ для любых $x_1,\, x_2 \in \mathbb{R}$;

(f2) Нормированность: $\displaystyle\iint\limits_{\mathbb{R}^2} f_{\xi_1,\xi_2}(x_1,\, x_2)\, dx_1\, dx_2 = 1.$

Справедливо и обратное: любая функция, обладающая этими свойствами, является плотностью некоторого совместного распределения. Доказательство этого факта ничем не отличается от одномерного случая.

По функции совместного распределения его плотность находится как смешанная частная производная:

(f3) $\displaystyle f_{\xi_1,\xi_2}(x_1,\, x_2) = \frac{\partial^2}{\partial x_1 \partial x_2} F_{\xi_1,\xi_2}(x_1,\, x_2)$ для почти всех $(x_1,\, x_2)$.

Из существования плотностей ξ_1 и ξ_2 не следует абсолютная непрерывность совместного распределения этих случайных величин. Например, вектор (ξ, ξ) принимает значения только на диагонали в $\mathbb{R}^2$ и уже поэтому не имеет плотности совместного распределения (его совместное распределение сингулярно). Обратное же свойство, как показывает следующая теорема, всегда верно: если совместное распределение абсолютно непрерывно, то и частные распределения тоже таковы.

Теорема 27. *Если случайные величины ξ_1 и ξ_2 имеют абсолютно непрерывное совместное распределение с плотностью $f(x_1, x_2)$, то ξ_1 и ξ_2 в отдельности также имеют абсолютно непрерывное распределение с плотностями:*

$$f_{\xi_1}(s_1) = \int\limits_{-\infty}^{\infty} f(s_1, s_2)\, ds_2; \quad f_{\xi_2}(s_2) = \int\limits_{-\infty}^{\infty} f(s_1, s_2)\, ds_1.$$

Для $n > 2$ плотности случайных величин $\xi_1, \ldots, \xi_n$ по плотности их совместного распределения $f(x_1, \ldots, x_n)$ находятся интегрированием функции f по всем «лишним» координатам.

Доказательство. Например, в силу равенств (16),

$$F_{\xi_1}(x_1) = \lim_{x_2 \to +\infty} F_{\xi_1, \xi_2}(x_1, x_2) = \int\limits_{-\infty}^{x_1} \left(\int\limits_{-\infty}^{\infty} f(s_1, s_2)\, ds_2 \right) ds_1 = \int\limits_{-\infty}^{x_1} f_{\xi_1}(s_1)\, ds_1.$$

Аналогично устанавливается и справедливость второго равенства. □

§ 3. Примеры многомерных распределений

Приведём два наиболее употребительных примера абсолютно непрерывных многомерных распределений.

Равномерное распределение. Пусть $S \subset \mathbb{R}^n$ — борелевское множество с конечной лебеговой мерой $\lambda(S)$. Говорят, что вектор $(\xi_1, \ldots, \xi_n)$ имеет равномерное распределение в области S, если плотность совместного распределения $f_{\xi_1, \ldots, \xi_n}(x_1, \ldots, x_n)$ постоянна в области S и равна нулю вне этой области:

$$f_{\xi_1, \ldots, \xi_n}(x_1, \ldots, x_n) = \begin{cases} \dfrac{1}{\lambda(S)}, & \text{если } (x_1, \ldots, x_n) \in S, \\ 0, & \text{если } (x_1, \ldots, x_n) \notin S. \end{cases} \tag{17}$$

Убедимся, что эта функция является плотностью распределения:

$$\int\limits_{\mathbb{R}^n} f_{\xi_1, \ldots, \xi_n}(x_1, \ldots, x_n)\, dx_1 \ldots dx_n = \frac{1}{\lambda(S)} \int\limits_{S} dx_1 \ldots dx_n = \frac{1}{\lambda(S)} \lambda(S) = 1.$$

Как и в одномерном случае, вектор $(\xi_1, \ldots, \xi_n)$ с равномерным распределением в области S есть просто вектор координат точки, брошенной наудачу в область S.

Многомерное нормальное распределение. Пусть $\Sigma > 0$ — положительно определённая симметричная матрица $(n \times n)$, матрица Σ^{-1} — обратная к Σ, и $\vec{a} \in \mathbb{R}^n$ — n-мерный вектор-столбец. Транспонированный вектор мы будем обозначать так: $\vec{a}^T = (a_1, \ldots, a_n)$.

Говорят, что вектор $(\xi_1, \ldots, \xi_n)$ имеет многомерное нормальное распределение $\mathrm{N}_{\vec{a}, \Sigma}$ с вектором средних $\vec{a}$ и матрицей ковариаций Σ, если плотность совместного распределения $f_{\xi_1, \ldots, \xi_n}(x_1, \ldots, x_n)$ равна

$$f_{\vec{\xi}}(\vec{x}) = \frac{1}{\sqrt{\det\Sigma}\,\left(\sqrt{2\pi}\right)^n} \exp\left\{-\frac{1}{2}(\vec{x}-\vec{a})^T \cdot \Sigma^{-1} \cdot (\vec{x}-\vec{a})\right\}.$$

Мы не будем проверять, что эта функция является плотностью совместного распределения, поскольку для этого требуется умение заменять переменные в многомерном интеграле. Выражение $(\vec{x}-\vec{a})^T\Sigma^{-1}(\vec{x}-\vec{a})$ в показателе экспоненты является квадратичной формой от переменных $(x_i - a_i)$: действительно, для матрицы $B = \Sigma^{-1}$ с элементами b_{ij}

$$(\vec{x}-\vec{a})^T B(\vec{x}-\vec{a}) = \sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij}(x_i - a_i)(x_j - a_j).$$

Подробно с многомерным нормальным распределением мы познакомимся в курсе математической статистики, и там же выясним, что означают слова «с вектором средних $\vec{a}$ и матрицей ковариаций Σ».

В частном случае, когда Σ — диагональная матрица с элементами $\sigma_1^2, \ldots, \sigma_n^2$ на диагонали, совместная плотность превращается в произведение плотностей нормальных случайных величин:

$$\begin{aligned} f_{\vec{\xi}}(\vec{x}) &= \frac{1}{\sigma_1 \ldots \sigma_n \left(\sqrt{2\pi}\right)^n} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\frac{1}{\sigma_i^2}(x_i - a_i)^2\right\} = \\ &= \frac{1}{\sigma_1\sqrt{2\pi}}\, e^{-\frac{(x-a)^2}{2\sigma_1^2}} \cdot \ldots \cdot \frac{1}{\sigma_n\sqrt{2\pi}}\, e^{-\frac{(x-a)^2}{2\sigma_n^2}}. \end{aligned}$$

Скоро мы увидим, что это равенство означает независимость случайных величин $\xi_1, \ldots, \xi_n$.

§ 4. Роль совместного распределения

Если нам известно совместное распределение двух или нескольких случайных величин, становится возможным отыскать распределение суммы, разности, произведения, частного, иных функций от этих случайных величин. Заметим (но не будем доказывать), что применение к набору случайных величин многих привычных нам функций не выводит нас из класса случайных величин. Интересующийся читатель может попробовать доказать, например, что сумма двух случайных величин есть снова случайная величина.

Следующие два простых примера показывают, что знания только частных распределений двух случайных величин недостаточно для отыскания распределения, например, суммы этих величин. Для этого необходимо знать их совместное распределение. Распределение суммы (и любой иной функции) не определяется, вообще говоря, распределениями слагаемых: при одних и тех же распределениях слагаемых распределение суммы может быть разным в зависимости от совместного распределения слагаемых.

Пример 33. Рассмотрим две случайные величины ξ и η с одним и тем же распределением Бернулли с параметром $p = 1/2$ и следующей таблицей совместного распределения: для $0 \leqslant r \leqslant 1/2$ положим

$$\mathsf{P}(\xi = 0,\, \eta = 0) = r, \qquad \mathsf{P}(\xi = 0,\, \eta = 1) = \frac{1}{2} - r,$$
$$\mathsf{P}(\xi = 1,\, \eta = 0) = \frac{1}{2} - r, \qquad \mathsf{P}(\xi = 1,\, \eta = 1) = r,$$

Если $r = 0$, то $\mathsf{P}(\xi + \eta = 1) = \mathsf{P}(\xi = 0,\, \eta = 1) + \mathsf{P}(\xi = 1,\, \eta = 0) = 1$, т. е. распределение $\xi + \eta$ вырождено в точке 1.

Если $r = 1/2$, то $\mathsf{P}(\xi{+}\eta = 0) = \mathsf{P}(\xi{+}\eta = 2) = 1/2$, т. е. $\xi{+}\eta$ имеет невырожденное дискретное распределение, принимая значения 0 и 2 с равными вероятностями.

Взяв $r = 1/4$, получим $\mathsf{P}(\xi + \eta = 0) = 1/4$, $\mathsf{P}(\xi + \eta = 2) = 1/4$ и $\mathsf{P}(\xi + {}+\eta = 1) = 1/2$, т. е. $\xi{+}\eta$ имеет биномиальное распределение с параметрами 2 и 1/2.

Если взять $r = 1/3$, получим уже $\mathsf{P}(\xi{+}\eta = 0) = 1/3$, $\mathsf{P}(\xi{+}\eta = 1) = 1/3$ и $\mathsf{P}(\xi + \eta = 2) = 1/3$, т. е. $\xi + \eta$ принимает значения 1, 2 и 3 с равными вероятностями (это не биномиальное распределение).

Ещё раз отметим, что частные распределения ξ и η от r не зависят. Распределение суммы меняется вместе с совместным распределением ξ и η при неизменных частных распределениях величин ξ и η.

Пример 34. Пусть случайная величина ξ имеет стандартное нормальное распределение.

Возьмём $\eta = -\xi$. Тогда η тоже имеет стандартное нормальное распределение, а сумма $\xi + \eta = 0$ имеет вырожденное распределение.

Возьмём теперь $\eta = \xi$. Тогда сумма $\xi + \eta = 2\xi$ имеет уже не вырожденное, а нормальное распределение $\mathrm{N}_{0,\,4}$ (проверить!).

Распределение функции от нескольких случайных величин может определяться их частными распределениями, если, например, потребовать независимости этих случайных величин. В этом случае совместное распределение также полностью определяется частными распределениями (а именно, как их произведение).

§ 5. Независимость случайных величин

Определение 37. Случайные величины $\xi_1, \ldots, \xi_n$ называют независимыми (в совокупности), если для любого набора борелевских множеств $B_1, \ldots, B_n \in \mathfrak{B}(\mathbb{R})$ имеет место равенство:

$$\mathsf{P}(\xi_1 \in B_1, \ldots, \xi_n \in B_n) = \mathsf{P}(\xi_1 \in B_1) \cdot \ldots \cdot \mathsf{P}(\xi_n \in B_n).$$

Определение 38. Случайные величины $\xi_1, \ldots, \xi_n$ называют попарно независимыми, если независимы любые две из них.

Оба этих определения годятся не только для конечного набора случайных величин, но и для их бесконечной последовательности.

Замечание 17. Независимость случайных величин в совокупности влечёт попарную независимость. Достаточно в определении независимости в качестве нескольких борелевских множеств взять числовую прямую $\mathbb{R}$.

Пример 35. Вспомним пример Бернштейна 25. Свяжем с событиями A, B и C случайные величины ξ_1, ξ_2 и ξ_3 — индикаторы этих событий. Например, $\xi_1 = 1$, если A произошло, и $\xi_1 = 0$, если A не произошло. Случайные величины ξ_1, ξ_2 и ξ_3 независимы попарно (проверить), но зависимы в совокупности:

$$\mathsf{P}(\xi_1 = 1, \xi_2 = 1, \xi_3 = 1) = \mathsf{P}(A \cap B \cap C) = \frac{1}{4},$$

$$\mathsf{P}(\xi_1 = 1)\,\mathsf{P}(\xi_2 = 1)\,\mathsf{P}(\xi_3 = 1) = \mathsf{P}(A)\,\mathsf{P}(B)\,\mathsf{P}(C) = \frac{1}{8}.$$

Попарная независимость случайных величин встречается редко. Поэтому всюду, где мы будем употреблять термин «независимы», будет подразумеваться независимость в совокупности.

Определение независимости можно сформулировать в терминах функций распределения:

Определение 39. Случайные величины $\xi_1, \ldots, \xi_n$ независимы (в совокупности), если для любых $x_1, \ldots, x_n$ имеет место равенство:

$$F_{\xi_1, \ldots, \xi_n}(x_1, \ldots, x_n) = F_{\xi_1}(x_1) \cdot \ldots \cdot F_{\xi_n}(x_n).$$

Для случайных величин с дискретным распределением эквивалентное определение независимости выглядит так:

Определение 40. Случайные величины $\xi_1, \ldots, \xi_n$ с дискретным распределением независимы (в совокупности), если для любых чисел $a_1, \ldots, a_n$ имеет место равенство:

$$\mathsf{P}(\xi_1 = a_1, \ldots, \xi_n = a_n) = \mathsf{P}(\xi_1 = a_1) \cdot \ldots \cdot \mathsf{P}(\xi_n = a_n).$$

Упражнение. Доказать, что из независимости в смысле определения 37 следует независимость в смысле определения 39 (доказательство в обратную сторону см. в § 4 гл. 3 учебника А. А. Боровкова «Теория вероятностей»).

Упражнение. Доказать, что для случайных величин с дискретным распределением определения 37 и 40 эквивалентны.

Для случайных величин с абсолютно непрерывным совместным распределением определение независимости можно сформулировать так:

Определение 41. Случайные величины $\xi_1, \ldots, \xi_n$ с абсолютно непрерывным совместным распределением независимы (в совокупности), если плотность совместного распределения равна произведению плотностей случайных величин $\xi_1, \ldots, \xi_n$, т. е. для любых $x_1, \ldots, x_n$ имеет место равенство: $f_{\xi_1,\ldots,\xi_n}(x_1, \ldots, x_n) = f_{\xi_1}(x_1) \cdot \ldots \cdot f_{\xi_n}(x_n)$.

Замечание 18. Плотность распределения определяется с точностью до её значений на множестве нулевой лебеговой меры (распределение не меняется от изменения плотности на множестве нулевой меры). Поэтому равенство плотности совместного распределения и произведения плотностей можно понимать тоже как равенство «почти всюду».

Доказательство. Докажем эквивалентность определений 39 и 41. По теореме 27, если совместное распределение $\xi_1, \ldots, \xi_n$ абсолютно непрерывно, то и в отдельности $\xi_1, \ldots, \xi_n$ также имеют абсолютно непрерывные распределения. Пусть случайные величины $\xi_1, \ldots, \xi_n$ независимы в смысле определения 39, т. е. для любых $x_1, \ldots, x_n$

$$F_{\xi_1,\ldots,\xi_n}(x_1, \ldots, x_n) = F_{\xi_1}(x_1) \cdot \ldots \cdot F_{\xi_n}(x_n).$$

Но функция совместного распределения равна

$$F_{\xi_1,\ldots,\xi_n}(x_1, \ldots, x_n) = \int\limits_{-\infty}^{x_1} \ldots \int\limits_{-\infty}^{x_n} f_{\xi_1,\ldots,\xi_n}(s_1, \ldots, s_n)\, ds_1 \ldots ds_n = I_1,$$

а произведение функций распределения записывается произведением интегралов, или одним n-мерным интегралом:

$$F_{\xi_1}(x_1) \cdot \ldots \cdot F_{\xi_n}(x_n) = \int\limits_{-\infty}^{x_1} f_{\xi_1}(s_1)\, ds_1 \cdot \ldots \cdot \int\limits_{-\infty}^{x_n} f_{\xi_n}(s_n)\, ds_n =$$

$$= \int\limits_{-\infty}^{x_1} \ldots \int\limits_{-\infty}^{x_n} f_{\xi_1}(s_1) \ldots f_{\xi_n}(s_n)\, ds_1 \ldots ds_n = I_2.$$

Равенство интегралов $I_1 = I_2$ при всех значениях $x_1, \ldots, x_n$ влечёт, после дифференцирования по $x_1, \ldots, x_n$, равенство подынтегральных выражений почти всюду (дифференцировать можно почти всюду), т. е. независимость в смысле определения 41. Для доказательства в обратную сторону можно использовать те же равенства, но в обратном порядке. $\square$

§ 6. Функции от двух случайных величин

Пусть ξ_1 и ξ_2 — случайные величины с плотностью совместного распределения $f_{\xi_1,\xi_2}(x_1, x_2)$, и задана борелевская функция $g : \mathbb{R}^2 \to \mathbb{R}$. Требуется найти функцию (а если существует, то и плотность) распределения случайной величины $\eta = g(\xi_1, \xi_2)$.

Пользуясь тем, что вероятность случайному вектору попасть в некоторую область можно вычислить как объем под графиком плотности распределения вектора над этой областью, сформулируем утверждение.

Теорема 28. *Пусть $x \in \mathbb{R}$, и область $D_x \subseteq \mathbb{R}^2$ состоит из точек (x_1, x_2) таких, что $g(x_1, x_2) < x$. Тогда случайная величина $\eta = g(\xi_1, \xi_2)$ имеет функцию распределения*

$$F_\eta(x) = \mathsf{P}\big(g(\xi_1, \xi_2) < x\big) = \mathsf{P}\big((\xi_1, \xi_2) \in D_x\big) = \iint\limits_{D_x} f_{\xi_1,\xi_2}(x_1, x_2)\, dx_1\, dx_2.$$

Далее в этой главе предполагается, что случайные величины ξ_1 и ξ_2 независимы, т. е. $f_{\xi_1,\xi_2}(x_1, x_2) \equiv f_{\xi_1}(x_1)\, f_{\xi_2}(x_2)$. В этом случае распределение величины $g(x_1, x_2)$ полностью определяется частными распределениями величин ξ_1 и ξ_2.

Следствие 9 (формула свёртки). *Если случайные величины ξ_1 и ξ_2 независимы и имеют абсолютно непрерывные распределения с плотностями $f_{\xi_1}(x_1)$ и $f_{\xi_2}(x_2)$, то плотность распределения суммы $\xi_1 + \xi_2$ равна «свёртке» плотностей f_{ξ_1} и f_{ξ_2}:*

$$f_{\xi_1+\xi_2}(t) = \int\limits_{-\infty}^{\infty} f_{\xi_1}(u)\, f_{\xi_2}(t-u)\, du = \int\limits_{-\infty}^{\infty} f_{\xi_2}(u)\, f_{\xi_1}(t-u)\, du \qquad (18)$$

Доказательство. Воспользуемся утверждением теоремы 28 для борелевской функции $g(x_1, x_2) = x_1 + x_2$. Интегрирование по области $D_x = \{(x_1, x_2) \mid x_1 + x_2 < x\}$ можно заменить последовательным вычислением двух интегралов: наружного — по переменной x_1, меняющейся в пределах от $-\infty$ до $+\infty$, и внутреннего — по переменной x_2, которая при каждом x_1 должна быть меньше, чем $x - x_1$. Здесь $D_x = \{(x_1, x_2) \mid x_1 \in \mathbb{R}, x_2 \in (-\infty, x - x_1)\}$. Поэтому

$$F_{\xi_1+\xi_2}(x) = \iint\limits_{D_x} f_{\xi_1}(x_1)\, f_{\xi_2}(x_2)\, dx_2\, dx_1 = \int\limits_{-\infty}^{\infty} \left(\int\limits_{-\infty}^{x-x_1} f_{\xi_1}(x_1) f_{\xi_2}(x_2)\, dx_2 \right) dx_1.$$

Сделаем в последнем интеграле замену переменной x_2 на t так: $x_2 = t - x_1$. При этом $x_2 \in (-\infty, x - x_1)$ перейдёт в $t \in (-\infty, x)$, $dx_2 = dt$. В полученном интеграле меняем порядок интегрирования: функция распределения

$F_{\xi_1+\xi_2}(x)$ равна

$$\int\limits_{-\infty}^{\infty}\left(\int\limits_{-\infty}^{x} f_{\xi_1}(x_1)\, f_{\xi_2}(t-x_1)\, dt\right) dx_1 = \int\limits_{-\infty}^{x}\left(\int\limits_{-\infty}^{\infty} f_{\xi_1}(x_1)\, f_{\xi_2}(t-x_1)\, dx_1\right) dt.$$

Итак, мы представили функцию распределения $F_{\xi_1+\xi_2}(x)$ в виде $\int\limits_{-\infty}^{x} f_{\xi_1+\xi_2}(t)\, dt$, где

$$f_{\xi_1+\xi_2}(t) = \int\limits_{-\infty}^{\infty} f_{\xi_1}(x_1)\, f_{\xi_2}(t-x_1)\, dx_1 = \int\limits_{-\infty}^{\infty} f_{\xi_1}(u)\, f_{\xi_2}(t-u)\, du.$$

Второе равенство получается либо из первого заменой переменных, либо просто из-за возможности поменять местами ξ_1 и ξ_2. □

Следствие 9 не только предлагает формулу для вычисления плотности распределения суммы, но и утверждает (заметьте!), что сумма двух независимых случайных величин с абсолютно непрерывными распределениями также имеет абсолютно непрерывное распределение.

Упражнение. Для тех, кто уже ничему не удивляется: привести пример двух случайных величин с абсолютно непрерывными распределениями таких, что их сумма имеет вырожденное распределение.

Если даже одна из двух независимых случайных величин имеет дискретное, а вторая — абсолютно непрерывное распределение, то их сумма тоже имеет абсолютно непрерывное распределение:

Упражнение. Пусть ξ имеет таблицу распределения $\mathsf{P}(\xi = a_i) = p_i$, а η имеет плотность распределения $f_\eta(x)$, и эти величины независимы. Доказать, что $\xi + \eta$ имеет плотность распределения $f_{\xi+\eta}(x) = \sum p_i f_\eta(x - a_i)$. Для вычисления функции распределения суммы использовать формулу полной вероятности.

§ 7. Примеры использования формулы свёртки

Пример 36. Пусть независимые случайные величины ξ и η имеют стандартное нормальное распределение. Докажем, что их сумма имеет нормальное распределение с параметрами $a = 0$ и $\sigma^2 = 2$.

Доказательство. По формуле свёртки, плотность суммы равна

$$f_{\xi+\eta}(x) = \int\limits_{-\infty}^{\infty} \frac{1}{2\pi}\, e^{-u^2/2}\, e^{-(x-u)^2/2}\, du = \int\limits_{-\infty}^{\infty} \frac{1}{2\pi}\, e^{-\left(u^2+\frac{x^2}{2}-xu\right)}\, du =$$

$$= e^{-x^2/4} \int\limits_{-\infty}^{\infty} \frac{1}{2\pi}\, e^{-\left(u-\frac{x}{2}\right)^2}\, du = \frac{1}{2\sqrt{\pi}}\, e^{-x^2/4} \int\limits_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}}\, e^{-v^2}\, dv = \frac{e^{-x^2/4}}{2\sqrt{\pi}}.$$

Последний интеграл равен единице, поскольку под интегралом стоит плотность нормального распределения с параметрами $a = 0$ и $\sigma^2 = \dfrac{1}{2}$.

Итак, мы получили, что плотность суммы есть плотность нормального распределения с параметрами 0 и 2. $\square$

Если сумма двух независимых случайных величин из одного и того же распределения (возможно, с разными параметрами) имеет такое же распределение, говорят, что это распределение устойчиво относительно суммирования.

В следующих утверждениях, доказать которые предлагается читателю, перечислены практически все устойчивые распределения.

Лемма 3. *Пусть случайные величины* $\xi \in \Pi_\lambda$ *и* $\eta \in \Pi_\mu$ *независимы. Тогда* $\xi + \eta \in \Pi_{\lambda+\mu}$.

Лемма 4. *Пусть случайные величины* $\xi \in \mathrm{B}_{n,p}$ *и* $\eta \in \mathrm{B}_{m,p}$ *независимы. Тогда* $\xi + \eta \in \mathrm{B}_{n+m,p}$.

Лемма 5. *Пусть случайные величины* $\xi \in \mathrm{N}_{a_1,\sigma_1^2}$ *и* $\eta \in \mathrm{N}_{a_2,\sigma_2^2}$ *независимы. Тогда* $\xi + \eta \in \mathrm{N}_{a_1+a_2,\,\sigma_1^2+\sigma_2^2}$.

Лемма 6. *Пусть случайные величины* $\xi \in \Gamma_{\alpha,\lambda_1}$ *и* $\eta \in \Gamma_{\alpha,\lambda_2}$ *независимы. Тогда* $\xi + \eta \in \Gamma_{\alpha,\lambda_1+\lambda_2}$.

Все эти утверждения мы докажем позднее, используя аппарат характеристических функций, хотя при некотором терпении можно попробовать доказать их и напрямую, как в примере 36.

Показательное распределение не устойчиво по суммированию, однако оно является частным случаем гамма-распределения, которое уже устойчиво относительно суммирования.

Лемма 7. *Пусть независимые случайные величины* $\xi_1, \ldots, \xi_n$ *имеют показательное распределение* $\mathrm{E}_\alpha = \Gamma_{\alpha,1}$. *Тогда* $\xi_1 + \ldots + \xi_n \in \Gamma_{\alpha,n}$.

Доказательство. Докажем утверждение по индукции. При $n = 1$ оно верно в силу равенства $\mathrm{E}_\alpha = \Gamma_{\alpha,\,1}$. Пусть утверждение леммы справедливо для $n = k - 1$. Докажем, что оно верно и для $n = k$. По предположению индукции $S_{k-1} = \xi_1 + \ldots + \xi_{k-1}$ имеет распределение $\Gamma_{\alpha,\,k-1}$, т. е. плотность распределения величины S_{k-1} равна

$$f_{S_{k-1}}(x) = \begin{cases} 0, & \text{если } x \leqslant 0, \\ \dfrac{\alpha^{k-1}}{(k-2)!}\, x^{k-2} e^{-\alpha x}, & \text{если } x > 0. \end{cases}$$

Тогда по формуле свёртки плотность суммы $S_k = \xi_1 + \ldots + \xi_k$ равна

$$f_{S_k}(x) = \int\limits_{-\infty}^{\infty} f_{S_{k-1}}(u) f_{\xi_k}(x-u)\, du = \int\limits_{0}^{\infty} \frac{\alpha^{k-1}}{(k-2)!}\, u^{k-2} e^{-\alpha u} f_{\xi_k}(x-u)\, du.$$

Так как $f_{\xi_k}(x-u) = 0$ при $x-u < 0$, т. е. при $u > x$, то плотность под интегралом отлична от нуля, только если переменная интегрирования изменяется в пределах $0 \leqslant u \leqslant x$ при $x > 0$. При $x \leqslant 0$ подынтегральная функция, а вместе с ней и плотность $f_{S_k}(x)$, равна нулю. При $x > 0$ имеем:

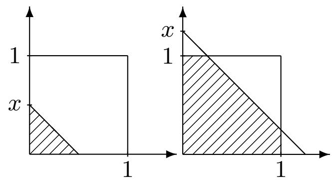
$$f_{S_k}(x) = \int\limits_0^x \frac{\alpha^{k-1}}{(k-2)!}\, u^{k-2} e^{-\alpha u}\, \alpha\, e^{-\alpha(x-u)}\, du = \frac{\alpha^k}{(k-1)!}\, x^{k-1} e^{-\alpha x}.$$

Поэтому $S_k \in \Gamma_{\alpha,\,k}$, что и требовалось доказать. $\square$

Пример 37. Равномерное распределение не является устойчивым относительно суммирования. Найдём функцию и плотность распределения суммы двух независимых случайных величин с одинаковым равномерным на отрезке $[0,\,1]$ распределением, но не по формуле свёртки, а используя геометрическую вероятность.

Пусть $\xi_1,\,\xi_2 \in \mathrm{U}_{0,\,1}$ — независимые случайные величины. Пару $(\xi_1,\,\xi_2)$ можно считать координатой точки, брошенной наудачу в единичный квадрат. Тогда $F_{\xi_1+\xi_2}(x) = \mathsf{P}(\xi_1+\xi_2 < x)$ равна площади области внутри квадрата под прямой $x_2 = x - x_1$. Эта область — заштрихованные треугольник при $0 < x \leqslant 1$, и пятиугольник при $1 < x \leqslant 2$. Окончательно получаем:

$$F_{\xi_1+\xi_2}(x) = \begin{cases} 0, & x \leqslant 0; \\ x^2/2, & 0 < x \leqslant 1; \\ 2x - x^2/2 - 1, & 1 < x \leqslant 2; \\ 1, & x > 2. \end{cases}$$



Плотность распределения суммы равна

$$f_{\xi_1+\xi_2}(x) = \begin{cases} 0, & x \notin (0,\,2); \\ x, & 0 < x \leqslant 1; \\ 2-x, & 1 < x \leqslant 2. \end{cases}$$

Это — плотность так называемого «треугольного» распределения Симпсона. Мы видим, что равномерное распределение не обладает устойчивостью относительно суммирования.

ГЛАВА 9

Числовые характеристики распределений

Если я имею одинаковые шансы на получение a или b, то цена моему ожиданию равна $(a+b)/2$.

Христиан Гюйгенс. О расчётах в азартной игре

§ 1. Математическое ожидание случайной величины

Определение 42. *Математическим ожиданием* $\mathsf{E}\,\xi$ (средним значением, первым моментом) случайной величины ξ с дискретным распределением, задаваемым таблицей $\mathsf{P}(\xi = a_i) = p_i$, где $i \in \mathbb{Z}$, называется *число*

$$\mathsf{E}\,\xi = \sum_i a_i p_i = \sum_i a_i \mathsf{P}(\xi = a_i),$$

если данный ряд абсолютно сходится, т. е. если $\sum |a_i| p_i < \infty$. В противном случае говорят, что математическое ожидание не существует.

Определение 43. *Математическим ожиданием* $\mathsf{E}\,\xi$ (средним значением, первым моментом) случайной величины ξ с абсолютно непрерывным распределением с плотностью распределения $f_\xi(x)$ называется *число*

$$\mathsf{E}\,\xi = \int\limits_{-\infty}^{\infty} x f_\xi(x)\,dx,$$

если этот интеграл абсолютно сходится, т. е. если $\int\limits_{-\infty}^{\infty} |x| f_\xi(x)\,dx < \infty$. В противном случае математическое ожидание не существует.

Математическое ожидание имеет простой физический смысл: если на прямой разместить единичную массу, поместив в точки a_i массу p_i (для дискретного распределения), или «размазав» её с плотностью $f_\xi(x)$ (для абсолютно непрерывного распределения), то точка $\mathsf{E}\,\xi$ будет координатой «центра тяжести» прямой.

Пример 38. Пусть случайная величина ξ равна числу очков, выпадающих при одном подбрасывании кубика. Тогда

$$\mathsf{E}\,\xi = \sum_{k=1}^{6} k \cdot \frac{1}{6} = \frac{1}{6}\,(1+2+3+4+5+6) = 3{,}5.$$

В среднем при одном подбрасывании кубика выпадает 3,5 очка.

Пример 39. Пусть случайная величина ξ — координата точки, брошенной наудачу на отрезок $[a,\, b]$. Тогда

$$\mathsf{E}\,\xi = \int\limits_a^b x \cdot \frac{1}{b-a}\,dx = \frac{x^2}{2(b-a)}\bigg|_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}$$

Центр тяжести равномерного распределения есть середина отрезка.

§ 2. Свойства математического ожидания

Во всех свойствах предполагается, что рассматриваемые математические ожидания существуют.

E1. Для произвольной борелевской функции $g : \mathbb{R} \to \mathbb{R}$

$$\mathsf{E}\,g(\xi) = \begin{cases} \sum\limits_k g(a_k)\mathsf{P}(\xi = a_k), & \text{если распределение } \xi \text{ дискретно;} \\ \int\limits_{-\infty}^{\infty} g(x) f_\xi(x)\,dx, & \text{если распределение } \xi \text{ абсолютно непрерывно.} \end{cases}$$

Доказательство. Мы докажем это свойство (как и почти все дальнейшие) только для дискретного распределения. Пусть $g(\xi)$ принимает значения $c_1,\, c_2,\, \ldots$ с вероятностями

$$\mathsf{P}(g(\xi) = c_m) = \sum_{k\colon g(a_k)=c_m} \mathsf{P}(\xi = a_k)\,.$$

Тогда

$$\mathsf{E}\,g(\xi) = \sum_m c_m \mathsf{P}(g(\xi) = c_m) = \sum_m c_m \sum_{k\colon g(a_k)=c_m} \mathsf{P}(\xi = a_k) =$$
$$= \sum_m \sum_{k\colon g(a_k)=c_m} g(a_k)\,\mathsf{P}(\xi = a_k) = \sum_k g(a_k)\,\mathsf{P}(\xi = a_k)\,. \qquad \square$$

Следствие 10. *Математическое ожидание ξ существует тогда и только тогда, когда* $\mathsf{E}\,|\xi| < \infty$.

Доказательство. Условием существование математического ожидания является абсолютная сходимость ряда или интеграла в определениях 42 и 43. По свойству (E1) это и есть условие $\mathsf{E}\,g(\xi) < \infty$ при $g(x) = |x|$. $\square$

E2. Математическое ожидание постоянной равно ей самой: $\mathsf{E}\,c = c$.

E3. Постоянную можно вынести за знак математического ожидания:

$$\mathsf{E}\,(c\xi) = c\,\mathsf{E}\,\xi.$$

Доказательство. Следует из свойства (E1) при $g(x) = c\,x$. □

E4. Математическое ожидание суммы любых случайных величин равно сумме их математических ожиданий, если только эти математические ожидания существуют:

$$\mathsf{E}\,(\xi + \eta) = \mathsf{E}\,\xi + \mathsf{E}\,\eta.$$

Доказательство. Пусть случайные величины ξ и η имеют дискретные распределения со значениями x_k и y_n соответственно. Для борелевской функции $g : \mathbb{R}^2 \to \mathbb{R}$ можно доказать свойство, аналогичное (E1) (сделать это). Воспользуемся этим свойством для $g(x, y) = x + y$:

$$\begin{aligned}
\mathsf{E}\,(\xi + \eta) &= \sum_{k,\,n}(x_k + y_n)\mathsf{P}(\xi = x_k,\, \eta = y_n) = \\
&= \sum_k x_k \sum_n \mathsf{P}(\xi = x_k,\, \eta = y_n) + \sum_n y_n \sum_k \mathsf{P}(\xi = x_k,\, \eta = y_n) = \\
&= \sum_k x_k\, \mathsf{P}(\xi = x_k) + \sum_n y_n\, \mathsf{P}(\eta = y_n) = \mathsf{E}\,\xi + \mathsf{E}\,\eta. \qquad \square
\end{aligned}$$

E5. Если $\xi \geqslant 0$ п. н., т. е. если $\mathsf{P}(\xi \geqslant 0) = 1$, то $\mathsf{E}\,\xi \geqslant 0$.

Упражнение. Доказать для дискретного и для абсолютно непрерывного распределений.

Замечание 19. Сокращение «п. н.» читается как «почти наверное» и означает «с вероятностью 1». По определению, математическое ожидание — это числовая характеристика распределения. Распределение же не изменится от изменения случайной величины на множестве нулевой вероятности. Поэтому, например, даже если $\xi(\omega) \geqslant 0$ не при всех ω, а на множестве единичной вероятности, математическое ожидание ξ всё равно неотрицательно.

E6. Если $\xi \geqslant 0$ п. н., и при этом $\mathsf{E}\,\xi = 0$, то $\xi = 0$ п. н., т. е. $\mathsf{P}(\xi = 0) = 1$.

Доказательство. Это свойство мы докажем, заранее предполагая, что ξ имеет дискретное распределение с неотрицательными значениями $a_i \geqslant 0$. Равенство $\mathsf{E}\,\xi = \sum a_i p_i = 0$ означает, что все слагаемые в этой сумме равны нулю, т. е. все вероятности p_i нулевые, кроме вероятности, соответствующей значению $a_i = 0$. □

Из свойств (E5) и (E6) вытекает множество полезных утверждений:

Следствие 11. *Если* $\xi \leqslant \eta$ *п. н., то* $\mathsf{E}\,\xi \leqslant \mathsf{E}\,\eta$.

Следствие 12. *Если* $\xi \leqslant \eta$ *п. н., но* $\mathsf{E}\,\xi = \mathsf{E}\,\eta$, *то* $\xi = \eta$ *п. н.*

Следствие 13. *Если* $a \leqslant \xi \leqslant b$ *п. н., то* $a \leqslant \mathsf{E}\,\xi \leqslant b$.

E7. Математическое ожидание произведения независимых случайных величин равно произведению их математических ожиданий: если ξ и η независимы и их математические ожидания существуют, то

$$\mathsf{E}\,(\xi\eta) = \mathsf{E}\,\xi\,\mathsf{E}\,\eta.$$

Доказательство. В дискретном случае:

$$\begin{aligned}\mathsf{E}\,(\xi\eta) &= \sum_{k,\,n} (x_k\, y_n)\, \mathsf{P}(\xi = x_k,\, \eta = y_n) = \\ &= \sum_k x_k\, \mathsf{P}(\xi = x_k) \sum_n y_n\, \mathsf{P}(\eta = y_n) = \mathsf{E}\,\xi\,\mathsf{E}\,\eta. \quad \square\end{aligned}$$

Замечание 20. Обратное утверждение к свойству (E7) неверно: из равенства $\mathsf{E}\,(\xi\eta) = \mathsf{E}\,\xi\,\mathsf{E}\,\eta$ не следует независимость величин ξ и η.

Пример 40. Пусть ξ принимает значения 0 и ± 1 с вероятностями по $1/3$ каждое, и $\eta = \xi^2$. Это зависимые случайные величины:

$$\mathsf{P}(\xi = 1,\, \eta = 0) = \mathsf{P}(\xi = 1,\, \xi^2 = 0) = 0 \neq \frac{1}{3} \cdot \frac{1}{3} = \mathsf{P}(\xi = 1)\, \mathsf{P}(\eta = 0).$$

Однако $\mathsf{E}\,\xi = 0$ и $\mathsf{E}\,(\xi\eta) = \mathsf{E}\,(\xi^3) = 0$, поэтому $\mathsf{E}\,(\xi\eta) = \mathsf{E}\,\xi\,\mathsf{E}\,\eta$.

Пример 41. Пусть $\varphi \in \mathrm{U}_{0,\,2\pi}$, и пусть $\xi = \cos\varphi$ и $\eta = \sin\varphi$ — заведомо зависимые случайные величины (доказать). Но математическое ожидание их произведения равно произведению их математических ожиданий из-за симметричности распределений ξ, η и $\xi\eta$ относительно нуля. Действительно, по свойству (E1) имеем:

$$\mathsf{E}\,\xi = \int_0^{2\pi} \frac{1}{2\pi}\, \cos x\, dx = 0, \quad \mathsf{E}\,\eta = \int_0^{2\pi} \frac{1}{2\pi}\, \sin x\, dx = 0,$$

$$\mathsf{E}\,\xi\eta = \int_0^{2\pi} \frac{1}{2\pi}\, \cos x \sin x\, dx = 0 = \mathsf{E}\,\xi\,\mathsf{E}\,\eta.$$

§ 3. Дисперсия и моменты старших порядков

Определение 44. Пусть $\mathsf{E}\,|\xi|^k < \infty$. Число $\mathsf{E}\,\xi^k$ называется моментом порядка k или k-м моментом случайной величины ξ, число $\mathsf{E}\,|\xi|^k$ называется абсолютным k-м моментом, $\mathsf{E}\,(\xi - \mathsf{E}\,\xi)^k$ называется центральным k-м моментом, и $\mathsf{E}\,|\xi - \mathsf{E}\,\xi|^k$ — абсолютным центральным k-м моментом случайной величины ξ. Число $\mathsf{D}\,\xi = \mathsf{E}\,(\xi - \mathsf{E}\,\xi)^2$ (центральный момент второго порядка) называется дисперсией случайной величины ξ.

Пример 42. Пусть, скажем, случайная величина ξ принимает значение 0 с вероятностью 0,99999, и значение 100 с вероятностью 0,00001. Посмотрим, как моменты разных порядков реагируют на большие, но маловероятные значения случайной величины:

$$\begin{aligned}
\mathsf{E}\,\xi &= 0\cdot 0{,}99999 + 100\cdot 0{,}00001 = 0{,}001,\\
\mathsf{E}\,\xi^2 &= 0^2\cdot 0{,}99999 + 100^2\cdot 0{,}00001 = 0{,}1,\\
\mathsf{E}\,\xi^4 &= 0^4\cdot 0{,}99999 + 100^4\cdot 0{,}00001 = 1\,000,\\
\mathsf{E}\,\xi^6 &= 0^6\cdot 0{,}99999 + 100^6\cdot 0{,}00001 = 10\,000\,000.
\end{aligned}$$

Пример 43. Дисперсия $\mathsf{D}\,\xi = \mathsf{E}\,(\xi - \mathsf{E}\,\xi)^2$ есть «среднее значение квадрата отклонения случайной величины ξ от своего среднего». Посмотрим, за что эта величина отвечает.

Пусть случайная величина ξ принимает значения ± 1 с равными вероятностями, а случайная величина η — значения ± 10 с равными вероятностями. Тогда $\mathsf{E}\,\xi = \mathsf{E}\,\eta = 0$, поэтому $\mathsf{D}\,\xi = \mathsf{E}\,\xi^2 = 1$, $\mathsf{D}\,\eta = \mathsf{E}\,\eta^2 = 100$. Говорят, что дисперсия характеризует степень разброса значений случайной величины вокруг её математического ожидания.

Определение 45. Число $\sigma = \sqrt{\mathsf{D}\,\xi}$ называют среднеквадратическим отклонением случайной величины ξ.

Чтобы прояснить связь моментов разных порядков, докажем несколько неравенств. Во-первых, очевидное утверждение, обеспечивающее существование моментов меньших порядков, если существуют моменты более высокого порядка:

Теорема 29. *Если существует момент порядка $t > 0$ случайной величины ξ, то существуют и её моменты порядка s, $0 < s < t$.*

Доказательство. Для любого числа x верно неравенство:

$$|x|^s \leqslant \max\{\,|x|^t,\ 1\} \leqslant |x|^t + 1.$$

Действительно, $|x|^s \leqslant |x|^t$ при $|x| > 1$, и $|x|^s \leqslant 1$ при $|x| \leqslant 1$. Из этого неравенства следует, что $|\xi(\omega)|^s \leqslant |\xi(\omega)|^t + 1$ для всех ω. Но следствие 11 позволяет из неравенства для случайных величин получить такое же неравенство для их математических ожиданий:

$$\mathsf{E}\,|\xi|^s \leqslant \mathsf{E}\,|\xi|^t + 1.$$

Момент порядка t существует, т. е. $\mathsf{E}\,|\xi|^t < \infty$. Поэтому и $\mathsf{E}\,|\xi|^s < \infty$. □

Докажем ещё одно чрезвычайно полезное неравенство.

Теорема 30 (неравенство Йенсена[16]). *Пусть функция $g : \mathbb{R} \to \mathbb{R}$ выпукла («выпукла вниз», т. е. её надграфик есть выпуклое множество). Тогда для любой случайной величины ξ с конечным первым моментом верно неравенство: $\mathsf{E}\,g(\xi) \geqslant g(\mathsf{E}\,\xi)$.*

Доказательство. Нам понадобится следующее свойство.

[16] Johan Ludwig William Valdemar Jensen (8.05.1859 — 5.03.1925, Denmark)

Лемма 8. *Пусть функция g выпукла. Тогда для всякого y найдётся число $c(y)$ такое, что при всех x*

$$g(x) \geqslant g(y) + c(y)(x - y).$$

Это свойство очевидно и означает, что график выпуклой функции лежит полностью выше любой из касательных к этому графику.

Возьмём в условиях леммы $y = \mathsf{E}\,\xi$, $x = \xi$. Тогда

$$g(\xi) \geqslant g(\mathsf{E}\,\xi) + c(\mathsf{E}\,\xi)(\xi - \mathsf{E}\,\xi).$$

Вычислим математическое ожидание обеих частей неравенства. Так как $\mathsf{E}\,(\xi - \mathsf{E}\,\xi) = 0$, и неравенство между математическими ожиданиями сохраняется по следствию 11, то $\mathsf{E}\,g(\xi) \geqslant g(\mathsf{E}\,\xi)$. □

Следующее неравенство связывает моменты разных порядков.

Следствие 14. *Если $\mathsf{E}\,|\xi|^t < \infty$, то для любого $0 < s < t$*

$$\sqrt[s]{\mathsf{E}\,|\xi|^s} \leqslant \sqrt[t]{\mathsf{E}\,|\xi|^t}$$

Доказательство. Поскольку $0 < s < t$, то $g(x) = |x|^{t/s}$ — выпуклая функция. По неравенству Йенсена для $\eta = |\xi|^s$,

$$(\mathsf{E}\,|\xi|^s)^{t/s} = (\mathsf{E}\,\eta)^{t/s} = g(\mathsf{E}\,\eta) \leqslant \mathsf{E}\,g(\eta) = \mathsf{E}\,|\eta|^{t/s} = \mathsf{E}\,|\xi|^{s\cdot t/s} = \mathsf{E}\,|\xi|^t.$$

Осталось извлечь из обеих частей корень степени t. □

§ 4. Свойства дисперсии

Свойства дисперсии следуют из соответствующих свойств математического ожидания. Заметим, что из существования второго момента следует существование математического ожидания случайной величины и конечность дисперсии. Во всех свойствах ниже предполагается существование вторых моментов случайных величин.

D1. Дисперсия может быть вычислена по формуле: $\mathsf{D}\,\xi = \mathsf{E}\,\xi^2 - (\mathsf{E}\,\xi)^2$.

Доказательство. Положим для удобства $a = \mathsf{E}\,\xi$. Тогда

$$\mathsf{D}\,\xi = \mathsf{E}\,(\xi - a)^2 = \mathsf{E}\,(\xi^2 - 2a\xi + a^2) = \mathsf{E}\,\xi^2 - 2a\mathsf{E}\,\xi + a^2 = \mathsf{E}\,\xi^2 - a^2.$$

D2. При умножении случайной величины на постоянную c дисперсия увеличивается в c^2 раз: $\mathsf{D}\,(c\xi) = c^2\,\mathsf{D}\,\xi$.

Упражнение. Доказать.

D3. Дисперсия всегда неотрицательна: $\mathsf{D}\,\xi \geqslant 0$. Дисперсия обращается в нуль лишь для вырожденного распределения: если $\mathsf{D}\,\xi = 0$, то $\xi = const$ п. н., и наоборот.

Доказательство. Дисперсия есть математическое ожидание почти наверное неотрицательной случайной величины $(\xi - \mathsf{E}\,\xi)^2$, и неотрицательность дисперсии следует из свойства (E5).

По свойству (E6), если $\mathsf{D}\,\xi = 0$, то $(\xi - \mathsf{E}\,\xi)^2 = 0$ п. н., т. е. $\xi = \mathsf{E}\,\xi$ п. н. И наоборот: если $\xi = c$ п. н., то $\mathsf{D}\,\xi = \mathsf{E}\,(c - \mathsf{E}\,c)^2 = 0$. □

D4. Дисперсия не зависит от сдвига случайной величины на постоянную: $\mathsf{D}\,(\xi + c) = \mathsf{D}\,\xi$.

Упражнение. Доказать.

D5. Если ξ и η независимы, то $\mathsf{D}\,(\xi + \eta) = \mathsf{D}\,\xi + \mathsf{D}\,\eta$.

Доказательство. Действительно,

$$\mathsf{D}\,(\xi + \eta) = \mathsf{E}\,(\xi + \eta)^2 - (\mathsf{E}\,(\xi + \eta))^2 =$$
$$= \mathsf{E}\,\xi^2 + \mathsf{E}\,\eta^2 + 2\mathsf{E}\,(\xi\eta) - (\mathsf{E}\,\xi)^2 - (\mathsf{E}\,\eta)^2 - 2\mathsf{E}\,\xi\mathsf{E}\,\eta = \mathsf{D}\,\xi + \mathsf{D}\,\eta,$$

так как математическое ожидание произведения независимых случайных величин равно произведению их математических ожиданий. □

Замечание 21. См. замечание 20.

Следствие 15. *Если ξ и η независимы, то*

$$\mathsf{D}\,(\xi - \eta) = \mathsf{D}\,(\xi + \eta) = \mathsf{D}\,\xi + \mathsf{D}\,\eta.$$

Доказательство. Из свойств (D5) и (D2) получим:

$$\mathsf{D}\,(\xi - \eta) = \mathsf{D}\,(\xi + (-\eta)) = \mathsf{D}\,\xi + \mathsf{D}\,(-\eta) = \mathsf{D}\,\xi + (-1)^2\mathsf{D}\,\eta = \mathsf{D}\,\xi + \mathsf{D}\,\eta. \quad \square$$

Следствие 16. *Для произвольных случайных величин ξ и η с конечными вторыми моментами имеет место равенство:*

$$\mathsf{D}\,(\xi + \eta) = \mathsf{D}\,\xi + \mathsf{D}\,\eta + 2\big(\mathsf{E}\,(\xi\eta) - \mathsf{E}\,\xi\,\mathsf{E}\,\eta\big).$$

D6. Минимум среднеквадратического отклонения случайной величины ξ от точек вещественной прямой есть среднеквадратическое отклонение ξ от своего математического ожидания: $\mathsf{D}\,\xi = \mathsf{E}\,(\xi - \mathsf{E}\,\xi)^2 = \min\limits_a \mathsf{E}\,(\xi - a)^2$.

Доказательство. Сравним величину $\mathsf{E}\,(\xi - a)^2$ с дисперсией:

$$\mathsf{E}\,(\xi - a)^2 = \mathsf{E}\,\big((\xi - \mathsf{E}\,\xi) + (\mathsf{E}\,\xi - a)\big)^2 =$$
$$= \mathsf{D}\,\xi + \big(\mathsf{E}\,\xi - a\big)^2 + 2(\mathsf{E}\,\xi - \mathsf{E}\,\xi)\,\big(\mathsf{E}\,\xi - a\big) = \mathsf{D}\,\xi + \big(\mathsf{E}\,\xi - a\big)^2 \geqslant \mathsf{D}\,\xi,$$

и последнее неравенство превращается в равенство лишь при $a = \mathsf{E}\,\xi$. □

§ 5. Математические ожидания и дисперсии стандартных распределений

Пример 44 (вырожденное распределение I_c). Математическое ожидание и дисперсию этого распределения мы знаем из свойств (E2) и (D3): $\mathsf{E}\,c = c$, $\mathsf{D}\,c = 0$.

Пример 45 (распределение Бернулли B_p). Вычислим два момента и дисперсию: $\mathsf{E}\,\xi = 1 \cdot p + 0 \cdot q = p$; $\mathsf{E}\,\xi^2 = 1^2 \cdot p + 0^2 \cdot q = p$; $\mathsf{D}\,\xi = \mathsf{E}\,\xi^2 - (\mathsf{E}\,\xi)^2 = p - p^2 = pq$.

Пример 46 (биномиальное распределение $\mathrm{B}_{n,\,p}$). Используем свойство устойчивости биномиального распределения относительно суммирования — лемму 4 (стр. 79). Возьмём на каком-нибудь вероятностном пространстве n независимых случайных величин $\xi_1, \ldots, \xi_n$ с распределением Бернулли $\mathrm{B}_p = \mathrm{B}_{1,\,p}$. Тогда их сумма $S_n = \xi_1 + \ldots + \xi_n$ имеет распределение $\mathrm{B}_{n,\,p}$, и по свойству (E4) имеем:

$$\mathsf{E}\,S_n = \sum_{i=1}^{n} \mathsf{E}\,\xi_i = n\mathsf{E}\,\xi_1 = np.$$

А поскольку ξ_i независимы, и дисперсия каждой равна pq, то

$$\mathsf{D}\,S_n = \sum_{i=1}^{n} \mathsf{D}\,\xi_i = n\mathsf{D}\,\xi_1 = npq.$$

Итак, $\mathsf{E}\,\xi = np$, $\mathsf{D}\,\xi = npq$ для $\xi \in \mathrm{B}_{n,\,p}$.

Пример 47 (геометрическое распределение G_p). Вычислим математическое ожидание ξ:

$$\begin{aligned}\mathsf{E}\,\xi &= \sum_{k=1}^{\infty} k\,p\,q^{k-1} = p\sum_{k=1}^{\infty} k\,q^{k-1} = p\sum_{k=1}^{\infty} \frac{dq^k}{dq} = \\ &= p\,\frac{d}{dq}\left(\sum_{k=1}^{\infty} q^k\right) = p\,\frac{d}{dq}\left(\frac{q}{1-q}\right) = p\,\frac{1}{(1-q)^2} = \frac{1}{p}.\end{aligned}$$

Вычислим так называемый «второй факториальный момент» ξ:

$$\begin{aligned}\mathsf{E}\,\xi(\xi-1) &= \sum_{k=1}^{\infty} k(k-1)\,p\,q^{k-1} = p\,q\sum_{k=0}^{\infty} \frac{d^2q^k}{dq^2} = p\,q\,\frac{d^2}{dq^2}\left(\sum_{k=0}^{\infty} q^k\right) = \\ &= p\,q\,\frac{d^2}{dq^2}\left(\frac{1}{1-q}\right) = p\,q\,\frac{2}{(1-q)^3} = \frac{2q}{p^2}.\end{aligned}$$

Найдём дисперсию через второй факториальный момент:

$$\mathsf{D}\,\xi = \mathsf{E}\,\xi(\xi-1) + \mathsf{E}\,\xi - (\mathsf{E}\,\xi)^2 = \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{2q-1+p}{p^2} = \frac{q}{p^2}.$$

Пример 48 (распределение Пуассона Π_λ). Вычислим математическое ожидание ξ:

$$\begin{aligned}\mathsf{E}\,\xi &= \sum_{k=0}^{\infty} k\,\frac{\lambda^k}{k!}\,e^{-\lambda} = e^{-\lambda}\sum_{k=1}^{\infty} k\,\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} = \\ &= \lambda e^{-\lambda}\sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda}\sum_{m=0}^{\infty} \frac{\lambda^m}{m!} = \lambda e^{-\lambda}e^{\lambda} = \lambda.\end{aligned}$$

Моменты более высоких порядков легко находятся через факториальные моменты $\mathsf{E}\,\xi^{[m]} = \mathsf{E}\,\xi(\xi-1)\ldots(\xi-m+1)$ порядка m. Так, второй факториальный момент ξ равен

$$\mathsf{E}\,\xi(\xi-1) = \sum_{k=0}^{\infty} k(k-1)\,\frac{\lambda^k}{k!}\,e^{-\lambda} = \lambda^2 e^{-\lambda}\sum_{k=2}^{\infty}\frac{\lambda^{k-2}}{(k-2)!} = \lambda^2 e^{-\lambda}e^{\lambda} = \lambda^2.$$

Поэтому $\mathsf{E}\,\xi^2 = \mathsf{E}\,\xi(\xi-1) + \mathsf{E}\,\xi = \lambda^2+\lambda$ и $\mathsf{D}\,\xi = \mathsf{E}\,\xi^2 - (\mathsf{E}\,\xi)^2 = \lambda$.

Пример 49 (равномерное распределение $\mathrm{U}_{a,b}$). Вычислим первые два момента:

$$\mathsf{E}\,\xi = \int\limits_{-\infty}^{\infty} x f_\xi(x)\,dx = \int\limits_a^b x\,\frac{1}{b-a}\,dx = \frac{a+b}{2};$$

$$\mathsf{E}\,\xi^2 = \int\limits_{-\infty}^{\infty} x^2 f_\xi(x)\,dx = \int\limits_a^b x^2\,\frac{1}{b-a}\,dx = \frac{b^3-a^3}{3(b-a)} = \frac{a^2+ab+b^2}{3}.$$

Дисперсия равна $\mathsf{D}\,\xi = \mathsf{E}\,\xi^2 - (\mathsf{E}\,\xi)^2 = (b-a)^2/\,12$.

Пример 50 (стандартное нормальное распределение $\mathrm{N}_{0,1}$). Математическое ожидание этого распределения существует в силу конечности $\mathsf{E}\,|\xi|$:

$$\mathsf{E}\,|\xi| = \frac{2}{\sqrt{2\pi}}\int\limits_0^{\infty} xe^{-x^2/2}\,dx = \frac{2}{\sqrt{2\pi}}\int\limits_0^{\infty} e^{-x^2/2}\,d(x^2/2) = \frac{2}{\sqrt{2\pi}} < \infty.$$

Математическое ожидание ξ равно

$$\mathsf{E}\,\xi = \int\limits_{-\infty}^{\infty} x f_\xi(x)\,dx = \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty} x\,e^{-x^2/2}\,dx = 0,$$

так как под сходящимся интегралом стоит нечётная функция. Далее,

$$\begin{aligned}\mathsf{E}\,\xi^2 &= \frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty} x^2\,e^{-x^2/2}\,dx = \frac{2}{\sqrt{2\pi}}\int\limits_0^{\infty} x^2\,e^{-x^2/2}\,dx = -\frac{2}{\sqrt{2\pi}}\int\limits_0^{\infty} x\,de^{-x^2/2} = \\ &= -\frac{2x}{\sqrt{2\pi}}\,e^{-x^2/2}\Big|_0^{\infty} + 2\int\limits_0^{\infty}\frac{1}{\sqrt{2\pi}}\,e^{-x^2/2}\,dx = 0 + \int\limits_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\,e^{-x^2/2}\,dx = 1.\end{aligned}$$

Поэтому $\mathsf{D}\,\xi = \mathsf{E}\,\xi^2 - (\mathsf{E}\,\xi)^2 = 1 - 0 = 1$.

Пример 51 (нормальное распределение N_{a,σ^2}). Если $\xi \in \mathrm{N}_{a,\sigma^2}$, то $\eta = (\xi-a)\,/\,\sigma \in \mathrm{N}_{0,1}$. Мы только что вычислили $\mathsf{E}\eta = 0$, $\mathsf{D}\eta = 1$.

Тогда (над каждым равенством подписать, каким свойствам оно обязано)

$$\mathsf{E}\,\xi = \mathsf{E}\,(\sigma\eta + a) = \sigma\mathsf{E}\,\eta + a = a; \quad \mathsf{D}\,\xi = \mathsf{D}\,(\sigma\eta + a) = \sigma^2\mathsf{D}\,\eta = \sigma^2.$$

Пример 52 (показательное распределение E_α). Найдём для произвольного $k \in \mathbb{N}$ момент порядка k.

$$\mathsf{E}\,\xi^k = \int\limits_{-\infty}^{\infty} x^k f_\xi(x)\,dx = \int\limits_0^{\infty} x^k\,\alpha\,e^{-\alpha x}\,dx = \frac{1}{\alpha^k}\int\limits_0^{\infty} (\alpha x)^k\,e^{-\alpha x}\,d(\alpha x) = \frac{k!}{\alpha^k}.$$

В последнем равенстве мы воспользовались гамма-функцией Эйлера:

$$\Gamma(k+1) = \int\limits_0^{\infty} u^k\,e^{-u}\,du = k!$$

Тогда

$$\mathsf{E}\,\xi = \frac{1}{\alpha}, \quad \mathsf{E}\,\xi^2 = \frac{2}{\alpha^2}, \quad \mathsf{D}\,\xi = \mathsf{E}\,\xi^2 - (\mathsf{E}\,\xi)^2 = \frac{1}{\alpha^2}.$$

Пример 53 (стандартное распределение Коши $\mathrm{C}_{0,\,1}$). Математическое ожидание распределения Коши не существует, так как расходится интеграл

$$\mathsf{E}\,|\xi| = \int\limits_{-\infty}^{\infty} |x|\,\frac{1}{\pi(1+x^2)}\,dx = \int\limits_0^{\infty} \frac{1}{\pi(1+x^2)}\,dx^2 = \lim_{x\to+\infty}\frac{1}{\pi}\ln(1+x^2) = +\infty.$$

Расходится он из-за того, что подынтегральная функция ведёт себя на бесконечности как $1/x$. Поэтому не существуют ни дисперсия, ни моменты более высоких порядков этого распределения. То же самое можно сказать про распределение Коши $\mathrm{C}_{a,\,\sigma}$.

Пример 54 (распределение Парето). У распределения Парето существуют только моменты порядка $t < \alpha$, поскольку

$$\mathsf{E}\,|\xi|^t = \int\limits_1^{\infty} x^t\,\alpha\,\frac{1}{x^{\alpha+1}}\,dx = \int\limits_1^{\infty} \alpha\,\frac{1}{x^{\alpha-t+1}}\,dx$$

сходится при $t < \alpha$, когда подынтегральная функция на бесконечности ведёт себя как $1\,/\,x^{s+1}$, где $s = \alpha - t > 0$.

Упражнение. Посчитать момент порядка $t < \alpha$ распределения Парето. При каких α у этого распределения существует дисперсия? А две тысячи триста семнадцатый момент?

ГЛАВА 10

Числовые характеристики зависимости

> Кажется, нельзя сомневаться ни в истине того, что всё в мире может быть представлено числами; ни в справедливости того, что всякая в нём перемена и отношение выражается аналитической функцией. Между тем обширный взгляд теории допускает существование зависимости только в том смысле, чтобы числа, одни с другими в связи, принимать как бы данными вместе.
>
> *Н. И. Лобачевский, Об исчезании тригонометрических строк*

§ 1. Ковариация двух случайных величин

Мы знаем, что для независимых случайных величин с конечными вторыми моментами дисперсия их суммы равна сумме их дисперсий. В общем случае дисперсия суммы равна

$$\mathsf{D}\,(\xi+\eta) = \mathsf{D}\,\xi + \mathsf{D}\,\eta + 2\big(\mathsf{E}\,(\xi\eta) - \mathsf{E}\,\xi\,\mathsf{E}\,\eta\big). \tag{19}$$

Величина $\mathsf{E}\,(\xi\eta) - \mathsf{E}\,\xi\,\mathsf{E}\,\eta$ равняется нулю, если случайные величины ξ и η независимы (свойство (E7) математического ожидания). С другой стороны, из равенства её нулю вовсе не следует независимость, как показывают примеры 40 и 41. Эту величину часто используют как «индикатор наличия зависимости» между двумя случайными величинами.

Определение 46. *Ковариацией* $\operatorname{cov}(\xi, \eta)$ случайных величин ξ и η называется *число*

$$\operatorname{cov}(\xi, \eta) = \mathsf{E}\,\big((\xi - \mathsf{E}\,\xi)(\eta - \mathsf{E}\,\eta)\big).$$

Свойство 13. *Справедливы равенства:* $\operatorname{cov}(\xi, \eta) = \mathsf{E}\,(\xi\eta) - \mathsf{E}\,\xi\,\mathsf{E}\,\eta$; $\operatorname{cov}(\xi, \xi) = \mathsf{D}\,\xi$; $\operatorname{cov}(\xi, \eta) = \operatorname{cov}(\eta, \xi)$; $\operatorname{cov}(c\,\xi, \eta) = c\operatorname{cov}(\xi, \eta)$.

Упражнение. Доказать свойство 13.

Упражнение. Доказать следующее свойство 14, пользуясь равенствами

$$(a+b)^2 = a^2 + b^2 + ab + ba = a^2 + b^2 + 2ab = aa + bb + ab + ba$$

и получив аналогичные равенства для квадрата суммы n слагаемых.

Свойство 14. *Дисперсия суммы нескольких случайных величин вычисляется по любой из следующих формул:*

$$\begin{aligned}\mathsf{D}\,(\xi_1+\ldots+\xi_n) &= \sum_{i=1}^{n}\mathsf{D}\,\xi_i+\sum_{i\neq j}\operatorname{cov}(\xi_i,\xi_j)= \\ &= \sum_{i=1}^{n}\mathsf{D}\,\xi_i+2\sum_{i<j}\operatorname{cov}(\xi_i,\xi_j)=\sum_{i,j}\operatorname{cov}(\xi_i,\xi_j).\end{aligned}$$

Обсудим достоинства и недостатки ковариации, как величины, характеризующей зависимость двух случайных величин.

1. Если ковариация $\operatorname{cov}(\xi,\,\eta)$ отлична от нуля, то величины ξ и η зависимы. Чтобы судить о наличии зависимости согласно любому из определений независимости, требуется знать совместное распределение пары ξ и η. Но найти совместное распределение часто бывает сложнее, чем посчитать математическое ожидание произведения ξ и η. Если нам повезёт, и математическое ожидание произведения ξ и η не будет равняться произведению их математических ожиданий, мы скажем, что ξ и η зависимы, не находя их совместного распределения. Это очень хорошо.

Пример 55. Покажем, что с помощью ковариации можно судить о зависимости даже тогда, когда для вычисления совместного распределения недостаточно данных. Пусть ξ и η — независимые случайные величины, и дисперсия ξ отлична от нуля (что это значит?). Покажем, что ξ и $\xi+\eta$ зависимы:

$$\mathsf{E}\,\bigl(\xi(\xi+\eta)\bigr)=\mathsf{E}\,\xi^2+\mathsf{E}\,\xi\,\mathsf{E}\,\eta,\qquad \mathsf{E}\,\xi\,\mathsf{E}\,(\xi+\eta)=(\mathsf{E}\,\xi)^2+\mathsf{E}\,\xi\,\mathsf{E}\,\eta,$$

поэтому $\operatorname{cov}(\xi,\xi+\eta)=\mathsf{E}\,\xi^2+\mathsf{E}\,\xi\,\mathsf{E}\,\eta-\bigl((\mathsf{E}\,\xi)^2+\mathsf{E}\,\xi\,\mathsf{E}\,\eta\bigr)=\mathsf{D}\,\xi>0$. Следовательно, ξ и $\xi+\eta$ зависимы.

Упражнение. Доказать, что величины ξ и $\xi+\eta$ независимы, если $\mathsf{D}\,\xi=0$.

2. Величина $\operatorname{cov}(\xi,\,\eta)$ не является «безразмерной»: если ξ — объем газа в сосуде, а η — давление этого газа, то ковариация измеряется в м$^3\times$ Па. Иначе говоря, при умножении ξ или η на какое-нибудь число ковариация тоже умножается на это число. Но умножение на число не сказывается на «степени зависимости» величин (они от этого «более зависимыми» не становятся), так что большое значение ковариации не означает более сильной зависимости. Это очень плохо.

Нужно как-то нормировать ковариацию, получив из неё «безразмерную» величину, абсолютное значение которой:

а) не менялось бы при умножении случайных величин на число;

б) свидетельствовало бы о «силе зависимости» случайных величин.

Замечание 22. Говоря о «силе» зависимости между случайными величинами, мы имеем в виду следующее. Самая сильная зависимость — функциональная, а из функциональных — линейная зависимость, когда $\xi = a\eta + b$ п. н. Бывают гораздо более слабые зависимости. Так, если по последовательности независимых случайных величин $\xi_1, \xi_2, \ldots$ построить величины $\xi = \xi_1 + \ldots + \xi_{24} + \xi_{25}$ и $\eta = \xi_{25} + \xi_{26} + \ldots + \xi_{90}$, то эти величины зависимы, но очень «слабо»: через единственное общее слагаемое ξ_{25}. Сильно ли зависимы число гербов в первых 25 подбрасываниях монеты и число гербов в испытаниях с 25-го по 90-е?

Итак, следующая величина есть всего лишь ковариация, нормированная нужным образом.

§ 2. Коэффициент корреляции

Определение 47. Коэффициентом корреляции $\rho(\xi,\eta)$ случайных величин ξ и η, дисперсии которых существуют и отличны от нуля, называется число

$$\rho(\xi,\eta) = \frac{\operatorname{cov}(\xi,\,\eta)}{\sqrt{\mathsf{D}\,\xi}\sqrt{\mathsf{D}\,\eta}}.$$

Замечание 23. Чтобы разглядеть «устройство» коэффициента корреляции, распишем по определению числитель и знаменатель:

$$\rho(\xi,\eta) = \frac{\mathsf{E}\left((\xi-\mathsf{E}\,\xi)(\eta-\mathsf{E}\,\eta)\right)}{\sqrt{\mathsf{E}\,(\xi-\mathsf{E}\,\xi)^2}\sqrt{\mathsf{E}\,(\eta-\mathsf{E}\,\eta)^2}}.$$

Здесь математикам уместно провести аналогии с «косинусом угла» между двумя элементами $\xi - \mathsf{E}\,\xi$ и $\eta - \mathsf{E}\,\eta$ гильбертова пространства, образованного случайными величинами с нулевым математическим ожиданием и конечным вторым моментом, снабженного скалярным произведением $\operatorname{cov}(\xi,\,\eta)$ и «нормой», равной корню из дисперсии, или корню из скалярного произведения $\operatorname{cov}(\xi,\xi)$.

Пример 56. Рассмотрим продолжение примера 55, но пусть ξ и η будут не только независимыми, но и одинаково распределёнными случайными величинами, и их дисперсия отлична от нуля. Найдём коэффициент корреляции величин ξ и $\xi+\eta$:

$$\rho(\xi,\xi+\eta) = \frac{\operatorname{cov}(\xi,\xi+\eta)}{\sqrt{\mathsf{D}\,\xi}\sqrt{\mathsf{D}\,(\xi+\eta)}} = \frac{\mathsf{D}\,\xi}{\sqrt{\mathsf{D}\,\xi}\sqrt{\mathsf{D}\,\xi+\mathsf{D}\,\eta}} = \frac{\mathsf{D}\,\xi}{\sqrt{\mathsf{D}\,\xi}\sqrt{2\mathsf{D}\,\xi}} = \frac{1}{\sqrt{2}}.$$

Коэффициент корреляции величин ξ и $\xi+\eta$ равен косинусу угла 45°, образованного «векторами» ξ и $\xi+\eta$, когда «$\xi \perp \eta$» и их «длина» одинакова.

Упражнение. Чтобы аналогия не заходила слишком далеко, и у читателя не возникло искушения любые случайные величины рисовать стрелочками на плоскости и вместо подсчёта математических ожиданий измерять углы, предлагаю убедиться, например, что коэффициент корреляции величин ξ и ξ^2 равен:

а) нулю, если ξ имеет нормальное распределение с нулевым средним;

б) $2/\sqrt{5}$, если ξ имеет показательное распределение с любым параметром.

§ 3. Свойства коэффициента корреляции

Предполагается, что коэффициент корреляции существует.

Теорема 31. *Коэффициент корреляции обладает свойствами:*

1) *если* ξ *и* η *независимы, то* $\rho(\xi,\eta) = 0$*;*

2) *всегда* $|\rho(\xi,\eta)| \leqslant 1$*;*

3) $|\rho(\xi,\eta)| = 1$ *тогда и только тогда, когда* ξ *и* η *п. н. линейно связаны, т. е. существуют числа* $a \neq 0$ *и* b *такие, что* $\mathsf{P}(\eta = a\xi + b) = 1$.

Доказательство.

1) Свойство (1) мы уже много раз (сколько?) упоминали и один раз доказали. Более того, при рассмотрении свойств математического ожидания мы привели примеры 40 и 41 — два из многих возможных примеров того, что свойство (1) в обратную сторону неверно.

2) Обозначим через σ_ξ^2 и σ_η^2 дисперсии ξ и η соответственно, и рассмотрим неотрицательную (почему?) дисперсию любой из двух случайных величин $\varphi^{\pm} = \sigma_\eta\xi \pm \sigma_\xi\eta$:

$$\begin{aligned} 0 \leqslant \mathsf{D}\,\varphi^{\pm} &= \mathsf{D}\,(\sigma_\eta\xi) + \mathsf{D}\,(\pm\sigma_\xi\eta) + 2\mathrm{cov}(\sigma_\eta\xi,\, \pm\sigma_\xi\eta) = \\ &= \sigma_\eta^2\sigma_\xi^2 + \sigma_\xi^2\sigma_\eta^2 \pm 2\sigma_\eta\sigma_\xi\mathrm{cov}(\xi,\, \eta) = 2\sigma_\xi^2\sigma_\eta^2(1 \pm \rho(\xi,\, \eta)). \end{aligned}$$

Мы получили два полезных соотношения:

$$1 + \rho(\xi,\, \eta) = \frac{\mathsf{D}\,\varphi^+}{2\sigma_\xi^2\sigma_\eta^2} \geqslant 0, \qquad 1 - \rho(\xi,\, \eta) = \frac{\mathsf{D}\,\varphi^-}{2\sigma_\xi^2\sigma_\eta^2} \geqslant 0. \tag{20}$$

Из них сразу следует, что $-1 \leqslant \rho(\xi,\, \eta) \leqslant 1$.

3) В одну сторону утверждение проверяется непосредственно:

Упражнение. Воспользоваться свойствами математического ожидания и дисперсии и доказать, что

$$\rho(\xi,\, a\xi + b) = \begin{cases} 1, & a > 0; \\ -1, & a < 0. \end{cases}$$

Не забудьте, что $\sqrt{a^2} = |a|$, а не просто a!

Докажем вторую часть свойства (3): если $|\rho(\xi,\, \eta)| = 1$, то существуют числа $a \neq 0$ и b такие, что $\mathsf{P}(\eta = a\xi + b) = 1$.

Рассмотрим сначала случай $\rho(\xi,\, \eta) = 1$. Это возможно только если второе неравенство в формуле (20) превращается в равенство:

$$0 = 1 - \rho(\xi,\, \eta) = \frac{\mathsf{D}\,\varphi^-}{2\sigma_\xi^2\sigma_\eta^2},$$

т. е. $\mathsf{D}\,\varphi^- = 0$. Тогда, по свойству (D3), $\varphi^- = c$ п. н., где c — некоторое число. Иначе говоря, $\sigma_\eta\xi - \sigma_\xi\eta = c$ п. н., или

$$\eta = \frac{\sigma_\eta}{\sigma_\xi}\,\xi - \frac{c}{\sigma_\xi} = a\xi + b \quad \text{п. н.}$$

В случае $\rho(\xi,\eta)=-1$ нужно рассмотреть первое неравенство в формуле (20) и повторить рассуждения. Тем самым теорема 31 доказана. □

Полезно знать следующие часто употребляемые термины.

Определение 48. Говорят, что ξ и η отрицательно коррелированы, если $\rho(\xi,\eta)<0$, положительно коррелированы, если $\rho(\xi,\eta)>0$, и некоррелированы, если $\rho(\xi,\eta)=0$.

Смысл знака $\rho(\xi,\eta)$ хорошо виден в случае $\rho(\xi,\eta)=\pm 1$. Тогда знак ρ равен знаку a в равенстве $\eta=a\xi+b$ п. н. Так, $\rho(\xi,\eta)=1$ означает, что чем больше ξ, тем больше и η. Напротив, $\rho(\xi,\eta)=-1$ означает, что чем больше ξ, тем меньше η. Похожим образом можно трактовать знак коэффициента корреляции и в случае, когда $|\rho(\xi,\eta)|<1$, помня при этом, что зависимость между ξ и η теперь уже не линейная и, возможно, даже не функциональная.

Так, величины ξ и $\xi+\eta$ в примерах 55 и 56 положительно коррелированы, но их зависимость не функциональная.

Следующее свойство показывает, что модуль коэффициента корреляции не меняется при линейных преобразованиях случайных величин.

Свойство 15. *Для любых случайных величин ξ и η с конечной и ненулевой дисперсией при любых постоянных $a\neq 0$ и b имеет место равенство:*

$$\rho(a\xi+b,\eta)=\operatorname{sgn}(a)\cdot\rho(\xi,\eta),\quad \textit{где}\quad \operatorname{sgn}(a)=\begin{cases}1, & a>0;\\ -1, & a<0.\end{cases}$$

Доказательство. Запишем $\rho(a\xi+b,\eta)$, не забывая про свойства дисперсии:

$$\rho(a\xi+b,\eta)=\frac{\operatorname{cov}(a\xi+b,\eta)}{\sqrt{\mathsf{D}\,(a\xi+b)}\,\sqrt{\mathsf{D}\,\eta}}=\frac{a\operatorname{cov}(\xi,\eta)}{\sqrt{a^2\mathsf{D}\,\xi}\,\sqrt{\mathsf{D}\,\eta}}=\frac{a}{|a|}\cdot\rho(\xi,\eta).$$

Осталось заметить, что знак a как раз и равен $\operatorname{sgn}(a)=a/|a|$. □

§ 4. Примеры

Пример 57. Если ξ и η суть координаты точки, брошенной наудачу в треугольник D с вершинами $(2,0)$, $(0,0)$ и $(0,1)$, то их коэффициент корреляции $\rho(\xi,\eta)$ отрицателен. Это можно объяснить так: чем больше ξ, тем меньше у η возможностей быть большой.

Предлагаю убедиться в этом, проверив справедливость следующих высказываний. Во-первых,

$$f_\xi(x)=\begin{cases}1-\dfrac{x}{2}, & 0\leqslant x\leqslant 2;\\ 0, & \text{иначе};\end{cases}\qquad f_\eta(y)=\begin{cases}2-2y, & 0\leqslant y\leqslant 1;\\ 0, & \text{иначе},\end{cases}$$

и вычисленные по этим плотностям средние (вычислить) равны соответственно $\mathsf{E}\,\xi = 2/3$ и $\mathsf{E}\,\eta = 1/3$.

Во-вторых, по определению многомерного равномерного распределения в области D,

$$\mathsf{E}\,(\xi\eta) = \iint\limits_D x \cdot y \cdot 1\,dx\,dy = \int\limits_0^2 \int\limits_0^{1-x/2} x\,y\,dy\;dx = \frac{1}{6}\ \textit{(кажется)}.$$

Т. е. ковариация (а с ней и коэффициент корреляции) отрицательна.

Упражнение. А почему коэффициент корреляции в примере 57 существует? Какие свойства случайных величин гарантируют конечность второго момента? А из их ограниченности следует существование каких-нибудь моментов?

Пример 58. Найдём коэффициент корреляции между числом выпадений единицы и числом выпадений шестерки при n подбрасываниях правильной игральной кости.

Обозначим для $i \in \{1, \ldots, 6\}$ через ξ_i случайную величину, равную числу выпадений грани с i очками при n подбрасываниях кубика. Посчитаем $\mathrm{cov}(\xi_1, \xi_6)$. Каждая из случайных величин ξ_i имеет биномиальное распределение с параметрами n и $1/6$, поэтому $\mathsf{E}\,\xi_i = n/6$, $\mathsf{D}\,\xi_i = 5n/36$.

Далее заметим, что $\xi_1 + \ldots + \xi_6 = n$. Из-за симметрии кубика математические ожидания $\mathsf{E}\,\xi_1\xi_2$, $\mathsf{E}\,\xi_1\xi_3$, $\ldots$, $\mathsf{E}\,\xi_1\xi_6$ одинаковы (но, надо думать, отличаются от $\mathsf{E}\,\xi_1\xi_1 = \mathsf{E}\,\xi_1^2 = \mathsf{D}\,\xi_1 + (\mathsf{E}\,\xi_1)^2 = 5n/36 + n^2/36$).

Посчитаем $\mathsf{E}\,\xi_1(\xi_1 + \cdots + \xi_6)$. С одной стороны, это равно

$$\mathsf{E}\,\xi_1(\xi_1 + \ldots + \xi_6) = \mathsf{E}\,\xi_1 \cdot n = n^2/6,$$

с другой стороны,

$$\mathsf{E}\,\xi_1(\xi_1 + \ldots + \xi_6) = \mathsf{E}\,\xi_1^2 + 5\mathsf{E}\,\xi_1\xi_6 = 5n/36 + n^2/36 + 5\mathsf{E}\,\xi_1\xi_6.$$

Отсюда $5\mathsf{E}\,\xi_1\xi_6 = n^2/6 - 5n/36 - n^2/36$, т. е. $\mathsf{E}\,\xi_1\xi_6 = (n^2 - n)/36$.

Следовательно, искомый коэффициент корреляции равен

$$\rho(\xi_1, \xi_6) = \frac{\mathsf{E}\,\xi_1\xi_6 - \mathsf{E}\,\xi_1\mathsf{E}\,\xi_6}{\sqrt{\mathsf{D}\,\xi_1\mathsf{D}\,\xi_6}} = \frac{(n^2 - n)/36 - n^2/36}{5n/36} = -\frac{1}{5}.$$

Интересно, что полученный коэффициент корреляции не зависит от n.

Упражнение. Объяснить, почему коэффициент корреляции $\rho(\xi_1, \xi_6)$ отрицателен. Найти коэффициенты корреляции $\rho(\xi_1, \xi_2)$ и $\rho(\xi_1, \xi_1)$.

Пример 59. Вычислим математическое ожидание и дисперсию гипергеометрического распределения. Мы не могли сделать это раньше, так как очень не хотели вычислять следующие суммы:

$$\mathsf{E}\,\xi = \sum_k k\,\frac{C_K^k C_{N-K}^{n-k}}{C_N^n}, \qquad \mathsf{E}\,\xi^2 = \sum_k k^2\,\frac{C_K^k C_{N-K}^{n-k}}{C_N^n},$$

где, напомним (чтобы читатель окончательно отказался от мысли вычислить эти суммы напрямую), суммирование ведётся по целым k таким, что $0 \leqslant k \leqslant K$ и $0 \leqslant n-k \leqslant N-K$.

Рассмотрим урну, содержащую K белых шаров и $N-K$ не белых, и пусть из неё наудачу и без возвращения выбирают по одному n шаров. Свяжем случайную величину ξ, равную числу белых шаров среди n выбранных, с результатами отдельных извлечений шаров.

Обозначим через ξ_i, где $i = 1, \ldots, n$, «индикатор» того, что i-й по счёту вынутый шар оказался белым: $\xi_i = 1$, если при i-м извлечении появился белый шар, иначе $\xi_i = 0$. Тогда $\xi = \xi_1 + \ldots + \xi_n$ — число появившихся белых шаров, и математическое ожидание считается просто:

$$\mathsf{E}\,\xi = \mathsf{E}\,(\xi_1 + \ldots + \xi_n) = \mathsf{E}\,\xi_1 + \ldots + \mathsf{E}\,\xi_n.$$

Убедимся, что случайные величины $\xi_1, \ldots, \xi_n$ имеют одно и то же распределение Бернулли B_p, где $p = K\,/\,N$.

Пронумеруем шары: белые — номерами от одного до K, остальные — номерами от $K+1$ до N. Элементарным исходом опыта является набор из n номеров шаров в схеме выбора n элементов из N без возвращения и с учётом порядка. Общее число исходов равно $|\Omega| = A_N^n$ по теореме 2.

Вычислим вероятность события $A_i = \{\xi_i = 1\}$. Событие A_i включает в себя элементарные исходы (наборы), в которых на i-м месте стоит любой из номеров белых шаров, а остальные $n-1$ место занимают любые из оставшихся $N-1$ номеров. По теореме 1 о перемножении шансов число благоприятных событию A_i исходов есть произведение K и A_{N-1}^{n-1}. Здесь K есть число способов поставить на i-е место один из номеров белых шаров, A_{N-1}^{n-1} — число способов после этого разместить на оставшихся $n-1$ местах остальные $N-1$ номеров шаров. Но тогда

$$p = \mathsf{P}(\xi_i = 1) = \mathsf{P}(A_i) = \frac{|A_i|}{|\,\Omega\,|} = \frac{K\,A_{N-1}^{n-1}}{A_N^n} = \frac{K}{N}\,,$$

что совершенно очевидно: вероятность двадцатому шару быть белым, если мы ничего не знаем про первые девятнадцать, точно такая же, как вероятность первому шару быть белым и равна отношению числа белых шаров к числу всех.

Вернёмся к математическому ожиданию:

$$\mathsf{E}\,\xi = \mathsf{E}\,\xi_1 + \ldots + \mathsf{E}\,\xi_n = n\mathsf{E}\,\xi_1 = np = \frac{nK}{N}\,.$$

Вычислим дисперсию ξ. До сих пор мы не интересовались совместным распределением $\xi_1, \ldots, \xi_n$: для вычисления математического ожидания их суммы нам было достаточно знания маргинальных распределений этих величин. Но дисперсия суммы уже не всегда равна сумме дисперсий. Зависимость величин $\xi_1, \ldots, \xi_n$ очевидна: если, скажем, случилось событие $A_1 = \{\xi_1 = 1\}$, то вероятность второму шару быть белым уже не равна $K\,/\,N$:

$$\mathsf{P}(\xi_2 = 1 \mid \xi_1 = 1) = \frac{K-1}{N-1} \neq \frac{K}{N} = \mathsf{P}(\xi_2 = 1).$$

Поэтому при вычислении дисперсии будем пользоваться свойством 14. Вычислим ковариацию величин ξ_i и ξ_j, $i \neq j$. Для этого сначала посчитаем $\mathsf{E}\,(\xi_i\xi_j)$. Произведение $\xi_i\xi_j$ снова имеет распределение Бернулли: $\xi_i\xi_j = 1$, если при i-м и j-м извлечениях появились белые шары. Вероятность этого события равна

$$\mathsf{P}(\xi_i\xi_j = 1) = \mathsf{P}(A_i \cap A_j) = \frac{|A_i \cap A_j|}{|\,\Omega\,|} = \frac{K(K-1)A_{N-2}^{n-2}}{A_N^n} = \frac{K(K-1)}{N(N-1)}.$$

Тогда

$$\mathrm{cov}(\xi_i,\, \xi_j) = \mathsf{E}\,(\xi_i\xi_j) - \mathsf{E}\,\xi_i \mathsf{E}\,\xi_j = \frac{K(K-1)}{N(N-1)} - \frac{K}{N}\,\frac{K}{N} = -\frac{K(N-K)}{N^2(N-1)}.$$

Подставляя одинаковые дисперсии $\mathsf{D}\,\xi_i = p(1-p)$ и эти не зависящие от i и j ковариации в формулу дисперсии суммы, получим:

$$\begin{aligned}
\mathsf{D}\,\xi &= \mathsf{D}\,(\xi_1 + \ldots + \xi_n) = \sum_{i=1}^{n} \mathsf{D}\,\xi_i + \sum_{i\neq j} \mathrm{cov}(\xi_i, \xi_j) = \\
&= np(1-p) + n(n-1)\mathrm{cov}(\xi_1,\, \xi_2) = \\
&= n\,\frac{K}{N}\Big(1 - \frac{K}{N}\Big) - n(n-1)\frac{K(N-K)}{N^2(N-1)} = n\,\frac{K}{N}\Big(1 - \frac{K}{N}\Big)\Big(1 - \frac{n-1}{N-1}\Big).
\end{aligned}$$

Заметим любопытнейшую вещь: если вынимать шары с в о з в р а щ е н и е м, то испытания станут независимыми испытаниями в схеме Бернулли, а ставшие независимыми величины ξ_i в сумме дадут число белых шаров, имеющее биномиальное распределение с параметрами n и $p = K\,/\,N$ и точно такое же математическое ожидание $np = nK\,/\,N$, как и у числа белых шаров при выборе б е з в о з в р а щ е н и я.

Дисперсия же у числа белых шаров при выборе без возвращения меньше, чем при выборе с возвращением — за счёт отрицательной коррелированности слагаемых ξ_i и ξ_j при $i \neq j$.

ГЛАВА 11

Куда и как сходятся последовательности случайных величин

Откуда, наконец, вытекает то удивительное, по-видимому, следствие, что, если бы наблюдения над всеми событиями продолжать всю вечность, причём вероятность, наконец, перешла бы в полную достоверность, то было бы замечено, что в мире всё управляется точными отношениями и постоянным законом изменений, так что даже в вещах, в высшей степени случайных, мы принуждены были бы признать как бы некоторую необходимость и, скажу я, рок.

Якоб Бернулли, Искусство предположений (1713)

§ 1. Сходимости «почти наверное» и «по вероятности»

Напомню, что случайная величина есть (измеримая) функция из некоторого непустого множества Ω в множество действительных чисел. Последовательность случайных величин $\{\xi_n\}_{n=1}^{\infty}$ есть, тем самым, последовательность функций, определённых на одном и том же множестве Ω. Существуют разные виды сходимости последовательности ф у н к ц и й. Давать определение любой сходимости мы будем, опираясь на сходимость ч и с л о в ы х последовательностей, как на уже известное основное понятие.

В частности, при каждом новом $\omega \in \Omega$ мы имеем новую ч и с л о в у ю последовательность $\xi_1(\omega)$, $\xi_2(\omega)$, $\xi_3(\omega)$, ... Поэтому, во-первых, можно говорить о сходимости последовательности значений функций в данной точке ω, а также во всех остальных точках $\omega \in \Omega$. В теории вероятностей можно не обращать внимание на неприятности, происходящие с нулевой вероятностью. Поэтому вместо сходимости «всюду» принято рассматривать сходимость «почти всюду», или «почти наверное».

О п р е д е л е н и е 49. Говорят, что последовательность $\{\xi_n\}$ сходится п о ч т и н а в е р н о е к случайной величине ξ при $n \to \infty$, и пишут: $\xi_n \to \xi$ п. н., если $\mathsf{P}\{\omega : \xi_n(\omega) \to \xi(\omega) \text{ при } n \to \infty\} = 1$. Иначе говоря, если $\xi_n(\omega) \to \xi(\omega)$ при $n \to \infty$ для всех $\omega \in \Omega$, кроме, возможно, $\omega \in A$, где A — событие, имеющее нулевую вероятность.

Заметим сразу: определение сходимости «почти наверное» требует знания того, как устроены отображения $\omega \mapsto \xi_n(\omega)$. В задачах же теории вероятностей, как правило, известны не сами случайные величины, а лишь их *распределения*.

Можем ли мы, обладая только информацией о распределениях, говорить о какой-либо сходимости последовательности случайных величин $\{\xi_n\}$ к случайной величине ξ?

Можно, например, потребовать, чтобы вероятность тех элементарных исходов ω, для которых $\xi_n(\omega)$ не попадает в «ε-окрестность» числа $\xi(\omega)$, уменьшалась до нуля с ростом n. Такая сходимость в функциональном анализе называется сходимостью «по мере», а в теории вероятностей — сходимостью «по вероятности».

Определение 50. Говорят, что последовательность случайных величин $\{\xi_n\}$ сходится *по вероятности* к случайной величине ξ при $n \to \infty$, и пишут: $\xi_n \xrightarrow{\text{p}} \xi$, если для любого $\varepsilon > 0$

$$\mathsf{P}\left(|\xi_n - \xi| \geqslant \varepsilon\right) \to 0 \text{ при } n \to \infty \quad (\text{или } \mathsf{P}\left(|\xi_n - \xi| < \varepsilon\right) \to 1 \text{ при } n \to \infty).$$

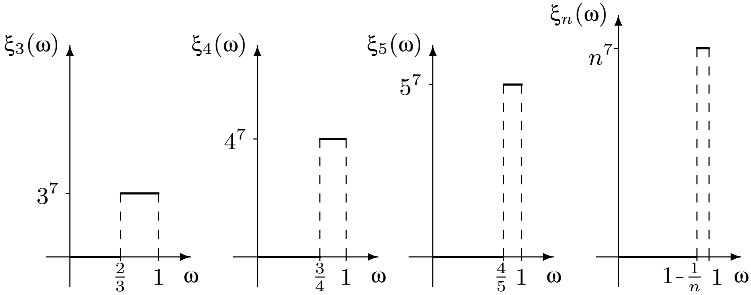
Пример 60. Рассмотрим последовательность $\xi_1, \xi_2, \ldots$, в которой все величины имеют *разные* распределения: величина ξ_n принимает значения 0 и n^7 с вероятностями $\mathsf{P}\big(\xi_n = n^7\big) = 1/n = 1 - \mathsf{P}(\xi_n = 0)$. Докажем, что эта последовательность сходится по вероятности к нулю.

Зафиксируем произвольное $\varepsilon > 0$. Для всех n, начиная с некоторого n_0 такого, что $n_0^7 > \varepsilon$, верно равенство $\mathsf{P}(\xi_n \geqslant \varepsilon) = \mathsf{P}(\xi_n = n^7) = 1/n$. Поэтому

$$\mathsf{P}\big(|\xi_n - 0| \geqslant \varepsilon\big) = \mathsf{P}\big(\xi_n \geqslant \varepsilon\big) = \mathsf{P}\big(\xi_n = n^7\big) = \frac{1}{n} \to 0 \text{ при } n \to \infty.$$

Итак, случайные величины ξ_n с ростом n могут принимать всё бóльшие и бóльшие значения, но со всё меньшей и меньшей вероятностью.

Например, последовательность $\{\xi_n\}$ можно задать на вероятностном пространстве $\langle \Omega, \mathcal{F}, \mathsf{P}\rangle = \langle [0, 1], \mathfrak{B}([0, 1]), \lambda\rangle$ так:



А именно, положим $\xi_n(\omega) = 0$ для $\omega \in [0, 1 - 1/n]$ и $\xi_n(\omega) = n^7$ для $\omega \in (1 - 1/n, 1]$. Заметим, что сходимость по вероятности имеет место совершенно независимо от того, как именно заданы случайные величины на Ω, поскольку определяется лишь их распределениями.

Замечание 24. Иное дело — сходимость «почти наверное». Если, скажем, задать случайные величины как показано на графиках, то сходимость «почти наверное» будет иметь место. Действительно, для всякого $\omega \in [0, 1)$ найдётся такое n_0, что $\omega \in [0, 1 - 1/n_0]$, и поэтому для всех $n \geqslant n_0$ все $\xi_n(\omega)$ равны нулю.

Можно попробовать задать случайные величины ξ_n на $[0, 1]$ как-нибудь иначе, чтобы не было сходимости почти наверное. Для этого нужно заставить отрезок длины $1/n$, на котором $\xi_n(\omega) = n^7$, «бегать» по отрезку $[0, 1]$, чтобы любая точка $\omega \in [0, 1]$ попадала внутрь этого отрезка бесконечное число раз, и, тем самым, для любого ω существовала подпоследовательность $\xi_{n_k}(\omega) \to \infty$.

Однако заметим, что если вероятности $\mathsf{P}(\xi_n = n^7)$ сходятся к нулю достаточно быстро (например, равны $1/n^2$ — чтобы ряд из них сходился), то сходимость п. н. к нулю всегда имеет место (см. теорему 2 § 1 гл. 6 на стр. 134 учебника А. А. Боровкова «Теория вероятностей»).

Сходимость по вероятности не обязательно сопровождается сходимостью математических ожиданий или моментов других порядков: из $\xi_n \xrightarrow{\text{p}} \xi$ не следует, что $\mathsf{E}\xi_n \to \mathsf{E}\xi$. Действительно, в примере 60 имеет место сходимость $\xi_n \xrightarrow{\text{p}} \xi = 0$, но $\mathsf{E}\,\xi_n = n^6 \not\to \mathsf{E}\,\xi = 0$. При этом вообще последовательность $\mathsf{E}\,\xi_n$ неограниченно возрастает.

А если вместо значения n^7 взять n (с той же вероятностью $1/n$), то получим $\mathsf{E}\,\xi_n = 1 \not\to \mathsf{E}\,\xi = 0$. Но теперь хотя бы предел у последовательности математических ожиданий конечен.

Если же ξ_n принимает значения 0 и $\sqrt{n}$ с теми же вероятностями, что и в примере 60, то $\mathsf{E}\,\xi_n = 1/\sqrt{n} \to \mathsf{E}\,\xi = 0$, но уже вторые моменты сходиться ко второму моменту ξ не будут: $\mathsf{E}\,\xi_n^2 = 1 \not\to \mathsf{E}\,\xi^2 = 0$.

Сходимость по вероятности обладает обычными свойствами пределов числовых последовательностей — например, такими:

Свойство 16. *Если* $\xi_n \xrightarrow{\text{p}} \xi$ *и* $\eta_n \xrightarrow{\text{p}} \eta$, *то:*

1. $\xi_n + \eta_n \xrightarrow{\text{p}} \xi + \eta$; 2. $\xi_n \cdot \eta_n \xrightarrow{\text{p}} \xi \cdot \eta$.

Доказательство. При первом прочтении его можно пропустить.

1. В доказательстве мы будем пользоваться свойством монотонности вероятности: если из события A следует событие B (A влечёт B), то вероятность A не превосходит вероятности B: если $A \subseteq B$, то $\mathsf{P}(A) \leqslant \mathsf{P}(B)$.

Остановимся и ответим на следующие «глупые вопросы». Верно ли, что:

а) модуль суммы не превосходит суммы модулей?

б) если $a > b$ и $c > a$, то $c > b$?

в) если $a + b > 2$, то хоть одно из чисел a, b больше единицы?

г) вероятность объединения событий не превосходит суммы их вероятностей?

д) вероятность пересечения событий не превосходит вероятности каждого?

Если на все вопросы вы ответили «да», можно двигаться дальше. Если не на все — ваш контрпример ошибочен. Если вы вообще не поняли, о чём это, лучше вернуться к началу курса.

Докажем, что для сходимости по вероятности предел суммы равен сумме пределов. Зафиксируем $\varepsilon > 0$. Требуется доказать, что $\mathsf{P}(|\xi_n + \eta_n - \xi - \eta| \geqslant \varepsilon) \to 0$ при $n \to \infty$. Но (ср. с вопросами выше):

$$|\xi_n + \eta_n - \xi - \eta| \overset{(а)}{\leqslant} |\xi_n - \xi| + |\eta_n - \eta|,$$

поэтому

$$\{|\xi_n+\eta_n-\xi-\eta| \geqslant \varepsilon\} \overset{(б)}{\subseteq} \{|\xi_n-\xi|+|\eta_n-\eta| \geqslant \varepsilon\} \overset{(в)}{\subseteq} \{|\xi_n-\xi| \geqslant \varepsilon/2\} \cup \{|\eta_n-\eta| \geqslant \varepsilon/2\}.$$

Тогда по свойству монотонности вероятности

$$\begin{aligned}\mathsf{P}(|\xi_n + \eta_n - \xi - \eta| \geqslant \varepsilon) &\leqslant \mathsf{P}\big(|\xi_n - \xi| \geqslant \varepsilon/2 \text{ или } |\eta_n - \eta| \geqslant \varepsilon/2\big) \overset{(г)}{\leqslant} \\ &\overset{(г)}{\leqslant} \mathsf{P}(|\xi_n - \xi| \geqslant \varepsilon/2) + \mathsf{P}(|\eta_n - \eta| \geqslant \varepsilon/2) \to 0\end{aligned}$$

при $n \to \infty$ в силу того, что $\xi_n \xrightarrow{\text{p}} \xi$ и $\eta_n \xrightarrow{\text{p}} \eta$.

2. Для доказательства второго утверждения нам понадобится, кроме положительных ответов на «глупые вопросы» (а)–(д), следующее «хорошее свойство»: для любой случайной величины ζ, согласно свойству (F2) функций распределения, вероятность $\mathsf{P}(|\zeta| \geqslant M)$ стремится к нулю при $M \to \infty$.

Представим $|\xi_n\eta_n - \xi\eta|$ как $|(\eta_n - \eta)(\xi_n - \xi) + \xi(\eta_n - \eta) + \eta(\xi_n - \xi)|$. Затем получим из (а) и монотонности вероятности, что

$$\begin{aligned}\mathsf{P}(|\xi_n\eta_n - \xi\eta| \geqslant \varepsilon) &\leqslant \mathsf{P}(|\eta_n - \eta|\,|\xi_n - \xi| \geqslant \varepsilon/3) + \mathsf{P}(|\xi|\,|\eta_n - \eta| \geqslant \varepsilon/3) + \\ &+ \mathsf{P}(|\eta|\,|\xi_n - \xi| \geqslant \varepsilon/3).\end{aligned}$$

Подумайте, что делать с первым слагаемым в правой части, а мы пока рассмотрим второе слагаемое (третье ничем принципиально от второго не отличается). Обозначим через $A_n = \{|\xi|\,|\eta_n - \eta| \geqslant \varepsilon/3\}$ событие под знаком второй вероятности. Зафиксируем некоторое $M > 0$ и разобьем событие A_n по полной группе событий $B = \{|\xi| \geqslant M\}$ и $\overline{B} = \{|\xi| < M\}$:

$$\mathsf{P}(A_n) = \mathsf{P}(AB) + \mathsf{P}(A\overline{B}) = \mathsf{P}\big(A_n,\, |\xi| \geqslant M\big) + \mathsf{P}\big(A_n,\, |\xi| < M\big).$$

Первую вероятность оцениваем в соответствии с вопросом (д), вторую — пользуясь тем, что из неравенств $|\xi|\,|\eta_n - \eta| \geqslant \varepsilon/3$ и $|\xi| < M$ следует неравенство $M\,|\eta_n - \eta| \geqslant \varepsilon/3$. Получаем:

$$\mathsf{P}(A_n) \leqslant \mathsf{P}(|\xi| \geqslant M) + \mathsf{P}\Big(M|\eta_n - \eta| \geqslant \frac{\varepsilon}{3}\Big) = \mathsf{P}(|\xi| \geqslant M) + \mathsf{P}\Big(|\eta_n - \eta| \geqslant \frac{\varepsilon}{3M}\Big).$$

Осталось для любого фиксированного $M > 0$ устремить n к бесконечности, получив для верхнего предела оценку $\overline{\lim\limits_{n\to\infty}}\, \mathsf{P}(A_n) \leqslant \mathsf{P}(|\xi| \geqslant M)$, после чего мы можем устремить к бесконечности M, пользуясь «хорошим свойством». □

Сходимость по вероятности, так же как и любая другая сходимость, не портится под действием непрерывной функции.

Свойство 17.

1. *Если* $\xi_n \xrightarrow{\text{p}} \xi$ *и* $g(x)$ *— непрерывная функция, то* $g(\xi_n) \xrightarrow{\text{p}} g(\xi)$.
2. *Если* $\xi_n \xrightarrow{\text{p}} c$ *и* $g(x)$ *непрерывна в точке* c*, то* $g(\xi_n) \xrightarrow{\text{p}} g(c)$.

Доказательство. Простое доказательство первого утверждения можно предложить в двух случаях, которыми мы и ограничимся: *если* $\xi = c = const$ (и тогда достаточно, чтобы g была непрерывна в точке c) или *если функция* g *равномерно непрерывна* (а что это значит?).

И в том, и в другом случае для любого $\varepsilon > 0$ найдётся такое $\delta > 0$, что для любого ω, удовлетворяющего условию $|\xi_n(\omega) - \xi(\omega)| < \delta$, выполняется неравенство $|g(\xi_n(\omega)) - g(\xi(\omega))| < \varepsilon$.

Т. е. событие $\bigl\{|\xi_n - \xi| < \delta\bigr\}$ влечёт событие $\bigl\{|g(\xi_n(\omega)) - g(\xi(\omega))| < \varepsilon\bigr\}$. Следовательно, вероятность первого не больше вероятности второго. Но, какое бы ни было $\delta > 0$, вероятность первого события стремится к единице по определению сходимости по вероятности:

$$1 \longleftarrow \mathsf{P}\bigl(|\xi_n - \xi| < \delta\bigr) \leqslant \mathsf{P}\bigl(|g(\xi_n(\omega)) - g(\xi(\omega))| < \varepsilon\bigr) \leqslant 1.$$

Следовательно, вероятность второго события также стремится к единице.

Предлагаю поразмышлять над тем, как доказывать первую часть свойства 17 в общем случае. □

Сходимость «почти наверное» сильнее сходимости по вероятности.

Свойство 18. *Если* $\xi_n \to \xi$ *п. н., то* $\xi_n \xrightarrow{\text{p}} \xi$.

Доказательство. При первом прочтении его можно пропустить. Ограничимся для простоты случаем, когда $\xi_n(\omega) \to \xi(\omega)$ для любого $\omega \in \Omega$.

Зафиксируем $\omega \in \Omega$. По определению предела, $\xi_n(\omega) \to \xi(\omega)$ при $n \to \infty$, если для всякого $\varepsilon > 0$ найдётся $N = N(\omega, \varepsilon) \geqslant 0$ такое, что для всех $n > N$ выполняется неравенство $|\xi_n(\omega) - \xi(\omega)| < \varepsilon$. Ничто не мешает нам дополнительно потребовать, чтобы $|\xi_N(\omega) - \xi(\omega)|$ было не меньше ε, т. е. чтобы среди всех возможных $N(\omega, \varepsilon)$ мы заранее выбрали наименьшее (примем $N = 0$, если $|\xi_n(\omega) - \xi(\omega)| < \varepsilon$ при всех $n \geqslant 1$).

Итак, событие $A = \{n > N(\omega, \varepsilon)\}$ влечёт событие $B = \{|\xi_n(\omega) - \xi(\omega)| < \varepsilon\}$. Но тогда

$$1 \geqslant \mathsf{P}(B) \geqslant \mathsf{P}(A) = \mathsf{P}\bigl(N(\omega, \varepsilon) < n\bigr) = F_{N(\varepsilon,\omega)}(n) \to 1 \text{ при } n \to \infty$$

по свойству (F2) функций распределения. Тогда и $\mathsf{P}(B) = \mathsf{P}\bigl(|\xi_n(\omega) - \xi(\omega)| < \varepsilon\bigr)$ стремится к единице. В силу произвольности выбора ε, это означает сходимость $\xi_n \xrightarrow{\text{p}} \xi$.

Осталось только проверить, является ли $A = \{n > N(\omega, \varepsilon)\}$ событием, т. е. является ли при каждом ε функция $N(\omega, \varepsilon)$ измеримым отображением из Ω в $\mathbb{N}$ (случайной величиной). Для этого достаточно установить, что $\{N(\omega, \varepsilon) = n\}$ — событие. Имеем при $n \geqslant 1$:

$$\{N(\omega, \varepsilon) = n\} = \bigl\{|\xi_n(\omega) - \xi(\omega)| \geqslant \varepsilon\bigr\} \ \cap \bigcap_{k=n+1}^{\infty} \bigl\{|\xi_k(\omega) - \xi(\omega)| < \varepsilon\bigr\} \in \mathcal{F},$$

поскольку $|\xi_k(\omega) - \xi(\omega)|$ — случайная величина. При $n = 0$ первый сомножитель $\{|\xi_n(\omega) - \xi(\omega)| \geqslant \varepsilon\}$ просто отсутствует. Итак, $\{N(\omega, \varepsilon) = n\} \in \mathcal{F}$, что и требовалось доказать. $\square$

Чтобы доказывать сходимость по вероятности, требуется уметь вычислять $\mathsf{P}\left(|\xi_n - \xi| \geqslant \varepsilon\right)$ при больших n. Но для этого нужно знать распределение ξ_n, что не всегда возможно. Скажем, ξ_n может быть суммой (или ещё хуже!) нескольких других случайных величин, распределения которых не устойчивы по суммированию, и вычислить распределение их суммы по формуле свертки или как-то ещё будет слишком сложно.

Если бы мы имели неравенства, позволяющие оценить $\mathsf{P}\left(|\xi_n - \xi| \geqslant \varepsilon\right)$ сверху чем-либо, что мы умеем устремлять к нулю и что проще вычисляется, то сходимость по вероятности мы получили бы по свойству зажатой последовательности: $0 \leqslant \mathsf{P}(\ldots) \leqslant \ldots \to 0$. Итак, неравенства Чебышёва[17].

§ 2. Неравенства Чебышёва

Все неравенства в этом параграфе принято относить к одному классу «неравенств Чебышёва». Следующее неравенство часто называют собственно неравенством Чебышёва, хотя в такой форме оно появилось впервые, видимо, в работах Маркова[18].

Теорема 32 (неравенство Маркова). *Если* $\mathsf{E}\,|\xi| < \infty$, *то для любого* $x > 0$

$$\mathsf{P}\big(|\xi| \geqslant x\big) \leqslant \frac{\mathsf{E}\,|\xi|}{x}\,.$$

Доказательство. Нам потребуется следующее понятие.

Определение 51. Назовём **индикатором** события A случайную величину $I(A)$, равную единице, если событие A произошло, и нулю, если A не произошло.

По определению, величина $I(A)$ имеет распределение Бернулли с параметром $p = \mathsf{P}(I(A) = 1) = \mathsf{P}(A)$, и её математическое ожидание равно вероятности успеха $p = \mathsf{P}(A)$. Индикаторы прямого и противоположного событий связаны равенством $I(A) + I(\overline{A}) = 1$. Поэтому

$$|\xi| = |\xi| \cdot I(|\xi| < x) + |\xi| \cdot I(|\xi| \geqslant x) \geqslant |\xi| \cdot I(|\xi| \geqslant x) \geqslant x \cdot I(|\xi| \geqslant x).$$

Тогда

$$\mathsf{E}\,|\xi| \geqslant \mathsf{E}\left(x \cdot I(|\xi| \geqslant x)\right) = x \cdot \mathsf{P}\big(|\xi| \geqslant x\big). \tag{21}$$

Осталось разделить обе части неравенства (21) на положительное x. $\square$

Следующее неравенство мы будем называть обобщённым неравенством Чебышёва.

---

[17] Пафнутий Львович Чебышёв (16.05.1821 — 8.12.1894)

[18] Андрей Андреевич Марков (14.06.1856 — 20.07.1922)

Следствие 17 (обобщённое неравенство Чебышёва). *Пусть функция g не убывает и неотрицательна на $\mathbb{R}$. Если $\mathsf{E}\, g(\xi) < \infty$, то для любого $x \in \mathbb{R}$*

$$\mathsf{P}\big(\xi \geqslant x\big) \leqslant \frac{\mathsf{E}\, g(\xi)}{g(x)}.$$

Доказательство. Заметим, что $\mathsf{P}\big(\xi \geqslant x\big) \leqslant \mathsf{P}\big(g(\xi) \geqslant g(x)\big)$, поскольку функция g не убывает. Оценим последнюю вероятность по неравенству Маркова, которое можно применять в силу неотрицательности g:

$$\mathsf{P}\big(g(\xi) \geqslant g(x)\big) \leqslant \frac{\mathsf{E}\, g(\xi)}{g(x)}. \qquad \square$$

Бьенеме[19] и Чебышёв прямыми методами доказали неравенство, которое нам будет удобно получить как следствие неравенства Маркова.

Следствие 18 (неравенство Чебышёва — Бьенеме). *Если $\mathsf{D}\,\xi$ существует, то для любого $x > 0$*

$$\mathsf{P}\big(|\xi - \mathsf{E}\,\xi| \geqslant x\big) \leqslant \frac{\mathsf{D}\,\xi}{x^2}.$$

Доказательство. Для $x > 0$ неравенство $|\xi - \mathsf{E}\,\xi| \geqslant x$ равносильно неравенству $(\xi - \mathsf{E}\,\xi)^2 \geqslant x^2$, поэтому

$$\mathsf{P}\big(|\xi - \mathsf{E}\,\xi| \geqslant x\big) = \mathsf{P}\big((\xi - \mathsf{E}\,\xi)^2 \geqslant x^2\big) \leqslant \frac{\mathsf{E}\,\big(\xi - \mathsf{E}\,\xi\big)^2}{x^2} = \frac{\mathsf{D}\,\xi}{x^2}. \quad \square$$

В качестве следствия получим так называемое «правило трёх сигм», которое означает, что *вероятность случайной величине отличаться от своего математического ожидания более чем на три корня из дисперсии, мала*. Разумеется, для каждого распределения величина этой вероятности своя: для нормального распределения, например, 0,0027 — см. свойство 12. Мы получим верную для всех распределений с конечной дисперсией оценку сверху для вероятности случайной величине отличаться от своего математического ожидания более чем на три корня из дисперсии.

Следствие 19. *Если $\mathsf{E}\,\xi^2 < \infty$, то $\mathsf{P}\big(|\xi - \mathsf{E}\,\xi| \geqslant 3\sqrt{\mathsf{D}\,\xi}\big) \leqslant \frac{1}{9}$.*

Доказательство. Согласно следствию 18,

$$\mathsf{P}\big(|\xi - \mathsf{E}\,\xi| \geqslant 3\sqrt{\mathsf{D}\,\xi}\big) \leqslant \frac{\mathsf{D}\,\xi}{\big(3\sqrt{\mathsf{D}\,\xi}\big)^2} = \frac{1}{9}. \qquad \square$$

Упражнение. Найти $\mathsf{P}\big(|\xi - \mathsf{E}\,\xi| \geqslant 3\sqrt{\mathsf{D}\,\xi}\big)$, если ξ имеет

а) равномерное распределение на каком-нибудь отрезке;

б) показательное распределение с каким-нибудь параметром;

в) распределение Бернулли с параметром $1/2$.

[19] Irénée-Jules Bienaymé (28.08.1796 — 19.10.1878, Paris)

§ 3. Законы больших чисел

Определение 52. Говорят, что последовательность случайных величин $\xi_1, \xi_2, \ldots$ с конечными первыми моментами удовлетворяет закону больших чисел (ЗБЧ), если

$$\frac{\xi_1 + \ldots + \xi_n}{n} - \frac{\mathsf{E}\,\xi_1 + \ldots + \mathsf{E}\,\xi_n}{n} \xrightarrow{\text{p}} 0 \text{ при } n \to \infty. \tag{22}$$

Законами больших чисел принято называть утверждения о том, при каких условиях последовательность случайных величин удовлетворяет закону больших чисел.

Выясним сначала, когда выполнен ЗБЧ для последовательности независимых и одинаково распределённых случайных величин.

Теорема 33 (ЗБЧ Чебышёва). *Для любой последовательности $\xi_1, \xi_2, \ldots$ попарно независимых и одинаково распределённых случайных величин с конечным вторым моментом $\mathsf{E}\,\xi_1^2 < \infty$ имеет место сходимость:*

$$\frac{\xi_1 + \ldots + \xi_n}{n} \xrightarrow{\text{p}} \mathsf{E}\,\xi_1. \tag{23}$$

Заметим, что если величины одинаково распределены, то их математические ожидания одинаковы (и равны, например, $\mathsf{E}\,\xi_1$), поэтому свойство (22) можно записать в виде (23).

ЗБЧ утверждает, что среднее арифметическое большого числа случайных слагаемых «стабилизируется» с ростом этого числа. Как бы сильно каждая случайная величина не отклонялась от своего среднего значения, при суммировании эти отклонения «взаимно гасятся», так что среднее арифметическое приближается к постоянной величине.

В дальнейшем мы увидим, что требование конечности второго момента (или дисперсии) связано исключительно со способом доказательства, и что утверждение останется верным, если требовать существования только первого момента.

Доказательство. Обозначим через $S_n = \xi_1 + \ldots + \xi_n$ сумму первых n случайных величин. Из линейности математического ожидания получим:

$$\mathsf{E}\left(\frac{S_n}{n}\right) = \frac{\mathsf{E}\,\xi_1 + \ldots + \mathsf{E}\,\xi_n}{n} = \frac{n\,\mathsf{E}\,\xi_1}{n} = \mathsf{E}\,\xi_1.$$

Пусть $\varepsilon > 0$. Воспользуемся неравенством Чебышёва (следствие 18):

$$\mathsf{P}\left(\left|\frac{S_n}{n} - \mathsf{E}\left(\frac{S_n}{n}\right)\right| \geqslant \varepsilon\right) \leqslant \frac{\mathsf{D}\left(\frac{S_n}{n}\right)}{\varepsilon^2} = \frac{\mathsf{D}\,S_n}{n^2\varepsilon^2} = \frac{\mathsf{D}\,\xi_1 + \ldots + \mathsf{D}\,\xi_n}{n^2\varepsilon^2} =$$

$$= \frac{n\,\mathsf{D}\,\xi_1}{n^2\varepsilon^2} = \frac{\mathsf{D}\,\xi_1}{n\varepsilon^2} \to 0 \text{ при } n \to \infty, \tag{24}$$

так как $\mathsf{D}\,\xi_1 < \infty$. Заметим, что дисперсия суммы превратилась в сумму дисперсий в силу попарной независимости слагаемых, из-за которой все ковариации $\mathrm{cov}(\xi_i, \xi_j)$ в свойстве 14 обратились в нуль при $i \neq j$. Сумма же дисперсий слагаемых равняется $n\,\mathsf{D}\,\xi_1$ из-за их одинаковой распределённости. □

Замечание 25. Мы не только доказали сходимость, но и получили оценку для вероятности среднему арифметическому любого числа попарно независимых и одинаково распределённых величин отличаться от $\mathsf{E}\,\xi_1$ более чем на заданное ε:

$$\mathsf{P}\left(\left|\frac{\xi_1 + \ldots + \xi_n}{n} - \mathsf{E}\,\xi_1\right| \geqslant \varepsilon\right) \leqslant \frac{\mathsf{D}\,\xi_1}{n\varepsilon^2}. \tag{25}$$

Легко видеть, что попарную независимость слагаемых в ЗБЧ Чебышёва можно заменить их попарной некоррелированностью, ничего не меняя в доказательстве. ЗБЧ может выполняться и для последовательности зависимых и разнораспределённых слагаемых. Предлагаю читателям, проследив за равенствами и неравенствами (24), получить доказательство следующего утверждения, предлагающего достаточные условия выполнения ЗБЧ для последовательности произвольных случайных величин.

Теорема 34 (ЗБЧ Маркова). *Последовательность случайных величин $\xi_1, \xi_2, \ldots$ с конечными вторыми моментами удовлетворяет ЗБЧ при выполнении любого из следующих условий:*

а) если $\mathsf{D}\,S_n = o(n^2)$, т. е. если $\dfrac{\mathsf{D}\,S_n}{n^2} \to 0$ при $n \to \infty$;

б) если $\xi_1, \xi_2, \ldots$ независимы и $\mathsf{D}\,\xi_1 + \ldots + \mathsf{D}\,\xi_n = o(n^2)$, т. е. если

$$\frac{\mathsf{D}\,\xi_1 + \ldots + \mathsf{D}\,\xi_n}{n^2} \to 0 \text{ при } n \to \infty;$$

в) если $\xi_1, \xi_2, \ldots$ независимы, одинаково распределены и имеют конечную дисперсию (ЗБЧ Чебышёва).

Теорема Маркова утверждает, что ЗБЧ выполнен, если дисперсия суммы n слагаемых растёт не слишком быстро с ростом n.

Сильная зависимость слагаемых приводит обычно к невыполнению ЗБЧ. Если, например, $\mathsf{D}\,\xi_1 \neq 0$ и $\xi_n \equiv \xi_1$, то $S_n = n\xi_1$, и свойство (23) не выполнено (убедиться в этом!). В этом случае не выполнено и условие (а) теоремы Маркова: $\mathsf{D}\,S_n = \mathsf{D}\,(n\xi_1) = cn^2$. Для одинаково распределённых слагаемых дисперсия суммы ещё быстрее расти уже не может.

Следующее утверждение мы докажем чуть позже. Сравните его условия с условиями ЗБЧ Чебышёва.

Теорема 35 (ЗБЧ Хинчина[20]). *Для любой последовательности $\xi_1, \xi_2, \ldots$ независимых (в совокупности) и одинаково распределённых случайных величин с конечным* первым *моментом $\mathsf{E}\,|\xi_1| < \infty$*

имеет место сходимость:

$$\frac{\xi_1 + \ldots + \xi_n}{n} \xrightarrow{\text{p}} \mathsf{E}\,\xi_1.$$

Более того, в условиях теоремы 35 имеет место и сходимость п. н. последовательности $(\xi_1 + \ldots + \xi_n)/n$ к $\mathsf{E}\,\xi_1$. Это утверждение называется усиленным законом больших чисел (УЗБЧ) Колмогорова, и его мы доказывать не будем.

Получим в качестве следствия из ЗБЧ Чебышёва закон больших чисел Я. Бернулли. В отличие от ЗБЧ Чебышёва, описывающего предельное поведение среднего арифметического случайных величин с произвольными распределениями, ЗБЧ Бернулли — утверждение только для схемы Бернулли.

Теорема 36 (ЗБЧ Бернулли). *Пусть событие A может произойти в любом из n независимых испытаний с одной и той же вероятностью p, и пусть $\nu_n(A)$ — число осуществлений события A в n испытаниях. Тогда* $\dfrac{\nu_n(A)}{n} \xrightarrow{\text{p}} p$. *При этом для любого* $\varepsilon > 0$

$$\mathsf{P}\left(\left|\frac{\nu_n(A)}{n} - p\right| \geqslant \varepsilon\right) \leqslant \frac{p(1-p)}{n\varepsilon^2}.$$

Доказательство. Заметим, что $\nu_n(A)$ есть сумма независимых, одинаково распределённых случайных величин, имеющих распределение Бернулли с параметром $p = \mathsf{P}(A)$ (индикаторов того, что в соответствующем испытании произошло A): $\nu_n(A) = \xi_1 + \ldots + \xi_n$, где

$$\xi_i = I_i(A) = \begin{cases} 1, & \text{если } A \text{ произошло в } i\text{-м испытании;} \\ 0, & \text{если } A \text{ не произошло в } i\text{-м испытании;} \end{cases}$$

и $\mathsf{E}\,\xi_1 = \mathsf{P}(A) = p$, $\mathsf{D}\,\xi_1 = p(1-p)$. Осталось воспользоваться ЗБЧ в форме Чебышёва и неравенством (25). □

§ 4. Примеры использования ЗБЧ Чебышёва

Пример 61. Монета подбрасывается 10 000 раз. Оценим вероятность того, что частота выпадения герба отличается от 1/2 на 0,01 или более.

Пусть $\xi_1, \ldots, \xi_n$ — независимые случайные величины, каждая из которых имеет распределение Бернулли с параметром $p = 1/2$ и равна единице, если при соответствующем подбрасывании выпал герб, и нулю иначе. Нужно оценить $\mathsf{P}\left(\left|\frac{\nu_n}{n} - \frac{1}{2}\right| \geqslant 0{,}01\right)$, где $n = 10^4$, а $\nu_n = \sum_{i=1}^{n} \xi_i$ — число выпадений герба. Поскольку $\mathsf{D}\,\xi_1 = 1/2 \cdot 1/2 = 1/4$, искомая оценка сверху

[20] Александр Яковлевич Хинчин (19.07.1894 — 18.11.1959)

выглядит так:

$$\mathsf{P}\left(\left|\frac{\nu_n}{n}-\frac{1}{2}\right|\geqslant 0{,}01\right)\leqslant\frac{\mathsf{D}\,\xi_1}{n\cdot 0{,}01^2}=\frac{1}{4\cdot 10^4\cdot 10^{-4}}=\frac{1}{4}.$$

Итак, неравенство Чебышёва позволяет заключить, что в среднем не более чем в четверти случаев при 10 000 подбрасываниях монеты частота выпадения герба будет отличаться от $1/2$ на одну сотую или больше. Мы увидим, насколько это грубая оценка, когда познакомимся с *центральной предельной теоремой*.

Пример 62. Пусть $\xi_1, \xi_2, \ldots$ — случайные величины, дисперсии которых ограничены одной и той же постоянной C, а ковариации любых двух ξ_i и ξ_j $(i \neq j)$, не являющихся соседними в последовательности, равны нулю. Удовлетворяет ли эта последовательность ЗБЧ?

Воспользуемся подходящим неравенством в (24) и свойством 14:

$$\mathsf{P}\left(\left|\frac{S_n}{n}-\mathsf{E}\left(\frac{S_n}{n}\right)\right|\geqslant\varepsilon\right)\leqslant\frac{\mathsf{D}\,S_n}{n^2\varepsilon^2};$$

$$\mathsf{D}\,(\xi_1+\ldots+\xi_n)=\sum_{i=1}^{n}\mathsf{D}\,\xi_i+2\sum_{i<j}\operatorname{cov}(\xi_i,\,\xi_j). \tag{26}$$

Но при $i<j$ по условию $\operatorname{cov}(\xi_i,\,\xi_j)=0$ для $i\neq j-1$. Следовательно, все ковариации в равенстве (26) равны нулю, кроме, может быть, $\operatorname{cov}(\xi_1,\,\xi_2)$, $\operatorname{cov}(\xi_2,\,\xi_3), \ldots, \operatorname{cov}(\xi_{n-1},\,\xi_n)$ (их ровно $n-1$ штука).

Оценим каждую из них, используя тот факт, что коэффициент корреляции по модулю не превосходит единицы:

$$\operatorname{cov}(\xi_i,\xi_j)\leqslant\sqrt{\mathsf{D}\,\xi_i}\sqrt{\mathsf{D}\,\xi_j}\leqslant\sqrt{C}\sqrt{C}=C,$$

$$\mathsf{D}\,S_n=\sum_{i=1}^{n}\mathsf{D}\,\xi_i+2\sum_{i=1}^{n-1}\operatorname{cov}(\xi_i,\,\xi_{i+1})\leqslant nC+2(n-1)C.$$

Получаем, что последовательность $\xi_1, \xi_2, \ldots$ удовлетворяет ЗБЧ, так как выполнено условие (а) теоремы Маркова: $\mathsf{D}\,S_n=o(n^2)$.

ГЛАВА 12

Центральная предельная теорема

> Из этой первой лекции по теории вероятностей я запомнил только полузнакомый термин «математическое ожидание». Незнакомец употреблял этот термин неоднократно, и каждый раз я представлял себе большое помещение, вроде зала ожидания, с кафельным полом, где сидят люди с портфелями и бюварами и, подбрасывая время от времени к потолку монетки и бутерброды, сосредоточенно чего-то ожидают. До сих пор я часто вижу это во сне. Но тут незнакомец оглушил меня звонким термином «предельная теорема Муавра — Лапласа» и сказал, что всё это к делу не относится.
>
> *Аркадий и Борис Стругацкие, Стажёры*

§ 1. Как быстро среднее арифметическое сходится к математическому ожиданию?

Пусть, как в законе больших чисел Чебышёва, $S_n = \xi_1 + \ldots + \xi_n$ — сумма n независимых и одинаково распределённых величин с конечной дисперсией. Тогда по ЗБЧ $\dfrac{S_n}{n} \xrightarrow{\text{p}} \mathsf{E}\,\xi_1$ с ростом n. Или, после приведения к общему знаменателю,

$$\frac{S_n - n\,\mathsf{E}\,\xi_1}{n} \xrightarrow{\text{p}} 0.$$

Если при делении на n мы получили в пределе нуль (в смысле некоторой, всё равно какой, сходимости), резонно задать себе вопрос: а не слишком ли на большую величину мы поделили? Нельзя ли поделить на что-нибудь, растущее к бесконечности медленнее, чем n, чтобы получить в пределе не нуль (и не бесконечность, само собой)?

Можно поставить этот вопрос по-другому. Вот последовательность, стремящаяся (как-то) к нулю. Можно ли её домножить на что-либо растущее, чтобы «погасить» это стремление к нулю? Получив, тем самым, что-нибудь конечное и отличное от нуля в пределе?

Оказывается, что уже последовательность случайных величин

$$\frac{S_n - n\,\mathsf{E}\,\xi_1}{\sqrt{n}} = \sqrt{n} \cdot \frac{S_n - n\,\mathsf{E}\,\xi_1}{n}$$

не сходится к нулю. Распределение членов этой последовательности становится всё более похожим на нормальное распределение! Можно считать, что такая последовательность сходится к случайной величине, имеющей нормальное распределение, но сходится никак не по вероятности, а только в смысле сходимости распределений, или «слабой сходимости».

§ 2. Слабая сходимость

Пусть задана последовательность случайных величин $\xi_1, \xi_2, \ldots$, задано некоторое распределение $\mathcal{F}$ с функцией распределения F_ξ и пусть ξ — произвольная случайная величина, имеющая распределение $\mathcal{F}$.

Определение 53. Говорят, что последовательность случайных величин $\xi_1, \xi_2, \ldots$ сходится слабо или по распределению к случайной величине ξ и пишут: $\xi_n \Rightarrow \xi$, если для любого x такого, что функция распределения F_ξ непрерывна в точке x, имеет место сходимость $F_{\xi_n}(x) \to F_\xi(x)$ при $n \to \infty$.

Итак, слабая сходимость — это сходимость функций распределения во всех точках непрерывности предельной функции распределения.

Замечание 26. Необходимо заметить, что запись $\xi_n \Rightarrow \xi$ удобна, но не всегда разумна: если «предельную» величину ξ заменить на другую величину η с тем же распределением, ничего не изменится: в том же смысле $\xi_n \Rightarrow \eta$. Поэтому слабая сходимость всё же не есть сходимость случайных величин, и ею нельзя оперировать как сходимостями п. н. и по вероятности, для которых предельная случайная величина единственна (с точностью до значений на множестве нулевой вероятности).

Поэтому в определении 53 часто говорят и пишут так: ξ_n слабо сходится к распределению $\mathcal{F}$, т. е. $\xi_n \Rightarrow \mathcal{F}$, либо даже так: распределения ξ_n слабо сходятся к распределению $\mathcal{F}$: $F_{\xi_n} \Rightarrow \mathcal{F}$.

Следующее свойство очевидно. Если нет — нужно вернуться к разделу 6 и вспомнить, что такое функция распределения.

Свойство 19. *Если $\xi_n \Rightarrow \xi$, и функция распределения F_ξ непрерывна в точках a и b, то $\mathsf{P}(\xi_n \in (a, b)) \to \mathsf{P}(\xi \in (a, b))$. Наоборот, если во всех точках a и b непрерывности функции распределения F_ξ имеет место сходимость $\mathsf{P}(\xi_n \in (a, b)) \to \mathsf{P}(\xi \in (a, b))$, то $\xi_n \Rightarrow \xi$.*

Вместо открытого интервала (a, b) можно взять $[a, b]$, $(a, b]$ или $[a, b)$.

Следующее свойство уточняет отношения между сходимостями.

Свойство 20. 1. *Если $\xi_n \xrightarrow{\mathrm{p}} \xi$, то $\xi_n \Rightarrow \xi$.*

2. *Если $\xi_n \Rightarrow c = const$, то $\xi_n \xrightarrow{\mathrm{p}} c$.*

Итак, сходимость по вероятности влечёт слабую сходимость. Обратное утверждение в общем случае смысла не имеет (см. замечание 26). Однако из слабой сходимости к постоянной вытекает сходимость по вероятности.

Доказательство. Свойство (1) мы докажем чуть позже.

Докажем (2): слабая сходимость к постояннной влечёт сходимость по вероятности. Пусть $\xi_n \Rightarrow c$, т. е.

$$F_{\xi_n}(x) \to F_c(x) = \begin{cases} 0, & x \leqslant c; \\ 1, & x > c \end{cases}$$

при любом x, являющемся точкой непрерывности предельной функции $F_c(x)$, т. е. при всех $x \neq c$.

Возьмём произвольное $\varepsilon > 0$ и докажем, что $\mathsf{P}(|\xi_n - c| < \varepsilon) \to 1$:

$$\mathsf{P}(-\varepsilon < \xi_n - c < \varepsilon) = \mathsf{P}(c - \varepsilon < \xi_n < c + \varepsilon) \geqslant \mathsf{P}(c - \varepsilon/2 \leqslant \xi_n < c + \varepsilon) =$$
$$= F_{\xi_n}(c + \varepsilon) - F_{\xi_n}(c - \varepsilon/2) \ \to \ F_c(c + \varepsilon) - F_c(c - \varepsilon/2) = 1 - 0 = 1,$$

поскольку в точках $c+\varepsilon$ и $c-\varepsilon/2$ функция F_c непрерывна, и, следовательно, имеет место сходимость последовательностей $F_{\xi_n}(c + \varepsilon)$ к $F_c(c + \varepsilon) = 1$ и $F_{\xi_n}(c - \varepsilon/2)$ к $F_c(c - \varepsilon/2) = 0$.

Осталось заметить, что $\mathsf{P}(|\xi_n - c| < \varepsilon)$ не бывает больше 1, так что по свойству предела зажатой последовательности $\mathsf{P}(|\xi_n - c| < \varepsilon) \to 1$. □

Следующее свойство приводит пример операций, которые можно применять к слабо сходящимся последовательностям — домножать их на последовательности, сходящиеся по вероятности к постоянным величинам.

Замечание 27. Желание написать «если $\xi_n \Rightarrow \xi$ и $\eta_n \Rightarrow \eta$, то $\xi_n + \eta_n \Rightarrow \xi + \eta$» сразу проходит, стоит перевести это «свойство» на язык функций распределения и задуматься — что такое «функция распределения суммы $\xi + \eta$», когда вместо них можно брать любые другие $\tilde{\xi}$ и $\tilde{\eta}$ с теми же распределениями, как угодно зависимые. Иное дело — когда одно из предельных распределений вырождено. Тогда функция распределения суммы или произведения определена однозначно.

Свойство 21. 1. *Если* $\xi_n \xrightarrow{\mathrm{p}} c = const$ *и* $\eta_n \Rightarrow \eta$, *то* $\xi_n \cdot \eta_n \Rightarrow c\eta$.

2. *Если* $\xi_n \xrightarrow{\mathrm{p}} c = const$ *и* $\eta_n \Rightarrow \eta$, *то* $\xi_n + \eta_n \Rightarrow c + \eta$.

Доказательство. Нелюбознательный читатель может пропустить это доказательство, вернувшись к нему при втором прочтении.

Заметим вначале, что если $\eta_n \Rightarrow \eta$, то $c\eta_n \Rightarrow c\eta$ и $c + \eta_n \Rightarrow c + \eta$ (доказать). Поэтому достаточно доказать первое утверждение свойства 21 при $c = 1$, а второе утверждение — при $c = 0$.

Рассмотрим второе утверждение, оставив первое любознательному читателю. Пусть $\xi_n \xrightarrow{\mathrm{p}} 0$ и $\eta_n \Rightarrow \eta$. Докажем, что тогда $\xi_n + \eta_n \Rightarrow \eta$.

Пусть x_0 — точка непрерывности функции распределения $F_\eta(x)$. Требуется доказать, что имеет место сходимость $F_{\xi_n+\eta_n}(x_0) \to F_\eta(x_0)$. Зафиксируем достаточно маленькое $\varepsilon > 0$ такое, что $F_\eta(x)$ непрерывна в точках $x_0 \pm \varepsilon$.

События $H_1 = \{|\xi_n| \geqslant \varepsilon\}$ и $H_2 = \{|\xi_n| < \varepsilon\}$ образуют полную группу, поэтому

$$\begin{aligned} F_{\xi_n+\eta_n}(x_0) &= \mathsf{P}(\xi_n + \eta_n < x_0) = \\ &= \mathsf{P}(\xi_n + \eta_n < x_0,\ H_1) + \mathsf{P}(\xi_n + \eta_n < x_0,\ H_2) = P_1 + P_2. \end{aligned}$$

Оценим $P_1 + P_2$ сверху и снизу. Для P_1 имеем:

$$0 \leqslant P_1 = \mathsf{P}(\xi_n + \eta_n < x_0,\ H_1) \leqslant \mathsf{P}(H_1) = \mathsf{P}(|\xi_n| \geqslant \varepsilon),$$

и последняя вероятность может быть выбором n сделана сколь угодно малой.

Для P_2, с одной стороны,

$$P_2 = \mathsf{P}(\xi_n + \eta_n < x_0,\ -\varepsilon < \xi_n < \varepsilon) \leqslant \mathsf{P}(-\varepsilon + \eta_n < x_0) = \mathsf{P}(\eta_n < x_0 + \varepsilon).$$

Неравенство следует из простого соображения: если $\xi_n > -\varepsilon$ и $\xi_n + \eta_n < x_0$, то, тем более, $-\varepsilon + \eta_n < x_0$.

С другой стороны,

$$\begin{aligned} P_2 &= \mathsf{P}(\xi_n + \eta_n < x_0,\ -\varepsilon < \xi_n < \varepsilon) \geqslant \mathsf{P}(\varepsilon + \eta_n < x_0,\ -\varepsilon < \xi_n < \varepsilon) \geqslant \\ &\geqslant \mathsf{P}(\varepsilon + \eta_n < x_0) - \mathsf{P}(|\xi_n| \geqslant \varepsilon) = \mathsf{P}(\eta_n < x_0 - \varepsilon) - \mathsf{P}(|\xi_n| \geqslant \varepsilon). \end{aligned}$$

Здесь первое неравенство объясняется включением

$$\{\varepsilon + \eta_n < x_0,\ -\varepsilon < \xi_n < \varepsilon\} \subseteq \{\xi_n + \eta_n < x_0,\ -\varepsilon < \xi_n < \varepsilon\},$$

которое получилось заменой в событии $\{\varepsilon + \eta_n < x_0\}$ числа ε на меньшую величину ξ_n, $\xi_n < \varepsilon$. Второе неравенство следует из свойств:

$$\mathsf{P}(A\overline{B}) \leqslant \mathsf{P}(\overline{B}), \text{ поэтому } \mathsf{P}(AB) = \mathsf{P}(A) - \mathsf{P}(A\overline{B}) \geqslant \mathsf{P}(A) - \mathsf{P}(\overline{B}).$$

Мы получили оценки снизу и сверху для $P_1 + P_2$, т. е. для $F_{\xi_n+\eta_n}(x_0)$:

$$\mathsf{P}(\eta_n < x_0 - \varepsilon) - \mathsf{P}(|\xi_n| \geqslant \varepsilon) \leqslant F_{\xi_n+\eta_n}(x_0) \leqslant \mathsf{P}(|\xi_n| \geqslant \varepsilon) + \mathsf{P}(\eta_n < x_0 + \varepsilon),$$

или

$$F_{\eta_n}(x_0 - \varepsilon) - \mathsf{P}(|\xi_n| \geqslant \varepsilon) \leqslant F_{\xi_n+\eta_n}(x_0) \leqslant \mathsf{P}(|\xi_n| \geqslant \varepsilon) + F_{\eta_n}(x_0 + \varepsilon).$$

Устремляя n к бесконечности, и вспоминая, что $x_0 \pm \varepsilon$ — точки непрерывности функции распределения F_η, получим

$$F_\eta(x_0 - \varepsilon) \leqslant \underline{\lim}\, F_{\xi_n+\eta_n}(x_0) \leqslant \overline{\lim}\, F_{\xi_n+\eta_n}(x_0) \leqslant F_\eta(x_0 + \varepsilon). \tag{27}$$

У любой функции распределения не более чем счётное множество точек разрыва. Поэтому можно выбрать такую уменьшающуюся до нуля последовательность ε, что в точках $x_0 \pm \varepsilon$ функция распределения F_η будет непрерывной, и, следовательно, останутся верны неравенства (27). Переходя к пределу по такой последовательности $\varepsilon \to 0$ и помня, что x_0 — точка непрерывности функции F_η, получим, что нижний и верхний пределы $F_{\xi_n+\eta_n}(x_0)$ при $n \to \infty$ совпадают и равны $F_\eta(x_0)$. □

Доказательство утверждения (1) из свойства 20. В качестве простого следствия из только что доказанного второго утверждения свойства 21 покажем, что сходимость $\xi_n \xrightarrow{\text{p}} \xi$ по вероятности влечёт слабую сходимость $\xi_n \Rightarrow \xi$.

Представим ξ_n в виде суммы $\xi_n = (\xi_n - \xi) + \xi$. Здесь последовательность $\xi_n - \xi$ по вероятности стремится к нулю, а «последовательность» ξ слабо сходится к ξ. Поэтому их сумма слабо сходится к ξ. □

Получим ещё одно следствие из свойства 21. Для удобства ссылок назовём следующее утверждение «теоремой о двойном пределе».

Теорема 37. *Пусть $\xi_n \Rightarrow \xi$, причём функция распределения случайной величины ξ непрерывна всюду, и пусть $x_n \to x_0 \in [-\infty, \infty]$ — числовая последовательность. Тогда $F_{\xi_n}(x_n) \to F_\xi(x_0)$.*

В формулировке теоремы мы, краткости ради, использовали запись $F_\xi(\pm\infty)$, которую следует понимать так: $F_\xi(-\infty) = 0$, $F_\xi(+\infty) = 1$.

Доказательство. Если $x_0 \in \mathbb{R}$, то утверждение теоремы сразу следует из свойства 21. Действительно, из $x_n \to x_0$ следует, что $x_n \xrightarrow{\text{p}} x_0$. К тому же $\xi_n \Rightarrow \xi$. Тогда утверждение (2) свойства 21 позволяет заключить, что $\xi_n - x_n \Rightarrow \xi - x_0$. Функция распределения $F_{\xi - x_0}(x)$ отличается от $F_\xi(x)$ лишь сдвигом и тоже непрерывна всюду, поэтому имеет место сходимость функций распределения в любой точке. В частности, в точке $x = 0$ имеет место сходимость при $n \to \infty$

$$F_{\xi_n}(x_n) = \mathsf{P}(\xi_n - x_n < 0) = F_{\xi_n - x_n}(0) \to F_{\xi - x_0}(0) = \mathsf{P}(\xi - x_0 < 0) = F_\xi(x_0).$$

Пусть теперь $x_0 = -\infty$. Нужно доказать, что $F_{\xi_n}(x_n) \to F_\xi(-\infty) = 0$.

По определению, $x_n \to -\infty$ с ростом n, если для всякого $M > 0$ найдётся номер N такой, что для всех $n \geqslant N$ выполнено неравенство: $x_n \leqslant -M$. В силу монотонности функций распределения, $0 \leqslant F_{\xi_n}(x_n) \leqslant F_{\xi_n}(-M)$. В точке $-M$, как и в любой иной точке, имеет место сходимость функций распределения $F_{\xi_n}(-M) \to F_\xi(-M)$. Выбором M величина $F_\xi(-M)$ может быть сделана сколь угодно близкой к нулю. Тем самым верхний предел последовательности $F_{\xi_n}(x_n)$ оказывается зажат между нулём и сколь угодно малой величиной, т. е. равняется нулю.

Случай $x_0 = +\infty$ проверяется аналогично. □

Несколько содержательных примеров слабой сходимости мы рассмотрим в следующей главе. Но основной источник слабо сходящихся последовательностей и необычайно мощное и универсальное средство для асимптотического анализа распределений сумм независимых и одинаково распределённых случайных величин предоставляет нам центральная предельная теорема.

§ 3. Центральная предельная теорема

Мы будем называть следующее утверждение «ЦПТ Ляпунова[21]», но сформулируем и докажем теорему Ляпунова только в частном случае — для последовательности независимых и одинаково распределённых случайных величин. Как и ранее, через S_n обозначена сумма первых n случайных величин в последовательности: $S_n = \xi_1 + \ldots + \xi_n$.

[21] Александр Михайлович Ляпунов (6.06.1857 — 3.11.1918)

Теорема 38 (ЦПТ Ляпунова). *Пусть* $\xi_1, \xi_2, \ldots$ — *независимые и одинаково распределённые случайные величины с конечной и ненулевой дисперсией:* $0 < \mathsf{D}\,\xi_1 < \infty$. *Тогда имеет место слабая сходимость*

$$\frac{S_n - n\,\mathsf{E}\,\xi_1}{\sqrt{n\,\mathsf{D}\,\xi_1}} \Rightarrow \mathrm{N}_{0,1}$$

последовательности центрированных и нормированных сумм случайных величин к стандартному нормальному распределению.

Пользуясь определением и свойствами слабой сходимости и заметив, что функция распределения $\Phi_{a,\sigma^2}(x)$ любого нормального закона непрерывна всюду на $\mathbb{R}$ (почему?), утверждение ЦПТ можно сформулировать любым из следующих способов:

Следствие 20. *Пусть* $\xi_1, \xi_2, \ldots$ — *независимые и одинаково распределённые случайные величины с конечной и ненулевой дисперсией. Тогда выполнены утверждения:*

а) для любых вещественных $x < y$ *при* $n \to \infty$ *имеет место сходимость*

$$\mathsf{P}\left(x \leqslant \frac{S_n - n\,\mathsf{E}\,\xi_1}{\sqrt{n\,\mathsf{D}\,\xi_1}} \leqslant y\right) \to \Phi_{0,1}(y) - \Phi_{0,1}(x) = \int\limits_x^y \frac{1}{\sqrt{2\pi}}\, e^{-t^2/2}\, dt;$$

б) если η — *произвольная случайная величина со стандартным нормальным распределением, то*

$$\sqrt{n}\left(\frac{S_n}{n} - \mathsf{E}\,\xi_1\right) = \frac{S_n - n\,\mathsf{E}\,\xi_1}{\sqrt{n}} \Rightarrow \sqrt{\mathsf{D}\,\xi_1}\cdot\eta \in \mathrm{N}_{0,\mathsf{D}\,\xi_1}.$$

Замечание 28. Еще раз напомним, что функция распределения стандартного нормального закона ищется либо по соответствующей таблице в справочнике, либо с помощью какого-либо программного обеспечения, но никак не путем нахождения первообразной.

Мы докажем ЦПТ и ЗБЧ в форме Хинчина несколькими главами позднее. Нам потребуется для этого познакомиться с мощным математическим инструментом, который в математике обычно называют «преобразованиями Фурье», а в теории вероятностей — «характеристическими функциями».

§ 4. Предельная теорема Муавра — Лапласа

Получим в качестве следствия из ЦПТ предельную теорему Муавра[22] — Лапласа[23]. Подобно ЗБЧ Бернулли, теорема Муавра — Лапласа имеет дело со схемой Бернулли.

[22] Abraham de Moivre (26.05.1667 — 27.11.1754, France, England)
[23] Pierre-Simon Laplace (23.03.1749 — 5.03.1827, France)

Теорема 39 (предельная теорема Муавра — Лапласа). *Пусть событие A может произойти в любом из n независимых испытаний с одной и той же вероятностью p и пусть $\nu_n(A)$ — число осуществлений события A в n испытаниях. Тогда*

$$\frac{\nu_n(A) - np}{\sqrt{np(1-p)}} \Rightarrow \mathrm{N}_{0,1} \ \text{ при } \ n \to \infty,$$

т. е. для любых вещественных $x < y$ имеет место сходимость

$$\mathsf{P}\left(x \leqslant \frac{\nu_n(A) - np}{\sqrt{np(1-p)}} \leqslant y\right) \to \Phi_{0,1}(y) - \Phi_{0,1}(x) = \int\limits_x^y \frac{1}{\sqrt{2\pi}}\, e^{-t^2/2}\, dt;$$

Доказательство. Величина $\nu_n(A)$ есть сумма независимых, одинаково распределённых случайных величин, имеющих распределение Бернулли с параметром, равным вероятности успеха p: $\nu_n(A) = \xi_1 + \ldots + \xi_n$, где

$$\xi_i = I_i(A) = \begin{cases} 1, & \text{если } A \text{ произошло в } i\text{-м испытании;} \\ 0, & \text{если } A \text{ не произошло в } i\text{-м испытании;} \end{cases}$$

$$\mathsf{E}\,\xi_1 = p, \quad \mathsf{D}\,\xi_1 = p(1-p).$$

Осталось воспользоваться ЦПТ. □

§ 5. Примеры использования ЦПТ

Пример 63. Задача из примера 61. Монета подбрасывается 10 000 раз. Оценим вероятность того, что частота выпадения герба отличается от половины на одну сотую или больше.

Решение. Требуется найти $\mathsf{P}\left(\left|\frac{\nu_n}{n} - \frac{1}{2}\right| \geqslant 0{,}01\right)$, где $n = 10\,000$, $\nu_n = \xi_1 + \ldots + \xi_n = S_n$ — число выпадений герба, а ξ_i — независимые случайные величины, имеющие одно и то же распределение Бернулли с параметром $1/2$. Вычислим вероятность дополнительного события. Домножим обе части неравенства под знаком вероятности на $\sqrt{n} = 100$ и поделим на корень из дисперсии $\sqrt{\mathsf{D}\,\xi_1} = 1/2$.

$$\mathsf{P}\left(\left|\frac{\nu_n}{n} - \frac{1}{2}\right| < 0{,}01\right) = \mathsf{P}\left(\frac{\sqrt{n}}{\sqrt{\mathsf{D}\,\xi_1}}\left|\frac{S_n}{n} - \mathsf{E}\,\xi_1\right| < 0{,}01\frac{\sqrt{n}}{\sqrt{\mathsf{D}\,\xi_1}}\right) =$$

$$= \mathsf{P}\left(\frac{\sqrt{n}}{\sqrt{\mathsf{D}\,\xi_1}}\left|\frac{S_n}{n} - \mathsf{E}\,\xi_1\right| < 2\right) = \mathsf{P}\left(-2 < \frac{S_n - n\mathsf{E}\,\xi_1}{\sqrt{n\mathsf{D}\,\xi_1}} < 2\right) \approx$$

$$\approx \Phi_{0,1}(2) - \Phi_{0,1}(-2) = 1 - 2\Phi_{0,1}(-2) = 1 - 2 \cdot 0{,}0228 = 1 - 0{,}0456.$$

Искомая вероятность примерно равна 0,0456:

$$\mathsf{P}\left(\left|\frac{\nu_n}{n} - \frac{1}{2}\right| \geqslant 0{,}01\right) = 1 - \mathsf{P}\left(\left|\frac{\nu_n}{n} - \frac{1}{2}\right| < 0{,}01\right) \approx 0{,}0456.$$

Центральной предельной теоремой пользуются для приближённого вычисления вероятностей, связанных с суммами большого числа независимых и одинаково распределённых величин. При этом распределение центрированной и нормированной суммы заменяют на стандартное нормальное распределение. Насколько велика ошибка при такой замене (погрешность приближения)?

Упражнение. Какие ещё предельные теоремы для схемы Бернулли вы знаете? Что такое теорема Пуассона? Найти её. Какова погрешность пуассоновского приближения? Вычислить её. Объяснить, исходя из полученной величины, почему теорема Пуассона не применима в задаче из примера 63.

В примере 63 мы вычислили искомую вероятность тоже не точно, а приближённо — взгляните на равенство «≈» и спросите себя: насколько мы ошиблись? Стоит ли доверять ответу «0,0456»? Следующий результат позволяет оценить погрешность приближения в ЦПТ.

Теорема 40 (неравенство Берри — Эссéена). *В условиях ЦПТ для любого $x \in \mathbb{R}$ и для любого распределения ξ_1*

$$\left| \mathsf{P}\left(\frac{S_n - n\,\mathsf{E}\,\xi_1}{\sqrt{n\,\mathsf{D}\,\xi_1}} < x \right) - \Phi_{0,1}(x) \right| \leqslant C \cdot \frac{\mathsf{E}\,|\xi_1 - \mathsf{E}\,\xi_1|^3}{\sqrt{n}\left(\sqrt{\mathsf{D}\,\xi_1}\right)^3}.$$

Замечание 29. Про постоянную C известно, что:

а) в общем случае C не превышает 0,7655,

б) погрешность приближения наиболее велика, если слагаемые ξ_i имеют распределение Бернулли. Как показывают численные расчёты, даже в этом случае можно смело брать в качестве C число 0,4, особенно при малых n, когда и это значение постоянной даёт слишком грубую оценку.

Подробный обзор можно найти в монографии В.М.Золотарева «Современная теория суммирования независимых случайных величин», стр. 264–291.

Продолжение примера 63. Проверьте, что для с.в. ξ_1 с распределением Бернулли

$$\mathsf{E}\,|\xi_1 - \mathsf{E}\,\xi_1|^3 = |0-p|^3\,\mathsf{P}(\xi_1 = 0) + |1-p|^3\,\mathsf{P}(\xi_1 = 1) = pq(p^2+q^2).$$

Поэтому разница между левой и правой частями приближённого равенства «≈» в примере 63 при $n = 10^4$ и $p = q = 1/2$ не превышает величины

$$C \cdot \frac{pq(p^2+q^2)}{\sqrt{npq}\left(\sqrt{pq}\right)^3} = C \cdot \frac{p^2+q^2}{\sqrt{n}\,\sqrt{pq}} \leqslant 0{,}4 \cdot \frac{1}{100} = 0{,}004,$$

так что искомая вероятность $\mathsf{P}\left(\left|\frac{\nu_n}{n} - \frac{1}{2}\right| > 0{,}01\right)$ не больше, чем $0{,}0456 + {}$ $+0{,}004$. Уместно сравнить этот ответ с оценкой, полученной с помощью ЗБЧ в примере 61.

Следующая проблема связана с распространённейшим среди студентов заблуждением, которое выглядит так: при $n \to \infty$

$$\mathsf{P}\left(\frac{S_n}{n} < \mathsf{E}\,\xi_1\right) \to \mathsf{P}\left(\mathsf{E}\,\xi_1 < \mathsf{E}\,\xi_1\right) = 0,$$

но уже

$$\mathsf{P}\left(\frac{S_n}{n} \leqslant \mathsf{E}\,\xi_1\right) \to \mathsf{P}\left(\mathsf{E}\,\xi_1 \leqslant \mathsf{E}\,\xi_1\right) = 1.$$

Пример 64. Пусть $\xi_1, \xi_2, \ldots$ — независимые и одинаково распределённые случайные величины с конечной и ненулевой дисперсией $\sigma^2 = \mathsf{D}\,\xi_1$, $S_n = \xi_1 + \cdots + \xi_n$ — сумма первых n случайных величин. При каких c имеет или не имеет место сходимость

$$\mathsf{P}\left(\frac{S_n}{n} < c\right) \to \mathsf{P}\left(\mathsf{E}\,\xi_1 < c\right)?$$

Решение. Согласно ЗБЧ, последовательность S_n/n сходится по вероятности, а, следовательно, и слабо, к $\mathsf{E}\,\xi_1$.

Слабая сходимость означает, что последовательность функций распределения $F_n(c) = \mathsf{P}\left(S_n/n < c\right)$ сходится к функции распределения $F(c) = \mathsf{P}\left(\mathsf{E}\,\xi_1 < c\right)$, если $F(x)$ непрерывна в точке c (и ничего не означает, если $F(x)$ разрывна в точке c). Но

$$F(c) = \mathsf{P}\left(\mathsf{E}\,\xi_1 < c\right) = \begin{cases} 0, & c \leqslant \mathsf{E}\,\xi_1; \\ 1, & c > \mathsf{E}\,\xi_1 \end{cases}$$

есть функция распределения вырожденного закона и непрерывна в любой точке c, кроме $c = \mathsf{E}\,\xi_1$.

Итак, первый вывод: сходимость $\mathsf{P}\left(S_n/n < c\right) \to \mathsf{P}\left(\mathsf{E}\,\xi_1 < c\right)$ имеет место для любого c, кроме, возможно, $c = \mathsf{E}\,\xi_1$.

Убедимся, что для $c = \mathsf{E}\,\xi_1$ такой сходимости быть не может. Согласно ЦПТ,

$$\mathsf{P}\left(\frac{S_n}{n} < \mathsf{E}\,\xi_1\right) = \mathsf{P}\left(\frac{\sqrt{n}}{\sigma}\left(\frac{S_n}{n} - \mathsf{E}\,\xi_1\right) < 0\right) \to \Phi_{0,1}(0) = \frac{1}{2},$$

тогда как $\mathsf{P}\left(\mathsf{E}\,\xi_1 < \mathsf{E}\,\xi_1\right) = 0 \neq 1/2$. Аналогично, кстати, ведёт себя и вероятность $\mathsf{P}\left(S_n/n \leqslant \mathsf{E}\,\xi_1\right)$. Она тоже стремится к $1/2$, а не к $\mathsf{P}\left(\mathsf{E}\,\xi_1 \leqslant \mathsf{E}\,\xi_1\right) = 1$. Совершенно, кстати, понятно появление $1/2$ в пределе. Согласно ЦПТ, разность $S_n/n - \mathsf{E}\,\xi_1$ становится при больших n всё более и более симметрично распределённой относительно нуля.

И изящное упражнение на ту же тему.

Упражнение. Доказать, что

$$\lim_{n\to\infty} \int\limits_0^{0{,}999999n} \frac{1}{(n-1)!}\, y^{n-1}\, e^{-y}\, dy \;=\; 0;$$

$$\lim_{n\to\infty} \int\limits_0^{n} \frac{1}{(n-1)!}\, y^{n-1}\, e^{-y}\, dy \;=\; \frac{1}{2};$$

$$\lim_{n\to\infty} \int\limits_0^{1{,}000001n} \frac{1}{(n-1)!}\, y^{n-1}\, e^{-y}\, dy \;=\; 1.$$

Указание. Каждый из интегралов равен значению в некоторой точке функции распределения суммы независимых случайных величин с каким-то показательным распределением. Вспомнить, что такое гамма-распределение и «устойчивость относительно суммирования».

Пример 65. Пусть $\xi_1,\ \xi_2,\ \ldots$ — последовательность независимых и одинаково распределённых случайных величин с конечной дисперсией $\mathsf{D}\,\xi_1 = \sigma^2 > 0$. Докажем, что для любых вещественных чисел $a < b$

$$\lim_{n\to\infty} \mathsf{P}(a \leqslant S_n \leqslant b) = 0.$$

Преобразуем событие под знаком вероятности:

$$\mathsf{P}(a \leqslant S_n \leqslant b) = \mathsf{P}\Big(\frac{a - n\mathsf{E}\,\xi_1}{\sigma\sqrt{n}} \leqslant \frac{S_n - n\mathsf{E}\,\xi_1}{\sigma\sqrt{n}} \leqslant \frac{b - n\mathsf{E}\,\xi_1}{\sigma\sqrt{n}}\Big).$$

По ЦПТ последовательность $(S_n - n\mathsf{E}\,\xi_1)\,/\,(\sigma\sqrt{n})$ слабо сходится к стандартному нормальному распределению. Предельная функция распределения всюду непрерывна, и можно применять теорему 37.

Если $\mathsf{E}\,\xi_1 > 0$, то $x_n = \dfrac{a - n\mathsf{E}\,\xi_1}{\sigma\sqrt{n}}$ и $y_n = \dfrac{b - n\mathsf{E}\,\xi_1}{\sigma\sqrt{n}}$ стремятся к $-\infty$ с ростом n. Если $\mathsf{E}\,\xi_1 < 0$, то $x_n \to +\infty$ и $y_n \to +\infty$. Если же $\mathsf{E}\,\xi_1 = 0$, то $x_n \to 0$ и $y_n \to 0$ при $n \to \infty$. По теореме 37,

$$\mathsf{P}\Big(x_n \leqslant \frac{S_n - n\mathsf{E}\,\xi_1}{\sigma\sqrt{n}} \leqslant y_n\Big) \to \Phi_{0,\,1}\Big(\lim_{n\to\infty} y_n\Big) - \Phi_{0,\,1}\Big(\lim_{n\to\infty} x_n\Big) = 0,$$

поскольку при любой величине $\mathsf{E}\,\xi_1$ значения $\Phi_{0,\,1}(x)$ в точках $\lim x_n$ и $\lim y_n$ совпадают (либо оба равны нулю, либо оба равны единице, либо $1/2$).

Упражнение. Верно ли утверждение данного примера, если $\sigma^2 = 0$, т. е. если ξ_1 имеет вырожденное в точке c распределение? Рассмотрите отдельно случаи $c = 0$ и $c \neq 0$.

ГЛАВА 13

Характеристические функции

Я напрямик спросил, какую пользу можно извлечь от изучения его работ о покере. «Примерно такую же, как от чтения персидской поэзии»,— ответил фон Нейман.

Д. Мак-Дональд, Игра называется бизнес

В этой главе $i = \sqrt{-1}$ — мнимая единица, t — вещественная переменная, $e^{it} = \cos t + i \sin t$ — формула Эйлера, $\mathsf{E}\,(\eta + i\zeta) = \mathsf{E}\,\eta + i\,\mathsf{E}\,\zeta$ — способ вычисления математического ожидания комплекснозначной случайной величины $\eta + i\zeta$, если математические ожидания её действительной (η) и мнимой (ζ) частей существуют.

Как всегда, модулем комплексного числа $z = x + iy$ называется $|z| = \sqrt{x^2 + y^2}$, так что $\left|e^{it}\right| = 1$.

Определение 54. Функция $\varphi_\xi(t) = \mathsf{E}\,e^{it\xi}$ вещественной переменной t называется характеристической функцией случайной величины ξ.

§ 1. Примеры вычисления

Пример 66. Пусть случайная величина ξ имеет распределение Бернулли с параметром p. Её характеристическая функция равна

$$\varphi_\xi(t) = \mathsf{E}\,e^{it\xi} = e^{it\cdot 0}\,\mathsf{P}(\xi = 0) + e^{it\cdot 1}\,\mathsf{P}(\xi = 1) = 1 - p + pe^{it}.$$

Пример 67. Пусть случайная величина ξ имеет биномиальное распределение с параметрами n и p. Её характеристическая функция равна

$$\begin{aligned}\varphi_\xi(t) &= \mathsf{E}\,e^{it\xi} = \sum_{k=0}^{n} e^{it\cdot k}\,\mathsf{P}(\xi = k) = \sum_{k=0}^{n} e^{it\cdot k}\,C_n^k\,p^k\,(1-p)^{n-k} = \\ &= \sum_{k=0}^{n} C_n^k\left(pe^{it}\right)^k(1-p)^{n-k} = \left(1 - p + pe^{it}\right)^n.\end{aligned}$$

Последнее равенство есть бином Ньютона.

Пример 68. Пусть случайная величина ξ имеет распределение Пуассона с параметром λ. Её характеристическая функция равна

$$\begin{aligned}\varphi_\xi(t) &= \mathsf{E}\, e^{it\xi} = \sum_{k=0}^{\infty} e^{it\cdot k}\, \mathsf{P}(\xi = k) = \sum_{k=0}^{\infty} e^{it\cdot k}\, \frac{\lambda^k}{k!}\, e^{-\lambda} = \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{\left(\lambda e^{it}\right)^k}{k!} = e^{-\lambda} e^{\lambda e^{it}} = \exp\{\lambda \left(e^{it} - 1\right)\}.\end{aligned}$$

Пример 69. Пусть случайная величина ξ имеет показательное распределение с параметром $\alpha > 0$. Её характеристическая функция равна

$$\begin{aligned}\varphi_\xi(t) &= \mathsf{E}\, e^{it\xi} = \int_0^{\infty} e^{it\cdot x}\, f_\xi(x)\, dx = \int_0^{\infty} e^{itx}\, \alpha e^{-\alpha x}\, dx = \int_0^{\infty} \alpha e^{-x(\alpha - it)}\, dx = \\ &= \frac{\alpha}{\alpha - it} \left(-e^{-x(\alpha-it)} \Big|_0^{\infty} \right) = \frac{\alpha}{\alpha - it},\end{aligned}$$

поскольку при $x \to +\infty$ модуль величины $e^{-x(\alpha-it)} = e^{-\alpha x} e^{itx}$ стремится к нулю: $\left|e^{-x(\alpha-it)}\right| = e^{-\alpha x} \to 0$.

Пример 70. Пусть случайная величина ξ имеет гамма-распределение с параметрами α и λ. Её характеристическая функция равна

$$\begin{aligned}\varphi_\xi(t) &= \mathsf{E}\, e^{it\xi} = \int_0^{\infty} e^{it\cdot x}\, f_\xi(x)\, dx = \int_0^{\infty} e^{itx}\, \frac{\alpha^\lambda}{\Gamma(\lambda)}\, x^{\lambda-1}\, e^{-\alpha x}\, dx = \\ &= \frac{\alpha^\lambda}{\Gamma(\lambda)} \int_0^{\infty} x^{\lambda-1}\, e^{-x(\alpha-it)}\, dx = \left(\frac{\alpha}{\alpha - it}\right)^{\lambda} = \left(1 - \frac{it}{\alpha}\right)^{-\lambda}.\end{aligned}$$

Интеграл мы вычислили с помощью гамма-функции: замена $y = x(\alpha - it)$ даёт

$$\int_0^{\infty} x^{\lambda-1}\, e^{-x(\alpha-it)}\, dx = \frac{1}{(\alpha - it)^\lambda} \int_0^{\infty} y^{\lambda-1}\, e^{-y}\, dy = \frac{\Gamma(\lambda)}{(\alpha - it)^\lambda}.$$

Пример 71. Пусть случайная величина ξ имеет стандартное нормальное распределение. Её характеристическая функция равна

$$\begin{aligned}\varphi_\xi(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx}\, e^{-x^2/2}\, dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-t^2/2}\, e^{-(x-it)^2/2}\, dx = \\ &= e^{-t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-it)^2/2}\, d(x - it) \;=\; e^{-t^2/2}.\end{aligned}$$

При интегрировании мы выделили полный квадрат в показателе экспоненты и вспомнили, чему равен интеграл по $\mathbb{R}$ от функции $\dfrac{1}{\sqrt{2\pi}}\,e^{-u^2/2}$ (а чему он равен?).

Самое время остановиться и спросить: «Ну и что? Зачем нам эти функции и какой от них прок?» Приглашаю читателя познакомиться с замечательными свойствами характеристических функций.

§ 2. Свойства характеристических функций

(Ф1). Характеристическая функция всегда существует:

$$|\varphi_\xi(t)| = \left|\mathsf{E}\,e^{it\xi}\right| \leqslant 1.$$

Доказательство. Воспользуемся свойством $\mathsf{D}\eta \geqslant 0$, равносильным неравенству $\big(\mathsf{E}\,\eta\big)^2 \leqslant \mathsf{E}\,\eta^2$:

$$\begin{aligned}|\varphi_\xi(t)|^2 &= \left|\mathsf{E}\,\cos(t\xi) + i\mathsf{E}\,\sin(t\xi)\right|^2 = \big(\mathsf{E}\,\cos(t\xi)\big)^2 + \big(\mathsf{E}\,\sin(t\xi)\big)^2 \leqslant \\ &\leqslant \mathsf{E}\,\cos^2(t\xi) + \mathsf{E}\,\sin^2(t\xi) = \mathsf{E}\,\big(\cos^2(t\xi) + \sin^2(t\xi)\big) = \mathsf{E}\,1 = 1. \quad \square\end{aligned}$$

(Ф2). По характеристической функции однозначно восстанавливается распределение (функция распределения, плотность или таблица распределения). Т. е. если две случайные величины имеют одинаковые характеристические функции, то и распределения этих величин совпадают.

Формулы, с помощью которых по характеристической функции восстанавливается распределение, в анализе называют формулами «обратного преобразования Фурье». Например, если модуль характеристической функции интегрируем на всей прямой, то у случайной величины есть плотность распределения, и она находится по формуле

$$f_\xi(x) = \frac{1}{2\pi}\int\limits_{-\infty}^{\infty} e^{-itx}\,\varphi_\xi(t)\,dt.$$

Ни одна из формул обратного преобразования Фурье нам не понадобится.

(Ф3). Характеристическая функция случайной величины $a + b\xi$ связана с характеристической функцией случайной величины ξ равенством:

$$\varphi_{a+b\xi}(t) = \mathsf{E}\,e^{it(a+b\xi)} = e^{ita}\mathsf{E}\,e^{i(tb)\xi} = e^{ita}\,\varphi_\xi(tb).$$

Пример 72. Вычислим характеристическую функцию случайной величины ξ, имеющей нормальное распределение с параметрами a и σ^2. Мы знаем, что у «стандартизованной» случайной величины $\eta = (\xi - a)/\sigma$ характеристическая функция равна $\varphi_\eta(t) = e^{-t^2/2}$. Тогда характеристическая функция величины $\xi = a + \sigma\eta$ равна

$$\varphi_\xi(t) = \varphi_{a+\sigma\eta}(t) = e^{ita}\,\varphi_\eta(t\sigma) = e^{ita}\,e^{-(t\sigma)^2/2}.$$

(Ф4). Характеристическая функция суммы независимых случайных величин равна произведению характеристических функций слагаемых: если случайные величины ξ и η независимы, то, по свойству (E7) математических ожиданий,

$$\varphi_{\xi+\eta}(t) = \mathsf{E}\, e^{it(\xi+\eta)} = \mathsf{E}\, e^{it\xi}\, \mathsf{E}\, e^{it\eta} = \varphi_\xi(t)\, \varphi_\eta(t).$$

Замечание 30. Чтобы характеристическая функция суммы n случайных величин распадалась в произведение их характеристических функций, попарной независимости слагаемых не хватит. То же самое можно сказать про свойство (E7) математических ожиданий.

Замечательным свойством (Ф4) мы сразу же воспользуемся, как обещали, для доказательства леммы 5, утверждающей устойчивость нормального распределения относительно суммирования.

Доказательство леммы 5. Пусть $\xi \in \mathrm{N}_{a_1,\sigma_1^2}$ и $\eta \in \mathrm{N}_{a_2,\sigma_2^2}$ независимы. Характеристическая функция суммы $\xi+\eta$ равна

$$\varphi_{\xi+\eta}(t) = \varphi_\xi(t)\, \varphi_\eta(t) = e^{ita_1} e^{-t^2\sigma_1^2/2} e^{ita_2} e^{-t^2\sigma_2^2/2} = e^{it(a_1+a_2)} e^{-t^2(\sigma_1^2+\sigma_2^2)/2}.$$

Видим, что характеристическая функция суммы есть характеристическая функция нормального распределения с параметрами a_1+a_2 и $\sigma_1^2+\sigma_2^2$. Следовательно, $\xi+\eta \in \mathrm{N}_{a_1+a_2,\,\sigma_1^2+\sigma_2^2}$ по свойству (Ф2). $\square$

Доказательство лемм 3, 4, 7. Докажем свойства устойчивости по суммированию биномиального распределения, распределения Пуассона и гамма-распределения, используя характеристические функции из примеров 66— 70.

Для независимых случайных величин с распределениями Пуассона Π_λ и Π_μ характеристическая функция суммы

$$\varphi_{\xi+\eta}(t) = \exp\left\{\lambda\left(e^{it}-1\right)\right\} \exp\left\{\mu\left(e^{it}-1\right)\right\} = \exp\left\{(\lambda+\mu)\left(e^{it}-1\right)\right\}$$

равна характеристической функции распределения Пуассона $\Pi_{\lambda+\mu}$.

Для независимых случайных величин с биномиальными распределениями $\mathrm{B}_{n,p}$ и $\mathrm{B}_{m,p}$ характеристическая функция суммы

$$\varphi_{\xi+\eta}(t) = \varphi_\xi(t)\, \varphi_\eta(t) = \left(1-p+pe^{it}\right)^n \left(1-p+pe^{it}\right)^m = \left(1-p+pe^{it}\right)^{n+m}$$

равна характеристической функции биномиального распределения с параметрами $n+m$ и p.

Для n независимых (в совокупности) случайных величин с показательным распределением E_α характеристическая функция суммы

$$\varphi_{\xi_1+\ldots+\xi_n}(t) = \left(\varphi_{\xi_1}(t)\right)^n = \left(\frac{\alpha}{\alpha-it}\right)^n = \left(1-\frac{it}{\alpha}\right)^{-n}$$

равна характеристической функции гамма-распределения $\Gamma_{\alpha,\,n}$. $\square$

(Ф5.) Пусть существует момент порядка $k \in \mathbb{N}$ случайной величины ξ, т. е. $\mathsf{E}\,|\xi|^k < \infty$. Тогда характеристическая функция $\varphi_\xi(t)$ непрерывно дифференцируема k раз, и её k-я производная в нуле связана с моментом порядка k равенством:

$$\varphi_\xi^{(k)}(0) = \left(\frac{d^k}{d\,t^k}\,\mathsf{E}\,e^{it\xi}\right)\Bigg|_{t=0} = \left(\mathsf{E}\,i^k\,\xi^k\,e^{it\xi}\right)\Bigg|_{t=0} = i^k\,\mathsf{E}\,\xi^k.$$

Существование и непрерывность k-й производной, равно как и законность переноса производной под знак математического ожидания мы доказывать не будем.

Упражнение. Доказать, что для случайной величины ξ со стандартным нормальным распределением момент чётного порядка $2k$ равен

$$\mathsf{E}\,\xi^{2k} = (2k-1)!! \;=\; (2k-1)\cdot(2k-3)\cdot\ldots\cdot 3\cdot 1.$$

Доказать по определению, что все моменты нечётных порядков стандартного нормального распределения существуют и равны нулю.

Как только появились производные высших порядков, самое время разложить функцию в ряд Тейлора[24].

(Ф6). Пусть существует момент порядка $k \in \mathbb{N}$ случайной величины ξ, т. е. $\mathsf{E}\,|\xi|^k < \infty$. Тогда характеристическая функция $\varphi_\xi(t)$ в окрестности точки $t = 0$ разлагается в ряд Тейлора

$$\begin{aligned}\varphi_\xi(t) \;&=\; \varphi_\xi(0) + \sum_{j=1}^{k}\frac{t^j}{j!}\,\varphi_\xi^{(j)}(0) + o(|t^k|) = 1 + \sum_{j=1}^{k}\frac{i^j t^j}{j!}\,\mathsf{E}\,\xi^j + o(|t^k|) = \\ &=\; 1 + it\,\mathsf{E}\,\xi - \frac{t^2}{2}\,\mathsf{E}\,\xi^2 + \ldots + \frac{i^k t^k}{k!}\,\mathsf{E}\,\xi^k + o(|t^k|).\end{aligned}$$

Ряды Тейлора бывают особенно полезны в теории пределов. Следующее основное свойство характеристических функций потребуется нам для доказательства предельных теорем, и это свойство — последняя теорема, оставленная нами без доказательства.

Теорема 41 (теорема о непрерывном соответствии[25]). *Случайные величины ξ_n слабо сходятся к случайной величине ξ тогда и только тогда, когда для любого t характеристические функции $\varphi_{\xi_n}(t)$ сходятся к характеристической функции $\varphi_\xi(t)$.*

Сформулированная теорема устанавливает непрерывное соответствие между классами $\langle F_\xi, \Rightarrow\rangle$ функций распределения со слабой сходимостью и $\langle \varphi_\xi, \to\rangle$ характеристических функций со сходимостью в каждой точке.

[24] Brook Taylor (18.08.1685 — 29.12.1731, England)
[25] Paul Pierre Lévy (15.09.1886 — 15.12.1971, France)

«Непрерывность» этого соответствия — в том, что пределу в одном классе относительно заданной в этом классе сходимости соответствует предел в другом классе относительно сходимости, заданной в этом другом классе.

Осталось воспользоваться теоремой о непрерывном соответствии и доказать ЗБЧ в форме Хинчина и ЦПТ.

§ 3. Доказательство ЗБЧ Хинчина

Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых в совокупности и одинаково распределённых случайных величин с конечным первым моментом $\mathsf{E}\,|\xi_1| < \infty$. Обозначим через a математическое ожидание $\mathsf{E}\,\xi_1$. Требуется доказать, что

$$\frac{S_n}{n} = \frac{\xi_1 + \cdots + \xi_n}{n} \xrightarrow{\text{p}} a.$$

По свойству 20 сходимость по вероятности к постоянной эквивалентна слабой сходимости. Так как a — постоянная, достаточно доказать слабую сходимость S_n/n к a. По теореме о непрерывном соответствии, эта сходимость имеет место тогда и только тогда, когда для любого $t \in \mathbb{R}$ сходятся характеристические функции

$$\varphi_{S_n/n}(t) \to \varphi_a(t) = \mathsf{E}\,e^{ita} = e^{ita}.$$

Найдём характеристическую функцию случайной величины S_n/n. Пользуясь свойствами (Ф3) и (Ф4), получим

$$\varphi_{S_n/n}(t) = \varphi_{S_n}\left(\frac{t}{n}\right) = \left(\varphi_{\xi_1}\left(\frac{t}{n}\right)\right)^n.$$

Вспомним, что первый момент ξ_1 существует, поэтому свойство (Ф6) позволяет разложить $\varphi_{\xi_1}(t)$ в ряд Тейлора в окрестности нуля:

$$\varphi_{\xi_1}(t) \ = \ 1 + it\,\mathsf{E}\,\xi_1 + o(|t|) = 1 + ita + o(|t|).$$

В точке t/n, соответственно,

$$\varphi_{\xi_1}\left(\frac{t}{n}\right) \ = \ 1 + \frac{ita}{n} + o\left(\left|\frac{t}{n}\right|\right),$$

$$\varphi_{S_n/n}(t) \ = \ \left(\varphi_{\xi_1}\left(\frac{t}{n}\right)\right)^n = \left(1 + \frac{ita}{n} + o\left(\left|\frac{t}{n}\right|\right)\right)^n.$$

При $n \to \infty$, пользуясь «замечательным пределом» $\left(1 + \dfrac{x}{n}\right)^n \to e^x$, получим

$$\varphi_{S_n/n}(t) \ = \ \left(1 + \frac{ita}{n} + o\left(\left|\frac{t}{n}\right|\right)\right)^n \ \to \ e^{ita},$$

что и требовалось доказать. □

§ 4. Доказательство центральной предельной теоремы

Пусть $\xi_1, \xi_2, \ldots$ — последовательность независимых в совокупности и одинаково распределённых случайных величин с конечной и ненулевой дисперсией. Обозначим через a математическое ожидание $\mathsf{E}\,\xi_1$ и через σ^2 — дисперсию $\mathsf{D}\,\xi_1$. Требуется доказать, что

$$\frac{S_n - na}{\sigma\sqrt{n}} = \frac{\xi_1 + \ldots + \xi_n - na}{\sigma\sqrt{n}} \Rightarrow \mathrm{N}_{0,1}.$$

Введём «стандартизованные» случайные величины $\zeta_i = (\xi_i - a)/\sigma$ — независимые случайные величины с нулевыми математическими ожиданиями и единичными дисперсиями (проверить). Пусть Z_n есть их сумма $Z_n = \zeta_1 + \cdots + \zeta_n = (S_n - na)/\sigma$. Требуется доказать, что $Z_n/\sqrt{n} \Rightarrow \mathrm{N}_{0,1}$.

Характеристическая функция величины $Z_n/\sqrt{n}$ равна

$$\varphi_{Z_n/\sqrt{n}}(t) = \varphi_{Z_n}\left(\frac{t}{\sqrt{n}}\right) = \left(\varphi_{\zeta_1}\left(\frac{t}{\sqrt{n}}\right)\right)^n. \qquad (28)$$

Характеристическую функцию случайной величины ζ_1 можно разложить в ряд Тейлора, в коэффициентах которого использовать известные моменты $\mathsf{E}\,\zeta_1 = 0,\ \mathsf{E}\,\zeta_1^2 = \mathsf{D}\,\zeta_1 = 1$. Получим

$$\varphi_{\zeta_1}(t) = 1 + it\,\mathsf{E}\,\zeta_1 - \frac{t^2}{2}\,\mathsf{E}\,\zeta_1^2 + o(t^2) = 1 - \frac{t^2}{2} + o(t^2).$$

Подставим это разложение, взятое в точке $t/\sqrt{n}$, в равенство (28) и устремим n к бесконечности. Ещё раз воспользуемся замечательным пределом.

$$\varphi_{Z_n/\sqrt{n}}(t) = \left(\varphi_{\zeta_1}\left(\frac{t}{\sqrt{n}}\right)\right)^n = \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n \to e^{-t^2/2} \text{ при } n \to \infty.$$

В пределе получили характеристическую функцию стандартного нормального распределения. По теореме о непрерывном соответствии можно сделать вывод о слабой сходимости

$$\frac{Z_n}{\sqrt{n}} = \frac{S_n - na}{\sigma\sqrt{n}} \Rightarrow \mathrm{N}_{0,1}. \qquad \square$$

Попробуйте теперь сами:

Упражнение. Пусть при любом $\lambda > 0$ случайная величина ξ_λ имеет распределение Пуассона с параметром λ. Используя теорему о непрерывном соответствии, доказать, что случайные величины $(\xi_\lambda - \lambda)/\sqrt{\lambda}$ слабо сходятся к стандартному нормальному распределению при $\lambda \to \infty$. Характеристическая функция случайной величины ξ_λ вычислена в примере 68.

Приложение

В этом разделе, который никогда не будет прочитан на лекциях, поскольку эта тема подробно разбирается на практических занятиях, мы поговорим о максимуме и минимуме из n случайных величин. Вдумчивый читатель уже догадался, что ничего общего с клубом «Максимин» ЭФ НГУ эта тема не имеет. Нам необходимо уметь обращаться с минимумом и максимумом из нескольких случайных величин хотя бы потому, что при изучении математической статистики мы не раз о них вспомним.

Пусть случайные величины $\xi_1, \xi_2, \ldots$ независимы в совокупности и одинаково распределены, $F_{\xi_1}(x)$ — их общая функция распределения.

Определение № N. Случайную величину $\varphi_n = \max\{\xi_1, \ldots, \xi_n\}$ назовём максимумом, а случайную величину $\psi_n = \min\{\xi_1, \ldots, \xi_n\}$ — минимумом из n случайных величин $\xi_1, \ldots, \xi_n$.

Замечание № N. Заметим, что $\varphi_n(\omega) = \max\{\xi_1(\omega), \ldots, \xi_n(\omega)\}$, т. е. φ_n на каждом элементарном исходе совпадает с одной из ξ_i, $1 \leqslant i \leqslant n$, но ни с одной из них не совпадает при всех ω (если величины независимы).

Упражнение № N. Доказать, что вероятность максимуму из первых n независимых и одинаково распределённых случайных величин, имеющих абсолютно непрерывное распределение, равняться первой из них, равно как и любой другой, есть $1/n$:

$$\mathsf{P}(\max\{\xi_1, \ldots, \xi_n\} = \xi_1) = \frac{1}{n} = \mathsf{P}(\xi_1 > \xi_2, \ldots, \xi_1 > \xi_n).$$

Для доказательства воспользоваться соображениями симметрии, разбив пространство Ω на несколько равновероятных событий вида $\{\xi_1 > \xi_2, \ldots, \xi_1 > \xi_n\}$ и несколько событий нулевой вероятности, включающих возможные равенства. Вспомнить, с какой вероятностью две (или больше) из $\xi_1, \ldots, \xi_n$ совпадают (нарисовать событие $\{\xi_1 = \xi_2\}$ на плоскости).

Теорема № N. *Функции распределения случайных величин* $\varphi_n = \max\{\xi_1, \ldots, \xi_n\}$ *и* $\psi_n = \min\{\xi_1, \ldots, \xi_n\}$ *равны соответственно*

$$F_{\varphi_n}(x) = \bigl(F_{\xi_1}(x)\bigr)^n \quad \textit{и} \quad F_{\psi_n}(x) = 1 - \bigl(1 - F_{\xi_1}(x)\bigr)^n.$$

Доказательство. Найдём функцию распределения $F_{\varphi_n}(x)$. Максимум из n величин меньше x тогда и только тогда, когда каждая из этих величин меньше x. Поэтому событие $\{\varphi_n < x\}$ равносильно пересечению n независимых событий $\{\xi_1 < x\}, \ldots, \{\xi_n < x\}$, имеющих одну и ту же вероятность $F_{\xi_1}(x)$:

$$\begin{aligned} F_{\varphi_n}(x) &= \mathsf{P}\big(\max\{\xi_1, \ldots, \xi_n\} < x\big) = \mathsf{P}\big(\xi_1 < x, \ldots, \xi_n < x\big) = \\ &= \mathsf{P}\big(\xi_1 < x\big) \cdot \ldots \cdot \mathsf{P}\big(\xi_n < x\big) = \Big(\mathsf{P}\big(\xi_1 < x\big)\Big)^n = \big(F_{\xi_1}(x)\big)^n. \end{aligned}$$

Найдём функцию распределения $F_{\psi_n}(x)$. Минимум из n величин не меньше x тогда и только тогда, когда каждая из этих величин не меньше x:

$$\begin{aligned} F_{\psi_n}(x) &= \mathsf{P}\big(\min\{\xi_1, \ldots, \xi_n\} < x\big) = 1 - \mathsf{P}\big(\min\{\xi_1, \ldots, \xi_n\} \geqslant x\big) = \\ &= 1 - \mathsf{P}\big(\xi_1 \geqslant x, \ldots, \xi_n \geqslant x\big) = 1 - \mathsf{P}\big(\xi_1 \geqslant x\big) \cdot \ldots \cdot \mathsf{P}\big(\xi_n \geqslant x\big) = \\ &= 1 - \Big(\mathsf{P}\big(\xi_1 \geqslant x\big)\Big)^n = 1 - \big(1 - F_{\xi_1}(x)\big)^n. \qquad \square \end{aligned}$$

Пример № N. Пусть случайные величины $\xi_1, \xi_2, \ldots$ независимы в совокупности и имеют равномерное распределение на отрезке $[0, 1]$. Докажем, что последовательность случайных величин $\varphi_1 = \xi_1$, $\varphi_2 = \max\{\xi_1, \xi_2\}$, $\varphi_3 = \max\{\xi_1, \xi_2, \xi_3\}, \ldots$ сходится по вероятности к правому концу отрезка — к единице. Можно произнести это утверждение так: «максимум из первых n случайных величин с ростом n сходится к единице по вероятности». Есть как минимум два способа доказательства.

Способ 1. По определению. Зафиксируем произвольное число $\varepsilon > 0$. Заметим, что $\varphi_n \leqslant 1$, поскольку это максимум из случайных величин, принимающих значения на отрезке $[0, 1]$ п. н. Поэтому

$$\mathsf{P}\big(|\varphi_n - 1| \geqslant \varepsilon\big) = \mathsf{P}\big(1 - \varphi_n \geqslant \varepsilon\big).$$

Для того, чтобы установить сходимость последней вероятности к нулю, можно её либо найти, либо оценить с помощью неравенства Маркова. Сделаем и то, и другое.

(A) Найдём эту вероятность.

$$\mathsf{P}\big(1 - \varphi_n \geqslant \varepsilon\big) = \mathsf{P}\big(\varphi_n \leqslant 1 - \varepsilon\big) = F_{\varphi_n}(1 - \varepsilon).$$

Для равномерного распределения на отрезке $[0, 1]$

$$F_{\xi_1}(x) = \begin{cases} 0, & x < 0, \\ x, & 0 \leqslant x \leqslant 1, \\ 1, & x > 1; \end{cases} \qquad F_{\varphi_n}(x) = \big(F_{\xi_1}(x)\big)^n = \begin{cases} 0, & x < 0, \\ x^n, & 0 \leqslant x \leqslant 1, \\ 1, & x > 1. \end{cases}$$

А если ещё заметить, что $1 - \varepsilon < 1$, то

$$F_{\varphi_n}(1 - \varepsilon) = \begin{cases} 0, & 1 - \varepsilon < 0, \text{ т. е. } \varepsilon > 1, \\ (1 - \varepsilon)^n, & 0 \leqslant 1 - \varepsilon < 1, \text{ т. е. } 0 < \varepsilon \leqslant 1. \end{cases}$$

Видно, что $\mathsf{P}\big(|\varphi_n - 1| \geqslant \varepsilon\big) = F_{\varphi_n}(1 - \varepsilon) \to 0$ при $n \to \infty$.

(Б) Оценим вероятность сверху. Поскольку $1 - \varphi_n \geqslant 0$ п. н. и $\varepsilon > 0$, то по неравенству Маркова

$$\mathsf{P}\big(1 - \varphi_n \geqslant \varepsilon\big) \leqslant \frac{\mathsf{E}\,(1-\varphi_n)}{\varepsilon} = \frac{1 - \mathsf{E}\,\varphi_n}{\varepsilon}. \tag{29}$$

Найдём плотность распределения φ_n и математическое ожидание $\mathsf{E}\,\varphi_n$:

$$f_{\varphi_n}(x) = \big(F_{\varphi_n}(x)\big)' = \begin{cases} 0, & x < 0; \\ nx^{n-1}, & 0 \leqslant x \leqslant 1 \\ 0, & x > 1; \end{cases}$$

$$\mathsf{E}\,\varphi_n = \int\limits_0^1 x\, nx^{n-1}\,dx = \frac{n}{n+1}.$$

Подставляя математическое ожидание в неравенство (29), получим

$$\mathsf{P}\big(1 - \varphi_n \geqslant \varepsilon\big) \leqslant \frac{1 - \mathsf{E}\,\varphi_n}{\varepsilon} = \frac{1 - \dfrac{n}{n+1}}{\varepsilon} = \frac{1}{(n+1)\,\varepsilon} \to 0 \text{ при } n \to \infty.$$

Способ 2. Используем связь со слабой сходимостью. Сходимость по вероятности к постоянной равносильна слабой сходимости (свойство 20). Докажем, что φ_n слабо сходится к единице. Требуется доказать, что функция распределения $F_{\varphi_n}(x)$ сходится к $F_1(x) = \mathsf{P}(1 < x)$ для любого $x \neq 1$ (почему кроме 1?).

При любом $x < 0$ имеем: $F_{\varphi_n}(x) = 0 \to F_1(x) = 0$ при $n \to \infty$. При любом $0 \leqslant x < 1$ имеем: $F_{\varphi_n}(x) = x^n \to F_1(x) = 0$ при $n \to \infty$. При любом $x > 1$ имеем: $F_{\varphi_n}(x) = 1 \to F_1(x) = 1$. И только при $x = 1$ сходимости нет: $F_{\varphi_n}(1) = 1$, тогда как $F_1(1) = 0$. Но сходимости в точке $x = 1$ и не требуется — в этой точке предельная функция распределения терпит разрыв:

$$F_1(x) = \begin{cases} 0, & x \leqslant 1; \\ 1, & x > 1; \end{cases}$$

Таким образом, φ_n слабо сходится к единице, и, следовательно, сходится к ней же по вероятности.

Упражнение № N+1. Доказать (способами (1А), (1Б) и (2), что, в условиях примера (N), последовательность $\psi_1, \psi_2, \psi_3, \ldots$ сходится по вероятности к нулю (мы будем говорить «минимум из первых n случайных величин с ростом n сходится к нулю по вероятности»).

Красивых задач, связанных с максимумом и минимумом, слишком много. Предлагаю вам решить, например, следующие:

Пусть случайные величины $\xi_1, \xi_2, \ldots$ независимы в совокупности и имеют равномерное распределение на отрезке $[a, b]$, $\varphi_n = \max\{\xi_1, \ldots, \xi_n\}$, $\psi_n = \min\{\xi_1, \ldots, \xi_n\}$. Доказать, что:

1) последовательность $n(b - \varphi_n)/(b - a)$ при $n \to \infty$ слабо сходится к показательному распределению с параметром 1;

2) точно так же себя ведёт последовательность $n(\psi_n - a)/(b - a)$;

3) это не удивительно, поскольку случайные величины $b - \varphi_n$ и $\psi_n - a$ одинаково распределены;

4) посчитав вероятность $\mathsf{P}(\psi_n \geqslant x,\ \varphi_n < y) = \big(\mathsf{P}(x \leqslant \xi_1 < y)\big)^n$, можно легко найти функцию совместного распределения случайных величин ψ_n, φ_n, и с её помощью, например, доказать зависимость этих величин. Зависимость, впрочем, и так очевидна, достаточно рассмотреть пустое пересечение двух событий $\{\varphi_n < (a + b)/2\}$ и $\{\psi_n > (a + b)/2\}$, вероятность которых положительна.

Простые и непростые задачи

1. Построить какую-нибудь вероятность на множестве натуральных чисел как на дискретном пространстве элементарных исходов.

2. Вероятность события A равна нулю. Верно ли, что тогда A — невозможное событие?

3. Являются ли события «выпал герб при первом броске монеты» и «выпал герб при втором броске монеты» несовместными?

4. Пусть $\Omega = \{1, 2, 3, 4\}$. Является ли алгеброй набор множеств $\{\varnothing, \{1, 2, 3, 4\}, \{1, 2\}, \{2, 3, 4\}, \{2\}\}$?

5. Задать какую-нибудь алгебру на множестве $\Omega = \{0, 1, \ldots, 10\}$.

6. Пусть $\Omega = \{1, 2\}$. На множестве 2^Ω задана функция: $\mu\{1, 2\} = 3$, $\mu\{1\} = 1$, $\mu\{2\} = 1$. Является ли μ мерой?

7. Задать какую-нибудь вероятностную меру на множестве всех подмножеств множества $\Omega = \{0, 1, \ldots, 10\}$.

8. Является ли σ-алгеброй декартово произведение $\mathfrak{B}(\mathbb{R}) \times \mathfrak{B}(\mathbb{R})$ двух борелевских σ-алгебр?

9. Принадлежит ли множество $\mathbb{Q}$ рациональных чисел σ-алгебре, порождённой множеством всех одноточечных подмножеств $\mathbb{R}$?

10. На борелевской σ-алгебре в $\mathbb{R}$ задана функция: $\mu(B) = 1$ для любого B. Является ли μ вероятностной мерой?

11. Пусть функция μ на множестве $2^\mathbb{R}$ задана так: $\mu(B) = 1$, если $0 \in B$, и $\mu(B) = 0$, если $0 \notin B$. Является ли функция μ мерой на множестве всех подмножеств $\mathbb{R}$?

12. Привести пример какой-нибудь меры, отличной от меры Лебега, на σ-алгебре $\mathfrak{B}(\mathbb{R})$ борелевских множеств на прямой.

См. определения 28, 29, 30 и т.д.

13. Доказать, что если $\mathsf{P}(A_i) = 0$ для всех $i = 1, 2, \ldots$, то

$$\mathsf{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = 0 \text{ и } \mathsf{P}\left(\bigcap_{i=1}^{\infty} A_i\right) = 0.$$

См. свойства вероятности 4 (стр. 30) и 7 (стр. 31).

14. Доказать, что если $\mathsf{P}(A_i) = 1$ для всех $i = 1, 2, \ldots$, то

$$\mathsf{P}\Bigl(\bigcup_{i=1}^{\infty} A_i\Bigr) = 1 \text{ и } \mathsf{P}\Bigl(\bigcap_{i=1}^{\infty} A_i\Bigr) = 1.$$

См. свойства вероятности 4 (стр. 30) и 7 (стр. 31).

15. Найти количество всех 2005-значных чисел, в записи которых используются все цифры от 1 до 9, но никакие соседние цифры в записи этих чисел не совпадают.

См. формулу включения-исключения (стр. 31).

16. Найти вероятность того, что при раздаче колоды в 52 карты четверым игрокам поровну хотя бы у одного из игроков соберутся все карты одной масти.

См. формулу включения-исключения (стр. 31).

17. Доказать, что $\{4\}$ является борелевским множеством.

18. Доказать, что $[1, 2]$ является борелевским множеством.

19. Доказать, что канторовское совершенное множество является борелевским множеством.

См. определения 13 и 11.

20. Могут ли два независимых события образовать полную группу событий?

21. Из полной колоды карт вынимают одну. Будут ли независимыми события «вынутая карта — король» и «вынутая карта — туз»?

22. Что означает независимость в совокупности событий A, B, C и D?

23. Следует ли из равенства

$$\mathsf{P}(A \cap B \cap C) = \mathsf{P}(A)\mathsf{P}(B)\mathsf{P}(C)$$

независимость событий A, B и C в совокупности?

24. Следует ли из того же равенства попарная независимость A, B, C?

25. Пусть $\langle \Omega, \mathcal{F}, \mathsf{P} \rangle = \langle [0, 1], \mathfrak{B}([0, 1]), \lambda \rangle$, где λ — мера Лебега. Построить на этом вероятностном пространстве три случайных величины с равномерными на $[0, 1]$ распределениями.

26. Пусть $\Omega = [0, 1]$, $\mathcal{F} = 2^{\Omega}$, $\xi(\omega) = \omega$. Задать вероятностную меру P так, чтобы функция ξ оказалась случайной величиной с вырожденным распределением.

27. Пусть $\Omega = [0, 1]$, $\mathcal{F} = 2^{\Omega}$, $\xi(\omega) = \omega$. Задать вероятностную меру P так, чтобы функция ξ оказалась случайной величиной с распределением Бернулли.

28. Привести пример вероятностного пространства и на нем трёх независимых попарно, но зависимых в совокупности случайных величин.

29. Доказать, что случайная величина с вырожденным распределением независима с любой другой случайной величиной.

30. Пусть случайная величина ξ имеет стандартное нормальное распределение, а случайная величина η равна знаку ξ. Проверить, независимы ли случайные величины $|\xi|$ и η.

31. Доказать, что из независимости в совокупности n случайных величин следует их попарная независимость.

32. Доказать, что случайные величины ξ и η независимы тогда и только тогда, когда для любых двух борелевских функций f и g независимы случайные величины $f(\xi)$ и $g(\eta)$.

33. Может ли какое-нибудь из стандартных дискретных распределений быть устойчивым относительно: а) умножения на некоторую постоянную? б) умножения на произвольную постоянную? в) произвольного линейного преобразования?

Здесь устойчивость — сохранение того же вида распределения.

34. Какие из стандартных абсолютно-непрерывных распределений устойчивы относительно: а) умножения на некоторую постоянную? б) умножения на произвольную постоянную? в) произвольного линейного преобразования?

35. Привести пример случайных величин ξ и η с одинаковым нормальным распределением, сумма которых имеет вырожденное распределение.

36. Привести пример случайных величин ξ и η с пуассоновскими распределениями, сумма которых имеет распределение, отличное от пуассоновского.

37. Привести пример случайных величин ξ и η с пуассоновскими распределениями с параметрами $\lambda \neq \mu$, сумма которых имеет распределение, отличное от пуассоновского.

См. пример 55.

38. Пусть $\xi \in N_{1,9}$ и $\eta \in N_{1,1}$ — независимые случайные величины. Какое распределение имеет $\xi - \eta$?

39. Когда возможно равенство $\mathsf{E}\,|\xi| = 0$?

40. Пусть четвёртый момент случайной величины ξ конечен, и имеет место равенство $\mathsf{E}\,\xi^2 = \mathsf{E}\,\xi^3 = \mathsf{E}\,\xi^4$. Доказать, что

$$\mathsf{P}(\xi = 0) + \mathsf{P}(\xi = 1) = 1.$$

Распределение Бернулли, и только оно, обладает свойством $\xi^2 = \xi$.

41. Привести пример случайной величины с дискретным распределением, у которой существует первый момент, но не существует дисперсия.

42. Сравнить $\mathsf{E}\,(\xi^4)$ и $(\mathsf{E}\,\xi)^4$.

43. Привести пример случайной величины с абсолютно непрерывным распределением, у которой существует второй момент, но не существует третий.

44. Привести пример случайных величин ξ и η таких, что $\mathsf{E}\,(\xi+\eta)$ существует, но ни $\mathsf{E}\,\xi$, ни $\mathsf{E}\,\eta$ не существуют.

45. Привести пример случайных величин ξ и η таких, что $\mathsf{E}\,\xi$ и $\mathsf{E}\,\eta$ существуют, но не существует $\mathsf{E}\,(\xi\eta)$.

46. Привести пример одинаково распределённых случайных величин ξ, η и ζ таких, что все три величины $\xi+\eta$, $\xi+\zeta$ и $\eta+\zeta$ имеют различные распределения.

Достаточно, если каждая будет принимать три значения.

47. Привести пример одинаково распределённых случайных величин ξ и η таких, что величины $\dfrac{\xi}{\xi+\eta}$ и $\dfrac{\eta}{\xi+\eta}$ имеют разные распределения.

48. Привести пример, показывающий что для одинаково распределённых случайных величин ξ и η не обязательно $\mathsf{E}\left(\dfrac{\xi}{\eta}\right) = \mathsf{E}\left(\dfrac{\eta}{\xi}\right)$, даже если эти математические ожидания существуют.

49. Привести пример, показывающий, что следующее утверждение неверно: «Для любых унимодальных распределений медиана всегда лежит между математическим ожиданием и модой».

Попробуйте распределение с плотностью $f_\xi(t) = p\beta e^{\beta t}$ *при* $t<0$ *и* $f_\xi(t) = (1-p)\alpha e^{-\alpha t}$ *при* $t \geqslant 0$*, где* $0<p<1$, $\alpha>0$, $\beta>0$ *— параметры.*

50. Привести пример того, что в ЗБЧ Хинчина существенно условие независимости.

51. Проверить, имеет ли место сходимость $\dfrac{\xi_1+\ldots+\xi_n}{n}$ к нулю по вероятности для последовательности $\xi_1, \xi_2, \ldots$ независимых случайных величин со стандартным распределением Коши.

52. Пусть ξ_1, ξ_2, $\ldots$ — последовательность независимых и одинаково распределённых случайных величин с невырожденным распределением. Доказать, что не существует случайной величины ξ, к которой данная последовательность сходилась бы по вероятности. Сходится ли эта последовательность по распределению?

53. Доказать, что в условиях ЦПТ последовательность $\dfrac{S_n - n\mathsf{E}\,\xi_1}{\sqrt{n\mathsf{D}\xi_1}}$ не сходится по вероятности ни к какой случайной величине.

Рассмотрите отдельно и вместе сумму первых и следующих n *слагаемых.*

Предметный указатель

Литература

1. Гнеденко Б. В. Курс теории вероятностей. М., 1988.
2. Чистяков В. П. Курс теории вероятностей. М., 1982.
3. Боровков А. А. Теория вероятностей. М., 1986.
4. Колемаев В. А., Калинина В. Н. Теория вероятностей и математическая статистика. М., Инфра-М, 1997.
5. Коршунов Д. А., Фосс С. Г. Сборник задач и упражнений по теории вероятностей. Новосибирск, 1997.
6. Севастьянов Б. А., Чистяков В. П., Зубков А. М. Сборник задач по теории вероятностей. М., 1986.
7. Вентцель Е. С., Овчаров Л. А. Теория вероятностей (Избранные главы высшей математики для инженеров и студентов втузов, задачи и упражнения). М., 1973.