

Методическое пособие
Теория вероятностей

Нижний Новгород, 2018

Содержание

I	События и их вероятности	3
1	Элементы комбинаторики. Схемы шансов	4
1.1	Эксперименты выбора шариков	4
1.2	Схема шансов без возвращения и с учетом порядка	6
1.3	Схема шансов без возвращения и без учёта порядка	6
1.4	Схема шансов с возвращением и с учётом порядка	6
1.5	Схема шансов с возвращением и без учёта порядка	7
2	События, операции над ними и σ -алгебры событий	8
3	Вероятность и её свойства	15
4	Способы задания и подсчёта вероятности	18
4.1	Экспериментальное нахождение вероятности	18
4.2	Вероятность на конечном пространстве.	19
4.3	Классическая вероятность	21
4.4	Вероятность на счётном пространстве	22
4.5	Геометрическая вероятность	25
5	Независимые события	28
6	Условная вероятность	30
7	Формула полной вероятности и формулы Байеса	31
8	Биномиальное распределение	33
9	k -номиальное распределение	35
10	Гипергеометрическое распределение	37
II	Теория случайных величин	38
11	Случайные величины	38
12	Абсолютно непрерывные случайные величины	43
13	Функции Хевисайда и Дирака	47

14	Функции одной случайной величины	53
15	Случайные векторы и их распределения	55
16	Функции от двух случайных величин	58
17	Математическое ожидание	62
18	Дисперсия	67
19	Числовые характеристики зависимости случайных величин	69
III Законы больших чисел		75
20	Неравенство Бьенеме–Чебышёва и неравенство Маркова . .	76
21	Последовательности случайных величин	78
22	Законы больших чисел	81
23	Предельные теоремы для биномиального распределения . .	85
24	Характеристические функции	90
25	Вычисление характеристических функций	94
26	Центральная предельная теорема	96
27	Сферическое, ξ^2 -распределение и распределение Стьюдента	101
27.1	Сферическое распределение	101
27.2	χ - и χ^2 -распределения	102
27.3	О распределении Стьюдента	104
28	Цепи Маркова	105

Исторические сведения

Возникновение теории вероятностей как науки относят к средним векам, когда появилась возможность и возникла необходимость изучения математическими методами азартных игр (таких как орлянка, кости, рулетка).

Самые ранние работы учёных в области теории вероятностей относятся к XVII веку. Первоначально её основные понятия не имели строго математического описания. Задачи, из которых позже выросла теория вероятностей представляли набор некоторых эмпирических фактов о свойствах реальных событий, которые формулировались с помощью наглядных описаний.

Исследуя прогнозирование выигрыша при бросании костей в письмах друг другу, Блез Паскаль и Пьер Ферма открыли первые вероятностные закономерности. Решением тех же задач занимался и Христиан Гюйгенс. При этом с перепиской Паскаля и Ферма он знаком не был и методику решения изобрёл самостоятельно.

Его статья, в которой он ввёл основные понятия теории вероятностей (понятие вероятности как величину шанса; математическое ожидание для дискретных случаев в виде цены шанса). В своей статье он использует (не сформулированные ещё в явном виде) теоремы сложения и умножения вероятностей. Статья была опубликована в печатном виде на двадцать лет раньше (1657 г.) издания писем Паскаля и Ферма (1679 г.).

Важный вклад в теорию вероятностей внёс Якоб Бернулли, он дал доказательство закона больших чисел в простейшем случае независимых испытаний. В первой половине XIX века теория вероятностей начинает применяться к анализу ошибок наблюдений; Лаплас и Пуассон доказали первые предельные теоремы.

Во второй половине XIX века основной вклад внесли русские учёные П. Л. Чебышёв, А. А. Марков и А. М. Ляпунов. В это время были доказаны закон больших чисел, центральная предельная теорема, а также разработана теория цепей Маркова.

Современный вид теория вероятностей получила благодаря аксиоматике, предложенной Андреем Николаевичем Колмогоровым. В результате теория вероятностей приобрела строгий математический вид и окончательно стала восприниматься как один из разделов математики.

Википедия, статья «Теория вероятностей».

Часть I

События и их вероятности

1. Элементы комбинаторики. Схемы шансов

В этом параграфе мы подсчитываем число элементарных событий или, проще говоря, исходов, шансов, которые могут возникать в результате эксперимента.

Например, при подбрасывании монеты могут произойти 2 исхода, при подбрасывании игрального кубика могут произойти 6 исходов, при извлечении карты из колоды в 54 листа могут произойти 53 исхода. Такие подсчёты изучают в разделе математики, называемом *комбинаторикой*.

Пусть A и B — два непересекающихся конечных множества с числом элементов m и n соответственно. Очевидны следующие две леммы.

Лемма 1.1. *(О сумме) Число шансов выбрать один элемент либо из A либо из B , т.е. из объединения $A \cup B$, равно $m + n$.*

Лемма 1.2. *(О произведении) Число шансов выбрать пару элементов, один из A , а другой из B , равно mn , т.е. числу элементов в декартовом произведении $A \times B$.*

Непосредственным обобщением предыдущей леммы является следующая теорема.

Теорема 1.3. *Пусть A_1, A_2, \dots, A_k — конечные непересекающиеся множества, имеющие n_1, n_2, \dots, n_k элементов соответственно. Выберем из каждого множества по одному элементу. Тогда общее число способов, которыми можно осуществить такой выбор, равно $n_1 n_2 \dots n_k$.*

Доказательство. Ясно, что число способов такого выбора равно числу точек (элементов) в декартовом произведении $A_1 \times A_2 \times \dots \times A_k$, т.е. равно $n_1 \cdot n_2 \dots n_k$. \square

1.1. Эксперименты выбора шариков

Рассмотрим ящик, содержащий n одинаковых шариков, на которых написаны числа $1, 2, \dots, n$. Эксперимент состоит в том, что из ящика, не глядя, по одному вынимают k шариков, где $k \leq n$. Обозначим через

$$(n_1, n_2, \dots, n_k)$$

упорядоченный набор чисел, где n_1 — номер 1-го вынутого шарика, n_2 — номер 2-го шарика, ..., n_k — номер k -го шарика.

Например, из 5 занумерованных шариков выбрали 3 шарика и получился набор (4, 2, 1).

Сколько имеется различных способов вынуть из ящика k шариков? На этот вопрос нельзя дать однозначный ответ, потому что такой эксперимент определён неоднозначно.

Во-первых, не определено, возвращают ли извлеченный шарик обратно в ящик. Во-вторых, не определено, какие наборы номеров считать различными и какие наборы считать одинаковыми.

Рассмотрим следующие возможные условия проведения эксперимента.

1. *Эксперимент с возвращением.* Каждый извлечённый шарик возвращается в ящик. В этом случае в наборе могут появляться одинаковые номера. Например, при выборе трёх шариков из ящика, содержащего пять шариков с номерами 1, 2, 3, 4 и 5, могут появиться наборы (3, 3, 5), (1, 2, 4) и (4, 2, 1).
2. *Эксперимент без возвращений.* Извлечённые шарики в ящик не возвращаются. В этом случае в наборе не могут встречаться одинаковые номера. В рассмотренном выше примере набор (3,3,5) не может появиться, а наборы (1,2,4) и (4,2,1) могут.

Опишем теперь, какие наборы номеров мы будем считать различными. Существуют ровно две возможности.

1. *Эксперимент с учётом порядка.* Два набора номеров считаются различными, если они отличаются либо составом, либо порядком. В рассмотренном выше примере все наборы (3,3,5), (1,2,4) и (4,2,1) считаются различными.
2. *Эксперимент без учёта порядка.* Два набора номеров считаются различными, если они отличаются только составом.

В рассмотренном выше примере наборы (1,2,4) и (4,2,1) доставляют одно и то же элементарное событие, а набор (3,3,5) — другое.

Подсчитаем теперь, сколько получится различных исходов для каждого из четырёх экспериментов. Заметим, что в литературе такие эксперименты часто называют схемами выбора или схемами шансов. Схема шансов — это условия (с возвратом или без, какие наборы различны и т.д.), при которых проводится эксперимент.

1.2. Схема шансов без возвращения и с учетом порядка

Теорема 1.4. В эксперименте без возвращения и с учётом порядка число способов выбрать k элементов из n -элементного множества равно

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

Число A_n^k называется числом размещений элементов k на n местах. Читается: « A из n по k ».

Доказательство. При выборе первого шарика имеется n возможностей. При выборе второго шарика остаётся $n-1$ возможностей, и т.д. При выборе последнего k -го шарика остаётся $n-k+1$ возможностей. По теор. 1.3 общее число наборов равно $n(n-1)\dots(n-k+1)$, что и требовалось доказать. \square

Следствие 1.5. Число перестановок из n элементов равно $n!$.

Доказательство. Очевидно, что перестановка есть результат выбора по схеме без возвращения и с учётом порядка всех n элементов из n , т.е. общее число перестановок равно $A_n^n = n!$. \square

1.3. Схема шансов без возвращения и без учёта порядка

Теорема 1.6. В эксперименте без возвращения и без учёта порядка число способов извлечь k из n -элементного множества равно

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

Число C_n^k называется числом сочетаний k элементов из n элементов. Читается: « C из n по k »

Доказательство. По следствию 1.5 из k элементов можно образовать $k!$ упорядоченных наборов. Поэтому количество сочетаний (неупорядоченных наборов) в $k!$ раз меньше, чем число размещений. Поделив A_n^k на $k!$, получим требуемый результат. \square

1.4. Схема шансов с возвращением и с учётом порядка

Теорема 1.7. В эксперименте с возвращением и с учётом порядка число способов извлечь k элементов из n -элементного множества равно n^k .

Доказательство. При выборе каждого из k шариков имеется n возможностей. По теореме 1.3 общее число наборов равно $n \cdot n \dots n = n^k$. \square

1.5. Схема шансов с возвращением и без учёта порядка

Замечание 1.8. Рассмотрим для примера ящик с двумя шариками 1 и 2, из которого мы вынимаем последовательно два шарика. Без учёта порядка имеется 3 исхода:

$$\{1, 1\}, \{1, 2\} = \{2, 1\}, \{2, 2\}$$

Теорема 1.9. В эксперименте с возвращением и без учёта порядка число способов извлечь k элементов из n -элементного множества равно C_{n+k-1}^k .

Доказательство. Т.к. порядок появления шариков не учитывается, то мы учитываем лишь только то, сколько раз в наборе появится i -й шарик для каждого $i = 1, 2, \dots, n$. Обозначим через k_i число появлений i -го шарика в наборе. Во-первых, $0 \leq k_i \leq k$, а во-вторых,

$$k_1 + k_2 + \dots + k_n = k.$$

Поставим каждому исходу в соответствие набор чисел (k_1, k_2, \dots, k_n) . Легко видеть, что это соответствие является взаимно однозначным. Такое соответствие можно рассматривать как способ нумерации наборов. (Например, исходам из замеч. 1.8 ставятся в соответствие следующие номера: $\{1, 1\} \leftrightarrow (2, 0)$, $\{1, 2\} \leftrightarrow (1, 1)$ и $\{2, 2\} \leftrightarrow (0, 2)$.) Рассмотрим теперь другой эксперимент. Пусть теперь имеется n урн с номерами $i = 1, 2, \dots, n$, в которых размещаются k неразличимых шариков. Сколько существует способов разложить шарики по урнам? Нас интересует только количество шариков в i -й урне для каждого i . Обозначим через k_i число шариков в i -й урне. Ясно, что $0 \leq k_i \leq k$, и что числа k_i и в этом эксперименте тоже удовлетворяют уравнению

$$k_1 + k_2 + \dots + k_n = k.$$

Исходы этого эксперимента тоже взаимно однозначно описываются наборами чисел (k_1, k_2, \dots, k_n) . Т.о., исходы в эксперименте с урнами и исходы предыдущего эксперимента с ящиком занумерованы одним и тем же набором чисел, поэтому число исходов в обоих экспериментах одно и то же и равно числу решений этого уравнения. Вычислим это число для эксперимента с урнами. Изобразим расположение шариков в урнах с помощью схематичного рисунка. Вертикальными линиями обозначим перегородки между урнами, а кружками — шарики, находящиеся в них. Например,

$$|\bullet\bullet| \bullet\bullet\bullet || \bullet || \bullet\bullet | \bullet |.$$

На рисунке показаны 9 шариков, рассыпанные по 7 урнам: 1-я и 6-я урны содержат по 2 шарика, 2-я урна содержит 3 шарика, 3-я и 5-я урны — пустые и, наконец, 4-я и 7-я урны содержат по одному шарiku.

Меняя местами шарики и стенки, можно получить все возможные расположения шариков в урнах. Другими словами, все расположения можно получить, расставляя k шариков и $n - 1$ стенок на $n - 1 + k$ местах. Число $n - 1 + k$ получается следующим образом. Число стенок у n урн равно $n + 1$, и т.к. две крайние стенки двигать нельзя, то число стенок, которые можно двигать равно $n - 1$. Поэтому шарики могут занимать k мест, а стенки урн — оставшиеся $n - 1$ место. По теореме 1.6 число способов расставить k шариков на $n - 1 + k$ местах и затем расставить стенки на оставшихся $n - 1$ местах равно C_{n+k-1}^k . Что и требовалось доказать. \square

2. События, операции над ними и σ -алгебры событий

Теория вероятностей, как и любая современная математическая теория, начинается с аксиоматических (неопределяемых) понятий. Такими являются следующие понятия.

1. Понятие: *эксперимент* = *опыт* = *испытание*. Считается, что все три слова означают одно и то же, т.е. являются синонимами.
2. Понятие: *произойти* = *возникнуть* = *появиться*.
3. Понятие: *элементарное событие* = *элементарный исход* = *результат* = *шанс*.

При этом считается, что в результате опыта происходит одно и только одно элементарное событие.

Определение 2.1. Множество всех элементарных событий данного эксперимента называется *пространством элементарных событий*, или часто короче *пространством*.

Будем обозначать его через Ω . Ясно, что пространство элементарных событий не пусто, $\Omega \neq \emptyset$.

Как множество пространство Ω может быть либо конечным¹, либо счёт-

¹Конечное множество — множество, количество элементов которого конечно, то есть, существует неотрицательное целое число n , равное количеству элементов этого множества. В противном случае множество называется бесконечным.

ным², либо несчётным³ множеством.

Определение 2.2. Конечные и счётные пространства элементарных событий называются *дискретными*.

Примеры.

1. Эксперимент: подбрасывание монеты. Элементарные события: o – выпадение орла, p – выпадение решётки. Пространство $\Omega = \{o, p\}$ является конечным множеством.
2. Эксперимент: одновременное подбрасывание двух монет одного достоинства. Пространство элементарных событий $\Omega = \{(o, o), (o, p), (p, p)\}$ является конечным множеством.
3. Эксперимент: подбрасывание одной монеты до выпадения первого орла. Пространство элементарных событий

$$\Omega = \{o, ro, rro, rrrro, rrrrro, rrrrrro, \dots\}$$

– счётное множество.

4. Эксперимент: на стол $\Omega = I^2 = I \times I$ садится мыльный пузырь и лопается, оставляя под собой точку (x, y) . Элементарное событие: появление на плоскости точки (x, y) . Пространство элементарных событий Ω – несчётное множество.

Определение 2.3. Если пространство Ω содержит только одно элементарное событие, то эксперимент называется с детерминированным (определённым) исходом; в противном случае эксперимент называется со случайным исходом.

Определение 2.4. Любое подмножество $A \subset \Omega$ пространства элементарных событий называется случайным событием или просто событием. Считается, что событие A произошло, если произошло любое элементарное событие ω , содержащееся в A , см. рис. 1.

Определение 2.5. 1) Событие Ω называется *достоверным*.

2) Событие $\emptyset \subset \Omega$ называется *невозможным*.

Каждое элементарное событие $\omega \in \Omega$ можно рассматривать как одноэлементное подмножество достоверного события Ω , т.е. $\{\omega\} \subset \Omega$. Для изображения событий можно использовать диаграммы Венна⁴.

Пусть A и B – события, они показаны на рис. 1.

²Счётное множество есть бесконечное множество, элементы которого можно пронумеровать натуральными числами. Более формально: множество Ω является счётным, если существует биекция $\Omega \leftrightarrow \mathbb{N}$, где \mathbb{N} обозначает множество всех натуральных чисел.

³Множество, не являющееся конечным или счетным, называется несчетным.

⁴Джон Венн (John Venn, 1834 – 1923), английский логик.

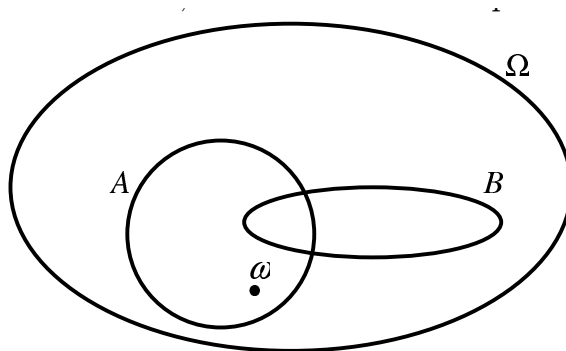


Рис. 1: События в достоверном событии Ω

Определение 2.6. 1) Событие $A \cup B$ называется *объединением* событий и состоит в том, что произошло хотя бы одно из событий A или B .

2) Событие $A \cap B$ называется *пересечением* событий A и B и состоит в том, что произошли оба события A и B .

3) Событие $A \setminus B$ называется *разностью* событий A и B и состоит в том, что событие A произошло, а B – нет.

4) Событие $\bar{A} = \Omega \setminus A$ называется *противоположным событием* A и состоит в том, что событие A не произошло. Ясно, что $A = \overline{\bar{A}}$ (событие, противоположное к противоположному, является исходным событием). Т.к. $\overline{\emptyset} = \Omega \setminus \emptyset$ и $\bar{\Omega} = \Omega \setminus \Omega = \emptyset$, то невозможное событие \emptyset и достоверное событие Ω являются взаимно противоположными (друг другу).

5) Говорят, что событие A *влечёт* событие B , и пишут $A \subset B$, если при наступлении события A происходит и событие B .

Определение 2.7. 1) События A и B называются *несовместными*, если $A \cap B$ является невозможным событием, т.е. $A \cap B = \emptyset$.

2) События A_1, \dots, A_n называются *попарно несовместными*, если для любых $1 \leq i < j \leq n$ события A_i и A_j несовместны.

Определение 2.8. Рассмотрим множество \mathfrak{A} , элементами которого являются события пространства Ω (не обязательно все!). Множество \mathfrak{A} называется алгеброй событий, если достоверное событие Ω и любые события $A, B \in \mathfrak{A}$ удовлетворяют аксиомам:

Акс. $\mathcal{A}1$. $\Omega \in \mathfrak{A}$.

Акс. $\mathcal{A}2$. $A \cup B \in \mathfrak{A}$.

Акс. $\mathcal{A}3$. $A \cap B \in \mathfrak{A}$.

Акс. $\mathcal{A}4$. $A \setminus B \in \mathfrak{A}$.

Из аксиомы $\mathcal{A}1$ следует, что алгебра событий не может быть пустой; она всегда содержит достоверное событие Ω . А т.к. $\Omega \setminus \Omega = \emptyset$, то из аксиомы $\mathcal{A}4$ следует, что $\emptyset \in \mathfrak{A}$, т.е. любая алгебра событий \mathfrak{A} содержит вместе с достоверным и невозможное событие.

Примеры

1) Для любого пространства элементарных событий Ω набор из двух множеств $\mathfrak{A}_0 = \{\emptyset, \Omega\}$ удовлетворяет аксиомам $\mathcal{A}1$ – $\mathcal{A}4$, поэтому $\mathfrak{A}_0 = \{\emptyset, \Omega\}$ является алгеброй событий. Алгебра событий \mathfrak{A}_0 называется тривиальной. Это самая маленькая алгебра событий.

2) В этом примере эксперимент – подбрасывание игральной кости. Пространство элементарных событий есть $\Omega = \{1, 2, 3, 4, 5, 6\}$. Пусть событие $A = \{1, 3, 5\}$ – выпадение нечётного числа очков, а событие $B = \{2, 4, 6\}$ – выпадение чётного числа очков. Множество событий $\mathfrak{A}_1 = \{\emptyset, \Omega, A, B\}$ удовлетворяет аксиомам $\mathcal{A}1$ – $\mathcal{A}4$, поэтому является алгеброй событий.

3) Для любого пространства элементарных событий Ω множество $\mathfrak{B}(\Omega)$ = множество всех подмножеств Ω удовлетворяет аксиомам $\mathcal{A}1$ – $\mathcal{A}4$, поэтому является алгеброй событий. Эта алгебра событий является самой большой на Ω . Для $\Omega = \{1, 2, 3, 4, 5, 6\}$ эта алгебра событий содержит 64 события.

4) Задайте ещё какую-нибудь алгебру событий на $\Omega = \{1, 2, 3, 4, 5, 6\}$. Сколько различных алгебр событий можно задать на этом пространстве элементарных событий?

Замечание 2.9. Из аксиом $\mathcal{A}1$ – $\mathcal{A}4$ следует, что если к конечному набору событий из любой алгебры событий применить операции объединения, пересечения и вычитания конечное число раз, то полученное событие тоже содержится в этой алгебре.

Замечание 2.10. Если пространство элементарных событий конечно, $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$, то любая его алгебра событий тоже конечна. Это следует из того, что множество $\mathfrak{B}(\Omega)$ всех возможных событий, содержащихся в Ω , тоже конечно и содержит 2^n событий.

Определение 2.11. Алгебра событий \mathfrak{A} называется σ -алгеброй, если для любого счётного набора событий $A_1, A_2, \dots \in \mathfrak{A}$ выполнена пятая аксиома:

$$\text{Акс. } \mathcal{A}5. \quad \bigcup_{n=1}^{\infty} A_n \in \mathfrak{A}.$$

Замечание 2.12. Из аксиом $\mathcal{A}1$ – $\mathcal{A}4$ не следует, что объединение несчётного количества событий является событием из σ -алгебры. Рассмотрение несчётных объединений событий приводит к построению т.н. неизмеримых событий, вероятность наступления которых не существует. Первый пример такого неизмеримого события построил Витали⁵. При построении неизмеримого множества Витали используется аксиома теории множеств – аксиома выбора.

⁵Джузеппе Витали (Giuseppe Vitali, 26.08.1875 – 29.02.1932, Italy) – итальянский математик.

Аксиома выбора. Для любого произвольного набора непустых непересекающихся множеств можно составить множество, выбрав в него по одному элементу из каждого множества этого набора.

Теорема 2.13 (Теорема Витали – построение неизмеримого множества). Существуют множества, длина которых не может быть выражена никаким числом.

Доказательство. Для доказательства нам понадобятся лишь следующие очевидные свойства длины:

- длина дуги остается неизменной при повороте окружности вокруг центра;
- длина дуги, которая представляет собой объединение счетного количества попарно непересекающихся дуг, равна сумме длин этих дуг.

Рассмотрим стандартную (единичного радиуса) окружность S^1 . Она эквивалентна отрезку $[0, 2\pi)$, т.е. её длина равна 2π . На этой окружности центральный угол в радианах равен длине дуги на которую он опирается.

Для любого рационального числа $\frac{p}{q}$, где $q \neq 0$, рассмотрим дугу длины $\frac{2\pi p}{q}$.

Если отложить её на окружности S^1 последовательно q раз, то полученная дуга замкнётся, т.е. начало 1-ой дуги совпадёт с концом q -ой.

Для любого иррационального числа α рассмотрим дугу длины $2\pi\alpha$. Ясно, что если отложить эту дугу n раз на окружности S^1 , то начало и конец дуги $2\pi\alpha n$ не совпадут ни при каком целом n . Это означает, что множество

$$A_0 = \{\dots, -2\pi\alpha n, \dots, -4\pi\alpha, -2\pi\alpha, 0, 2\pi\alpha, 4\pi\alpha, \dots, 2\pi\alpha n, \dots\}$$

является счётным. Заметим, что для любого целого n поворот на угол $2\pi\alpha n$ переводит множество A_0 в себя. Если взять произвольную точку (угол) $\varphi \in [0, 2\pi)$ на окружности, такую чтобы $\varphi \notin A_0$, то множество

$$A_\varphi = \{\dots, \varphi - 2\pi\alpha n, \dots, \varphi - 4\pi\alpha, \varphi - 2\pi\alpha, \varphi, \varphi + 2\pi\alpha, \varphi + 4\pi\alpha, \dots, \varphi + 2\pi\alpha n, \dots\}$$

тоже является счётным, и для любого целого n поворот множества A_φ на угол $2\pi n\alpha$ тоже переводит его в себя.

Заметим, что т.к. $\varphi \notin A_0$, то множества A_0 и A_φ не пересекаются: если бы они имели общую точку, то при $n = \pm 1, \pm 2, \dots$ все остальные их точки получились бы прибавлением дуг $2\pi\alpha n$, и тогда бы множества A_0 и A_φ совпадали.

Заметим, что для любого целого n поворот на угол $2\pi\alpha n$ переводит множество A_φ в себя.

Если теперь выбрать угол $\psi \in [0, 2\pi)$, такой чтобы $\psi \notin A_0$ и $\psi \notin A_\varphi$, то множество

$$A_\psi = \{\dots, \psi - 2\pi\alpha n, \dots, \psi - 4\pi\alpha, \psi - 2\pi\alpha, \psi, \psi + 2\pi\alpha, \psi + 4\pi\alpha, \dots, \psi + 2\pi\alpha n, \dots\}$$

тоже является счётным, и для любого целого n поворот множества A_ψ на угол $2\pi\alpha n$ тоже переводит его в себя. Заметим, что т.к. $\psi \notin A_0$ и $\psi \notin A_\varphi$, то множество A_ψ не пересекается ни с A_0 и ни с A_φ .

Заметим, что для любого целого n поворот на угол $2\pi\alpha n$ переводит множество A_ψ в себя.

Процесс построения таких непересекающихся множеств можно продолжить неограниченно⁶, пока не будут исчерпаны (выбраны) все точки окружности S^1 . В каждом таком множестве содержится счетное число точек, и все точки в одном множестве получаются друг из друга поворотами на угол $2\pi\alpha$.

Разные множества не пересекаются. Таким образом, окружность S^1 является объединением непересекающихся множеств $A_0, A_\varphi, A_\psi, \dots$ т.е.

$$S^1 = \bigcup_{\lambda \in \{0, \varphi, \psi, \dots\}} A_\lambda$$

Объединение счётного набора счётных множеств есть счётное множество. А т.к. окружность S^1 состоит из несчетного множества точек, то набор построенных множеств $A_0, A_\varphi, A_\psi, \dots$ является несчётным. Другими словами множество индексов $\{0, \varphi, \psi, \dots\}$ – несчётное.

Выберем из каждого множества $A_0, A_\varphi, A_\psi, \dots$ ровно по одной точке и по аксиоме выбора составим из этих точек множество B_0 .

⁶В математике различают два типа бесконечности: потенциальная и актуальная бесконечности.

Потенциальная бесконечность означает, что процесс построения какого-либо объекта может быть продолжен неограниченно. В нашем случае процесс построения множеств $A_\varphi, A_\xi, A_\psi, \dots$ может быть продолжен неограниченно (если не принимать во внимание, какое время может быть потрачено на это построение). Другой пример потенциальной бесконечности возникает при построении натурального ряда. Если мы выпишем все натуральные числа от 1 до n , то ничто не мешает нам написать число $n + 1$, и т.д. Потенциальная бесконечность есть бесконечный процесс построения объектов, у которого нет последнего шага. Например, в доказательстве по методу математической индукции.

Под актуальной бесконечностью понимается бесконечная совокупность, построение которой завершено, и все ее элементы наличествуют одновременно. Например, мы будем иметь дело с актуальной бесконечностью, если перечислим весь натуральный ряд полностью и имеем его в законченном виде $N = \{1, 2, \dots, n, \dots\}$. Актуальная бесконечность представляет собой весьма сильную идеализацию. В самом деле, она допускает не только возможность построения последующего объекта, если построен предыдущий, но и постулирует, что все возможные объекты уже построены и существуют одновременно. В нашем случае актуальная бесконечность означает, что процесс построения множеств $A_0, A_\varphi, A_\psi, \dots$ закончен, и мы имеем это множество множеств $\{A_0, A_\varphi, A_\psi, \dots\}$ налицо.

Построенное множество B_0 называется *множеством Витали*. Для каждого $n = \pm 1, \pm 2, \dots$ обозначим через B_n множество, которое получается в результате поворота множества B_0 на угол $2\pi n\alpha$.

Ясно, что все множества B_n , $n \in \mathbb{Z}$, имеют одну и ту же длину. Т.к. все точки множества A_λ , где $\lambda \in \{0, \varphi, \psi, \dots\}$, можно получить, поворачивая одну из них на углы $2\pi n\alpha$, и т.к. в множестве B_0 собраны по одной точке из каждого множества A_λ , то объединение всех B_n составляет окружность S^1 , т.е.

$$S^1 = \bigcup_{n=-\infty}^{\infty} B_n$$

Оставшуюся часть доказательства проведём методом от противного. Предположим, что длина множества Витали B_0 существует, и обозначим её через $\mu(B_0)$. Тогда все множества B_n имеют одну и ту же длину, т.к. получены из B_0 поворотом на угол кратный $2\pi\alpha$. Т.к. множества B_n не пересекаются, то сумма их длин равна длине окружности S^1 . Поэтому

$$\begin{aligned} 2\pi = \mu(S^1) &= \mu\left(\bigcup_{n=-\infty}^{\infty} B_n\right) = \sum_{n=-\infty}^{\infty} \mu(B_n) = \sum_{n=-\infty}^{\infty} \mu(B_0) = \\ &= \begin{cases} \infty, & \text{если } \mu(B_0) > 0 \\ 0, & \text{если } \mu(B_0) = 0 \end{cases} \end{aligned}$$

Полученное противоречие означает, что длины у множества Витали B_0 нет вообще никакой (ни нулевой, ни конечной, ни бесконечной). Такие множества и соответствующие им события называются *неизмеримыми*. \square

Замечание 2.14. 1) В предыдущем примере было доказано существование неизмеримого множества (множества Витали) и предьявлена конструкция его построения. Меняя иррациональное число α , можно получить несчётное количество неизмеримых множеств на окружности $S^1 = [0, 2\pi)$.

Если аксиому выбора не признавать, то неизвестно можно ли вообще построить неизмеримые множества. Можно доказать, что аксиома $\mathcal{A}5$ не допускает появление неизмеримых событий.

2) Если Ω – несчётное множество (отрезок, площадка, объёмное тело), то множество $\mathcal{P}(\Omega)$ всех подмножеств множества Ω не является σ -алгеброй, потому что оно содержит неизмеримые события.

Замечание 2.15. Все конечные и счётные алгебры событий являются σ -алгебрами.

Лемма 2.16. Если A_1, A_2, \dots есть счётный набор событий из σ -алгебры \mathcal{A} , то $\bigcap_{n=1}^{\infty} A_n \in \mathcal{A}$

Доказательство. Пусть $A_1, A_2, \dots \in \mathcal{A}$, Тогда из аксиомы $\mathcal{A}4$ следует, что

$$\overline{A_1}, \overline{A_2}, \dots \in \mathcal{A},$$

а из аксиомы $\mathcal{A}5$ следует, что

$$\bigcup_{n=1}^{\infty} \overline{A_n} \in \mathcal{A}.$$

Тогда по аксиоме 4 дополнение к этому множеству тоже принадлежит \mathcal{A} , т.е.

$$\overline{\bigcup_{n=1}^{\infty} \overline{A_n}} \in \mathcal{A}$$

по формулам двойственности де Моргана

$$\overline{\bigcup_{n=1}^{\infty} \overline{A_n}} = \bigcap_{n=1}^{\infty} A_n,$$

поэтому

$$\bigcap_{n=1}^{\infty} A_n \in \mathcal{A}$$

□

Следствие 2.17. Лемма 2.16 и аксиома $\mathcal{A}5$ эквивалентны.

3. Вероятность и её свойства

Определение 3.1. Пусть Ω — пространство элементарных событий, и \mathcal{A} — его σ -алгебра событий. Вероятностью называется функция множества $P : \mathcal{A} \rightarrow \mathbb{R}$, удовлетворяющая следующим аксиомам:

Акс. $\mathcal{P}1$ (неотрицательность вероятности). Для любого события $A \in \mathcal{A}$ выполнено равенство $P(A) \geq 0$.

Акс. $\mathcal{P}2$ (нормированность вероятности). $P(\Omega) = 1$

Акс. $\mathcal{P}3$ (счётная аддитивность вероятности как функции множества).

Для любого счётного набора попарно несовместных событий $A_1, A_2, \dots \in \mathfrak{A}$ имеет место равенство

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$$

В частности, эта аксиома справедлива и для конечного набора попарно несовместных событий. Акс. $\mathcal{P}1$ (непрерывность вероятности). Для любой убывающей последовательности событий $A_1 \supset A_2 \supset \dots \supset A_n \supset \dots$ из σ -алгебры \mathfrak{A} такой, что $\bigcap_{n=1}^{\infty} A_n = \emptyset$, выполняется равенство

$$\lim_{n \rightarrow \infty} P(A_n) = 0$$

Определение 3.2. Тройка $(\Omega, (\mathfrak{A}), P)$ называется *вероятностным пространством*.

Докажем теперь основные свойства вероятности пп. 3.3 – 3.12.

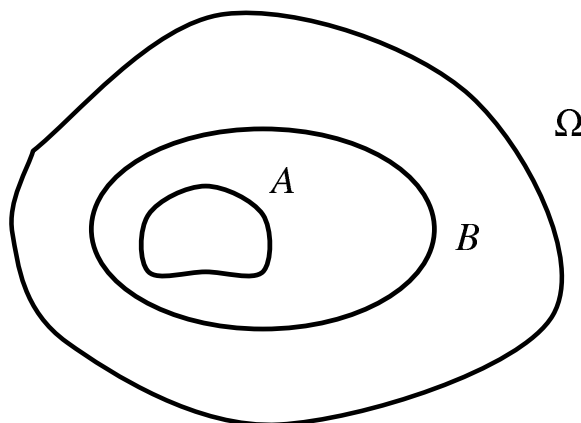


Рис. 2: Если $A \subseteq B$, то $P(B \setminus A) = P(B) - P(A)$.

Лемма 3.3. Если событие A влечёт событие B , т.е. $A \subseteq B$, то $P(B \setminus A) = P(B) - P(A)$ (см. рис. 2)

Доказательство. Т.к. $B = A \cup (B \setminus A)$, и события A и $B \setminus A$ несовместны, то по аксиоме P имеем $P(B) = P(A) + P(B \setminus A)$, что и требовалось доказать. \square

Следствие 3.4. Если $A \subseteq B$, то $P(A) \leq P(B)$.

Лемма 3.5. $P(\emptyset) = 0$.

Доказательство. $P(\emptyset) = P(\Omega \setminus \Omega) = P(\Omega) = P(\Omega) = 1 - 1 = 0$ \square

Лемма 3.6. Для любого события $A \in \mathfrak{A}$ выполнено неравенство

$$0 \leq P(A) \leq 1$$

Доказательство. Т.к. $\emptyset \subseteq A \subseteq \Omega$, то из след. 3.4 следует утверждение леммы. \square

Лемма 3.7. Для любого события $A \in \mathfrak{A}$ имеем $P(\bar{A}) = 1 - P(A)$

Доказательство. $P(\bar{A}) = P(\Omega \setminus A) = P(\Omega) - P(A) = 1 - P(A)$ \square

Лемма 3.8. Для любых событий $A, B \in \mathfrak{A}$ выполнено равенство

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

Доказательство. Т.к. $A \cup B = A \cup (B \setminus (A \cap B))$, и события A и $B \setminus (A \cap B)$ несовместны, то применяя сначала Акс. $\mathcal{P}3$, а затем лемму 3.3 получим $P(A \cup B) = P(A) + P(B \setminus (A \cap B)) = P(A) + P(B) - P(A \cap B)$. \square

Следствие 3.9. Для любых событий $A, B \in \mathfrak{A}$ выполнено равенство

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

Следствие 3.10. Если события A и B несовместны, т.е. $A \cap B = \emptyset$, то

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

Доказательство. По лемме 3.8 имеем $P(A \cup B) = P(A) + P(B) - P(A \cap B) = P(A) + P(B) - P(\emptyset) = P(A) + P(B)$. \square

Следствие 3.11. $P(A_1 \cup \dots \cup A_n) \leq \sum_{i=1}^n P(A_i)$.

Теорема 3.12.

$$\begin{aligned} P(A_1 \cup A_2 \cup \dots \cup A_n) &= \sum_{i=1}^n P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) \\ &+ \sum_{1 \leq i < j < m \leq n} P(A_i \cap A_j \cap A_m) - \dots + (-1)^{n-1} P(A_1 \cap A_2 \cap \dots \cap A_n). \end{aligned}$$

Доказательство. Применяя метод математической индукции и лемму 3.8 получим требуемый результат. \square

4. Способы задания и подсчёта вероятности

4.1. Экспериментальное нахождение вероятности

В этом пункте описан способ экспериментального определения вероятности наступления так называемых массовых событий.

Определение 4.1. Событие A называется *массовым*, если опыт (эксперимент, испытание), при котором событие A может произойти, можно повторить *неограниченное число раз при одних и тех же условиях*.

Методами теории вероятностей изучают (в основном) эксперименты, которые порождают массовые события. Всюду до конца лекций будут рассматриваться только массовые события.

Пример 4.2. 1) Опыт: однократное подбрасывание ломаного гроша. События выпадение орла O и выпадение решки P являются массовыми событиями. Заметим, что пространство элементарных событий $\Omega = \{O, P\}$ состоит из конечного числа элементарных исходов.

2) Опыт: подбрасывание ломаного гроша до появления первого орла. События $P, OP, OOP, \dots, O \dots OP, \dots$ являются массовым. Заметим, что в этом примере пространство $\Omega = \{P, OP, OOP, \dots, O \dots OP, \dots\}$ состоит из счётного количества элементарных исходов: положение кольца однозначно определяется положением его центра.

3) Опыт: бросание обручального кольца на клетчатую скатерть (диаметр кольца меньше стороны клетки). Событие A : кольцо падает внутрь какой-нибудь клетки, не пересекая её границы. Это событие является массовым. Заметим, что в этом примере пространство состоит из несчётного количества элементарных исходов.

Типичность этих трёх примеров состоит в том, что в теории вероятностей встречаются три типа пространств элементарных событий: конечные, счётные и несчётные.

Определение 4.3. Если в результате n испытаний массовое событие A произошло $\mu(A)$ раз, то число $\frac{\mu(A)}{n}$ называется *относительной частотой* появления события A .

Для каждого примера 4.2 можно провести n опытов, сосчитать число $\mu(A)$ появлений этого события и подсчитать относительную частоту $\frac{\mu(A)}{n}$.

Массовые события обладают свойством «статистической устойчивости», а именно: *при увеличении числа экспериментов относительная частота $\frac{\mu(A)}{n}$ появления события A имеет тенденцию стабилизироваться, стремясь к некоторому числу $P(A)$.*

Определение 4.4. Число $P(A) = \lim_{n \rightarrow \infty} \frac{\mu(A)}{n}$ называется *статистической вероятностью* события A и находится при больших n по приближённой формуле $P(A) \approx \mu(A)$.

4.2. Вероятность на конечном пространстве.

Пусть задано конечное пространство $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$, состоящее из n элементарных событий (исходов), и заданы вероятности наступления этих событий $P(\omega_1) = p_1, P(\omega_2) = p_2, \dots, P(\omega_n) = p_n$ так, что $p_1 + p_2 + \dots + p_n = 1$. Обозначим через ω переменную величину, принимающую значения из Ω .

Определение 4.5. Такое соответствие записывается в виде таблицы

ω	ω_1	ω_2	\dots	ω_n
$P(\xi)$	p_1	p_2	\dots	p_n

которая называется *рядом распределения* случайных исходов или *законом распределения*. Если $n = 2$, то ряд называется *схемой Бернулли*. Если $n \geq 3$, то ряд называется *схемой независимых испытаний с несколькими исходами*.

Ряд распределения задаёт функцию в виде таблицы, которая каждому элементарному исходу ставит в соответствие вероятность его наступления.

Потребуем теперь выполнение аксиомы $\mathcal{P}3$.

Лемма-определение 4.6 (Формула вероятности на конечном пространстве). Если вероятность, определяемая рядом распределения из опред. 4.5, подчиняется аксиоме $\mathcal{P}3$, то вероятность события $A = \{\omega_{i_1}, \omega_{i_2}, \dots, \omega_{i_k}\} \subset \Omega$ вычисляется по формуле

$$P(A) = p_{i_1} + p_{i_2} + \dots + p_{i_k},$$

которая называется *формулой вероятности на конечном пространстве*.

Доказательство.

$$\begin{aligned} P(A) &= P(\omega_{i_1}, \omega_{i_2}, \dots, \omega_{i_k}) = \\ &= P\left(\bigcup_{j=1}^k \{\omega_{i_j}\}\right) \stackrel{\mathcal{P}3}{=} \sum_{j=1}^k P(\omega_{i_j}) = \sum_{j=1}^k p_{i_j} = p_{i_1} + p_{i_2} + \dots + p_{i_k}. \end{aligned}$$

□

Пример 4.7. Простейший ряд распределения имеет эксперимент с детерминированным исходом, $\Omega = \{\omega\}$ (см. опред. 2.3):

ξ	ω
P	1

Этот тривиальный случай удовлетворяет всем аксиомам и определениям, однако никакого значения в теории вероятностей не имеет. Придётся с этим мириться.

Определение 4.8. Пусть в результате опыта могут возникнуть только два события: «успех», который обозначается единицей — 1 и наступает с вероятностью p , и «неудача», которая обозначается нулём — 0 и наступает с вероятностью $q = 1 - p$; $\Omega = \{0, 1\}$. Такой опыт называется *схемой Бернулли*⁷ и имеет ряд распределения

ξ	0	1
P	$q = 1 - p$	p

Схема Бернулли является первым нетривиальным примером вероятности на конечном пространстве.

Схема Бернулли имеет простую интерпретацию. Рассмотрим ломаный грош с вероятностями выпадения орла p и решки $q = 1 - p$. Обозначим появление орла через 1, а его не выпадение через 0, получим схему Бернулли.

Пример 4.9 (Геометрическая интерпретация схемы независимых испытаний с несколькими исходами). Рассмотрим произвольный выпуклый многогранник с n гранями, на которых написаны символы $\omega_1, \omega_2, \dots, \omega_n$.

При этом многогранник должен быть таким, чтобы перпендикуляр, опущенный из его центра тяжести на любую грань, пересекал эту грань в её внутренней точке (чтобы многогранник, падая на эту грань, не перекатывался на другую).

Пусть вероятности выпадения многогранника гранью вниз равны соответственно p_1, p_2, \dots, p_n , где естественно $p_1 + p_2 + \dots + p_n = 1$.

Легко видеть, что вероятность p_i пропорциональна телесному углу α_i с вершиной C в центре тяжести многогранника, опирающегося на грань ω_i .

Так как сумма всех телесных углов с вершиной C равна 4π стерadian, т.е. $\alpha_1 + \alpha_2 + \dots + \alpha_n = 4\pi$, то

$$\frac{\alpha_1}{4\pi} + \frac{\alpha_2}{4\pi} + \dots + \frac{\alpha_n}{4\pi} = 1,$$

поэтому $p_i = \frac{\alpha_i}{4\pi}$.

⁷Якоб Бернулли (Jakob Bernoulli, 1654-1705), швейцарский математик

4.3. Классическая вероятность

Классическая вероятность является частным случаем вероятности на конечном пространстве, когда вероятность подчинена принципу равной вероятности. Исторически классическая вероятность применялась в теории азартных игр и появилась раньше вероятности на конечном пространстве.

Принцип равной вероятности. Если на конечном пространстве Ω вероятность наступления его элементарных исходов $\omega_1, \omega_2, \dots, \omega_n$ одна и та же,

$$p_1 = p_2 = \dots = p_n = p,$$

то говорят, что вероятность на конечном пространстве удовлетворяет принципу равной вероятности (или равной возможности).

Выпадение орла и решки при подбрасывании симметричной монеты; выпадение граней при подбрасывании правильного многогранника (тетраэдра, куба, октаэдра, додекаэдра или икосаэдра); появление к.-л. карты из полной колоды карт удовлетворяют принципу равной возможности.

Классическая вероятность характеризуется только числом элементарных исходов n в пространстве Ω . Она имеет ряд распределения

ω	ω_1	ω_2	\dots	ω_n
$P(\omega)$	p	p	\dots	p

где $np = 1$, поэтому $p = \frac{1}{n}$.

Лемма 4.10 (Формула классической вероятности). Если вероятность P удовлетворяет принципу равной возможности, то вероятность наступления события $A = \{\omega_{i_1}, \omega_{i_2}, \dots, \omega_{i_k}\} \subset \Omega$ определяется по формуле, называемой формулой классической вероятности,

$$P(A) = \frac{k}{n}$$

Доказательство. $P(A) = p_{i_1} + p_{i_2} + \dots + p_{i_k} = kp = \frac{k}{n}$ □

Если обозначить число k элементов множества A через $\mu(A)$, а число n элементов пространства Ω через $\mu(\Omega)$, то формулу классической вероятности можно записать в виде

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

Число $\mu(A) = k$ называется числом исходов, благоприятствующих наступлению события A .

Замечание 4.11. 1) Вероятности элементарных исходов в экспериментах с шариками, описанные в теоремах 1.4, 1.6 и 1.7, удовлетворяют принципу равной возможности. Эти вероятности соответственно $1/A_n^k$, $1/C_n^k$ и $1/n^k$.

2) Вероятности элементарных исходов в теореме 1.9 (выбор с возвращением и без учёта порядка) не удовлетворяют этому принципу. Например, при $n = 2$ и $k = 2$ вероятности элементарных событий имеют ряд распределения

ω	(1, 1)	(1, 2)	(2, 2)
$P(\omega)$	1/4	1/2	1/4

3) Следует отметить, что при рассмотрении подобных вопросов ошибались даже такие великие математики, как, например, Д'Аламбер.

Так, однажды у Даламбера спросили, с какой вероятностью монета, брошенная дважды, хотя бы один раз выпадет гербом. Ответ учёного был $\frac{2}{3}$, т.к. он считал, что есть 3 возможных исхода (герб-герб, герб-решка, решка-решка) и среди них 2 благоприятствующих.

Д'Аламбер пренебрегал тем, что эти три возможных исхода не равновозможны.

Правильным ответом является $\frac{3}{4}$, поскольку из четырёх равновозможных исходов (герб-герб, герб-решка, решка-герб, решка-решка) три благоприятствуют указанному событию.

Точка зрения Д'Аламбера была даже опубликована во Французской энциклопедии в 1754 г. в статье «Герб и решётка» («Croix on pile»).

4.4. Вероятность на счётном пространстве

Пусть теперь $\Omega = \{\omega_1, \omega_2, \dots, \omega_n, \dots\}$ – счётное пространство. Пусть $p_1 + p_2 + \dots + p_n + \dots = 1$ – сходящийся (к единице) числовой ряд, удовлетворяющий для всех $i \in \mathbb{N}$ условию $0 < p_i < 1$. Последнее условие позволяет трактовать члены этого ряда как вероятности элементарных событий пространства Ω .

Определение 4.12. Вероятность элементарных исходов, заданная в виде таблицы

ω	ω_1	ω_2	\dots	ω_i	\dots
$P(\omega)$	p_1	p_2	\dots	p_i	\dots

называется *рядом распределения на счётном пространстве*. Потребуем теперь выполнение аксиомы $\mathcal{P}3$.

Лемма 4.13 (Формула вероятности на счётном пространстве). Если события из пространства Ω подчиняются аксиоме $\mathcal{P3}$, то вероятность события $A = \{\omega_{i_1}, \omega_{i_2}, \dots, \omega_{i_k}, \dots\} \subset \Omega$ вычисляется по формуле

$$P(A) = p_{i_1} + p_{i_2} + \dots + p_{i_k} + \dots$$

Доказательство.

$$\begin{aligned} P(A) &= P(\omega_{i_1}, \omega_{i_2}, \dots, \omega_{i_k}, \dots) = \\ &= P\left(\bigcup_{j=1}^{\infty} \{\omega_{i_j}\}\right) \stackrel{\mathcal{P3}}{=} \sum_{j=1}^{\infty} P(\omega_{i_j}) = \sum_{j=1}^{\infty} p_{i_j} = p_{i_1} + p_{i_2} + \dots + p_{i_k} + \dots \end{aligned}$$

□

Замечание 4.14. Каждому степенному ряду на той части его интервала сходимости, на которой его члены положительны, можно поставить в соответствие параметрическое семейство распределений со счётным пространством. Рассмотрим, например, экспоненту e^λ . Её ряд МакЛорена

$$1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \dots + \frac{\lambda^k}{k!} + \dots = e^\lambda$$

сходится при всех значениях $-\infty < \lambda < \infty$ и имеет положительные члены при $\lambda > 0$. Умножим обе части этого тождества на $e^{-\lambda}$, получим тождество

$$e^{-\lambda} + \frac{\lambda}{1!}e^{-\lambda} + \frac{\lambda^2}{2!}e^{-\lambda} + \dots + \frac{\lambda^k}{k!}e^{-\lambda} + \dots = 1$$

Составим следующий ряд распределения:

k	0	1	2	...	k	...
$P_\lambda(k)$	$e^{-\lambda}$	$\frac{\lambda}{1!}e^{-\lambda}$	$\frac{\lambda^2}{2!}e^{-\lambda}$...	$\frac{\lambda^k}{k!}e^{-\lambda}$...

Определение 4.15. Распределение, реализуемое этим рядом, называется *распределением Пуассона с параметром λ* .

На рис. 25 показаны функции $P_\lambda(k)$ для параметра $\lambda = 0, 1; 1; 10$, где пунктиром показаны огибающие. В каждом случае сумма длин вертикальных отрезков равна 1.

Распределение Пуассона описывает, например, вероятность $P_\lambda(k) = \frac{\lambda^k e^{-\lambda}}{k!}$ поступления на телефонную станцию k звонков за какой-нибудь фиксированный промежуток времени, где число звонков $k = 0, 1, 2, \dots$.

Замечание 4.16. Другим источником построения распределений со счётным пространством являются сходящиеся числовые ряды с положительными членами. Например, известно (Л. Эйлер), что

$$\frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \dots + \frac{1}{k^2} + \dots = \frac{\pi^2}{6}.$$

Разделив обе части этого числового тождества на $\frac{\pi^2}{6}$, можно получить (безымянный) ряд распределения, задаваемый формулой $P(k) = \frac{1}{k^2} \cdot \frac{6}{\pi^2} = \frac{6}{\pi^2 k^2}$, со счётным пространством $\Omega = \mathbb{N}$. Среди числовых рядов наиболее встречающимися в теории вероятностей являются геометрические прогрессии с положительными членами.

Определение 4.17. Если члены ряда $p_1 + p_2 + \dots + p_n + \dots = 1$ положительны и являются членами убывающей геометрической прогрессией с первым членом $p_1 = p$ и с знаменателем $q = 1 - p$, то вероятность на счётном пространстве называется *геометрическим распределением*. Геометрическое распределение имеет ряд

τ_1	ω_1	ω_2	\dots	ω_k	\dots
P	q	qp	\dots	$q^{k-1}p$	\dots

Пример 4.18. Рассмотрим схему Бернулли (ломаного гроша) с вероятностью «успеха» (выпадения орла – события 1) равной p . Она имеет ряд распределения

ξ	0	1
P	$1 - p$	p

Пусть эксперимент состоит в том, что испытания по схеме Бернулли проводятся неограниченное число раз. С таким экспериментом связаны две случайные величины: τ_1 — номер первого выпавшего орла и τ_0 — число выпавших решек, появившихся до первого орла. Легко подсчитать, что случайные величины τ_0 и τ_1 имеют ряды распределения

τ_0	0	1	\dots	k	\dots
P	q	qp	\dots	$q^k p$	\dots

и

τ_1	1	2	\dots	k	\dots
P	q	qp	\dots	$q^{k-1} p$	\dots

В строке вероятностей эти ряды содержат одну и ту же геометрическую прогрессию; они называются соответственно τ_0 - и τ_1 -геометрическими распределениями. Эти случайные величины связаны очевидной формулой $\tau_1 = \tau_0 + 1$.

Формула вероятности для τ_0 – геометрического распределения определена по формуле

$$P(\tau_0 = k) = q^k p,$$

где $0 < p < 1$ и $q = p - 1$. Формула вероятности для τ_1 – геометрического распределения:

$$P(\tau_1 = k) = q^{k-1} p,$$

где $0 < p < 1$ и $q = p - 1$.

4.5. Геометрическая вероятность

Рассмотрим ограниченную измеримую область $\Omega \subset \mathbb{R}^n$, состоящую из несчётного множества точек.

Измеримость области означает: на прямой область $\Omega \subset \mathbb{R}^1$ имеет конечную ненулевую длину, на плоскости область $\Omega \subset \mathbb{R}^2$ имеет конечную ненулевую площадь, в пространстве область $\Omega \subset \mathbb{R}^3$ имеет конечный ненулевой объём и т.д.

Пусть на Ω определена σ -алгебра \mathfrak{A} . Для любого события $A \in \mathfrak{A}$ обозначим через $\mu(A)$ его меру (длину, площадь, объём и т.д. соответственно).

Пусть эксперимент состоит в том, что в область Ω бросают точку.

Принцип равномерности⁸. Если для любого события $A \in \mathfrak{A}$ его вероятность задаётся по формуле

$$P(A) = \alpha \mu(A),$$

где α — постоянное число, независящее от выбора события A (т.е. не зависящее от формы A и его расположения в Ω), то говорят, что вероятность $P(A)$ удовлетворяет принципу равномерности.

Лемма 4.19. Если вероятность $P(A)$ удовлетворяет принципу равномерности, то

$$\alpha = \frac{1}{\mu(\Omega)}$$

Доказательство. Подставим в формулу $P(A) = \alpha \mu(A)$ достоверное событие $\Omega \in \mathfrak{A}$, получим: $1 = \alpha \mu(\Omega)$. □

Определение 4.20. Полученная формула

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

⁸Вероятности, которые не подчиняются этому принципу, являются главным предметом изучения теории вероятностей. Они будут изучены в гл. 2. «Теория случайных величин».

называется формулой *геометрической* вероятности. (Не путать с геометрическим рядом распределения!)

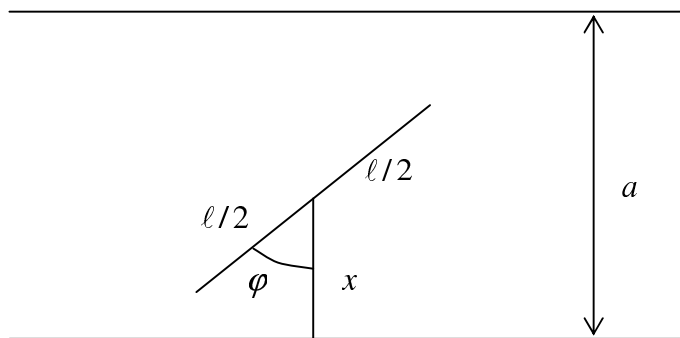


Рис. 3: Задача Бюффона

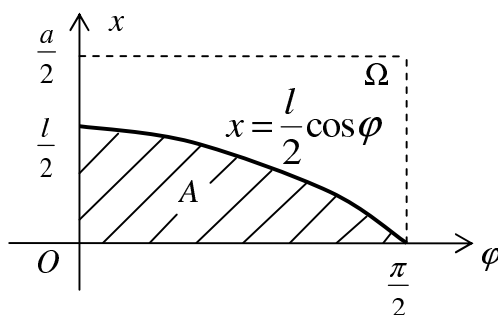


Рис. 4: Пространство элементарных событий Ω в задаче Бюффона

Пример 4.21 (Задача Бюффона (1777 г.)). На плоскости нарисовано счётное множество параллельных прямых. Расстояние между соседними прямыми равно a . На плоскость брошена игла длины $l < a$. Какова вероятность того, что игла пересечёт одну из прямых?

Решение. Возможные положения иглы на плоскости полностью определяются двумя координатами: расстоянием x от середины иглы до ближайшей прямой и острым углом φ между иглой и перпендикуляром к параллельным прямым, см. рис. 3. Ясно, что $x \in [0, \frac{a}{2}]$ и $\varphi \in [0, \frac{\pi}{2}]$. Поэтому множество $\Omega = [0, \frac{\pi}{2}] \times [0, \frac{a}{2}]$ есть пространство элементарных исходов этого эксперимента, см. рис. 4, и $\mu(\Omega) = \frac{\pi a}{4}$.

Событие $A = \{\text{игла пересечёт одну из прямых}\}$ эквивалентно неравенству $A = \{x \leq 2l \cos \varphi\}$, поэтому множество благоприятных исходов располагается в пространстве Ω под графиком $x = \frac{l}{2} \cos \varphi$. Вычислим

$$\mu(A) = \int_0^{\frac{\pi}{2}} \frac{l}{2} \cos \varphi d\varphi = \frac{l}{2} \sin \varphi \Big|_0^{\frac{\pi}{2}} = \frac{l}{2}$$

Отсюда получаем $P(A) = \frac{\mu(A)}{\mu(\Omega)} = \frac{2l}{\pi a}$

Пример 4.22 («Парадокс» Бертрана⁹, 1888). В круге наудачу выбирается хорда. Какова вероятность того, что её длина будет больше, чем длина стороны вписанного в круг правильного треугольника?

Решение. Обозначим через A событие, состоящее в том, что длина хорды будет больше, чем длина стороны вписанного в круг правильного треугольника. Существует по крайней мере три способа «выбрать наудачу» хорду в круге.

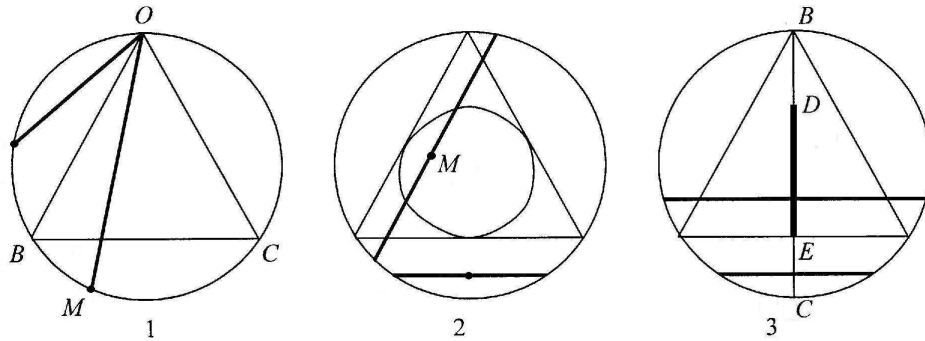


Рис. 5: Парадокс Бертрана

1-й способ. Зафиксируем один конец хорды O на окружности. Положение другого конца хорды M будем считать равномерно распределённым на окружности, см. рис. 5.1. Пусть координата конца M хорды есть длина дуги OM окружности, проходимой против часовой стрелки, тогда пространство элементарных событий Ω есть окружность, т.е. $\mu(\Omega) = 2\pi R$. Благоприятным исходом является положение конца хорды на дуге BC . Событие A всех благоприятных исходов есть дуга BC , которая составляет третью часть окружности, т.е. $\mu(A) = \frac{2\pi}{3}R$. Поэтому $P(A) = \frac{\mu(A)}{\mu(\Omega)} = \frac{1}{3}$.

2-й способ. Для каждой точки M внутри круга (кроме его центра¹⁰) существует единственная хорда, для которой точка M является её серединой, см. рис. 5.2.

Поэтому, бросая точку M в круг радиуса R , можно по ней однозначно восстановить хорду. Середину M хорды будем считать равномерно распределённой в круге. Пространство элементарных событий Ω есть круг радиуса R , его площадь $\mu(\Omega) = \pi R^2$.

Благоприятными событию A являются положения середины M хорды внутри окружности, вписанной в треугольник. Легко подсчитать, что радиус вписанной окружности равен $\frac{R}{2}$. Поэтому $\mu(A) = \frac{\pi R^2}{4}$ и $P(A) = \frac{\mu(A)}{\mu(\Omega)} = \frac{1}{4}$.

⁹ Жозеф Луи Франсуа Бертран (Joseph Louis Francois Bertrand, 1822 — 1900), французский математик.

¹⁰ Т.к. вероятность попадания точки в центр круга равна нулю, то наступление этого события не влияет на вероятность какого-либо события.

3-й способ. Можно ограничиться рассмотрением только хорд, которые перпендикулярны какому-нибудь диаметру, например $С$ (остальные положения могут быть получены поворотом), см. рис. 5.3.

Середину хорды будем считать равномерно распределённой на диаметре BC . Пространство элементарных событий Ω есть диаметр BC , которому перпендикулярны хорды, его длина $\mu(\Omega) = 2R$.

Благоприятными событию A являются положения середины хорды на отрезке DE , лежащем на диаметре BC , так что центр отрезка DE совпадает с центром круга. Можно видеть, что длина отрезка DE есть $\mu(A) = R$. Поэтому $P(A) = \frac{\mu(A)}{\mu(\Omega)} = \frac{1}{2}$.

Причина разных ответов заключается в том, что условие в круге наудачу выбирается хорда определяет эксперимент не однозначно. В решении мы провели три разных эксперимента по выбору хорды, и поэтому в каждом случае был получен правильный ответ.

5. Независимые события

Определение 5.1. События A и B называются *независимыми*, если выполняется тождество

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

в противном случае они называются *зависимыми*.

Примеры.

1) Рассмотрим колоду карт в 52 листа. Пусть событие A означает вытянуть пику \spadesuit , а B означает вытянуть даму D . Тогда событие $A \cap B$ означает вытянуть пиковую даму $D\spadesuit$. Легко подсчитать вероятности этих событий $P(A \cap B) = 1/52$; $P(A) = 1/4$ и $P(B) = 1/13$. Подставив эти вероятности в равенство опред. 5.1, получим тождество $1/52 = 1/4 \cdot 1/13$; т.е. по опред. 5.1 события A и B независимы (в этой колоде).

2) Рассмотрим полную колоду карт в 54 листа (с двумя шутами). Пусть события A и B — те же как в предыдущем пункте. Легко подсчитать, что в этой колоде $P(A \cap B) = 1/54$, $P(A) = 13/54$ и $P(B) = 2/27$. Таким образом $1/54 \neq 13/54 \cdot 2/27$, и по опред. 5.1 события A и B являются зависимыми в полной колоде. Получилось, что свойство быть или не быть независимыми зависит не от самих событий, а от строения пространства Ω (52 или 54).

Лемма 5.2. Если события A и B независимы, то пары событий A и \overline{B} , \overline{A} и B , \overline{A} и \overline{B} тоже являются независимыми.

Доказательство. Докажем, что события A и \overline{B} независимы. Так как $A = (A \cap B) \cup (A \cap \overline{B})$, и события $A \cap B$ и $A \cap \overline{B}$ несовместны, то $P(A) = P(A \cap$

$B) + P(A \cap \bar{B})$. Поэтому $P(A \cap \bar{B}) = P(A) - P(A \cap B) = P(A) - P(A)P(B) = P(A)(1 - P(B)) = P(A)P(\bar{B})$.

Независимость пар событий \bar{A} и B , A и \bar{B} доказывается аналогично. Доказать самостоятельно. \square

Определение 5.3. События A_1, \dots, A_n называются независимыми в совокупности, если для $1 \leq i_1 < \dots < i_k \leq n$ выполнено равенство

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \cdot \dots \cdot P(A_{i_k}).$$

Замечание 5.4. Если события A_1, \dots, A_n независимыми в совокупности, то они попарно независимы. Чтобы это увидеть, достаточно в последнем равенстве положить $k = 2$. Обратное, как показывает следующий пример, не верно.

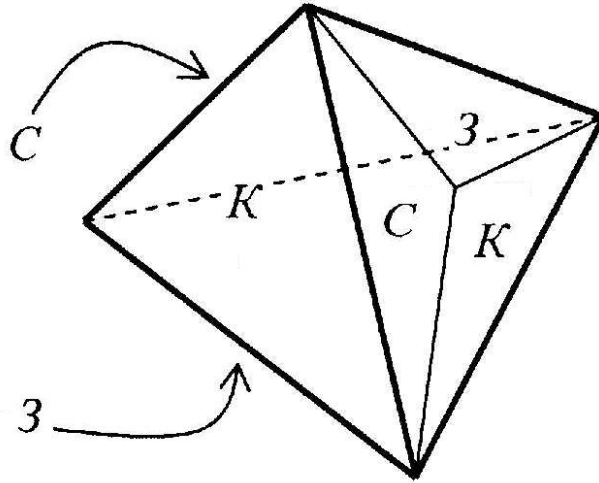


Рис. 6: Тетраэдр Бернштейна

Пример 5.5 (Тетраэдр Бернштейна¹¹). Рассмотрим правильный тетраэдр, три грани которого окрашены в эти синий, зелёный, и красный цвет, см. рис. 6, а четвёртая грань разделена на три треугольника, и эти треугольники окрашены в те же три цвета. Обозначим через B , G и R события означающие выпадение снизу грани, содержащей соответственно синий (blue), зелёный (green), и красный (red) цвета. Легко видеть, что каждый цвет нарисован на двух из четырёх граней поэтому $P(B) = P(G) = P(R) = 1/2$. Также легко видеть, что появление любой пары из них имеет вероятности $P(B \cap G) = P(G \cap R) = P(B \cap R) = 1/4$. Т.о., для каждой пары из этих событий формула определения 5.1 выполнена, и следовательно события попарно независимы.

¹¹Сергей Натанович Бернштейн (1880 — 1968), советский математик.

Теперь проверим независимость событий B, G и R в совокупности. Легко видеть, что вероятность $P(B \cap G \cap R)$ выпадения трёхцветной грани равна $1/4$. Это левая часть равенства определения 5.3. Правая часть равна $P(B) \cdot P(G) \cdot P(R) = 1/8$. Т.е. равенство определяемое леммой 5.2 не выполнено, значит события B, G и R зависимы в совокупности.

6. Условная вероятность

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

Решение. $\Omega_1 = \{4, 5, 6\}$, $A = \{4, 6\}$, и $P(A) = \frac{\mu(A)}{\mu(\Omega)} = \frac{2}{3}$.

Замечание 6.1. Пусть Ω – пространство элементарных событий и $B \subset \Omega$ – событие, отличное от невозможного, т.е. $B \neq \emptyset$. Пусть $A \subset \Omega$ – другое событие. Какова вероятность того, что произойдёт событие A , при условии, что событие B произошло? Слова *событие B произошло означают*, что новым пространством элементарных событий становится событие $\Omega_1 = B$ и его мера равна $\mu(\Omega_1) = \mu(B)$.

Слова *произойдёт событие A , если событие B произошло* означают ту часть события A , которая содержится в B , т.е. означают *произойдёт событие $A \cap B$* . Ясно, что $A \cap B \subset \Omega$ и $A \cap B \subset B = \Omega_1$.

Определение 6.2. Для того, чтобы подчеркнуть, что событие $A \cap B$ есть событие из нового пространства элементарных событий $\Omega_1 = B$ его обозначают $A|B$ и называют *событие A при условии, что событие B произошло*.

Очевидно, что $P(A|B) = \frac{\mu(A|B)}{\mu(B)}$. Вероятность $P(A|B)$ называется *условной вероятностью*.

Лемма 6.3. $P(A|B) = \frac{P(A \cap B)}{P(B)}$.

Доказательство.

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

□

Следующая формула непосредственно следует из леммы 6.3 и традиционно называется *теоремой умножения*.

Теорема 6.4 (Теорема умножения для двух событий). Если $P(B) > 0$, $P(A) > 0$, то

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

Теорема 6.5 (Теорема умножения для n событий).

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \dots P(A_n|A_1 \cap A_2 \cap \dots \cap A_{n-1}),$$

если все условные вероятности определены.

Доказательство. Доказать методом математической индукции. □

7. Формула полной вероятности и формулы Байеса

Пример. Три завода производят одну и ту же машину Renault. При этом 1-й завод производит 20%, 2-й — 30%, 3-й — 50% автомобилей. Брак на 1-м заводе составляет 5%, на 2-м — 2%, на 3-м — 4% автомобилей. Автомашины случайным образом поступают в продажу. Найти

а) вероятность купить бракованную машину и

б) условную вероятность того, что машина изготовлена на 1-м заводе.

Решение. а) Т.к. сборка машин на заводах не зависима (замеч. 6.1), и сборка машин на них не совместна (аксиома $\mathcal{P3}$), то вероятность купить бракованную машину равна доле бракованных машин во всей продукции, то есть $0,05 \cdot 0,2 + 0,02 \cdot 0,3 + 0,04 \cdot 0,5 = 0,036$.

б) Во этом случае условная вероятность покупки бракованной машины с 1-го завода равна доле брака 1-го завода среди всего количества бракованных машин

$$\frac{0,05 \cdot 0,2}{0,05 \cdot 0,2 + 0,02 \cdot 0,3 + 0,04 \cdot 0,5} = \frac{0,01}{0,036} = 0,2(7).$$

Определение 7.1. Если события $H_1, H_2, \dots \subset \Omega$

1) попарно несовместны (т.е. $H_i \cap H_j = \emptyset$ при $i \neq j$),

2) их объединение составляет всё пространство элементарных событий, т.е.

$$\bigcup_{i=1}^{\infty} H_i = \Omega,$$

3) $P(H_i) > 0$ для всех $i \in \mathbb{N}$,

то совокупность $\{H_1, H_2, \dots\}$ называется *полной группой событий*, а события H_1, H_2, \dots называются *гипотезами*.

На рис. 7 показана полная группа событий и некоторое событие A в пространстве Ω .

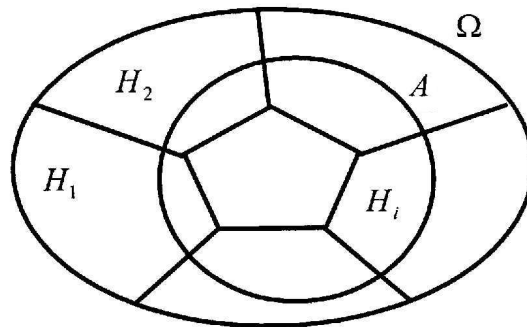


Рис. 7: Полная группа событий и некоторое событие A в пространстве Ω

Теорема 7.2 (Формула полной вероятности). Пусть H_1, H_2, \dots – полная группа событий. Тогда вероятность любого события $A \subset \Omega$ вычисляется по формуле

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

Доказательство. Заметим, что $A = A \cap \Omega = A \cap \left(\bigcup_{i=1}^{\infty} H_i \right) = \bigcup_{i=1}^{\infty} A \cap H_i$, где события $A \cap H_1, A \cap H_2, \dots$ – попарно несовместны. Используя аддитивность вероятности (аксиома $\mathcal{P}3$), а затем теорему умножения 6.3, получим требуемый результат

$$P(A) = \sum_{i=1}^{\infty} P(A \cap H_i) = \sum_{i=1}^{\infty} P(H_i)P(A|H_i).$$

□

Теорема 7.3 (формулы Байеса¹²). Пусть H_1, H_2, \dots – полная группа событий и A – некоторое событие положительной вероятности. Тогда условная вероятность события H_k при условии, что событие A произошло, для любого k вычисляется по формуле

$$P(H_k|A) = \frac{P(H_k)P(A|H_k)}{\sum_{i=1}^{\infty} P(H_i)P(A|H_i)}.$$

¹²Томас Байес (Reverend Thomas Bayes, 1702 – 1761), английский математик и пресвитерианский священник.

Доказательство. По определению условной вероятности имеем

$$P(H_k|A) = \frac{P(A \cap H_k)}{P(A)} = \frac{P(H_k)P(A|H_k)}{\sum_{i=1}^{\infty} P(H_i)P(A|H_i)}.$$

где последнее равенство следует из теоремы умножения и формулы полной вероятности \square

8. Биномиальное распределение

Рассмотрим эксперимент, состоящим в n -кратном повторении схемы Бернулли (т.е. в подбрасывании ломаного гроша n раз). Вопрос: «Какова вероятность того, что в результате этих n испытаний «успех» (орёл) выпадет ровно k раз, а остальные $n - k$ раз выпадет «неудача» (решка)?»

Обозначим искомую вероятность (выпадения орла) через $P_n(k)$. Ясно, что этот эксперимент имеет конечное пространство элементарных событий $\Omega = \{0, 1, 2, \dots, n\}$ – орёл выпал 0 раз, 1 раз, 2 раза, \dots , n раз.

Теорема 8.1. Для любого $k \in \Omega$ имеет место формула Бернулли

$$P_n(k) = C_n^k p^k (1 - p)^{n-k} = C_n^k p^k q^{n-k}.$$

Доказательство. Рассмотрим один из благоприятных элементарных исходов:

$$\underbrace{(y, y, \dots, y)}_{k \text{ раз}}, \underbrace{(H, H, \dots, H)}_{n-k \text{ раз}}$$

Здесь буквами «у» и «н» обозначены соответственно «успех» и «неудача». Поскольку испытания независимы, вероятность элементарного исхода (первые k испытаний завершились успехом, а остальные неудачей) по опред. 5.1 равна $p^k (1 - p)^{n-k}$.

Другие элементарные исходы, благоприятные событию A , отличаются от рассмотренного $\underbrace{(y, y, \dots, y)}_{k \text{ раз}}, \underbrace{(H, H, \dots, H)}_{n-k \text{ раз}}$ лишь перераспределением k успехов на n местах. По теор. 1.6 существует ровно C_n^k способов расположить k успехов на n местах. Поэтому событие A состоит из C_n^k элементарных исходов, вероятность каждого из них равна $p^k (1 - p)^{n-k}$. \square

Определение 8.2. Эксперимент, имеющий ряд распределения

k	$P_n(k)$
0	$C_n^0 p^0 q^n = q^n$
1	$C_n^1 p^1 q^{n-1} = npq^{n-1}$
2	$C_n^2 p^2 q^{n-2} = \frac{n(n-1)}{2} p^2 q^{n-2}$
...	...
k	$C_n^k p^k q^{n-k} = \frac{n!}{k!(n-k)!} p^k q^{n-k}$
...	...
n	$C_n^n p^n = p^n$

называется биномиальным распределением.

Название этого распределения связан с тем, что сумма всех вероятностей из второй строки ряда распределения может быть вычислены по формуле бинома Ньютона, т.е.

$$\sum_{k=0}^n C_n^k p^k q^{n-k} = (p + q)^n = 1$$

На рис. 8 показан график биномиального распределения $P_{16}(k) = C_{16}^k \cdot 0.65^k \cdot 0.35^{16-k}$. График является симметричным при $p = q = \frac{1}{2}$.

Если $p > q$, то максимум сдвигается вправо, и наоборот. Возникает естественный вопрос. Какое число успехов при n испытаниях наиболее вероятно? Другими словами, при каком k достигается максимум функции $P_n(k) = C_n^k p^k q^{n-k}$? Другими словами, при каком (каких) k вероятность достигает максимума?

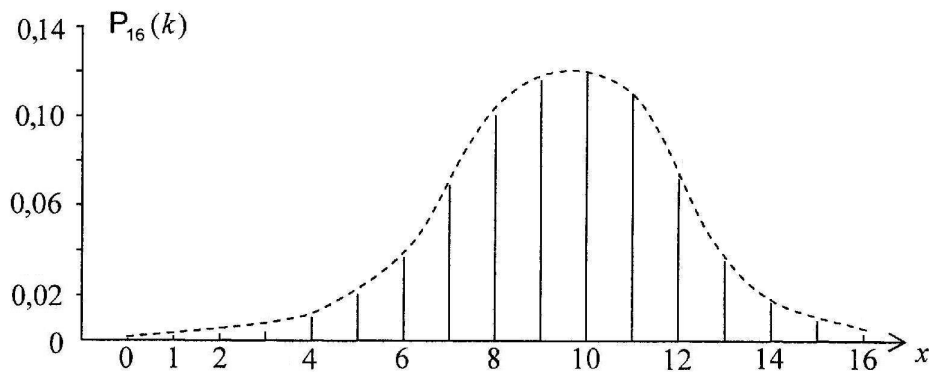


Рис. 8: График биномиального распределения при $n = 16$

Теорема 8.3. Если в схеме Бернулли вероятность появления «успеха» (орла) равна p , то в биномиальном распределении с n испытаниями наиболее вероятным числом «успехов» (орлов) является либо

- единственное число $[np + p]$, если число $np + p$ не целое, либо
- два числа $np + p$ и $np + p + 1$, если число $np + p$ целое.

Доказательство. Сравним отношение чисел $P_n(k)$ и $P_n(k - 1)$ с единицей.

$$\frac{P_n(k)}{P_n(k - 1)} = \frac{C_n^k p^k q^{n-k}}{C_n^{k-1} p^{k-1} q^{n-k+1}} = \frac{p(n - k + 1)}{kq} = 1 + \frac{np + p - k}{kq}$$

Видно, что

1. $P_n(k) > P_n(k - 1)$ при $np + p - k > 0$, т.е. при $k < np + p$;
2. $P_n(k) < P_n(k - 1)$ при $np + p - k < 0$, т.е. при $k > np + p$;
3. $P_n(k) = P_n(k - 1)$ при $np + p - k = 0$, что возможно лишь, если $np + p$ — целое число.

□

9. k -номинальное распределение

Полиномиальное распределение возникает в результате повторения n раз схемы Бернулли (см. §8). По аналогии k -номинальное распределение возникает в результате повторения n раз схемы независимых испытаний с $k \geq 3$ исходами (см. пп. 4.5 и 4.9).

Пример. Асимметричный тетраэдр, грани которого обозначены $\omega_1 - \omega_4$ подбрасывают 14 раз. Вероятности выпадения этих граней лицом вниз соответственно равны $\frac{1}{2}, \frac{1}{3}, \frac{1}{9}, \frac{1}{18}$. Найти вероятность того, что грани $\omega_1, \omega_2, \omega_3, \omega_4$ выпадут соответственно 5, 3, 4, 2 раза.

Ответ: $\frac{14!}{5!3!4!2!} (\frac{1}{2})^5, (\frac{1}{3})^3, (\frac{1}{9})^4, (\frac{1}{18})^2 \simeq 0,00137$

Замечание 9.1. Известно, что биномом Ньютона называют формулу

$$(p + q)^n = \sum_{m=0}^n C_n^m p^m q^{n-m},$$

где $C_n^m = \frac{n!}{m!(n-m)!}$

Обобщением бинома Ньютона является следующая формула

$$(p_1 + p_2 + \dots + p_k)^n = \sum_{m_1+m_2+\dots+m_k} \frac{n!}{m_1!m_2!\dots m_k!} p_1^{m_1} p_2^{m_2} \dots p_k^{m_k},$$

где $m_1, m_2, \dots, m_k \geq 0$. Эта формула может быть выведена из бинома Ньютона индукцией по k ; по аналогии будем называть её k -номом Ньютона.

Лемма 9.2. Сумма k -номиальных коэффициентов равна k^n , т.е.

$$\sum_{m_1+m_2+\dots+m_k} \frac{n!}{m_1!m_2!\dots m_k!} = k^n$$

Доказательство. Подставим в k -ном Ньютона $p_1 = p_2 = \dots = p_k = 1$, получим требуемый результат. \square

Лемма 9.3. Если $p_1 + p_2 + \dots + p_k = 1$, то сумма членов в правой части k -нома Ньютона равна единице, т.е.

$$\sum_{m_1+m_2+\dots+m_k=n} \frac{n!}{m_1!m_2!\dots m_k!} p_1^{m_1} p_2^{m_2} \dots p_k^{m_k} = 1$$

Доказательство. Подставим в k -ном Ньютона $p_1 + p_2 + \dots + p_k = 1$, получим требуемый результат. \square

Определение 9.3. Пусть эксперимент состоит в том, что

- 1) проводят n независимых испытаний в одинаковых условиях,
- 2) в каждом испытании появляется одно из k несовместных событий

$$A_1, A_2, \dots, A_k,$$

- 3) эти события происходят с вероятностями p_1, p_2, \dots, p_k соответственно,
- и 4) $p_1 + p_2 + \dots + p_k = 1$.

Тогда такой эксперимент называется схемой независимых испытаний.

Теорема 9.4. Для любой схемы n независимых испытаний и любых $m_1 \geq 0, m_2 \geq 0, \dots, m_k \geq 0$, таких что $m_1 + m_2 + \dots + m_k = n$, вероятность $P(m_1, m_2, \dots, m_k)$ того, что события A_1, A_2, \dots, A_k произойдут соответственно m_1, m_2, \dots, m_k раз, определяется по формуле

$$P(m_1, m_2, \dots, m_k) = \frac{n!}{m_1!m_2!\dots m_k!} p_1^{m_1} p_2^{m_2} \dots p_k^{m_k}.$$

Доказательство. Рассмотрим один элементарный исход схемы n независимых испытаний:

$$\underbrace{(A_1, \dots, A_1)}_{m_1 \text{ раз}}, \underbrace{(A_2, \dots, A_2)}_{m_2 \text{ раз}}, \dots, \underbrace{(A_k, \dots, A_k)}_{m_k \text{ раз}}$$

Это один из благоприятных исходов: сначала событие A_1 произошло m_1 раз, затем событие A_2 произошло m_2 раз, \dots , и, наконец, событие A_k произошло m_k раз. Вероятность этого элементарного исхода равна $p_1^{m_1} p_2^{m_2} \dots p_k^{m_k}$.

Все остальные благоприятные исходы отличаются лишь расположением событий из того же набора событий на n местах. Число таких исходов равно числу способов расставить на n местах m_1 событий A_1 , потом m_2 событий A_2, \dots , и, наконец, m_k событий A_k . По теор. 1.3 и 1.6 это число равно

$$C_n^{m_1} \cdot C_{n-m_1}^{m_2} \cdot C_{n-m_1-m_2}^{m_3} \cdot \dots \cdot C_{n-m_1-\dots-m_{k-1}}^{m_k} = [\text{проверьте это!}] = \frac{n!}{m_1!m_2!\dots m_k!}$$

□

Теперь становится ясно, откуда взялся ответ задачи про асимметричный тетраэдр в начале этого параграфа.

10. Гипергеометрическое распределение

Задача 10.1. Из урны, содержащей n_1 белых и $n - n_1$ чёрных шаров, вынимают без возвращения $k \leq n$ шаров. Найти вероятность события A , состоящего в том, что будет вынуто ровно k_1 белых и $k - k_1$ чёрных шаров. (Мы полагаем, что $k_1 \leq n_1$ и $k - k_1 \leq n - n_1$.)

Решение. Результатом эксперимента является набор из k шаров. Оказывается, что искомая вероятность не зависит от того, будем ли мы или не будем учитывать порядок следования шаров. Чтобы показать это, подсчитаем её для обоих экспериментов.

1. Эксперимент без учёта порядка. В этом эксперименте общее число $\mu(\Omega)$ элементарных исходов равно числу k -элементных подмножеств множества, состоящего из n элементов.

По теореме 1.6 оно равно $\mu(\Omega) = C_n^k$. Обозначим через A событие, состоящее в появлении набора, содержащего k_1 белых шаров и $k - k_1$ чёрных.

По теореме 1.3 число его благоприятных исходов равно произведению числа способов выбрать k_1 шаров из n_1 белых шаров и числа способов выбрать $k - k_1$ шаров из $n - n_1$ чёрных, т.е. $\mu(A) = C_{n_1}^{k_1} \cdot C_{n-n_1}^{k-k_1}$.

Получаем, что вероятность появления события A в

$$P(A) = \frac{C_{n_1}^{k_1} \cdot C_{n-n_1}^{k-k_1}}{C_n^k}$$

2. Эксперимент с учётом порядка. По теореме 1.4 общее число $\mu(\Omega)$ элементарных исходов равно числу способов разместить n элементов на k местах, т.е. $\mu(\Omega) = A_n^k = n(n-1)\dots(n-k+1)$.

В этом эксперименте при подсчёте числа благоприятных исходов $\mu(A)$ надо учесть как число способов выбрать нужное число k шаров, так и число способов расположить белые и чёрные шары среди выбранных k шаров.

Во-первых, мы подсчитываем число способов расположить k_1 шаров на k местах. Оно равно $C_k^{k_1}$.

Затем мы подсчитываем число способов расположить k_1 белых шаров на n_1 местах. Учитывая их порядок, получаем, что это число равно $A_{n_1}^{k_1}$.

И наконец, мы подсчитываем число способов разместить $k - k_1$ чёрных шаров на оставшихся $n - n_1$ местах. Оно равно $A_{n-n_1}^{k-k_1}$.

Перемножая эти числа, по теореме 1.4 получим

$$\mu(A) = C_k^{k_1} \cdot A_{n_1}^{k_1} \cdot A_{n-n_1}^{k-k_1}$$

Подсчитывая искомую вероятность, получим

$$P(A) = \frac{C_k^{k_1} \cdot A_{n_1}^{k_1} \cdot A_{n-n_1}^{k-k_1}}{A_n^k} = \frac{C_{n_1}^{k_1} \cdot C_{n-n_1}^{k-k_1}}{C_n^k}$$

Определение 10.2. Ряд распределения, определённый соответствием

$$k_1 \mapsto P(A) = \frac{C_{n_1}^{k_1} \cdot C_{n-n_1}^{k-k_1}}{C_n^k},$$

где $0 \leq k_1 \leq \min(k, n_1)$ и $k - k_1 \leq n - n_1$ называется *гипергеометрическим распределением*.

Часть II

Теория случайных величин

11. Случайные величины

На практике элементарные случайные события являются либо числами, либо наборами чисел (случайными векторами). В §§ 11–13 мы изучим одномерные случайные величины, у которых $\Omega \subset \mathbb{R}^1$ — числовое множество. Начиная с §13 изучим случайные векторы, у которых $\Omega \subset \mathbb{R}^n$, при $n \geq 2$.

Дадим общее определение случайной величины. Пусть $(\Omega, \mathcal{A}, \mathcal{P})$ — произвольное вероятностное пространство.

Определение 11.1. Функция $\xi : \Omega \rightarrow \mathbb{R}$ называется случайной величиной на σ -алгебре событий \mathcal{A} , если для любого числа $x \in \mathbb{R}$ прообраз луча $\xi^{-1}((-\infty, x]) = \{\omega \in \Omega | \xi(\omega) \leq x\}$ является событием из σ -алгебры \mathcal{A} .

Замечание 11.2. 1) Самая простая функция $\xi : \Omega \rightarrow \mathbb{R}$ — это постоянная функция, заданная для любого $\omega \in \Omega$ по формуле $\xi(\omega) = c$. Она принимает одно значение и является не случайной, а детерминированной. Она рассматривается в теории вероятностей как частный тривиальный случай.

2) Первая нетривиальная случайная величина (с $\Omega \in \mathbb{R}$) задаётся с помощью тождественной функции $\xi(\omega) = \omega$. В этом случае событие $\xi^{-1}((-\infty, x]) = \{\omega \in \Omega | \xi(\omega) \leq x\}$ обозначают короче: $\{\xi \leq x\}$.

3) Оказывается, что другие случайные величины ($\xi(\omega) \neq \omega$) могут быть описаны через случайную величину $\xi(\omega) = \omega$. Это будет показано в § 13.

Определение 11.3. Функцией распределения случайной величины ξ называется функция $F_\xi(x) : \mathbb{R} \rightarrow [0, 1]$, определённая по формуле

$$F_\xi(x) = P(\xi \leq x)$$

Очевидно, что $0 \leq F_\xi(x) \leq 1$

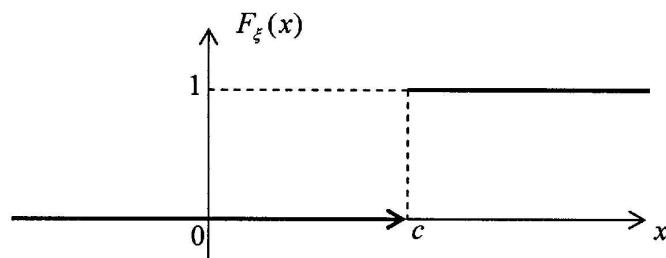


Рис. 9: Функция распределения детерминированной величины.

Примеры

1) Детерминированная (вырожденная случайная) величина $\xi : \Omega \rightarrow \mathbb{R}$ имеет пространство элементарных событий, состоящее из одного элементарного события, $\Omega = \{\omega\}$. Она принимает только одно значение $\xi(\omega) = c = \text{const} \in \mathbb{R}$ с вероятностью равной 1, т.е. имеет ряд распределения

ξ	c
P	1

. Её функция распределения показана на рис. 9.

2) Случайная величина ξ , имеющая распределение Бернулли, и принимающая значения 1 (успех) и 0 (неудача) с вероятностями соответственно p и $1 - p$, имеет ряд распределения

ξ	0	1
P	$1 - p$	p

 и имеет график, показанный на рис. 10.

3) Случайная величина ξ — номер грани при подбрасывании игральной кости имеет функцию распределения, показанную на рис. 11.

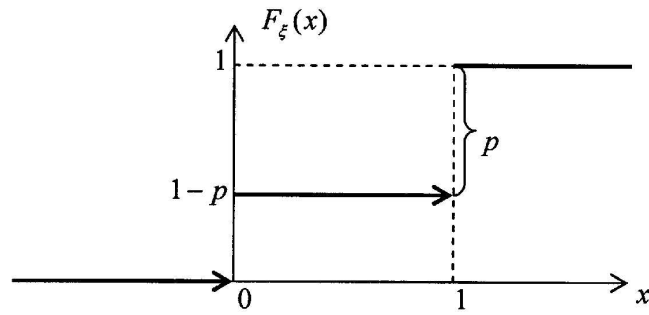


Рис. 10: Функция распределения Бернулли.

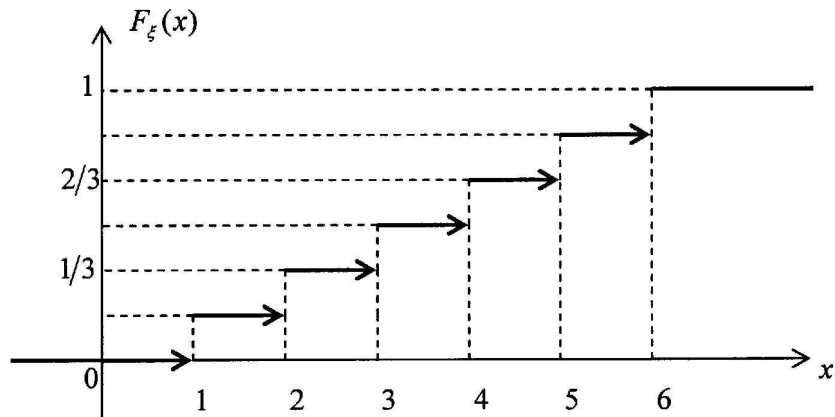


Рис. 11: Функция распределения выпадения числа на игральной кости.

4) Случайная величина ξ — номер появления успеха в геометрическом распределении имеет ряд распределения

ξ	1	2	...	k	...
P	p	pq	...	pq^{k-1}	...

и имеет функцию распределения, показанную на рис. 12. В дальнейшем мы будем использовать следующие обозначения для пределов функций слева и справа соответственно:

$$\lim_{\varepsilon \rightarrow 0} h(x - \varepsilon) \stackrel{\text{опр}}{=} h(x - 0) \text{ и } \lim_{\varepsilon \rightarrow 0} h(x + \varepsilon) \stackrel{\text{опр}}{=} h(x + 0),$$

где всегда $\varepsilon > 0$.

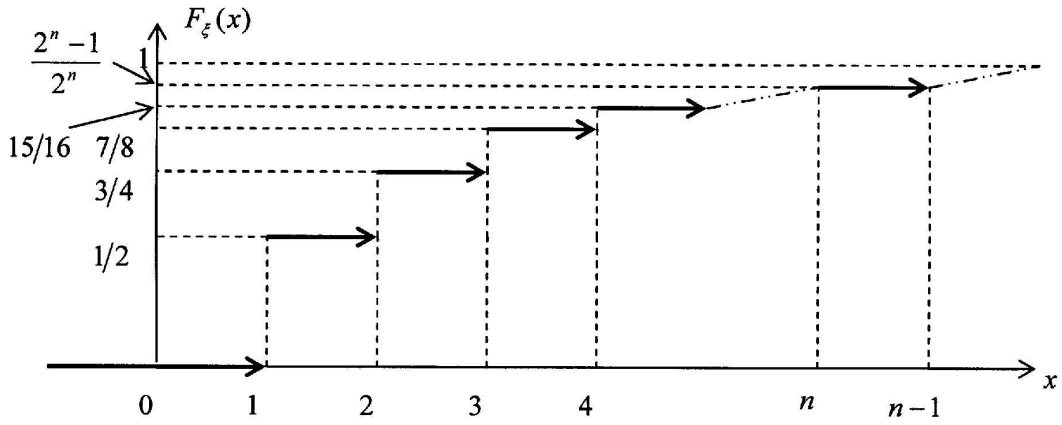


Рис. 12: Функция распределения выпадения числа на игральной кости.

Функция распределения $F_\xi(x)$ обладает следующими свойствами.

Теорема 11.4. 1. $F_\xi(x)$ — неубывающая функция, другими словами, если $x_1 < x_2$, то $F_\xi(x_1) \leq F_\xi(x_2)$.

2. $F_\xi(-\infty) = \lim_{x \rightarrow -\infty} F_\xi(x) = 0$ и $F_\xi(+\infty) = \lim_{x \rightarrow +\infty} F_\xi(x) = 1$

3. $F_\xi(x)$ непрерывна справа: для любой точки $x_0 \in \mathbb{R}$ имеем

$$F_\xi(x + x_0) = F_\xi(x_0)$$

Доказательство. По пп. 11.3, 3.3, 3.1 имеем

1. $F_\xi(x_2) - F_\xi(x_1) = P(\xi \leq x_2) - P(\xi \leq x_1) = P\{\xi \leq x_2\} \setminus P\{\xi \leq x_1\} = P(x_1 < \xi \leq x_2) \geq 0.$

2. $F_\xi(-\infty) = \lim_{x \rightarrow -\infty} F_\xi(x) = \lim_{x \rightarrow -\infty} P(\xi \leq x) = P(\xi \leq -\infty) = 0$ и $F_\xi(+\infty) = \lim_{x \rightarrow +\infty} F_\xi(x) = \lim_{x \rightarrow +\infty} P(\xi \leq x) = P(\xi \leq +\infty) = 1.$

3. $F_\xi(x_0 + 0) = \lim_{\varepsilon \rightarrow 0} F_\xi(x_0 + \varepsilon) = \lim_{x \rightarrow \varepsilon} P(\xi \leq x_0 + \varepsilon) = P(\xi \leq x_0) = F_\xi(x_0).$

□

Теорема 11.5. Функция распределения имеет не более чем счётное множество скачков.

Доказательство. Разобьём множество \mathbb{N} всех скачков на подмножества скачков по убыванию высоты h :

1) H_1 — множество скачков высоты $h \in (\frac{1}{2}, 1]$,

- 2) H_2 – множество скачков высоты $h \in (\frac{1}{4}, \frac{1}{2}]$,
 3) H_2 – множество скачков высоты $h \in (\frac{1}{8}, \frac{1}{4}]$,
 ...
 n) – множество скачков высоты $h \in (\frac{1}{2^n}, \frac{1}{2^{n-1}}]$
 ...

Видно, что множества H_n для всех $n \in \mathbb{Z}$ попарно непересекающиеся, т.е. $H_i \cap H_j = \emptyset$ при $i \neq j$; и $\bigcup_{i=1}^{\infty} H_i = H$. Функция распределения может иметь

Во-первых, не более одного скачка высоты $h \in (\frac{1}{2}, 1]$, т.е. $\mu(H_1) \leq 1$;

Во-вторых, не более трёх скачков высоты $h \in (\frac{1}{4}, \frac{1}{2}]$, т.е. $\mu(H_2) \leq 3$;

В-третьих, не более семи скачков высоты $h \in (\frac{1}{8}, \frac{1}{4}]$, т.е. $\mu(H_3) \leq 7$;

...

В- n -ых, не более $2^n - 1$ скачков высоты $h \in (\frac{1}{2^n}, \frac{1}{2^{n-1}}]$, т.е. $\mu(H_n) \leq 2^n - 1$;

...

Можно занумеровать все скачки по убыванию их высоты, нумеруя при этом скачки равной высоты, если такие повстречаются. Эта нумерация ставит множество всех скачков во взаимно однозначное соответствие с множеством \mathbb{Z} натуральных чисел. Что и требовалось доказать.

□

Теорема 11.6 (О связи вероятности событий-интервалов с функцией распределения). *Для любых точек $x, a, b \in \mathbb{R}$, где $a < b$, имеем*

1. $P(\xi < x) = F_\xi(x - 0)$,
2. $P(\xi = x) = F_\xi(x) - F_\xi(x - 0)$,
3. если F_ξ непрерывна в точке x , то $P(\xi = x) = 0$;
4. $P(a < \xi \leq b) = F_\xi(b) - F_\xi(a)$,
5. $P(a \leq \xi \leq b) = F_\xi(b) - F_\xi(a - 0)$,
6. $P(a < \xi < b) = F_\xi(b - 0) - F_\xi(a)$
7. $P(a \leq \xi < b) = F_\xi(b - 0) - F_\xi(a - 0)$.

Доказательство. 1. $P(\xi < x) = \lim_{\varepsilon \rightarrow 0} P(\xi \leq x - \varepsilon) = \lim_{\varepsilon \rightarrow 0} F_\xi(x - \varepsilon) = F_\xi(x - 0)$.

2. $P(\xi = x) = P(\{\xi \leq x\} \setminus \{\xi < x\}) = P(\xi \leq x) - P(\xi < x) = F_\xi(x) - F_\xi(x - 0)$.

3. $P(\xi = x) = F_\xi - F_\xi(x - 0) = F_\xi(x) - \lim_{\varepsilon \rightarrow 0} F_\xi(x - \varepsilon) = F_\xi(x) - F_\xi(x) = 0.$
4. $P(a < \xi \leq b) = P(\{\xi \leq b\}) \setminus \{\xi \leq a\} = P(\xi \leq b) - P(\xi \leq a) = F_\xi(b) - F_\xi(a).$
5. Самостоятельно
6. Самостоятельно
7. Самостоятельно

□

Следствие 11.7. Если функция распределения $F_\xi(x)$ непрерывна, то

$$P(a \leq \xi < b) = P(a < \xi < b) = P(a \leq \xi \leq b) = P(a < \xi \leq b) = F_\xi(b) - F_\xi(a).$$

12. Абсолютно непрерывные случайные величины

Определение 12.1. Случайная величина ξ называется абсолютно непрерывной, если существует такая неотрицательная функция $f_\xi(x)$ (возможно обобщённая, см. §13), что для любого $x \in \mathbb{R}$ функция распределения $F_\xi(x)$ представима в виде¹³

$$F_\xi(x) = \int_{-\infty}^x f_\xi(t) dt$$

При этом функция $f_\xi(x)$ называется плотностью вероятности случайной величины ξ

Замечание 12.2. Если $F_\xi(x)$ – дифференцируемая функция распределения, то название «плотность вероятности» имеет следующее объяснение. С одной стороны, используя формулу дифференцирования интеграла по верхнему пределу, имеем $\frac{dF_\xi(x)}{dx} = f_\xi(x)|_{t=x}$. С другой стороны, по определению производной имеем:

$$\begin{aligned} \frac{dF_\xi(x)}{dx} &= \frac{F_\xi(x + dx) - F_\xi(x)}{dx} = \frac{P(\xi \leq x + dx) - P(\xi \leq x)}{dx} = \\ &= \frac{P(\{\xi \leq x + dx\} \setminus \{\xi \leq x\})}{dx} = \frac{P(x < \xi \leq x + dx)}{dx} \end{aligned}$$

¹³В дальнейшем мы рассматриваем только такие случайные величины, для которых все связанные с ними несобственные интегралы и суммы бесконечных рядов сходятся абсолютно.

Поэтому

$$f_{\xi}(x) = \frac{P(x, \xi \leq x + dx)}{dx}$$

есть отношение вероятности попадания случайной величины ξ в интервал $(x, x + dx]$ к длине этого интервала dx , что физически означает плотность вероятности («массы») случайной величины ξ в точке x .

Теорема 12.3. *Плотность обладает свойствами:*

1) $f_{\xi}(x) \geq 0$ для любого x ;

2) $\int_{-\infty}^{\infty} f_{\xi}(t)dt = 1$

Доказательство. 1) По теор. 11.4.1 функция распределения $F_{\xi}(x)$ – неубывающая, поэтому $\frac{dF_{\xi}(x)}{dx} = f_{\xi}(x) \geq 0$

2) $\int_{-\infty}^{\infty} f_{\xi}(t)dt = \lim_{x \rightarrow \infty} \int_{-\infty}^x f_{\xi}(t)dt = \lim_{x \rightarrow \infty} F_{\xi}(x) = 1$ по свойству 11.4.2. \square

Определение 12.4. Пусть область $U \subseteq \mathbb{R}^n$ имеет меру $\mu(U)$. Говорят, что в области U функция $h : U \rightarrow \mathbb{R}$ удовлетворяет некоторому свойству \mathcal{P} почти всюду, если подмножество $A \subset U$, в котором свойство \mathcal{P} не выполняется, имеет меру 0, т.е. $\mu(A) = 0$. Докажем свойства абсолютно непрерывных случайных величин.

Теорема 12.5. *Если случайная величина ξ абсолютно непрерывна, то*

1) её функция распределения $F_{\xi}(x)$ непрерывна в \mathbb{R} и дифференцируема почти всюду в \mathbb{R} , т.е. равенство $f_{\xi}(x) = \frac{d}{dx}F_{\xi}(x)$ справедливо почти всюду,

2) $P(\xi = x) = 0$ для любого $x \in \mathbb{R}$,

3) $P(a \leq \xi < b) = P(a < \xi < b) = P(a \leq \xi \leq b) = P(a < \xi \leq b) = \int_a^b f_{\xi}(t)dt$.

Доказательство. 1) Во-первых, функция $F_{\xi}(x) = \int_{-\infty}^x f_{\xi}(t)dt$ непрерывна как функция верхнего предела интеграла. Во-вторых, по теор. 11.5 функция $F_{\xi}(x)$ имеет не более чем счётное множество скачков, поэтому она не дифференцируема не более чем в счётном множестве точек. Любое счётное множество точек на прямой имеет нулевую меру (длину). Поэтому $F_{\xi}(x)$ дифференцируема почти всюду в \mathbb{R} .

2) Это следует из 11.6.3).

3) Из 11.6.4) имеем

$$P(a < \xi \leq b) = F_{\xi}(b) - F_{\xi}(a) = \int_{-\infty}^b f_{\xi}(t)dt - \int_{-\infty}^a f_{\xi}(t)dt = \int_a^b f_{\xi}(t)dt$$

Остальные равенства следуют из 11.7.

Примеры абсолютно непрерывных случайных величин

Определение 12.6. Случайная величина ξ имеет равномерное распределение на отрезке a, b , если

$$F_{\xi}(x) = P(\xi \leq x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x > b \end{cases} \text{ и } f_{\xi}(x) = \begin{cases} 0, & x < a \\ \frac{1}{b-a}, & a \leq x \leq b \\ 1, & x > b \end{cases}$$

Графики функций $F_{\xi}(x)$ и $f_{\xi}(x)$ показаны на рис. 13. Заметим, что в точках a и b функция распределения $F_{\xi}(x)$ не дифференцируема, поэтому значение плотности вероятности в этих точках можно задать как угодно: либо $f_{\xi}(a) = 0$, либо $f_{\xi}(a) = \frac{1}{b-a}$, а на другом конце тоже либо $f_{\xi}(b) = 0$, либо $f_{\xi}(b) = \frac{1}{b-a}$.

Определение 12.7. Случайная величина ξ имеет показательное распределение с параметром λ (рис. 14), если

$$f_{\xi}(x) = \begin{cases} 0, & x < 0 \\ \lambda e^{-\lambda x}, & x \geq 0 \end{cases}$$

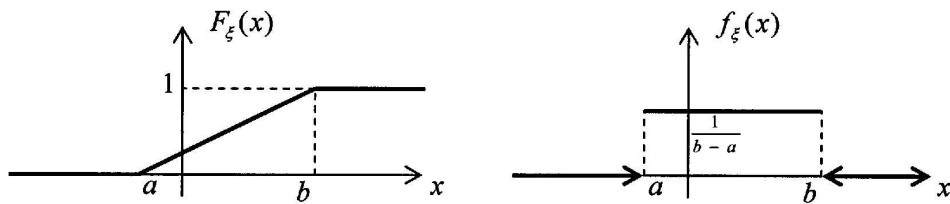


Рис. 13: Равномерное распределение

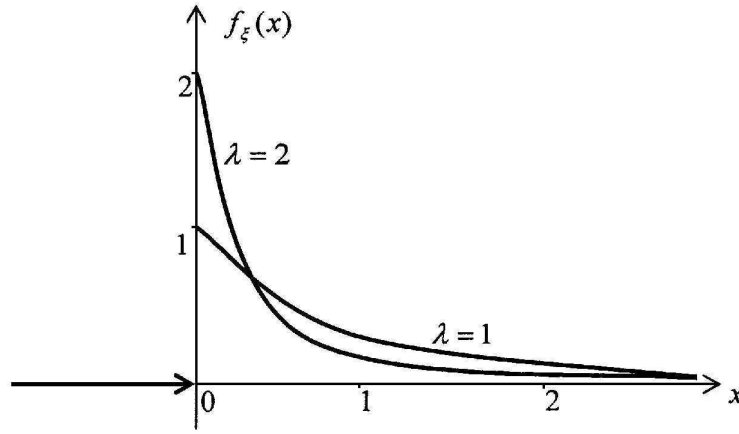


Рис. 14: Показательное распределение

Определение 12.8. Если для любого $x \in \mathbb{R}$ случайная величина ξ имеет плотность

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

то говорят, что ξ имеет нормальное (или Гаусса¹⁴) распределение с математическим ожиданием a и дисперсией σ^2 , где $a \in \mathbb{R}$ и $\sigma > 0$. (См. рис. 15) Очевидно, что $f_{\xi}(x) \geq 0$ для любого x ;

2) Используем табличный интеграл (Пуассона) $\int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = \sqrt{2\pi}$.

$$\begin{aligned} \int_{-\infty}^{\infty} f_{\xi}(x) dx &= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} dx = \left[\begin{array}{l} \text{замена переменных} \\ t = \frac{x-a}{\sigma}, dx = \sigma dt \end{array} \right] = \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = 1. \end{aligned}$$

Свойство 12.9. Легко сосчитать, что расстояние между точками перегиба равно 2σ , поэтому параметр 2σ является характеристикой ширины графика на уровне точек перегиба.

¹⁴Карл Фридрих Гаусс (Johann Carl Friedrich Gauß, 1777 — 1855), выдающийся немецкий математик, астроном и физик, считается одним из величайших математиков всех времён.

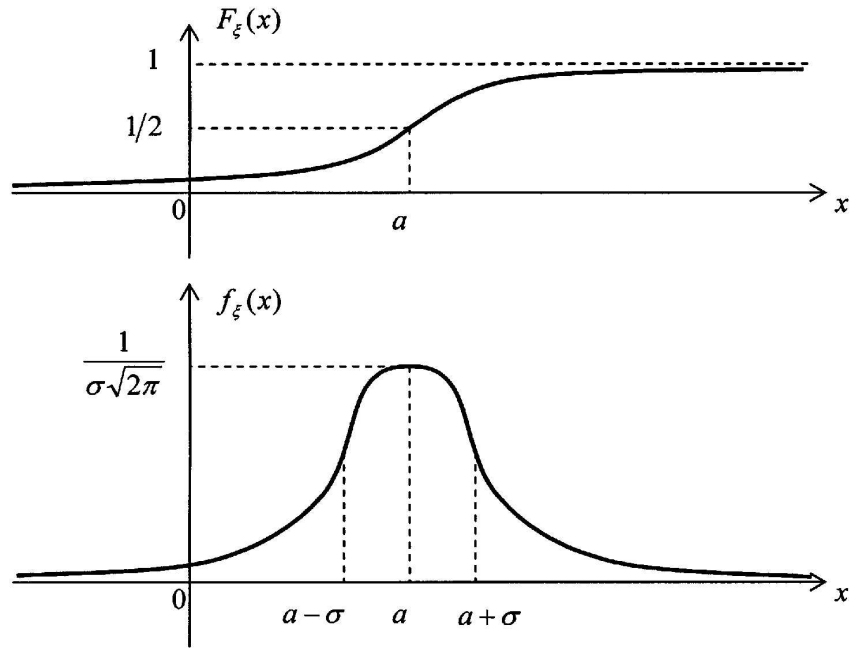


Рис. 15: Нормальное распределение (Гаусса)

Заметим, что вероятность попадания нормальной случайной величины в интервал между точками перегиба

$$P(a-\sigma \leq \xi \leq a+\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_{a-\sigma}^{a+\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}} dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-1}^1 e^{-\frac{t^2}{2}} dt = 2\Phi(1) \approx 0.6827$$

где $\Phi(1)$ приближённое значение функции $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$ взято из таблицы.

Свойство 12.10 («Правило трёх сигм»). Если ξ — нормальная случайная величина, то $P(|\xi - a| \leq 3\sigma) = 2\Phi(3) \approx 2 \cdot 0.49865 \approx 0,997$. Запоминать число 0,997 нет никакого смысла, а вот помнить, что почти вся вероятность («масса») нормального распределения сосредоточена в интервале $[a - 3\sigma, a + 3\sigma]$, всегда полезно.

13. Функции Хевисайда и Дирака

Теорема 13.1. *Функцией единичного скачка или короче функцией Хевисайда¹⁵ называется функция $U : \mathbb{R} \rightarrow \mathbb{R}$, заданная по формуле (см. рис. 16)*

¹⁵Оливер Хевисайд (англ. Oliver Heaviside; 18 мая 1850 — 3 февраля 1925) — английский учёный-самоучка, инженер, математик и физик. Впервые применил комплексные числа для изучения электрических цепей. Переписал уравнения Максвелла из их первоначальной формы, состоявшей из 20 уравнений с

$$u(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$$

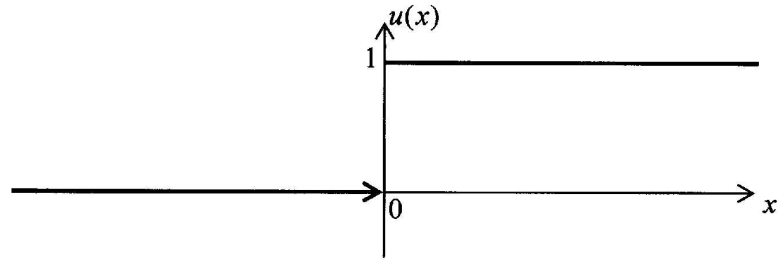


Рис. 16: Функция Хевисайда

Пример 13.2. Функция Хевисайда полезна для записи в строчку формул (многоэтажных) кусочно заданных функций.

1) Функция единичного импульса единичной длины см. рис. 17.

$$U(x) = \begin{cases} 0, & x < 0 \\ 1, & 0 \leq x < 1 \\ 0, & x \geq 1 \end{cases} = u(x) - u(x - 1).$$

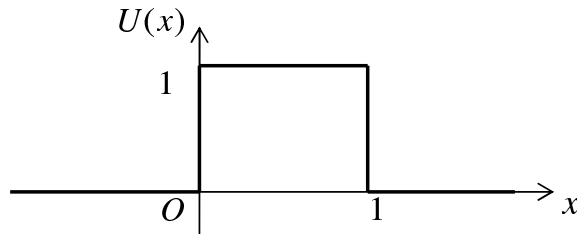


Рис. 17: $U(x) = u(x) - u(x - 1)$

2) Функция распределения Бернулли, см. рис. 10.

$$F_{\xi}(x) = \begin{cases} 0, & x < c \\ 1 - p, & 0 \leq x < 1 \\ p, & 1 \leq x \end{cases} = (1 - p)u(x) - pu(x - 1).$$

12 переменными, к современной форме, состоящей из 4 дифференциальных уравнений, выраженной в терминах современного векторного анализа. Предложил операционное исчисление (он ввёл обозначение D для дифференциального оператора) и метод решения дифференциальных уравнений с помощью сведения к обыкновенным алгебраическим уравнениям. Ввёл термины: «проводимость», «проницаемость», «индуктивность», «импеданс».

3) Функция равномерного распределения, см. рис. 13

$$F_{\xi}(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x < b \\ 1, & x \geq b \end{cases} = \frac{x-a}{b-a}u(x-a) - \frac{x-a}{b-a}u(x-b) + u(x-b).$$

4) Интеграл от функции Хевисайда.

$$\int_{-\infty}^x f(t)u(t)dt = \int_0^x f(t)dt$$

В частности,

$$\int_{-\infty}^x u(t)dt = xu(x).$$

Лемма 13.3. *Кусочная функция*

$$g(x) = \begin{cases} g_0(x), & x < x_0 \\ g_1(x), & x_0 \leq x < x_1 \\ \dots, & \dots \\ g_n(x), & x_{n-1} \leq x < x_n \\ g_{n+1}(x), & x_n \leq x \end{cases}$$

представима в виде

$$g(x) = g_0(x)u(x_0 - x) + g_1(x)[u(x - x_0) - u(x - x_1)] + \dots \\ \dots + g_n(x)[u(x - x_{n-1}) - u(x - x_n)] + g_{n+1}(x)u(x - x_n).$$

Доказательство. Функция $g(x)$ является суммой следующих функций

$$\tilde{g}_0(x) = \begin{cases} g_0(x), & x < x_0 \\ 0, & x_0 \leq x \end{cases} = g_0(x)u(x_0 - x),$$

$$\tilde{g}_1(x) = \begin{cases} 0(x), & x < x_0 \text{ или } x_1 \leq x \\ g_1(x), & x_0 \leq x < x_1 \end{cases} = g_1(x)u(x - x_0) - g_1(x)u(x - x_1), \\ \dots, \dots$$

$$\tilde{g}_n(x) = \begin{cases} 0(x), & x < x_{n-1} \text{ или } x_n \leq x \\ g_n(x), & x_{n-1} \leq x < x_n \end{cases} = g_n(x)u(x - x_{n-1}) - g_n(x)u(x - x_n),$$

$$\tilde{g}_{n+1}(x) = \begin{cases} 0(x), & x < x_n \\ g_n(x), & x_n \leq x < x_n \end{cases} = g_0(x)u(x - x_n).$$

□

Определение 13.4. δ -функцией или функцией Дирака называется обобщённая функция $\delta : \mathbb{R} \rightarrow \mathbb{R} \cup \{\infty\}$, заданная по формуле

$$\delta(x) = \begin{cases} 0, & x \neq 0 \\ \infty, & x = 0 \end{cases}$$

и для любого $\varepsilon > 0$ подчинённая условию $\int_{-\varepsilon}^{\varepsilon} \delta(x)dx = 1$. См. рис. 18.

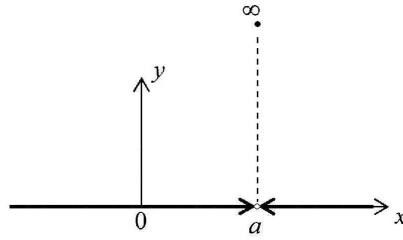


Рис. 18: Функция $y = \delta(x - a)$

Свойство 13.5. 1. Для любого $c \in \mathbb{R}$ и любого $\varepsilon > 0$, имеет место формула

$$\int_{c-\varepsilon}^{c+\varepsilon} \delta(x - c)dx = 1$$

2. Имеет место формула

$$u(x) = \int_{-\infty}^x \delta(t)dt$$

3. Имеет место формула

$$u'(x) = \delta(x)$$

4. Для любой непрерывной функции $f(x)$ любого $a \in \mathbb{R}$ имеет место тождество

$$f(x)\delta(x-a) = f(a)\delta(x-a).$$

5. Для любой непрерывной функции $f(x)$ любого $a \in \mathbb{R}$ имеет место тождество

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

6. Для любой непрерывной функции $f(x)$ любого $a \in \mathbb{R}$ имеет место тождество

$$\int_{-\infty}^x f(t)\delta(t-a)dt = f(a)u(x-a),$$

в частности

$$\int_{-\infty}^x \delta(t-a)dt = u(x-a)$$

7. Если функция $f(x)$ непрерывна и все её нули (корни) $x_1, x_2, \dots, x_k, \dots$ простые (т.е. имеют кратность 1), то

$$\delta(f(x)) = \sum_k \frac{\delta(x-x_k)}{|f'(x_k)|}.$$

8. Дельта-функция получается при вычислении преобразования Фурье от константы:

$$\int_{-\infty}^{\infty} e^{ixt}dt = 2\pi\delta(x)$$

Пример 13.6. 1. См. пример 13.2.3.

Дано

$$F_{\xi}(x) = \frac{x-a}{b-a}[u(x-a) - u(x-b)] + u(x-b).$$

Найти $f_{\xi}(x)$.

Решение.

$$f_{\xi}(x) = F'_{\xi}(x) =$$

$$\begin{aligned} & \left(\frac{x-a}{b-a} \right)' [u(x-a) - u(x-b)] + \frac{x-a}{b-a} [u(x-a) - u(x-b)] + \delta(x-b) = \\ & = \frac{1}{b-a} [u(x-a) - u(x-b)] + \frac{x-a}{b-a} [\delta(x-a) - \delta(x-b)] + \delta(x-b) = \\ & = \frac{1}{b-a} [u(x-a) - u(x-b)] + \frac{x-a}{b-a} \delta(x-a) - \frac{x-a}{b-a} \delta(x-b) + \delta(x-b) = \\ & = \frac{1}{b-a} [u(x-a) - u(x-b)] + \frac{a-a}{b-a} \delta(x-a) - \frac{b-a}{b-a} \delta(x-b) + \delta(x-b) = \\ & = \frac{1}{b-a} [u(x-a) - u(x-b)]. \end{aligned}$$

2. Найти плотность вероятности по данной функции распределения (см. рис. 19)

Дано.

$$F_{\xi}(x) \begin{cases} 0, & x < 0 \\ \frac{1}{24}(5x+8), & 0 \leq x < 2 \\ 1, & x \geq 2 \end{cases} = \frac{1}{24}(5x+8)[u(x) - x(x-2)] + u(x-2).$$

Решение.

$$\begin{aligned} f_{\xi}(x) &= F'_{\xi}(x) = \\ &= \frac{5}{24}[u(x) - u(x-2)] + \frac{1}{24}(5x+8)[\delta(x) - \delta(x-2)] + \delta(x-2) = \\ &= \frac{5}{24}[u(x) - u(x-2)] + \frac{1}{3}\delta(x) - \frac{3}{4}\delta(x-2) = \\ &= \frac{1}{3}\delta(x) + \frac{5}{24}[u(x) - u(x-2)] + \frac{1}{4}\delta(x-2). \end{aligned}$$

См. рис. 20

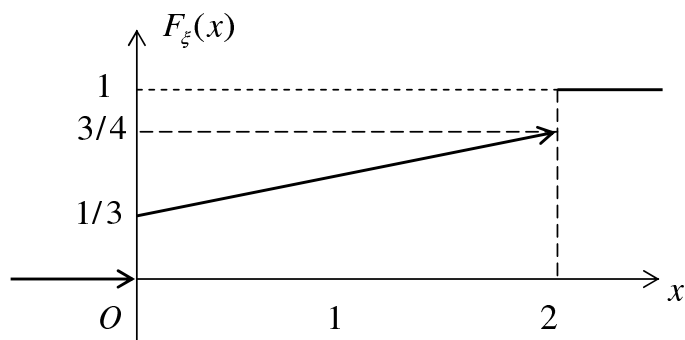


Рис. 19: Функция распределения с двумя скачками

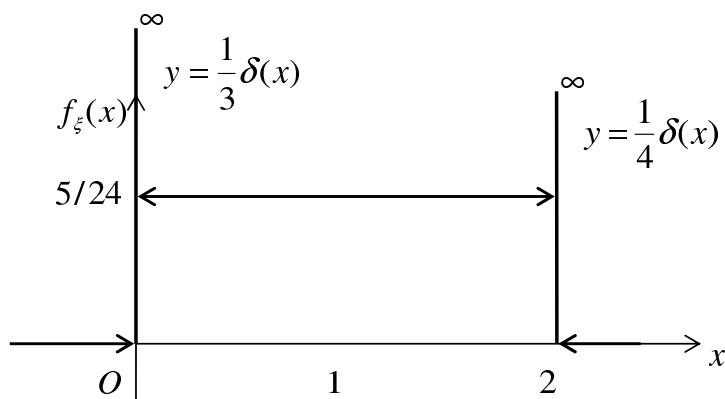


Рис. 20: Площадь под плотностью равна $\frac{1}{3} + \frac{5}{24} \cdot 2 + \frac{1}{4} = 1$.

Определение 13.7. Функции Хевисайда и Дирака и их композиции относятся к классу так называемых обобщённых функций.

14. Функции одной случайной величины

Замечание 14.1. Пусть $\xi : \Omega \rightarrow \mathbb{R}$ – абсолютно непрерывная случайная величина, имеющая плотность $f_\xi(x)$. Построим с помощью функции $g : \mathbb{R} \rightarrow \mathbb{R}$ новую случайную величину по формуле $\eta = g(\xi)$.

Требуется найти функцию распределения и плотность случайной величины η . Мы решим эту задачу сначала в предположении, что функция $y = g(x)$ дифференцируема и монотонна, т.е. когда во всех точках $x \in \mathbb{R}$ выполнено либо $g'(x) > 0$, либо $g'(x) < 0$.

Теорема 14.2. Если ξ – абсолютно непрерывная случайная величина, имеющая функцию распределения $F_\xi(x)$ и плотность $f_\xi(x)$, и если $g : \mathbb{R} \rightarrow \mathbb{R}$ – дифференцируемая и монотонная функция, то случайная величина $\eta = g(\xi)$

имеет плотность вероятности

$$f_\eta(y) = f_\xi(g^{-1}(y)) \left| \frac{d[g^{-1}(y)]}{dy} \right|$$

Доказательство. Заметим, что если $g : \mathbb{R} \rightarrow \mathbb{R}$ – монотонная функция, то существует её функция $g^{-1} : \mathbb{R} \rightarrow \mathbb{R}$, и выполнено тождество

$$g(g^{-1}(y)) \equiv y$$

Дифференцируя его, получим тождество

$$g'(g^{-1}(y)) \cdot (g^{-1}(y))' \equiv 1,$$

которое означает, что производные g' и $(g^{-1})'$ – одного знака, т.е. функции g и g^{-1} либо обе возрастающие, либо обе убывающие.

1) Пусть сначала g – возрастающая функция, т.е. $g' > 0$ и $g^{-1} > 0$. Это означает, что неравенство $g(\xi) \leq y$ можно записать в виде $\xi \leq g^{-1}(y)$.

$$\begin{aligned} F_\eta(y) &= F_{g(\xi)}(y) = \mathbf{P}(g(\xi) \leq y) = \mathbf{P}(\xi \leq g^{-1}(y)) = F_\xi(g^{-1}(y)) = \\ &= \int_{-\infty}^{g^{-1}(y)} f_\xi(t) dt = \left[\begin{array}{l} t = g^{-1}(\tau), \quad dt = (g^{-1}(\tau))' d\tau \\ t = -\infty \mapsto \tau = -\infty, \quad t = g^{-1}(y) \mapsto \tau = y \end{array} \right] = \\ &= \int_{-\infty}^y (g^{-1}(\tau))' f_\xi(g^{-1}(\tau)) d\tau. \end{aligned}$$

2) Пусть теперь g – убывающая функция, т.е. $g' < 0$ и $(g^{-1})' < 0$, тогда неравенство $g(\xi) \leq y$ можно записать в виде $\xi \geq g^{-1}(y)$.

$$\begin{aligned} F_\eta(y) &= F_{g(\xi)}(y) = \mathbf{P}(g(\xi) \leq y) = \mathbf{P}(\xi \geq g^{-1}(y)) = \\ &= \int_{g^{-1}(y)}^{\infty} f_\xi(t) dt = \left[\begin{array}{l} t = g^{-1}(\tau), \quad dt = (g^{-1}(\tau))' d\tau \\ t = \infty \mapsto \tau = -\infty, \quad t = g^{-1}(y) \mapsto \tau = y \end{array} \right] = \\ &= \int_y^{-\infty} (g^{-1}(\tau))' f_\xi(g^{-1}(\tau)) d\tau = \int_{-\infty}^y |(g^{-1}(\tau))'| f_\xi(g^{-1}(\tau)) d\tau. \end{aligned}$$

Объединяя оба случая в один, получим требуемую формулу. □

Определение 14.3. Пусть A – подмножество на прямой \mathbb{R} , т.е. $A \subset \mathbb{R}$. Функция $\mathbf{1}_A : \mathbb{R} \rightarrow \{0, 1\}$, определённая по формуле

$$\mathbf{1}_A(x) = \begin{cases} 1, & \text{если } x \in A \\ 0, & \text{если } x \notin A \end{cases}$$

называется выделяющей функцией множества A или индикатором A . Если $A = [0, \infty)$, то обозначение $\mathbf{1}_A(x)$ сокращают до $\mathbf{1}(x)$. Заметим, что $\mathbf{1}(x) = u(x)$, т.е. совпадает с функцией Хевисайда.

Замечание 14.4. Пусть теперь $g : \mathbb{R} \rightarrow \mathbb{R}$, $y = g(x)$ – дифференцируемая, кусочно монотонная функция, имеющая интервалы монотонности:

$$\mathcal{D} = \{D_1 = (-\infty, a_1], D_2 = (a_1, a_2], \dots, D_n = (a_{n-1}, \infty)\}$$

Ясно, что все ограничения $g|_{D_i} : D_i \rightarrow \mathbb{R}$, определённые по формулам $g|_{D_i}(x) = g(x)$ являются взаимно однозначными функциями и поэтому имеют обратные $(g|_{D_i})^{-1}(y) = g|_{D_i}^{-1}(y)$ с областями определения $g(D_i)$ соответственно.

Теорема 14.5 (Без доказательства). Если ξ – абсолютно непрерывная случайная величина, имеющая функцию распределения $F_\xi(x)$ и плотность $f_\xi(x)$, и если $g : \mathbb{R} \rightarrow \mathbb{R}$ – кусочно дифференцируемая и кусочно монотонная функция на интервалах \mathcal{D} , то случайная величина $\eta = g(\xi)$ имеет плотность вероятности

$$f_\eta(y) = \sum_{i=1}^n f_\xi(g|_{D_i}^{-1}(y)) \cdot \left| \frac{d[g|_{D_i}^{-1}(y)]}{dy} \right| \cdot \mathbf{1}_{g(D_i)}(y)$$

15. Случайные векторы и их распределения

Пусть $(\Omega, \mathfrak{A}, P)$ – произвольное вероятностное пространство.

Определение 15.1. Вектор (ξ_1, \dots, ξ_n) называется случайным вектором, если ξ_1, \dots, ξ_n являются случайными величинами, заданными на одном и том же вероятностном пространстве Ω .

Определение 15.2. Функцией совместного распределения случайных величин (ξ_1, \dots, ξ_n) (или случайного вектора (ξ_1, \dots, ξ_n)) называется функция $F_{\xi_1, \dots, \xi_n} : \mathbb{R}^n \rightarrow [0, 1]$, определённая по формуле

$$F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = P(\xi_1 \leq x_1, \dots, \xi_n \leq x_n).$$

Ясно, что $0 \leq F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) \leq 1$

Определение 15.3. Говорят, что случайные величины (ξ_1, \dots, ξ_n) имеют абсолютно непрерывное совместное распределение, если существует такая неотрицательная функция $f_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n)$, что для любой точки $(x_1, \dots, x_n) \in \mathbb{R}^n$ функция распределения $F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n)$ представима в виде

$$F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f_{\xi_1, \dots, \xi_n}(t_1, \dots, t_n) dt_1 \dots dt_n.$$

При этом функция $f_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n)$ называется плотностью вероятности совместного распределения случайных величин ξ_1, \dots, ξ_n .

Лемма 15.4. *Справедлива формула*

$$f_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = \frac{\partial^n F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n)}{\partial x_1 \dots \partial x_n}.$$

Доказательство. Формула получается в результате последовательного дифференцирования формулы опред. 15.3 по верхним пределам. \square

Определение 15.5. Случайные величины ξ_1, \dots, ξ_n называются независимыми, если для любого набора множеств $A_1, \dots, A_n \subset \mathbb{R}$, такого что $\xi_1^{-1}(A_1), \dots, \xi_n^{-1}(A_n) \in \mathfrak{A}$ иммет место равенство

$$P(\xi_1 \in A_1, \dots, \xi_n \in A_n) = P(\xi_1 \in A_1) \cdot \dots \cdot P(\xi_n \in A_n).$$

Лемма 15.6. 1. *Случайные величины независимы, если для любых*

$$x_1, \dots, x_n \in \mathbb{R}$$

имеет место равенство

$$F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = F_{\xi_1}(x_1) \cdot \dots \cdot F_{\xi_n}(x_n).$$

2. *Случайные величины независимы, если для любых $x_1, \dots, x_n \in \mathbb{R}$ имеет место равенство*

$$f_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = f_{\xi_1}(x_1) \cdot \dots \cdot f_{\xi_n}(x_n).$$

Доказательство.

$$\begin{aligned} 1. F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) &= P(\xi_1 \leq x_1, \dots, \xi_n \leq x_n) = \\ &= P(\xi_1 \leq x_1) \cdot \dots \cdot P(\xi_n \leq x_n) = F_{\xi_1}(x_1) \cdot \dots \cdot F_{\xi_n}(x_n). \end{aligned}$$

$$\begin{aligned}
2. f_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) &= \frac{\partial^n F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n)}{\partial x_1 \dots \partial x_n} = \frac{\partial^n [F_{\xi_1}(x_1) \cdot \dots \cdot F_{\xi_n}(x_n)]}{\partial x_1 \dots \partial x_n} = \\
&= \frac{\partial F_{\xi_1}(x_1)}{\partial x_1} \cdot \dots \cdot \frac{\partial F_{\xi_n}(x_n)}{\partial x_n} = f_{\xi_1}(x_1) \cdot \dots \cdot f_{\xi_n}(x_n).
\end{aligned}$$

□

В следующей теореме без потери общности и для простоты формулировки мы ограничимся двумерным случаем, $n = 2$.

Теорема 15.7. *Свойства функции распределения.*

1. Функция $F_{\xi_1, \xi_2}(x_1, x_2)$ является неубывающей по каждому аргументу x_1 и x_2 .

2. Существуют пределы

$$\begin{aligned}
F_{\xi_1, \xi_2}(-\infty, x_2) &= \lim_{x_1 \rightarrow -\infty} F_{\xi_1, \xi_2}(x_1, x_2) = 0 \\
F_{\xi_1, \xi_2}(x_1, -\infty) &= \lim_{x_2 \rightarrow -\infty} F_{\xi_1, \xi_2}(x_1, x_2) = 0
\end{aligned}$$

3. Существуют пределы

$$\begin{aligned}
F_{\xi_1, \xi_2}(\infty, x_2) &= \lim_{x_1 \rightarrow \infty} F_{\xi_1, \xi_2}(x_1, x_2) = F_{\xi_2}(x_2) \\
F_{\xi_1, \xi_2}(x_1, \infty) &= \lim_{x_2 \rightarrow \infty} F_{\xi_1, \xi_2}(x_1, x_2) = F_{\xi_1}(x_1)
\end{aligned}$$

4. Справедливо

$$\begin{aligned}
f_{\xi_1}(x_1) &= \int_{-\infty}^{\infty} f_{\xi_1, \xi_2}(x_1, x_2) dx_2, \\
f_{\xi_2}(x_2) &= \int_{-\infty}^{\infty} f_{\xi_1, \xi_2}(x_1, x_2) dx_1,
\end{aligned}$$

5. $F_{\xi, \eta}(x, y)$ непрерывна справа по каждой координате x и y .

(Без доказательства)

Пример 15.8. 1. Пусть $A \subset \mathbb{R}_n$ — множество, имеющее конечный объём $\mu(A)$, в пространстве с координатами $x = (x_1, \dots, x_n)$. Тогда функция

$$f_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = \begin{cases} \frac{1}{\mu(A)}, & \text{если } x \in A \\ 0, & \text{если } x \notin A \end{cases}$$

определяет плотность случайного вектора $\{\xi_1, \dots, \xi_n\}$ с равномерным распределением в A .

Функция $f_{\xi, \eta}(x, y) =$

$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \exp \left\{ -\frac{1}{2(1-r^2)} \left[\frac{(x-a)^2}{\sigma_1^2} - \frac{2r(x-a)(y-b)}{\sigma_1\sigma_2} + \frac{(y-b)^2}{\sigma_2^2} \right] \right\},$$

где r — коэффициент корреляции (см. §18), определяет двумерную нормальную плотность вероятности случайного вектора (ξ, η) , показанную на рис. 21. Там же показан эллипс равных вероятностей

$$\frac{(x-a)^2}{\sigma_1^2} - \frac{2r(x-a)(y-b)}{\sigma_1\sigma_2} + \frac{(y-b)^2}{\sigma_2^2} = \text{const}$$

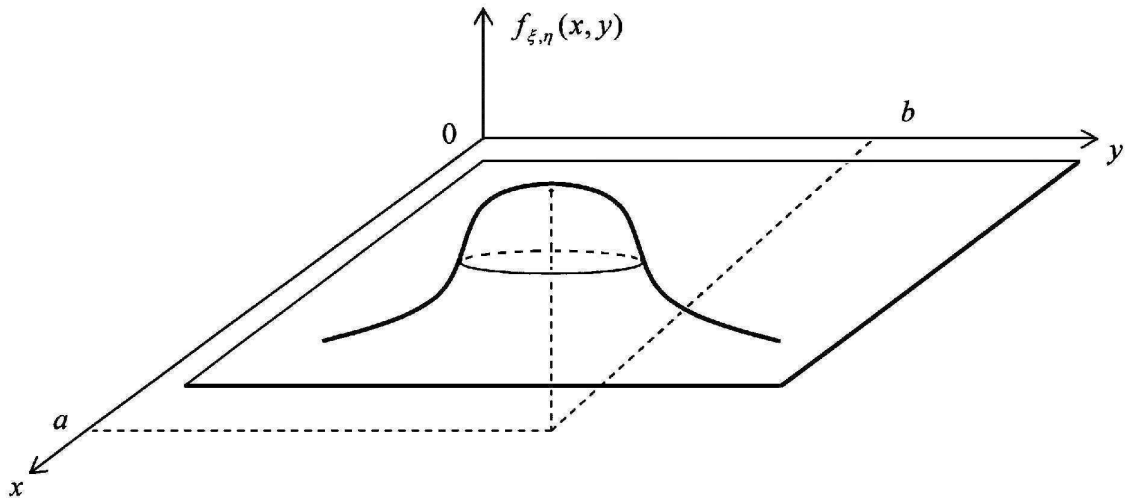


Рис. 21: Плотность вероятности двумерного нормального распределения.

16. Функции от двух случайных величин

Замечание 16.1. Пусть ξ_1, ξ_2 — случайные величины с совместной плотностью вероятности $f_{\xi_1, \xi_2}(x_1, x_2)$. Построим с помощью функции $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ случайную величину $\eta = g(\xi_1, \xi_2)$. Требуется найти функцию распределения $F_\eta(y)$ и плотность вероятности $f_\eta(y)$ случайной величины η .

Лемма 16.2. Пусть $y \in \mathbb{R}$, и область $D_y \subseteq \mathbb{R}^2$ состоит из точек (x_1, x_2) , удовлетворяющих неравенству $g(x_1, x_2) \leq y$, т.е.

$$D_y = \{(x_1, x_2) \in \mathbb{R}^2 | g(x_1, x_2) \leq y\}$$

. Тогда случайная величина $\eta = g(\xi_1, \xi_2)$ имеет функцию распределения

$$F_\eta(y) = \iint_{D_y} f_{\xi_1, \xi_2}(x_1, x_2) dx_1 dx_2$$

Доказательство.

$$F_\eta(y) = P(g(x_1, x_2) \leq y) = P((x_1, x_2) \in D_y) = \iint_{D_y} f_{\xi_1, \xi_2}(x_1, x_2) dx_1 dx_2$$

где последнее равенство следует из того, что вероятность попадания случайного вектора (x_1, x_2) в область D_y равна объёму цилиндра, ограниченного сверху графиком плотности вероятности, а снизу – областью D_y . \square

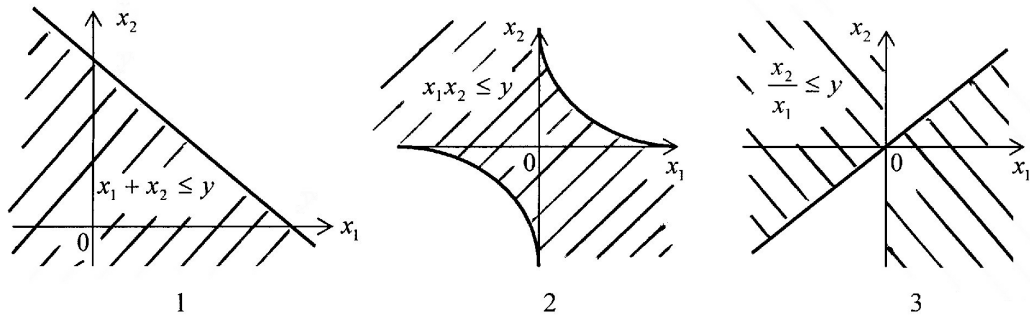


Рис. 22: Случайные функции суммы, произведения и частного

Лемма 16.3 (Распределение суммы). Пусть $\eta = \xi_1 + \xi_2$, тогда

$$F_\eta(y) = P(\eta \leq y) = P(x_1 + x_2 \leq y) = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{y-x_1} f_{\xi_1, \xi_2}(x_1, x_2) dx_2 \right) dx_1$$

Доказательство. Воспользуемся леммой 16.2 для функции $g(x_1, x_2) = x_1 + x_2$.

Заменяем двойной интеграл по области $D_y = \{(x_1, x_2) \in \mathbb{R}^2 | x_1 + x_2 \leq y\}$ (см. рис. 22.1) на повторный, получим требуемый результат. \square

Лемма 16.4 (Распределение произведения). Пусть $\eta = \xi_1 \xi_2$, тогда

$$F_\eta(y) = P(x_1 x_2 \leq y) = \int_{-\infty}^0 \left(\int_{y/x_1}^{\infty} f_{\xi_1, \xi_2}(x_1, x_2) dx_2 \right) dx_1 + \\ + \int_0^{\infty} \left(\int_{-\infty}^{y/x_1} f_{\xi_1, \xi_2}(x_1, x_2) dx_2 \right) dx_1$$

Доказательство. Воспользуемся леммой 16.2 для функции $g(x_1, x_2) = x_1 x_2$.

Область $D_y = \{(x_1, x_2) \in \mathbb{R} | x_1 x_2 \leq y\}$ показана на рис. 22.2. Переходя от двойного интеграла к повторному, получим требуемый результат. \square

Лемма 16.5 (Распределение частного). Пусть $\eta = \frac{\xi_2}{\xi_1}$, тогда

$$F_\eta(y) = P\left(\frac{x_2}{x_1} \leq y\right) = \int_{-\infty}^0 \left(\int_{yx_1}^{\infty} f_{\xi_1, \xi_2}(x_1, x_2) dx_2 \right) dx_1 + \\ + \int_0^{\infty} \left(\int_{-\infty}^{yx_1} f_{\xi_1, \xi_2}(x_1, x_2) dx_2 \right) dx_1$$

Доказательство. Доказательство. Воспользуемся леммой 16.2 для функции $g(x_1, x_2) = \frac{x_2}{x_1}$.

Область $D_y = \{(x_1, x_2) \in \mathbb{R}^2 | \frac{x_2}{x_1} \leq y\}$ показана на рис. 22.3. Переходя от двойного интеграла к повторному, получим требуемый результат. \square

Лемма 16.6 (О свёртке). Если случайные величины ξ_1 и ξ_2 независимы и имеют плотности вероятности соответственно $f_{\xi_1}(x_1)$ и $f_{\xi_2}(x_2)$, то плотность вероятности суммы $\eta = \xi_1 + \xi_2$ равна свёртке плотностей f_{ξ_1} и f_{ξ_2} :

$$f_\eta(t) = f_{\xi_1 + \xi_2}(t) = \int_{-\infty}^{\infty} f_{\xi_1}(u) f_{\xi_2}(t - u) du = \int_{-\infty}^{\infty} f_{\xi_2}(u) f_{\xi_1}(t - u) du$$

Доказательство. Воспользуемся леммой 16.3

$$F_\eta(y) = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{y-x_1} f_{\xi_1, \xi_2}(x_1, x_2) dx_2 \right) dx_1 = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{y-x_1} f_{\xi_1}(x_1) f_{\xi_2}(x_2) dx_2 \right) dx_1$$

Сделаем в последнем интеграле замену переменной: $x_2 = t - x_1$, $dx_2 = dt$. При этом $x_2 \in (-\infty, x - x_1)$ перейдёт в $t \in (-\infty, x)$. В последнем интеграле меняем порядок интегрирования и получаем:

$$\int_{-\infty}^{\infty} \left(\int_{-\infty}^x f_{\xi_1}(x_1) f_{\xi_2}(t - x_1) dt \right) dx_1 = \int_{-\infty}^x \overbrace{\left(\int_{-\infty}^{\infty} f_{\xi_1}(x_1) f_{\xi_2}(t - x_1) dx_1 \right)}^{f_{\xi_1+\xi_2}(t)} dt =$$

Итак, мы представили функцию распределения $F_{\eta}(x) = F_{\xi_1+\xi_2}(x)$ в виде $\int_{-\infty}^x f_{\xi_1+\xi_2}(t) dt$, где

$$f_{\xi_1+\xi_2}(t) = \int_{-\infty}^{\infty} f_{\xi_1}(x_1) f_{\xi_2}(t - x_1) dx_1 = \int_{-\infty}^{\infty} f_{\xi_1}(u) f_{\xi_2}(t - u) du$$

Второе равенство теоремы получается из первого, если всюду в доказательстве поменять местами индексы 1 и 2. \square

Определение 16.7. Если две независимые случайные величины имеют одно и то же распределение (возможно, с разными параметрами), и их сумма имеет то же самое распределение, то распределение называется устойчивым относительно суммирования.

Замечание 16.8. Следующие леммы могут быть доказаны непосредственно с помощью формулы свёртки, однако более короткие доказательства будут приведены в параграфе 25, используя характеристические функции.

Лемма 16.9. Пусть независимые случайные величины ξ_n и ξ_m имеют биномиальное распределение (см. опред 12.8), с параметрами n и m и вероятностью «успеха» p , тогда случайная величина $\xi_n + \xi_m$ имеет биномиальное распределение ξ_{n+m} . Другими словами, биномиальное распределение является устойчивым относительно суммирования.

Лемма 16.10. Пусть независимые случайные величины ξ и η имеют нормальное распределение (см. опред 12.8), и имеют соответственно математические ожидания a_1 и a_2 и дисперсии σ_1^2 и σ_2^2 , тогда случайная величина $\xi + \eta$ имеет нормальное распределение с математическим ожиданием $a_1 + a_2$ и дисперсией $\sigma_1^2 + \sigma_2^2$. Другими словами, нормальное распределение является устойчивым относительно суммирования.

Пример. Товарняк состоит из 100 вагонов. Масса (измеряемая в тоннах) произвольного вагона – случайная величина, распределенная по нормальному закону с средним значением 65 тонн и среднеквадратичным отклонением 9 тонн. Локомотив может везти состав массой не более 6600 тонн, в противном случае необходимо прицеплять второй локомотив. Найти вероятность того, что второй локомотив не потребуется.

Решение. По лемме 16.10 имеем $a = a_1 + \dots + a_{100} = 65 \dots 100 = 6500$, $\sigma = \sigma_1^2 + \dots + \sigma_{100}^2 = 81 \cdot 100 = 8100$.

Вероятность того, что второй локомотив не потребуется, составляет $P(\xi \leq 6600) = F_\xi(6600) = \frac{1}{2} + \Phi\left(\frac{x-a}{\sigma}\right) = \frac{1}{2} + \Phi\left(\frac{6600-6500}{90}\right) = \frac{1}{2} + \Phi\left(\frac{10}{9}\right) \approx \frac{1}{2} + \Phi(1.11) \approx 0.87$, где $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_0^z e^{-t^2/2} dt$ – табличный интеграл вероятности нормального распределения.

17. Математическое ожидание

Определение 17.1. 1. Начальным моментом n -го порядка дискретной случайной величины ξ называется число

$$M\xi^n = \sum_k x_k^n p_k.$$

2. Начальным моментом n -го порядка абсолютно непрерывной случайной величины ξ называется число

$$M\xi^n = \int_{-\infty}^{\infty} x^n f_\xi(x) dx.$$

Определение 17.2. 1. Математическим ожиданием или средним значением дискретной случайной величины ξ называется начальный момент 1-го порядка

$$M\xi = \sum_k x_k p_k$$

2. Математическим ожиданием или средним значением абсолютно непрерывной случайной величины ξ называется начальный момент 1-го порядка

$$M\xi = \int_{-\infty}^{\infty} x f_\xi(x) dx.$$

Замечание 17.3. Если ось x рассматривать как тонкую металлическую проволоку с линейной плотностью массы $f_\xi(x)$, то число $M\xi = \int_{-\infty}^{\infty} x f_\xi(x) dx$ есть координата центра тяжести оси x .

Примеры.

1. Пусть случайная величина ξ есть число очков, выпадающих при одном подбрасывании игральной кости. Тогда $M\xi = \sum_{k=1}^6 k \cdot \frac{1}{6} = \frac{1}{6} \cdot 21 = 3,5$ Т.е. в среднем выпадает 3,5 очка!
2. Пусть случайная величина ξ имеет равномерное распределение (см. опред. 12.6). Тогда $M\xi = \int_a^b x \frac{1}{b-a} dx = \frac{a+b}{2}$. Т.е. центр массы, равномерно распределённой на отрезке, находится посередине.
3. Пусть случайная величина ξ имеет нормальное распределение (см. опред. 12.8) $f_\xi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$. Тогда

$$M\xi = \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} dx = a$$

4. Если случайная величина ξ принимает конечное число значений

$$x_1, x_2, \dots, x_k,$$

классические вероятности которых соответственно равны $p_1 = \frac{n_1}{n}, p_2 = \frac{n_2}{n}, \dots, p_k = \frac{n_k}{n}$, где $n_1 + n_2 + \dots + n_k = n$, то математическое ожидание

$$M\xi = \frac{x_1 n_1 + \dots + x_k n_k}{n}$$

совпадает со средним арифметическим величин $\underbrace{(x_1, \dots, x_1)}_{n_1 \text{ раз}}, \dots, \underbrace{(x_k, \dots, x_k)}_{n_k \text{ раз}}$

Математическое ожидание функции случайной величины доставляет следующая теорема.

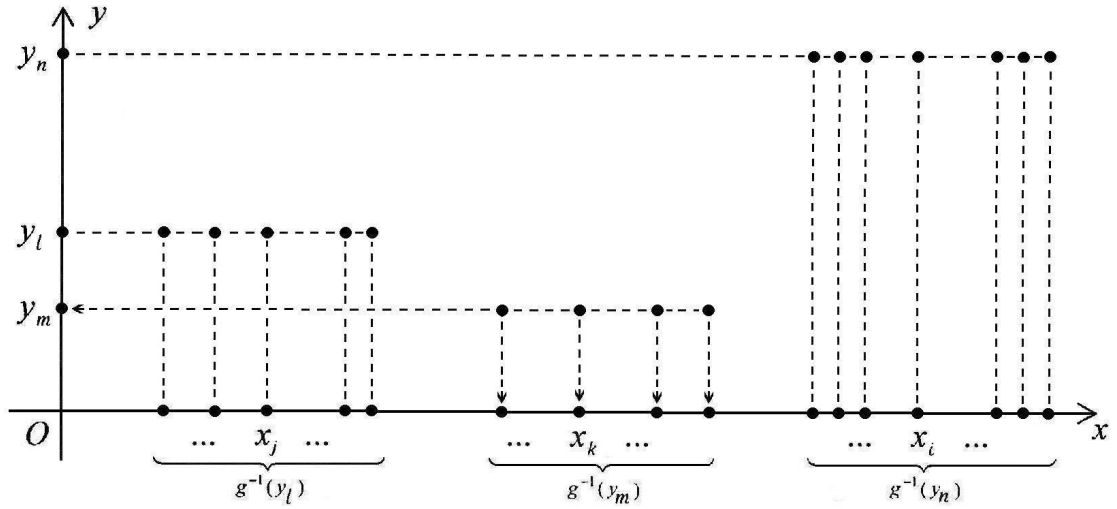


Рис. 23: К доказательству теор. 17.4

Теорема 17.4. 1. Для любой дискретной случайной величины ξ и произвольной функции $g : \mathbb{R} \rightarrow \mathbb{R}$, математическое ожидание случайной величины $\eta = g(\xi)$ вычисляется по формуле

$$\mathbb{M}\eta = \mathbb{M}g(\xi) = \sum_k g(x_k)p_k$$

2. Для любого абсолютно непрерывного случайного вектора $\xi = (\xi_1, \dots, \xi_n) \in \mathbb{R}^n$ и произвольной функции $g : \mathbb{R} \rightarrow \mathbb{R}$, математическое ожидание случайной величины $\eta = g(\xi)$ вычисляется по формуле

$$\mathbb{M}\eta = \mathbb{M}g(\xi) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(t_1, \dots, t_n) f_{\xi_1, \dots, \xi_n}(t_1, \dots, t_n) dt_1 \dots dt_n.$$

Доказательство. Мы докажем эту теорему только для дискретного случая. Пусть случайная величина ξ принимает значения $x_1, x_2, \dots, x_k, \dots$ соответственно с вероятностями $p_1, p_2, \dots, p_k, \dots$. Тогда $\eta = g(\xi)$ принимает значения $y_1, y_2, \dots, y_m, \dots$ соответственно с вероятностями $q_1, q_2, \dots, q_m, \dots$, при этом полный прообраз $g^{-1}(y_m)$ может состоять из одного или более чем из одного элемента. Поэтому

$$q_m = \mathbb{P}(g(\xi) = y_m) = \mathbb{P}(\xi \in g^{-1}(y_m)) = \sum_{x_k \in g^{-1}(y_m)} p_k.$$

Тогда

$$\begin{aligned} M\eta &= Mg(\xi) = \sum_m y_m q_m = \sum_m y_m \sum_{x_k \in g^{-1}(y_m)} p_k = \\ &= \sum_m \sum_{x_k \in g^{-1}(y_m)} g(x_k) p_k = \sum_k g(x_k) p_k. \end{aligned}$$

Последнее равенство справедливо потому, что суммирование по $x_k \in g^{-1}(y_m)$ означает суммирование вдоль горизонтального уровня $y = y_m$ (см. рис. 23), суммирование по m означает суммирование сумм на горизонтальных уровнях. Поэтому двойное суммирование в левой части эквивалентно суммированию по k в правой части. \square

Теорема 17.5. *Математическое ожидание имеет следующие свойства.*

1. Если $C = \text{const}$, то $MC = C$.
2. Для любой случайной величины ξ имеем $M(M\xi) = M\xi$.
3. Если $C = \text{const}$, то $M(C\xi) = CM\xi$.
4. Для любой случайной величины ξ имеем $|M\xi| \leq M|\xi|$.
5. Для любых случайных величин ξ_1 и ξ_2 имеем $M(\xi_1 + \xi_2) = M\xi_1 + M\xi_2$.
6. Если случайные величины ξ_1 и ξ_2 независимы, то $M(\xi_1 \xi_2) = M\xi_1 M\xi_2$.

Доказательство. 1. Константу C можно рассматривать как вырожденную случайную величину $\xi = C$, сосредоточенную в точке $x = C$. Её плотность есть $f_\xi(x) = \delta(x - C)$. Поэтому имеем $M\xi^n = MC^n = \int_{-\infty}^{\infty} x^n \delta(x - C) dx = C^n$ и, в частности $MC = C$

2. Т.к. $M\xi = \text{const}$, то $M(M\xi) = M\xi$.

$$3. M(C\xi)^n = \int_{-\infty}^{\infty} (Cx)^n f_\xi(x) dx = C^n \int_{-\infty}^{\infty} x^n f_\xi(x) dx = C^n M\xi^n$$

$$4. |M\xi^n| = \left| \int_{-\infty}^{\infty} x^n f_\xi(x) dx \right| \leq \int_{-\infty}^{\infty} |x^n| f_\xi(x) dx = M|\xi|^n \text{ и, в частности, } |M\xi| \leq M|\xi|.$$

5. По теор. 17.4 и 15.7 имеем

$$\begin{aligned}
M(\xi_1 + \xi_2) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 + x_2) f_{\xi_2, \xi_2}(x_1, x_2) dx_1 dx_2 = \\
&= \int_{-\infty}^{\infty} x_1 \left(\int_{-\infty}^{\infty} f_{\xi_2, \xi_2}(x_1, x_2) dx_1 dx_2 \right) + \int_{-\infty}^{\infty} x_2 \left(\int_{-\infty}^{\infty} f_{\xi_2, \xi_2}(x_1, x_2) dx_1 dx_2 \right) = \\
&= \int_{-\infty}^{\infty} x_1 f_{\xi}(x_1) dx_1 + \int_{-\infty}^{\infty} x_2 f_{\xi}(x_2) dx_2 = M\xi_1 + M\xi_2.
\end{aligned}$$

6. По теор. 17.4.2 и 15.7.4 имеем

$$\begin{aligned}
M(\xi_1 \xi_2) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 f_{\xi_2, \xi_2}(x_1, x_2) dx_1 dx_2 = \\
&= \int_{-\infty}^{\infty} x_1 f_{\xi_2, \xi_2}(x_1, x_2) dx_1 \int_{-\infty}^{\infty} x_2 f_{\xi_2, \xi_2}(x_1, x_2) dx_2 = M\xi_2 M\xi_2
\end{aligned}$$

□

Замечание 17.6. Обратное утверждение, к только что доказанному свойству б), неверно. Следующий пример показывает, что из равенства $M(\xi_1 \xi_2) = M\xi_2 M\xi_2$ не следует независимость величин ξ_1 и ξ_2 .

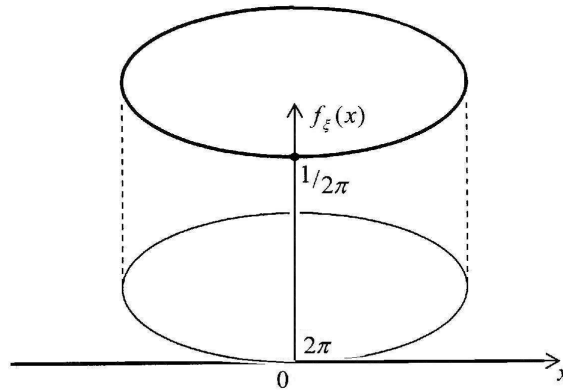


Рис. 24: Равномерно распределённый угол.

Пример 17.7. Пусть случайный угол φ равномерно распределён на окружности $S^1 = \{z = e^{i\varphi}\} \subset \mathbb{C}$, где $\varphi \in [0, 2\pi]$. Её плотность

$$f_\varphi(x) = \begin{cases} \frac{1}{2\pi}, & \text{если } x \in [0, 2\pi] \\ 0, & \text{если } x \notin [0, 2\pi] \end{cases}$$

показана на рис. 24. Случайные величины $\xi_1 = \cos \varphi$ и $\xi_2 = \sin \varphi$ являются заведомо зависимыми. Ясно, что $\xi_2^2 + \xi_1^2 = 1$. Однако математическое ожидание их произведения равно произведению их математических ожиданий. Действительно по теор. 17.4 имеем

$$\begin{aligned} M\xi_1 &= \int_0^{2\pi} \cos x \frac{1}{2\pi} dx = 0 \\ M\xi_2 &= \int_0^{2\pi} \sin x \frac{1}{2\pi} dx = 0 \\ M(\xi_1\xi_2) &= \int_0^{2\pi} \cos x \sin x \frac{1}{2\pi} dx = 0 \end{aligned}$$

18. Дисперсия

Определение 18.1. 1) Центральным моментом n -го порядка случайной величины ξ называется число

$$M(\xi - M\xi)^n = \int_{-\infty}^{\infty} (x - M\xi)^n f_\xi(x) dx$$

2) Дисперсией случайной величины ξ называется центральный момент 2-го порядка, т.е. число

$$D\xi = M(\xi - M\xi)^2 = \int_{-\infty}^{\infty} (x - M\xi)^2 f_\xi(x) dx.$$

Замечание 18.2. Для любой случайной величины ξ её центральный момент 1-го порядка равен нулю, $M(\xi - M\xi) = M\xi - M\xi = 0$.

Замечание 18.3. Если ось x рассматривать как тонкую металлическую проволоку с линейной плотностью массы $f_\xi(x)$, то дисперсия $D\xi$ есть в точности момент инерции стержня относительно его центра тяжести $M\xi$.

Лемма 18.4. 1) Для любой случайной величины ξ имеем $D\xi \geq 0$.

2) Если $C = \text{const}$, то $DC = 0$.

3) Если $C = \text{const}$, то для любой случайной величины ξ имеем $D(C\xi) = C^2 D\xi$.

Доказательство. 1) Т.к. $(x - M\xi)^2 f_\xi(x) \geq 0$, то интеграл в опред. 18.1.2) не меньше нуля.

2) $DC = M(C - MC)^2 = 0$.

3) $D(C\xi) = M[C\xi - M(C\xi)]^2 = M[C(\xi - M\xi)]^2 = C^2 M(\xi - M\xi)^2 = C^2 D\xi$.

□

Замечание 18.5. Дисперсия (по прочтению формулы $D\xi = M(\xi - M\xi)^2$) равна «среднему значению квадрата отклонения случайной величины от её среднего». То есть, дисперсия характеризует отклонение (разброс, кучность) значений случайной величины вокруг её среднего значения $M\xi$. В таком случае полезно следующее определение.

Определение 18.6. Число $\sigma = \sqrt{D\xi}$ называется средне квадратическим отклонением случайной величины ξ .

Теорема 18.7. Для любой случайной величины ξ справедлива формула

$$D\xi = M\xi^2 - (M\xi)^2.$$

Доказательство.

$$\begin{aligned} D\xi &= M(\xi - M\xi)^2 = M[\xi^2 - 2\xi M\xi + (M\xi)^2] = \\ &= M\xi^2 - 2M(\xi M\xi) + M[(M\xi)^2] = \\ &= M\xi^2 - 2(M\xi)^2 + (M\xi)^2 = M\xi^2 - (M\xi)^2. \end{aligned}$$

□

Теорема 18.8. Если случайные величины ξ_1 и ξ_2 независимы, то

$$D(\xi_1 + \xi_2) = D\xi_1 + D\xi_2.$$

Доказательство.

$$\begin{aligned} D(\xi_1 + \xi_2) &= M(\xi_1 + \xi_2)^2 - (M(\xi_1 + \xi_2))^2 = \\ &= M(\xi_1^2 + 2\xi_1\xi_2 + \xi_2^2) - (M\xi_1 + M\xi_2)^2 \stackrel{17.4.5)}{=} \\ &= M\xi_1^2 + 2M(\xi_1\xi_2) + M\xi_2^2 - (M\xi_1)^2 - 2M\xi_1 M\xi_2 - (M\xi_2)^2 \stackrel{17.4.6)}{=} \\ &= M\xi_1^2 - (M\xi_1)^2 + M\xi_2^2 - (M\xi_2)^2 = D\xi_1 + D\xi_2. \end{aligned}$$

□

Пример 18.9. 1) Пусть ξ — подчиняется распределению Бернулли, т.е. принимает два значения: 1 с вероятностью p и 0 с вероятностью $q = 1 - p$, тогда

$$\begin{aligned} M\xi &= 1 \cdot p + 0 \cdot q = p, & M\xi^2 &= 1^2 \cdot p + 0^2 \cdot q = p, \\ D\xi &= M\xi^2 - (M\xi)^2 = p - p^2 = pq. \end{aligned}$$

2) Вычислим математическое ожидание и дисперсию случайной величины ξ , имеющей биномиальное распределение, т.е. ξ принимает значения $0, 1, 2, \dots, n$ соответственно с вероятностями $P(\xi_n = k) = C_n^k p^k (1-p)^{n-k}$. Возьмем n независимых случайных величин ξ_1, \dots, ξ_n имеющих одно и то же распределение Бернулли, т.е. с $M\xi_i = p$ и $D\xi_i = pq$. Тогда их сумма $\xi = \xi_1 + \dots + \xi_n$ имеет биномиальное распределение и

$$M\xi = \sum_{i=1}^n M\xi_i = nM\xi_1 = np, \quad D\xi = D\xi_i = nD\xi_1 = npq.$$

19. Числовые характеристики зависимости случайных величин

Замечание 19.1. Мы знаем, что для независимых случайных величин дисперсия их суммы равна сумме их дисперсий

$$D(\xi_1 + \xi_2) = D\xi_2 + D\xi_1$$

Чему равна дисперсия суммы в общем случае?

$$\begin{aligned} D(\xi_1 + \xi_2) &= M(\xi_1 + \xi_2)^2 - [M(\xi_1 + \xi_2)]^2 = \\ &= M(\xi_1^2 + \xi_2^2 + 2\xi_1\xi_2) - (M\xi_1)^2 - (M\xi_2)^2 - 2M\xi_1M\xi_2 = \\ &= D\xi_1 + D\xi_2 + 2[M(\xi_1\xi_2) - M\xi_1M\xi_2]. \end{aligned}$$

Если случайные величины ξ_1 и ξ_2 независимы, то по теор. 17.5.6. величина $M(\xi_1\xi_2) - M\xi_1M\xi_2$ равна нулю. С другой стороны, из равенства её нулю вовсе не следует их независимость (см. пример 19.7). Поэтому эту величину можно использовать как «индикатор степени зависимости» случайных величин ξ_1 и ξ_2 .

Определение 19.2. Ковариацией $\text{cov}(\xi_1, \xi_2)$ случайных величин ξ_1 и ξ_2 называется число

$$\text{cov}(\xi_1, \xi_2) = M[(\xi_1 - M\xi_1)(\xi_2 - M\xi_2)]$$

Лемма 19.3. *Свойства ковариации*

1. $\text{cov}(\xi_1, \xi_2) = \mathbf{M}(\xi_1 \xi_2) - \mathbf{M}\xi_1 \mathbf{M}\xi_2$,
2. $\text{cov}(\xi, \xi) = \mathbf{D}\xi$
3. $\text{cov}(\xi_1, \xi_2) = \text{cov}(\xi_2, \xi_1)$
4. $\text{cov}(C\xi_1, \xi_2) = C \text{cov}(\xi_1, \xi_2)$.

Доказательство.

1. $\text{cov}(\xi_1, \xi_2) = \mathbf{M}[(\xi_1 - \mathbf{M}\xi_1)(\xi_2 - \mathbf{M}\xi_2)] = \xi_1 \xi_2 - \xi_1 \mathbf{M}\xi_2 - \xi_2 \mathbf{M}\xi_1 + \mathbf{M}\xi_1 \mathbf{M}\xi_2 =$
 $= \mathbf{M}(\xi_1 \xi_2) - \mathbf{M}(\xi_1) \mathbf{M}(\xi_2)$
2. Следует из опред. 19.1
3. Очевидно
4. $\text{cov}(C\xi_1, \xi_2) = \mathbf{M}(\xi_1 \xi_2) - \mathbf{M}(C\xi_1 \xi_2) - \mathbf{M}(C\xi_1) \mathbf{M}\xi_2 = C[\mathbf{M}(\xi_1 \xi_2) - \mathbf{M}\xi_1 \mathbf{M}\xi_2]$

□

Пусть ξ_1, \dots, ξ_n — случайные величины. Для сокращения формул обозначим через σ_{ij} ковариацию $\text{cov}(\xi_i, \xi_j)$, где $i, j \in \{1, \dots, n\}$.

Лемма 19.4. *(О дисперсии линейной комбинации)*

Для любых случайных величин ξ_1, \dots, ξ_n и любых чисел $c_1, \dots, c_n \in \mathbb{R}$ справедливо

$$\mathbf{D}(c_1 \xi_1 + \dots + c_n \xi_n) = \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij} c_i c_j.$$

Доказательство. Рассмотрим случайную величину $\eta = c_1 \xi_1 + \dots + c_n \xi_n$. Нетрудно увидеть, что

$$\eta - \mathbf{M}\eta = \sum_{i=1}^n c_i (\xi_i - \mathbf{M}\xi_i),$$

и что

$$(\eta - \mathbf{M}\eta)^2 = \sum_{i=1}^n \sum_{j=1}^n c_i c_j (\xi_i - \mathbf{M}\xi_i)(\xi_j - \mathbf{M}\xi_j).$$

Вычисляя математическое ожидание от обеих частей последнего равенства, получим требуемый результат. □

Замечание 19.5. Т.к. $D\eta \geq 0$, то квадратичная форма

$$\sum_{i=1}^n \sum_{j=1}^n \sigma_{ij} c_i c_j,$$

где $c_1, \dots, c_n \in \mathbb{R}$ рассматриваются как переменные, является неотрицательной. Из теории квадратичных форм известно, что все главные миноры матрицы неотрицательной квадратичной формы, в нашем случае матрицы

$$(\sigma_{ij}) = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2n} \\ \dots & \dots & \dots & \dots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_{nn} \end{pmatrix}$$

неотрицательны, и в частности, $\det \sigma_{ij} \geq 0$. При $n = 2$ это неравенство принимает вид

$$\begin{vmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{vmatrix} = \begin{vmatrix} D\xi_1 & \text{cov } \xi_1 \xi_2 \\ \text{cov } \xi_2 \xi_1 & D\xi_2 \end{vmatrix} = D\xi_1 D\xi_2 - \text{cov}^2(\xi_1, \xi_2) \geq 0$$

Замечание 19.6. Обсудим достоинства и недостатки ковариации, как величины, характеризующей зависимость двух случайных величин.

1. Если ковариация $\text{cov}(\xi_1, \xi_2)$ отлична от нуля, то величины ξ_1 и ξ_2 зависимы.
2. Если ковариация $\text{cov}(\xi_1, \xi_2)$ равна нулю, то зависимость случайных величин ξ_1 и ξ_2 приходится исследовать с помощью пп. 17.5 и 17.6.
3. Самая сильная зависимость между случайными величинами — функциональная, а из функциональных — линейная зависимость, когда $\eta = a\xi + b$. Встречаются более слабые зависимости. Например, если по последовательности независимых случайных величин ξ_1, ξ_2, \dots построить случайные величины $\xi = \xi_1 + \dots + \xi_{32} + \xi_{33}$ и $\eta = \xi_{33} + \xi_{34} + \dots + \xi_{101}$, то эти величины зависимы, но очень «слабо» и лишь через одно-единственное общее слагаемое ξ_{33} . (Представьте себе, сильно ли зависимы число появлений орла в первых 33-х подбрасываниях монеты и число появлений орла в той же серии, но в испытаниях с 33-го по 101-е.)
4. Ковариация не является «безразмерной». Если ξ — объём газа в сосуде, а η — давление этого газа, то ковариация $\text{cov}(\xi_1, \xi_2)$ измеряется в $m^3 \cdot P$. При переходе к другой размерности, т.е. при умножении одной из величин ξ, η на какое-нибудь число, ковариация тоже умножается на это

число. Но умножение на число не сказывается на «степени зависимости» величин ξ и η . Поэтому с физической точки зрения имеет смысл сравнивать зависимость случайных величин, ковариации которых имеют одну и ту же размерность. Из сказанного следует, что нужно как-то нормировать ковариацию, чтобы получить из неё безразмерную величину, абсолютное значение которой не менялось бы при умножении или сдвиге случайных величин на число и свидетельствовало бы о «силе» зависимости случайных величин. Такую величину доставляет следующее определение.

Определение 19.7. Коэффициентом корреляции $\rho(\xi, \eta)$ случайных величин ξ и η называется число

$$\rho(\xi, \eta) = \frac{\text{cov}(\xi, \eta)}{\sqrt{D\xi}\sqrt{D\eta}}.$$

Замечание 19.8. Чтобы увидеть «устройство» коэффициента корреляции, распишем по определению величины, стоящие в числителе и знаменателе

$$\rho(\xi, \eta) = \frac{M[(\xi - M\xi)(\eta - M\eta)]}{\sqrt{M(\xi - M)^2}\sqrt{M(\eta - M\eta)^2}}.$$

Для физиков уместно провести аналогию формулы косинуса угла между векторами в евклидовом пространстве и формулой коэффициента корреляции.

Во-первых, так как для любых случайных величин $\xi, \eta : \Omega \rightarrow \mathbb{R}$ определена их сумма $(\xi + \eta)(x) = \xi(x) + \eta(x)$ и умножение любой случайной величины на число $(\alpha \cdot \xi)(x) = \alpha \cdot \xi(x)$, где $\alpha \in \mathbb{R}$, то эти операции превращают множество $\{\xi : \Omega \rightarrow \mathbb{R}\}$ всех случайных величин в линейное (векторное) пространство.

Во-вторых, в этом пространстве ковариация $\text{cov}(\xi, \eta)$ является скалярным произведением двух «векторов» ξ и η и превращает его в евклидово.

И в-третьих, «длиной» случайной величины ξ является её среднеквадратичное отклонение, которая равна корню из скалярного произведения «вектора» ξ самого на себя $\sqrt{\text{cov}(\xi, \xi)} = \sqrt{D\xi}$. Поэтому коэффициент корреляции есть косинус угла между случайными величинами ξ и η в пространстве случайных величин. Надо только доказать, что $-1 \leq \rho(\xi, \eta) \leq 1$. Заметим, что т.к. пространство случайных величин бесконечномерное, то рисовать в виде векторов можно только те случайные величины, которые в фиксированном базисе представимы в виде конечных линейных комбинаций элементов этого базиса.

Определение 19.9. Случайные величины ξ и η называются некоррелированными, если $\rho(\xi, \eta) = 0$. (Некоррелированность является аналогом ортогональности.)

Теорема 19.10 (Свойства коэффициента корреляции). *Свойства:*

1. Если случайные величины ξ и η независимы, то $\rho(\xi, \eta) = \text{cov}(\xi, \eta) = 0$.
2. $|\rho(\xi, \eta)| \leq 1$.
3. $|\rho(\xi, \eta)| = 1$ тогда и только тогда, когда случайные величины ξ и η линейно зависимы, т.е. существуют числа $a, b \in \mathbb{R}$ и $a \neq 0$, такие что $\xi = a\eta + b$.

Доказательство. 1. Косвенно мы уже это доказали, см. пп. 19.5.6) и 19.3.1).

2. Требуемое неравенство следует из неравенства $D\xi_1 D\xi_2 - \text{cov}^2(\xi_1, \xi_2) \geq 0$, доказанного в п. 19.5.
3. Докажем сначала, что из $|\rho(\xi, \eta)| = 1$ следует, что $\xi = a\eta + b$. Перепишем опред. 19.7 в виде

$$\rho(\xi, \eta) = M \left(\frac{\xi - M\xi}{\sqrt{D\xi}} \cdot \frac{\eta - M\eta}{\sqrt{D\eta}} \right).$$

Рассмотрим сначала случай $\rho(\xi, \eta) = +1$. Воспользуемся неравенством $\alpha\beta \leq \frac{1}{2}(\alpha^2 + \beta^2)$

$$\begin{aligned} \rho(\xi, \eta) &= M \left(\frac{\xi - M\xi}{\sqrt{D\xi}} \cdot \frac{\eta - M\eta}{\sqrt{D\eta}} \right) \leq \\ &\leq \frac{1}{2} M \left[\left(\frac{\xi - M\xi}{\sqrt{D\xi}} \right)^2 + \left(\frac{\eta - M\eta}{\sqrt{D\eta}} \right)^2 \right] = \frac{1}{2} \cdot 2 = 1 \end{aligned}$$

Т.к. $\rho(\xi, \eta) = +1$, то последнее неравенство переходит в равенство

$$M \left(\frac{\xi - M\xi}{\sqrt{D\xi}} \cdot \frac{\eta - M\eta}{\sqrt{D\eta}} \right) = \frac{1}{2} M \left[\left(\frac{\xi - M\xi}{\sqrt{D\xi}} \right)^2 + \left(\frac{\eta - M\eta}{\sqrt{D\eta}} \right)^2 \right]$$

которое можно переписать в виде

$$M \left(\frac{\eta - M\eta}{\sqrt{D\eta}} - \frac{\xi - M\xi}{\sqrt{D\xi}} \right)^2 = 0$$

или в виде

$$M \left(\xi - \frac{\sqrt{D\xi}}{\sqrt{D\eta}} \eta - M\xi + \frac{\sqrt{D\xi} M\eta}{\sqrt{D\eta}} \right)^2 = 0$$

Обозначим $a = \frac{\sqrt{D\xi}}{\sqrt{D\eta}}$ и $b = \frac{\sqrt{D\xi}M\eta}{\sqrt{D\eta}}$, получим

$$M(\xi - a\eta - b)^2 = 0$$

По определению математического ожидания равенство нулю математического ожидания неотрицательной случайной величины означает, что эта величина равна нулю, поэтому $\xi = a\eta + b$.

В случае $\rho(\xi, \eta) = -1$ воспользуемся неравенством $\alpha\beta \geq -\frac{1}{2}(\alpha^2 + \beta^2)$ и по аналогии из неравенства

$$\begin{aligned} \rho &= M\left(\frac{\xi - M\xi}{\sqrt{D\xi}} \cdot \frac{\eta - M\eta}{\sqrt{D\eta}}\right) \leq \\ &\leq \frac{1}{2}M\left[\left(\frac{\xi - M\xi}{\sqrt{D\xi}}\right)^2 + \left[\left(\frac{\eta - M\eta}{\sqrt{D\eta}}\right)^2\right]\right] = \frac{1}{2} \cdot 2 = -1 \end{aligned}$$

получим требуемый результат.

Докажем теперь, что из $\eta = a\xi + b$ следует, что $|\rho(\xi, \eta)| = 1$. Воспользуемся свойствами математического ожидания и дисперсии.

$$\begin{aligned} \rho(\xi, a\xi + b) &= M\left(\frac{\xi - M\xi}{\sqrt{D\xi}} \cdot \frac{a\xi + b - M(a\xi + b)}{\sqrt{D(a\xi + b)}}\right) = \\ &= M\left(\frac{\xi - M\xi}{\sqrt{D\xi}} \cdot \frac{a\xi + b - M(a\xi + b)}{\sqrt{D(a\xi + b)}}\right) = M\left(\frac{\xi - M\xi}{\sqrt{D\xi}} \cdot \frac{a\xi - M(a\xi)}{\sqrt{D(a\xi)}}\right) = \\ &= aM\left(\frac{(\xi - M\xi)^2}{\sqrt{D\xi D(a\xi)}}\right) = \frac{aM(\xi - M\xi)^2}{\sqrt{a^2 D^2 \xi}} = \frac{a}{|a|} = \begin{cases} 1, & \text{если } a > 0 \\ -1, & \text{если } a < 0 \end{cases} \end{aligned}$$

□

Определение 19.11. Если $\rho(\xi, \eta) < 0$, то случайные величины ξ и η называются отрицательно коррелированными; если $\rho(\xi, \eta) > 0$, то — положительно коррелированными.

Замечание 19.12. Смысл знака коэффициента корреляции особенно ясен в случае $\rho(\xi, \eta) = \pm 1$. В этом случае знак ρ совпадает со знаком a в равенстве $\eta = a\xi + b$. Значение $\rho(\xi, \eta) = 1$ означает, что чем больше ξ , тем больше и η . Напротив, $\rho(\xi, \eta) = -1$ означает, что чем больше ξ , тем меньше и η . Аналогично можно трактовать знак коэффициента корреляции и в случае, когда $|\rho(\xi, \eta)| < 1$, помня при этом, что зависимость случайных величин ξ и η не линейная и может быть даже не функциональная.

Задача 19.13. Найти коэффициент корреляции между числом выпадений единицы и числом выпадений шестёрки при n подбрасываниях игральной кости.

Решение.

Для $i = 1, 2, 3, 4, 5, 6$ обозначим через ξ_i случайную величину, равную числу выпадений грани с i очками при n подбрасываниях кости. Посчитаем сначала ковариацию $\text{cov}(\xi_1, \xi_6) = M(\xi_1 \xi_6) - M\xi_1 M\xi_6$. Каждая из случайных величин ξ_i имеет биномиальное распределение и каждое из её значений может появиться в каждом из n подбрасываниях с вероятностью $p = 1/6$, и по п. 19.9.2) $M\xi_i = np = n/6$ и $D\xi_i = npq = 5n/36$.

Т.к. при каждом подбрасывании выпадает какая-то грань, то $\xi_1 + \dots + \xi_6 = n$. Т.к. игральная кость симметрична, то при $i \neq j$ все математические ожидания $M(\xi_i \xi_j)$ одинаковы. Заметим, что из той же симметрии следует, что $M(\xi_i \xi_i)$ тоже одинаковы и равны $M^2 \xi_i = D\xi_i + (M\xi_i)^2 = 5n/36 + n^2/36 = (n^2 + 5n)/36$.

Подсчитаем величину $M[\xi_1(\xi_1 + \dots + \xi_6)]$ двумя способами. Во-первых, она равна

$$M[\xi_1(\xi_1 + \dots + \xi_6)] = M(\xi_1 \cdot n) = n^2/6,$$

а во-вторых,

$$M[\xi_1(\xi_1 + \dots + \xi_6)] = M\xi_1^2 + 5M(\xi_1 \xi_6) = (n^2 + 5n)/36 + 5M(\xi_1 \xi_6).$$

Отсюда $5M(\xi_1 \xi_6) = n^2/6 - (n^2 + 5n)/36$, т.е.

$$M(\xi_1 \xi_6) = \frac{n^2 - n}{36}$$

Следовательно, искомый коэффициент корреляции равен

$$\rho(\xi_1, \xi_2) = \frac{\text{cov}(\xi_1, \xi_6)}{\sqrt{D\xi_1} \sqrt{D\xi_6}} = \frac{M(\xi_1, \xi_6) - M(\xi_1)M(\xi_6)}{\sqrt{D\xi_1} \sqrt{D\xi_6}} = \frac{(n^2 - n)/36 - n^2/36}{5n/36} = -\frac{1}{5}$$

Естественно, что коэффициент корреляции не зависит от числа подбрасываний игральной кости.

Часть III

Законы больших чисел

20. Неравенство Бьенеме—Чебышёва и неравенство Маркова

Напомним, что мы рассматриваем случайные величины, имеющие конечные начальные и центральные моменты до n -го порядка включительно. Метод доказательства неравенств, изучаемых в этом параграфе, принадлежит Чебышёву¹⁶.

Теорема 20.1 (Неравенство Маркова¹⁷, 1913 г.). *Для любой случайной величины ξ и для любого $\varepsilon \in (0, \infty)$, имеет место неравенство*

$$P(|\xi| \geq \varepsilon) \leq \frac{M|\xi|}{\varepsilon}$$

Доказательство. Если s – плотность случайной величины ξ , то

$$P(|\xi| \geq \varepsilon) = \int_{|x| \geq \varepsilon} f_{\xi}(t) dt$$

Ясно, что для всех $t \in \{|x| \geq \varepsilon\}$ выполнено неравенство $\frac{|t|}{\varepsilon} \geq 1$. Поэтому при замене $f_{\xi}(t)$ на $\frac{|t|}{\varepsilon} f_{\xi}(t)$ подынтегральное выражение не уменьшится, т.е.

$$P(|\xi| \geq \varepsilon) = \int_{|x| \geq \varepsilon} f_{\xi}(t) dt \leq \int_{|x| \geq \varepsilon} \frac{|t|}{\varepsilon} f_{\xi}(t) dt = \frac{1}{\varepsilon} \int_{|x| \geq \varepsilon} |t| f_{\xi}(t) dt.$$

Если теперь мы увеличим область интегрирования с $|x| \geq \varepsilon$ до $(-\infty, \infty)$, то интеграл справа тоже не уменьшится. Окончательно получаем

$$P(|\xi| \geq \varepsilon) \leq \frac{1}{\varepsilon} \int_{|x| \geq \varepsilon} |t| f_{\xi}(t) dt \leq \frac{1}{\varepsilon} \int_{-\infty}^{\infty} |t| f_{\xi}(t) dt = \frac{M|\xi|}{\varepsilon}$$

¹⁶Чебышёв, Пафнутий Львович (распространено неправильное произношение его фамилии с ударением на первый слог) (1821 — 1894), выдающийся русский математик и механик, внёсший большой вклад в теорию вероятностей, теорию приближений, теорию интерполирования функций, интегральное исчисление и картографию. Работая на «оборонку», он улучшил дальность и точность артиллерийской стрельбы, чем оказал большое влияние на развитие русской артиллерии.

¹⁷Марков, Андрей Андреевич (1856 — 1922), выдающийся русский математик, внёсший большой вклад в теорию вероятностей и математический анализ.

□

Следствие 20.2 (Двойственное неравенство Маркова). Для любой случайной величины ξ и для любого $\varepsilon \in (0, \infty)$ имеет место неравенство

$$P(|\xi| < \varepsilon) \geq 1 - \frac{M|\xi|}{\varepsilon}$$

Доказательство. По лемме 3.7 имеем $P(|\xi| < \varepsilon) = 1 - P(|\xi| \geq \varepsilon)$. Подставим это выражение в неравенство Маркова, получим требуемый результат. □

Теорема 20.3 (Обобщённое неравенство Маркова). Для любой случайной величины ξ , для любого $\varepsilon \in (0, \infty)$ и любой монотонно возрастающей функции $g : (0, \infty) \rightarrow (0, \infty)$ имеет место неравенство

$$P(|\xi| \geq \varepsilon) \leq \frac{Mg(|\xi|)}{g(\varepsilon)}$$

Доказательство. Поскольку функция g монотонно возрастает, то $P(|\xi| \geq \varepsilon) = P(g(|\xi|) \geq g(\varepsilon))$. Оценивая последнюю вероятность согласно неравенству Маркова, получим требуемый результат:

$$P(g(|\xi|) \geq g(\varepsilon)) \leq \frac{Mg(|\xi|)}{g(\varepsilon)}$$

□

Следствие 20.4. Для любой случайной величины ξ , для любого $\varepsilon \in (0, \infty)$, и любой монотонно возрастающей функции $g : (0, \infty) \rightarrow (0, \infty)$ имеет место неравенство

$$P(g(|\xi|) < g(\varepsilon)) > 1 - \frac{Mg(|\xi|)}{g(\varepsilon)}$$

(двойственное к обобщённому неравенству Маркова). Доказательство аналогично доказательству след. 20.2. В 1853 г. И.-Ж. Бьенеме¹⁸ и в 1866 г. независимо от него П.Л. Чебышёв доказали следующее неравенство.

Теорема 20.5 (Неравенство Бьенеме – Чебышёва). Для любой случайной величины ξ и любого $\varepsilon \in (0, \infty)$ имеет место неравенство

$$P(|\xi - M\xi| \geq \varepsilon) \leq \frac{D\xi}{\varepsilon^2}$$

¹⁸И.-Ж. Бьенеме (Irene-Jules Biemaume, 1796 – 1878), французский математик, основные работы по теории вероятностей и математической статистике.

Доказательство. Т.к. неравенство Маркова 20.1 справедливо для произвольной случайной величины, то перепишем его для случайной величины $\xi - M\xi$, получим

$$P(|\xi - M\xi| \geq \varepsilon) \leq \frac{M|\xi - M\xi|^2}{\varepsilon}$$

Зададим функцию $g : (0, \infty) \rightarrow (0, \infty)$ по формуле $g(\varepsilon) = \varepsilon^2$. Видно, что она является монотонно возрастающей и поэтому удовлетворяет условиям теор. 20.3, и следовательно, для неё выполнено обобщённое неравенство Маркова

$$P(|\xi - M\xi| \geq \varepsilon) \leq \frac{M(\xi - M\xi)^2}{\varepsilon^2} = \frac{D\xi}{\varepsilon^2}$$

□

Следствие 20.6. Для любой случайной величины ξ и любого $\varepsilon \in (0, \infty)$, имеет место неравенство

$$P(|\xi - M\xi| < \varepsilon) \geq 1 - \frac{D\xi}{\varepsilon^2}$$

(двойственное к неравенству Бьенеме – Чебышёва). Мы будем называть двойственное неравенство тоже неравенством Бьенеме – Чебышёва.

Следствие 20.7. Для любой случайной величины ξ вероятность того, что она примет значение, отличающееся от её среднего $M\xi$ более чем на три корня из её дисперсии, не превосходит $\frac{1}{9}$, т.е.

$$P(|\xi - M\xi| \geq 3\sqrt{D\xi}) \leq \frac{1}{9}$$

Доказательство. Из неравенства Бьенеме-Чебышёва имеем

$$P(|\xi - M\xi| \geq 3\sqrt{D\xi}) \leq \frac{D\xi}{(3\sqrt{D\xi})^2} = \frac{1}{9}$$

□

21. Последовательности случайных величин

Замечание 21.1. Напомним, что по опред. 11.1 случайная величина есть функция $\xi : \omega \rightarrow R$. Поэтому последовательность случайных величин

$$\xi_1, \xi_2, \dots, \xi_n, \dots = \{\xi_n\}_{n=1}^{\infty} \equiv \{\xi_n\}$$

есть на самом деле последовательность функций $\{\xi_n : \omega \rightarrow R\}_{n=1}^{infy}$, определённая на одном и том же пространстве элементарных событий ω .

В математическом анализе мы изучили следующие две сходимости последовательности функций.

Определение 21.2. Последовательность функций $\xi_1, \xi_2, \dots, \xi_n, \dots$ сходится к функции ξ поточечно, если для любой точки $\omega \in \omega$ последовательность чисел (значений) $\xi_1(\omega), \xi_2(\omega), \dots, \xi_n(\omega), \dots$ сходится к значению $\xi(\omega)$ функции ξ ; или короче¹⁹ $\lim_{n \rightarrow \infty} \xi_n(\omega) = \xi(\omega)$ для любой точки $\omega \in \omega$.

Определение 21.3. Последовательность функций $\xi_1, \xi_2, \dots, \xi_n, \dots$ сходится к функции ξ поточечно почти всюду в ω , если подмножество $A \subset \omega$, в которых поточечная сходимость не выполняется, имеет меру 0, т.е. $\mu(A) = 0$. В теории вероятностей объекты ω, A, ω являются событиями, а мерой их наступления является вероятность, поэтому в теории вероятностей сходимости «почти всюду» соответствует так называемая сходимость «почти наверное».

Определение 21.4. Говорят, что последовательность случайных величин $\{\xi_n\}$ сходится почти наверное к случайной величине ξ , если имеют место эквивалентные друг другу равенства

$$P \left\{ \omega \in \Omega \mid \lim_{n \rightarrow \infty} \xi_n(\omega) = \xi(\omega) \right\} = 1$$

и

$$P \left\{ \omega \in \Omega \mid \lim_{n \rightarrow \infty} \xi_n(\omega) \neq \xi(\omega) \right\} = 0$$

Замечание 21.5. Чтобы пользоваться на практике сходимостью «почти наверное» необходимо знать, как устроены отображения $\omega \mapsto \xi_n(\omega)$. Как правило в задачах теории вероятностей известны не сами случайные величины ξ_n , а их функции распределения, скажем $P(\xi_n \leq x) = F_{\xi_n}(x) = \int_{-\infty}^x f_{\xi_n}(t) dt$.

Можно ли в таком случае, обладая информацией только о функциях распределения, каким-нибудь образом исследовать сходимость последовательности случайных величин $\{\xi_n\}$? Ответ: да, можно, если потребовать, чтобы вероятность тех элементарных исходов ω , для которых значение $\xi_n(\omega)$ не попадает в « ε -окрестность» числа $\xi(\omega)$, сходилась к нулю при $n \rightarrow \infty$. В теории вероятностей эту идею реализуют с помощью так называемой «сходимости по вероятности».

Определение 21.6. Говорят, что последовательность случайных величин $\{\xi_n\}$ сходится по вероятности к случайной величине ξ , если для любого $\varepsilon \in$

¹⁹Используя определение сходимости числовой последовательности.

$(0, \infty)$ имеют место эквивалентные равенства

$$\begin{aligned}\lim_{n \rightarrow \infty} P(|\xi_n - \xi| \geq \varepsilon) &= 0 \\ \lim_{n \rightarrow \infty} P(|\xi_n - \xi| \leq \varepsilon) &= 1\end{aligned}\quad (*)$$

Если ξ_n сходится по вероятности к ξ , то вместо пределов (21.6) коротко пишут « $\xi_n \xrightarrow{P} \xi$ при $n \rightarrow \infty$,» а т.к. $n \rightarrow \infty$ всегда, то ещё короче: « $\xi_n \xrightarrow{P} \xi$.» Хотя на практике для проверки такой сходимости проверяют выполнение одного из равенств (21.6).

Пример 21.7. Рассмотрим последовательность случайных величин $\{\xi_n\}$ в которой для каждого n случайная величина задана следующим рядом распределения

ξ_n	c	n^3
P	$1 - 1/n$	$1/n$

Докажем, что эта последовательность сходится по вероятности к вырожденной случайной (детерминированной) величине ξ , имеющей ряд распределения

ξ	0
P	1

, а проще говоря, сходится по вероятности к нулю.

Действительно, зафиксируем произвольное $\varepsilon > 0$. Для всех n , начиная с некоторого n_0 такого, что $n_0^3 > \varepsilon$, выполнено равенство

$$P(|\xi_n - 0| \geq \varepsilon) = P(\xi_n \geq \varepsilon) = P(\xi_n = n^3) = \frac{1}{n}.$$

Применяя теперь опред. 21.6 получаем

$$\lim_{n \rightarrow \infty} P(|\xi_n - 0| \geq \varepsilon) = \lim_{n \rightarrow \infty} \frac{1}{n} = 0$$

т.е. последовательность случайных величин $\xi_1, \xi_2, \dots, \xi_n, \dots$ сходится по вероятности к ξ , т.е. $\xi_n \xrightarrow{P} \xi$.

В нашем примере случайные величины ξ_n с ростом n могут принимать всё б´ольшие и б´ольшие значения, но с всё меньшей и меньшей вероятностью.

Замечание 21.8. Сходимость по вероятности может не сопровождаться сходимостью математических ожиданий (или других начальных и центральных моментов): из $\xi_n \xrightarrow{P} \xi$ не следует, что $M\xi_n \rightarrow M\xi$. Например, в предыдущем примере $\xi_n \xrightarrow{P} \xi$, однако $M\xi_n^2 = n^2$ и $M\xi = 0$, и ясно, что последовательность $1, 4, 9, \dots$ не сходится к нулю. Сходимость по вероятности обладает теми же свойствами, как и поточечная сходимость.

Теорема 21.9 (Без доказательства). Если $\xi_n \xrightarrow{P} \xi$ и $\eta_n \xrightarrow{P} \eta$, то

1. $\xi_n + \eta_n \xrightarrow{P} \xi + \eta$,
2. $\xi_n \cdot \eta_n \xrightarrow{P} \xi \cdot \eta$,
3. Если g — непрерывная функция, то $g(\xi_n) \xrightarrow{P} g(\xi)$.
4. Если $\xi_n \xrightarrow{P} C$ и функция g непрерывна в точке C , то $g(\xi_n) \xrightarrow{P} g(C)$.

Другой тип сходимости случайных величин, который нам понадобится в дальнейшем, определяется через поточечную сходимость функций распределений. Это так называемая «слабая сходимость.» Пусть задана последовательность случайных величин $\{\xi_n\}_{n=1}^{\infty}$ с функциями распределения $\{F_{\xi_n}(x)\}_{n=1}^{\infty}$, и задана случайная величина ξ с функцией распределения $F_{\xi}(x)$.

Определение 21.10. Говорят, что последовательность случайных величин $\{\xi_n\}_{n=1}^{\infty}$ слабо сходится к случайной величине ξ , если последовательность их функций распределения $\{F_{\xi_n}(x)\}_{n=1}^{\infty}$ сходится поточечно к функции распределения $F_{\xi}(x)$, т.е. $\lim_{n \rightarrow \infty} F_{\xi_n}(x) = F_{\xi}(x)$.

Если $\{\xi_n\}$ слабо сходится к ξ , то коротко пишут $\xi_n \Rightarrow \xi$. На практике для проверки слабой сходимости проверяют выполнение равенства

$$\lim_{n \rightarrow \infty} F_{\xi_n}(x) = F_{\xi}(x)$$

для любого $x \in \mathbb{R}$.

Теорема 21.11.

1. Если $\xi_n \xrightarrow{P} \xi$, то $\xi_n \Rightarrow \xi$
2. Если $\xi_n \Rightarrow C = \text{const}$, то $\xi_n \xrightarrow{P} C$
3. Если $\xi_n \xrightarrow{P} C$ и $\eta_n \Rightarrow \eta$, то $\xi_n \cdot \eta_n \Rightarrow C\eta$
4. Если $\xi_n \xrightarrow{P} C$ и $\eta_n \Rightarrow \eta$, то $\xi_n + \eta_n \Rightarrow C + \eta$

22. Законы больших чисел

Одним из главных разделов теории вероятностей составляют так называемые «предельные теоремы». Они описывают условия возникновения вероятностных закономерностей при наличии большого числа случайных факторов.

Предельные теоремы можно условно разделить на три типа: 1) закон больших чисел, 2) предельные теоремы биномиального распределения и 3) центральные предельные теоремы.

Этот и следующие параграфы посвящены изучению таких теорем.

Определение 22.1. Говорят, что последовательность случайных величин $\{\xi_n\}_{n=1}^{\infty}$ удовлетворяет закону больших чисел (ЗБЧ), если для любого $\varepsilon \in (0, \infty)$ (при некоторых дополнительных условиях) имеют место эквивалентные друг другу равенства

$$\lim_{n \rightarrow \infty} P \left(\left| \frac{1}{n} \sum_{k=1}^n \xi_k - \frac{1}{n} \sum_{k=1}^n M\xi_k \right| < \varepsilon \right) = 1 \quad (1)$$

и

$$\lim_{n \rightarrow \infty} P \left(\left| \frac{1}{n} \sum_{k=1}^n \xi_k - \frac{1}{n} \sum_{k=1}^n M\xi_k \right| \geq \varepsilon \right) = 0 \quad (2)$$

Другими словами, среднее арифметическое первых n членов последовательности случайных величин сходится по вероятности к среднему арифметическому математических ожиданий её первых n членов при $n \rightarrow \infty$, или в символической форме:

$$\frac{1}{n} \sum_{k=1}^n \xi_k \xrightarrow{P} \frac{1}{n} \sum_{k=1}^n M\xi_k \text{ при } n \rightarrow \infty$$

Замечание 22.2. Закон больших чисел в обобщённом смысле – это совокупность теорем о том, что при некоторых условиях имеют место формулы типа (1) и (2) и их многочисленные обобщения.

Т.к. каждая такая теорема представляет собой конкретный закон больших чисел, то имеет смысл говорить о законах больших чисел.

Широкие условия применимости ЗБЧ были найдены впервые Чебышёвым в 1867 г. Эти условия затем были обобщены А.А. Марковым (старшим).

Окончательное решение проблемы о необходимых и достаточных условиях применимости ЗБЧ было найдено А.Н. Колмогоровым в 1928 г. В этом параграфе мы изучим три наиболее известных таких закона.

Теорема 22.3 (ЗБЧ в форме Чебышёва). Пусть $\xi_1, \xi_2, \dots, \xi_n, \dots$ – последовательность попарно независимых случайных величин, имеющих дисперсии, ограниченные одной и той же константой C , т.е.

$$D\xi_1 \leq C, \quad D\xi_2 \leq C, \quad \dots, \quad D\xi_n \leq C, \quad \dots$$

Тогда для любого $\varepsilon \in (0, \infty)$ при $n \rightarrow \infty$

$$\frac{1}{n} \sum_{k=1}^n \xi_k \xrightarrow{P} \frac{1}{n} \sum_{k=1}^n M\xi_k$$

Доказательство. По 18.4.3) и 18.8 имеем

$$D\left(\frac{1}{n} \sum_{k=1}^n \xi_k\right) = \frac{1}{n^2} D \sum_{k=1}^n \xi_k = \frac{1}{n^2} \sum_{k=1}^n D\xi_k \leq \frac{1}{n^2} Cn = \frac{C}{n}$$

По след. 22.6

$$P\left(\left|\frac{1}{n} \sum_{k=1}^n \xi_k - \frac{1}{n} \sum_{k=1}^n M\xi_k\right| < \varepsilon\right) \geq 1 - \frac{D\left(\frac{1}{n} \sum_{k=1}^n \xi_k\right)}{\varepsilon^2} \geq 1 - \frac{C}{n\varepsilon^2}$$

Переходя к пределу при $n \rightarrow \infty$ получим неравенство

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{1}{n} \sum_{k=1}^n \xi_k - \frac{1}{n} \sum_{k=1}^n M\xi_k\right| < \varepsilon\right) \geq 1$$

откуда по лемме 3.6 получаем требуемый результат. \square

Замечание 22.4. Из доказательства теоремы 22.3 следует, что двойственное неравенство Бьенеме – Чебышёва можно записать в виде:

$$P\left(\left|\frac{1}{n} \sum_{k=1}^n \xi_k - \frac{1}{n} \sum_{k=1}^n M\xi_k\right| \geq \varepsilon\right) \leq \frac{D\left(\frac{1}{n} \sum_{k=1}^n \xi_k\right)}{\varepsilon^2} \quad (3)$$

Теорема 22.5 (ЗБЧ в форме Хинчина²⁰ (1929)). Пусть $\xi_1, \xi_2, \dots, \xi_n, \dots$ – последовательность попарно независимых и одинаково распределённых случайных величин, то есть

$$M\xi_1 = M\xi_2 = \dots = M\xi_n = \dots \quad \text{и} \quad D\xi_1 = D\xi_2 = \dots = D\xi_n = \dots$$

Тогда среднее арифметическое $\frac{1}{n} \sum_{k=1}^n \xi_k$ сходится по вероятности к $M\xi_1$, другими словами

$$\frac{1}{n} \sum_{k=1}^n \xi_k \xrightarrow{P} M\xi_1$$

²⁰Александр Яковлевич Хинчин (1894 – 1959), один из наиболее значимых математиков в советской школе теории вероятностей. Им получены основополагающие результаты в теории функций действительного переменного, теории чисел, теории вероятностей и статистической физике.

При этом для любого $\varepsilon \in (0, \infty)$ неравенство Бьенеме – Чебышёва можно записать в виде:

$$\mathbf{P} \left(\left| \frac{1}{n} \sum_{k=1}^n \xi_k - \mathbf{M}\xi_1 \right| \geq \varepsilon \right) \leq \frac{\mathbf{D}\xi_1}{n\varepsilon^2}$$

Доказательство. Для доказательства достаточно подставить $\frac{1}{n} \sum_{k=1}^n \mathbf{M}\xi_k$ в ЗБЧ в форме Чебышёва и в неравенство (3). \square

Теорема 22.6 (ЗБЧ в форме Бернулли (1713)). Пусть A – событие, которое может произойти в любом из n независимых испытаний с одной и той же вероятностью p . Пусть $n(A)$ – число появлений события A в этих n испытаниях. Тогда относительная частота $\frac{n(A)}{n}$ сходится по вероятности к вероятности p , т.е.

$$\frac{n(A)}{n} \xrightarrow{\mathbf{P}} \mathbf{M}\xi_1$$

При этом для любого $\varepsilon \in (0, \infty)$ неравенство Бьенеме – Чебышёва можно записать в виде:

$$\mathbf{P} \left(\left| \frac{n(A)}{n} - p \right| \geq \varepsilon \right) \leq \frac{p(1-p)}{n\varepsilon^2}$$

Доказательство. Видно, что ЗБЧ в форме Бернулли удовлетворяет условиям ЗБЧ в форме Хинчина. Среднее арифметическое $\frac{1}{n} \sum_{k=1}^n \xi_k$ нулей и единиц, выпавших при n независимых испытаниях, равно частоте появления события A (см. опред 4.2). В примере 18.9.1) мы нашли, что $\mathbf{M}\xi_1 = p$ и $\mathbf{D}\xi_1 = p(1-p)$. Подставляя эти величины в ЗБЧ в форме Хинчина, получим ЗБЧ в форме Бернулли. \square

Пример 22.7. Монету подбрасывают 10 000 раз. Оценить вероятность того, что относительная частота выпадения герба отличается от классической вероятности $1/2$ не менее чем на 0,01.

Решение. Другими словами, требуется оценить $\mathbf{P} \left(\left| \frac{n(A)}{n} - \frac{1}{2} \right| \geq 0,01 \right)$ где $n = 10^4$, $n(A) = \sum_{k=1}^n \xi_k$ – число выпадений герба при условии, что ξ_k являются независимыми случайными величинами, имеющими распределение Бернулли с $p = \frac{1}{2}$, и равные единице, если выпал герб, и нулю – в противном случае.

Применим ЗБЧ в форме Бернулли. Поскольку $\mathbf{D}\xi_1 = p(1-p) = \frac{1}{4}$, то искомая оценка сверху выглядит следующим образом:

$$\mathbf{P} \left(\left| \frac{n(A)}{n} - \frac{1}{2} \right| \geq 0,01 \right) \leq \frac{\mathbf{D}\xi_1}{n \cdot 0,01^2} = \frac{1}{4 \cdot 10^4 \cdot 10^{-4}} = \frac{1}{4}.$$

Другими словами, неравенство Бьенеме – Чебышёва позволяет заключить, что в среднем не более чем в четверти случаев при 10 000 подбрасываниях монеты частота выпадения герба будет отличаться от $1/2$ более чем на 0,01. Мы увидим, что эта оценка достаточно грубая, когда изучим так называемую «центральную предельную теорему.»

В заключение параграфа приведём без доказательства ЗБЧ в форме Пуассона.

Теорема 22.8 (ЗБЧ в форме Пуассона²¹ (1837)). Пусть A – событие, которое может произойти в n независимых испытаниях с вероятностями p_1, p_2, \dots, p_n . Пусть $n(A)$ – число появлений события A в этих n испытаниях.

Тогда $\frac{n(A)}{n}$ сходится по вероятности к среднему арифметическому вероятностей p_k , т.е.

$$\frac{n(A)}{n} \xrightarrow{P} \frac{1}{n} \sum_{k=1}^n p_k$$

23. Предельные теоремы для биномиального распределения

Теорема Пуассона

Замечание 23.1. По теор. 9.2 вероятность появления k успехов при n испытаниях схемы Бернулли с вероятностью успеха равной p вычисляется по формуле

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

Если числа k и n большие и при этом одна из вероятностей p или q мала, то подсчёт по этой формуле становится очень трудным. Мы рассмотрим три случая, когда для упрощения вычислений применимы приближённые формулы.

Определение 23.2. Дискретная случайная величина ξ со счётным пространством $\Omega = \{0, 1, 2, \dots, k, \dots\}$, подчиняется распределению Пуассона с параметром $\lambda > 0$, если её ряд распределения определена соответствием $k \rightarrow P_\lambda(k) = \frac{\lambda^k}{k!} e^{-\lambda}$. (Ср. с опред 4.11.)

²¹Симеон-Дени Пуассон (Simeon-Denis Poisson, 1781 – 1840), знаменитый французский физик и математик. Его труды относятся к разным областям чистой математики, математической физики, теоретической и небесной механики.

Нетрудно подсчитать, что для распределения Пуассона $M\xi = \lambda$ и $D\xi = \lambda$. На рис. 25 показаны функции $P_\lambda(k)$ для $\lambda = 0, 1; 1; 10$, где пунктиром показаны огибающие. В каждом случае сумма длин вертикальных отрезков равна 1. Сравните график на рис. 25.3 с графиком биномиального распределения, показанным на рис. 8: обе случайные величины имеют близкие математические ожидания и похожие графики, только биномиальный график – без «хвоста». Эта внешняя схожесть наводит на мысль, что биномиальное распределение может быть аппроксимировано распределением Пуассона.

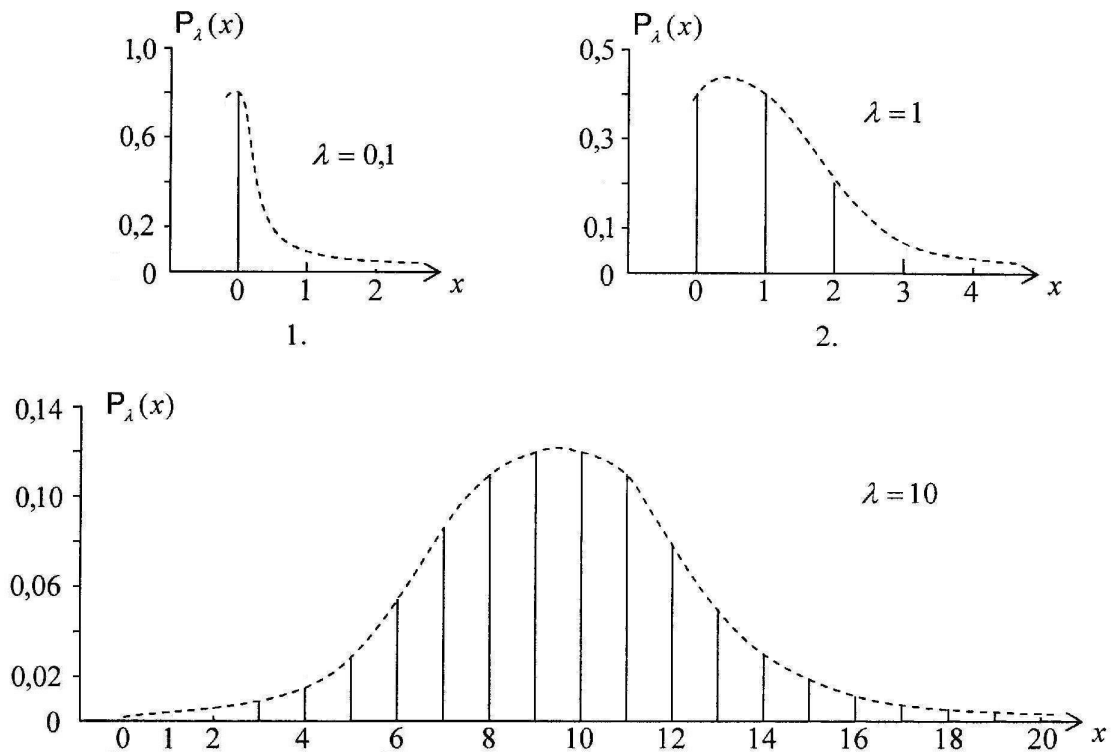


Рис. 25: Распределение Пуассона

Рассмотрим бесконечную таблицу событий:

$$\begin{array}{ccccccc}
 & & A_{11} & & & & \\
 & A_{21} & A_{22} & & & & \\
 & A_{31} & A_{32} & A_{33} & & & \\
 & \dots & \dots & \dots & \dots & & \\
 A_{n1} & A_{n2} & A_{n3} & \dots & A_{nn} & & \\
 \dots & \dots & \dots & \dots & \dots & &
 \end{array}$$

где 1) в n -ой строке, $n = 1, 2, \dots$, стоят n событий $A_{n1}, A_{n2}, A_{n3}, \dots, A_{nn}$; 2) в каждой строке события попарно независимы; 3) каждое событие в n -ой строке происходит с одной и той же вероятностью p_n , зависящей только от номера строки. Обозначим через ξ_n случайную величину, равную числу

событий A_{ni} , произошедших в n -ой строке. Ясно, что случайная величина ξ_n принимает значения $k = 0, 1, 2, \dots, n$ и имеет биномиальное распределение $k \rightarrow C_n^k p_n^k (1 - p_n)^{n-k}$.

Определение 23.3. Говорят, что последовательность $p_1, p_2, \dots, p_n, \dots$ удовлетворяет закону малых чисел, если $\lim_{n \rightarrow \infty} p_n = 0$ и существует предел

$$\lim_{n \rightarrow \infty} np_n = \lambda, \quad \text{где } 0 < \lambda < \infty.$$

Теорема 23.4 (Пуассона). Если последовательность $\{p_n\}_{n=1}^{\infty}$ удовлетворяет закону малых чисел, то для любого $k \geq 0$ имеет место равенство

$$\lim_{n \rightarrow \infty} C_n^k p_n^k (1 - p_n)^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda}$$

Доказательство. Обозначим $\lambda_n = n \cdot p_n$. Тогда

$$\begin{aligned} P &= (\xi_n = k) = C_n^k p_n^k (1 - p_n)^{n-k} = \\ &= \frac{n(n-1) \dots (n-k+1)}{k!} \left(\frac{\lambda_n}{n}\right)^k \left(1 - \frac{\lambda_n}{n}\right)^{n-k} = \\ &= \frac{\lambda_n^k}{k!} \cdot \frac{n(n-1) \dots (n-k+1)}{n^k} \left(1 - \frac{\lambda_n}{n}\right)^n \left(1 - \frac{\lambda_n}{n}\right)^{-k} = \\ &= \frac{\lambda_n^k}{k!} \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \dots \left(1 - \frac{k-1}{n}\right) \left(1 - \frac{\lambda_n}{n}\right)^n \left(1 - \frac{\lambda_n}{n}\right)^{-k}. \end{aligned}$$

Отсюда переходя к пределу при $n \rightarrow \infty$ получим требуемый результат. \square

Замечание 23.5. Теорема Пуассона доставляет приближённую формулу

$$P(\xi_n = k) = C_n^k p_n^k (1 - p_n)^{n-k} \approx \frac{(np)^k}{k!} e^{-np}$$

Заметим, что в примере 23.9.2) мы вычислили математическое ожидание и дисперсию для биномиального распределения $M\xi = np$ и $D\xi = npq$. Поэтому имеем

$$P(\xi_n = k) \approx \frac{(M\xi)^k}{k!} e^{-M\xi}.$$

Параметр $\lambda = np$ является математическим ожиданием для распределения Пуассона. Таким образом, биномиальное распределение аппроксимируется распределением Пуассона так, что оба распределения имеют одно и

то же математическое ожидание. На самом деле оба распределения имеют и достаточно близкие дисперсии. В области применения приближённой формулы вероятность p должна быть мала, а значит $q = 1 - p \approx 1$, поэтому $D\xi = npq \approx np = \lambda$.

Пример. Вероятность попадания в цель при каждом выстреле равна 0,001. Найти вероятность попадания в цель двумя и более пулями, если число выстрелов равно 5 000. Решение. Этот эксперимент реализует схему Бернулли, и использование точной формулы даёт значение искомой вероятности равное

$$P(\xi_n \geq 2) = \sum_{k=2}^{5000} P(\xi_{5000} = k) = \quad (4)$$

$$= 1 - P(\xi_{5000} = 0) - P(\xi_{5000} = 1) = \quad (5)$$

$$= 1 - 0,999^{5000} - 5000 \cdot 0,001 \cdot 0,999^{4999}. \quad (6)$$

Воспользуемся теоремой Пуассона $P(\xi_n = k) \approx \frac{(np)^k}{k!} e^{-np}$. Вычислим: $\lambda = np = 5000 \cdot 0,001 = 5$, $P(\xi_{5000} = 0) \approx e^{-5}$, $P(\xi_{5000} = 1) \approx 5e^{-5}$. Поэтому

$$P_n(\xi \geq 2) = 1 - P(\xi_{5000} = 0) - P(\xi_{5000} = 1) \approx 1 - 6e^{-5} \approx 0,9596.$$

Вычисление по точной формуле (4) с точностью до четвёртого знака даёт ответ $P_n(\xi \geq 2) = 0,9575$. Нетрудно подсчитать, что ошибка от применения приближённой формулы составляет меньше 0,25%.

Теорема Муавра. Эта предельная теорема обеспечивает приближённое вычисление вероятностей биномиального распределения,

$$P(\xi_n = k) = C_n^k \cdot p^k \cdot q^{n-k},$$

в случае, когда при $n \rightarrow \infty$ обе величины $k \rightarrow \infty$ и $n - k \rightarrow \infty$ так, что величина $x_k = \frac{k-np}{\sqrt{npq}}$ ограничена, т.е. существует $N > 0$, такое что $|x_k| < N$ для $k = 0, 1, 2, \dots$.

Теорема 23.6 (Локальная теорема Муавра²²). Если вероятность $0 < p < 1$ постоянна, и величина $x_k = \frac{k-np}{\sqrt{npq}}$ ограничена, то справедлива формула

$$\lim_{n \rightarrow \infty} \left[C_n^k p^k q^{n-k} - \frac{1}{\sqrt{2\pi npq}} \exp\left(-\frac{x_k^2}{2}\right) \right] = 0$$

(Без доказательства.)

²²Абрахам де Муавр (Abraham de Moivre, 1667 — 1754), английский математик французского происхождения. Основные труды по теории рядов, теории вероятностей, теории комплексных чисел.

Замечание 23.7. Локальная теорема Муавра доставляет приближённую формулу

$$P(\xi_n = k) = C_n^k p^k q^{n-k} \approx \frac{1}{\sqrt{2\pi npq}} \exp\left(-\frac{x_k^2}{2}\right)$$

Напомним, что в примере 18.9.2) мы вычислили математическое ожидание и дисперсию для биномиального распределения $M\xi = np$ и $D\xi = npq$. Поэтому имеем

$$x_k = \frac{k - np}{\sqrt{npq}} = \frac{k - M\xi}{\sqrt{D\xi}} = \frac{K - M\xi}{\sigma}$$

$$P(\xi_n = k) \approx \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(k - M\xi)^2}{2\sigma^2}\right)$$

Последняя формула означает, что биномиальное распределение аппроксимировано нормальным, причём оба распределения имеют одинаковые математические ожидания и дисперсии.

Пример. Пусть надо вычислить $P(\xi_n = k)$ при $n = 10000, k = 40, p = 0,005$. Сделаем подготовительные вычисления:

$$\sqrt{npq} = \sqrt{10000 \cdot 0,005 \cdot 0,995} = \sqrt{49,75} \approx 7,05$$

$$\frac{k - np}{\sqrt{npq}} \approx -1,42.$$

По теореме Муавра имеем

$$P(\xi_n = k) \approx \frac{1}{7,05\sqrt{2\pi}} \exp\left(-\frac{1,42^2}{2}\right).$$

Функция

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

табулирована (см., например, Таблицу 23.8.8 в [6]), по таблице находим

$$P(\xi_n = k) \approx \frac{0,1456}{7,05} = 0,0206.$$

Заметим, что точное значение с точностью до третьего знака $P(\xi_n = k) \approx 0,00197$. Полученная ошибка составляет 4,6 %.

Теорема 23.8 (Интегральная теорема Муавра(1730)–Лапласа²³ (1812)). Если случайная величина ξ имеет биномиальное распределение $P(\xi_n = k) = C_n^k p^k q^{n-k}$, тогда для любых вещественных k_1 и k_2 имеет место равенство

$$\lim_{n \rightarrow \infty} P(k_1 \leq \xi_n \leq k_2) = \int_x^y \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt,$$

где $x = \frac{k_1 - np}{\sqrt{npq}}$ и $y = \frac{k_2 - np}{\sqrt{npq}}$. Теорему Муавра-Лапласа мы докажем в §25 как частный случай центральной предельной теоремы.

24. Характеристические функции

Замечание 24.1. Решение многих задач теории вероятностей, особенно тех, которые связаны с суммированием независимых случайных величин, удаётся получить с помощью т.н. характеристических функций. Связь между преобразованием Фурье и характеристической функцией можно описать следующим образом. Если обозначить преобразование Фурье через Φ , то формулы прямого и обратного преобразования можно записать следующим образом

$$\begin{aligned}\varphi(t) &= \Phi[f] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-ixt} f(x) dx, \\ f(x) &= \Phi^{-1}[\varphi] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx} \varphi(t) dt.\end{aligned}$$

Если ξ — случайная величина, имеющая плотность $f_\xi(x)$, то в терминах преобразования Фурье её характеристическая функция есть

$$\Theta_\xi(t) = \sqrt{2\pi} \Phi^{-1}[f_\xi(x)] = \int_{-\infty}^{\infty} e^{itx} dx$$

²³Пьер-Симон Лаплас (Pierre-Simon Laplace, 1749 – 1827), французский математик, астроном, физик и философ; известен работами по небесной механике, дифференциальным уравнениям, является одним из создателей теории вероятностей. Заслуги Лапласа в чистой и прикладной математике и астрономии громадны: он усовершенствовал почти все разделы этих наук.

Определение 24.2. Пусть $f_\xi(x)$ – плотность случайной величины ξ , тогда функция

$$\Theta_\xi(t) = \mathbf{M}e^{it\xi} = \int_{-\infty}^{\infty} e^{itx} f_\xi(x) dx$$

называется характеристической функцией случайной величины ξ .

Замечание 24.3. 1) Пусть ξ и η – вещественные случайные величины. Составим комплексную случайную величину $\xi + i\eta$. Мы распространяем действие знака математического ожидания \mathbf{M} на любую комплексную случайную величину по свойству линейности:

$$\mathbf{M}(\xi + i\eta) = \mathbf{M}\xi + i\mathbf{M}\eta.$$

2) Зная характеристическую функцию $\Theta_\xi(t)$, можно однозначно восстановить функцию распределения, а также плотность вероятности или ряд распределения случайной величины ξ . Например, если модуль характеристической функции $\Theta_\xi(t)$ интегрируем на всей прямой, то плотность $f_\xi(x)$ случайной величины ξ находится по формуле

$$f_\xi(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{itx} \Theta_\xi(t) dt.$$

Свойства характеристических функций

Теорема 24.4. *Характеристическая функция $\Theta_\xi(t)$ любой случайной величины ξ равномерно непрерывна. (Без доказательства.)*

Лемма 24.5. 1. $\Theta_\xi(0) = 1$

2. $|\Theta_\xi(t)| \leq 1$ для $-\infty < t < \infty$.

Доказательство. 1. $\Theta_\xi(0) = \mathbf{M}e^0 = 1$.

2. $|\Theta_\xi(t)| = |\mathbf{M}e^{it\xi}| \leq \mathbf{M}|e^{it\xi}| = \mathbf{M}1 = 1$.

□

Лемма 24.6. *Если $\eta = a\xi + b$, где a и b – постоянные, то*

$$\Theta_\eta(t) = \Theta_\xi(at)e^{ibt}.$$

Доказательство. $\Theta_\eta(t) = \mathbf{M}e^{it\eta} = \mathbf{M}e^{it(a\xi+b)} = e^{ibt}\mathbf{M}e^{iat\xi} = e^{ibt}\Theta_\xi(at)$

□

Лемма 24.7. Если ξ_1 и ξ_2 — независимые случайные величины, то

$$\Theta_{\xi_1+\xi_2}(t) = \Theta_{\xi_1}(t) \cdot \Theta_{\xi_2}(t).$$

Доказательство. Т.к. ξ_1 и ξ_2 — независимые величины, то $e^{it\xi_1}$ и $e^{it\xi_2}$ тоже независимы. Тогда

$$\Theta_{\xi_1+\xi_2}(t) = \mathbf{M}e^{it(\xi_1+\xi_2)} = \mathbf{M}e^{it\xi_1} \cdot e^{it\xi_2} \stackrel{16.5.6)}{=} \mathbf{M}e^{it\xi_1} \cdot \mathbf{M}e^{it\xi_2} = \Theta_{\xi_1}(t) \cdot \Theta_{\xi_2}(t).$$

□

Следствие 24.8. Если ξ_1, \dots, ξ_n — независимые случайные величины, то

$$\Theta_{\xi_1+\dots+\xi_n}(t) = \Theta_{\xi_1}(t) \cdot \dots \cdot \Theta_{\xi_n}(t).$$

Лемма 24.9. $\Theta_{\xi}(-t) = \overline{\Theta_{\xi}(t)}$, где черта означает комплексное сопряжение.

Доказательство. $\Theta_{\xi}(-t) = \mathbf{M}e^{-it\xi} = \mathbf{M}\overline{e^{it\xi}} = \overline{\mathbf{M}e^{it\xi}} = \overline{\Theta_{\xi}(t)}$.

□

Лемма 24.10. Если ξ — дискретная случайная величина, заданная рядом распределения (конечным или бесконечным)

ξ	x_1	x_2	\dots	x_k
P	p_1	p_2	\dots	p_k

$$\text{то } \Theta_{\xi}(t) = \sum_k e^{itx_k} p_k$$

Доказательство. Это утверждение непосредственно следует из теор. 17.4.1).

□

Лемма 24.11. Пусть существует начальный момент n -го порядка, $n = 1, 2, \dots$, случайной величины ξ , т.е. $\mathbf{M}|\xi|^n < \infty$. Тогда её характеристическая функция $\Theta_{\xi}(t)$ является n раз непрерывно дифференцируемой, и её n -я производная в нуле равна

$$\Theta_{\xi}^{(n)}(0) = i^n \mathbf{M}\xi^n$$

Доказательство. Заметим сначала, что т.к. существует начальный момент $\mathbf{M}\xi^n$, то существуют все моменты $\mathbf{M}\xi^k$ при $k < n$.

По лемме 24.5.2) имеем $|\Theta_{\xi}(t)| \leq 1$, поэтому интеграл в правой части

$$\Theta_{\xi}(t) = \int_{-\infty}^{\infty} e^{itx} f_{\xi}(x) dx$$

равномерно сходится по параметру t и следовательно его можно дифференцировать по t :

$$\begin{aligned}\Theta'_\xi(t) &= i \int_{-\infty}^{\infty} x e^{itx} f_\xi(x) dx, \\ \Theta''_\xi(t) &= i^2 \int_{-\infty}^{\infty} x^2 e^{itx} f_\xi(x) dx, \\ &\dots\dots\dots \\ \Theta^{(n)}_\xi(t) &= i^n \int_{-\infty}^{\infty} x^n e^{itx} f_\xi(x) dx.\end{aligned}$$

Подставляя в эти равенства $t = 0$, получим требуемый результат: $\Theta'_\xi(0) = iM\xi$, $\Theta''_\xi(0) = i^2 M\xi^2$, \dots , $\Theta^{(n)}_\xi(0) = i^n M\xi^n$. \square

Лемма 24.12. Если существуют начальные моменты $M\xi$, $M\xi^2$, \dots , $M\xi^n$ случайной величины ξ , т.е. $M|\xi|^n < \infty$, тогда в окрестности нуля её характеристическая функция $\Theta_\xi(t)$ разлагается в ряд Тейлора²⁴

$$\Theta_\xi(t) = 1 + \frac{iM\xi}{1!}t + \frac{i^2 M\xi^2}{2!}t^2 + \dots + \frac{i^n M\xi^n}{n!}t^n + o(t^n)$$

Доказательство. В стандартное разложение функции $\Theta_\xi(t)$ в ряд Тейлора

$$\Theta_\xi(t) = 1 + \frac{\Theta'_\xi(0)}{1!}t + \frac{\Theta''_\xi(0)}{2!}t^2 + \dots + \frac{\Theta^{(n)}_\xi(0)}{n!}t^n + o(t^n)$$

подставим выражения для $\Theta'_\xi(0)$, $\Theta''_\xi(0)$, \dots , $\Theta^{(n)}_\xi(0)$, доставляемые теор. 24.11, получим требуемый результат. \square

В заключение параграфа сформулируем без доказательства теорему Леви, которая устанавливает «непрерывное» соответствие между слабо сходящимися последовательностями случайных величин и поточечно сходящимися последовательностями характеристических функций. «Непрерывность» этого соответствия состоит в том, что пределу при слабой сходимости соответствует предел при поточечной сходимости и наоборот.

²⁴Брук Тэйлор (Brook Taylor, 1685 — 1731), английский математик, именем которого назван ряд (опубликованный им в 1715—1717 гг.), однако этот ряд был известен и применялся ещё в XVII веке Грегори и Ньютоном.

Теорема 24.13 (Леви²⁵ о непрерывном соответствии). *Последовательность случайных величин $\{\xi_n\}_{n=1}^\infty$ слабо сходится к случайной величине ξ тогда и только тогда, когда последовательность их характеристических функций $\{\Theta_{\xi_n}(t)\}_{n=1}^\infty$ поточечно сходится к характеристической функции $\Theta_\xi(t)$.*

25. Вычисление характеристических функций

Пример 25.1. Пусть случайная величина ξ имеет распределение Бернулли

ξ	0	1
P	$1-p$	p

По лемме 24.10 её характеристическая функция равна

$$\Theta_\xi(t) = \mathbf{M}e^{it\xi} = e^{it \cdot 0}(1-p) + e^{it \cdot 1}p = 1-p + pe^{it}$$

Пример 25.2. Пусть случайная величина ξ имеет биномиальное распределение (см. опред. 9.3) $k \mapsto \mathbf{P}(\xi_n = k) = C_n^k p^k (1-p)^{n-k}$, где $k = 0, 1, 2, \dots, n$. Её характеристическая функция равна

$$\Theta_\xi(t) = \mathbf{M}e^{it\xi} = \sum_{k=0}^n e^{itk} C_n^k p^k (1-p)^{n-k} = \sum_{k=0}^n C_n^k (pe^{it})^k (1-p)^{n-k} = (1-p+pe^{it})^n,$$

где последнее равенство есть бином Ньютона.

Пример 25.3. Пусть случайная величина ξ имеет распределение Пуассона (см. опред. 23.2) $k \mapsto \mathbf{P}\lambda(k) = \frac{\lambda^k}{k!} e^{-\lambda}$, где $k = 0, 1, 2, \dots$

По лемме 24.10 её характеристическая функция равна

$$\Theta_\xi(t) = \mathbf{M}e^{it\xi} = \sum_{k=0}^{\infty} e^{itk} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^{it})^k}{k!} = e^{-\lambda} e^{\lambda e^{it}} = e^{\lambda(e^{it}-1)}$$

Пример 25.4. Пусть случайная величина ξ имеет показательное распределение (см. опред. 12.7), $f_\xi(x) = \begin{cases} 0, & x < 0 \\ \lambda e^{-\lambda x}, & x \geq 0 \end{cases}$. По лемме 24.10 её характеристическая функция равна

$$\begin{aligned} \Theta_\xi(t) &= \mathbf{M}e^{it\xi} = \int_0^\infty e^{itx} \lambda e^{-\lambda x} dx = \lambda \int_0^\infty e^{-(\lambda-it)x} dx = \\ &= \frac{\lambda}{\lambda-it} \left(-e^{-(\lambda-it)x} \Big|_0^\infty \right) = \frac{\lambda}{\lambda-it}, \end{aligned}$$

²⁵ Поль Пьер Леви (Paul Pierre Levy, 1886 — 1971), выдающийся французский математик, основные труды по теории вероятностей, функциональному анализу, теории функций и механике.

поскольку при $x \rightarrow \infty$ модуль величины $e^{-(\lambda-it)x} = e^{itx} \lambda e^{-\lambda x}$ стремится к нулю: $|e^{-(\lambda-it)x}| = e^{-\lambda x} \rightarrow 0$.

Пример 25.5. Пусть случайная величина ξ имеет нормальное распределение (см. опред. 12.8)

$$f_{\xi}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right)$$

Её характеристическая функция равна

$$\begin{aligned} \Theta_{\xi}(t) = \mathbf{M}e^{it\xi} &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx} e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-t^2/2} e^{-(x-it)^2/2} dx = \\ &= e^{-t^2/2} \int_{-\infty}^{\infty} e^{-(x-it)^2/2} d(x-it) = e^{-t^2/2} \end{aligned}$$

где в показателе экспоненты мы выделили полный квадрат и получили интеграл от функции $\frac{1}{\sqrt{2\pi}} e^{-u^2/2}$.

Напомним:

Определение 16.7. Если две независимые случайные величины имеют одно и то же распределение (возможно, с разными параметрами), и их сумма имеет то же самое распределение, то распределение называется устойчивым относительно суммирования.

Лемма 16.9. Биномиальное распределение является устойчивым относительно суммирования.

Доказательство.

$$\Theta_{\xi_n+\xi_m}(t) = \theta_{\xi_n}(t)\theta_{\xi_m}(t) = (1-p+pe^{it})^n(1-p+pe^{it})^m = (1-p+pe^{it})^{n+m},$$

следовательно

$$\mathbf{P}(\xi_{n+m} = k) = C_{n+m}^k p^k (1-p)^{n+m-k},$$

где $k = 0, 1, 2, \dots, n+m$. □

Лемма 16.10. Нормальное распределение является устойчивым относительно суммирования.

Доказательство.

$$\begin{aligned}\Theta_{\xi_n+\xi_m}(t) &= \theta_{\xi_n}(t)\theta_{\xi_m}(t) = \exp\left(ita_1 - \frac{t^2\sigma_1^2}{2}\right) \exp\left(ita_2 - \frac{t^2\sigma_2^2}{2}\right) = \\ &= \exp\left(it(a_1 + a_2) - \frac{t^2(\sigma_1^2 + \sigma_2^2)}{2}\right),\end{aligned}$$

следовательно

$$P(\xi_{n+m} = k) = C_{n+m}^k p^k (1-p)^{n+m-k},$$

где $k = 0, 1, 2, \dots, n+m$. □

26. Центральная предельная теорема

Замечание 26.1. Центральные предельные теоремы (ЦПТ) в теории вероятностей – это класс теорем, утверждающих, что сумма большого количества независимых случайных величин имеет распределение близкое к нормальному.

Так как многие случайные явления в природе и социальном обществе являются суммами большого числа случайных факторов, то центральные предельные теоремы обосновывают фундаментальность нормального закона распределения. История получения ЦПТ растянулась на два века – от первых работ Муавра 1730 г. до необходимых и достаточных условий, полученных в 30-х гг. XX века.

Самый общий классический случай ЦПТ для неодинаково распределённых последовательностей случайных величин принадлежит А.М. Ляпунову²⁶ (1901). Мы докажем ЦПТ в классической форме (теор. 26.3), а чтобы иметь представление о ЦПТ в форме Ляпунова (теор. 26.5) мы приведём её без доказательства.

Замечание 26.2. Вернёмся ещё раз к ЗБЧ в форме Хинчина. Пусть

$$\xi_1, \xi_2, \dots, \xi_n, \dots$$

– последовательность попарно независимых одинаково распределённых случайных величин, т.е. все имеющие математическое ожидание $M\xi_1$ и дисперсию $D\xi_1$. Обозначим $S_n = \sum_{k=1}^n \xi_k$.

²⁶Александр Михайлович Ляпунов (1857 – 1918), выдающийся русский математик, основные труды по устойчивости динамических систем, теории потенциала и механике. Занятие теорией вероятностей было кратким эпизодом в его научной работе, тем не менее в этой области он добился фундаментальных результатов. Он доказал ЦПТ при весьма широких условиях методом, который и сейчас является одним из основных в теории вероятностей.

Тогда по ЗБЧ в форме Хинчина среднее арифметическое $\frac{S_n}{n}$ сходится по вероятности к $M\xi_1$, или, что то же самое, величина $\frac{S_n - nM\xi_1}{n}$ сходится по вероятности к нулю.

Запишем это утверждение с помощью неравенства Бьенеме – Чебышёва:

$$P\left(\left|\frac{S_n - nM\xi_1}{n}\right| < \varepsilon\right) \geq 1 - \frac{D\xi_1}{\varepsilon^2}$$

Т.к. вероятность не превосходит 1, то это неравенство можно переписать в виде

$$1 - \frac{D\xi_1}{\varepsilon^2} \geq P\left(\left|\frac{S_n - nM\xi_1}{n}\right| < \varepsilon\right) \leq 1$$

Если теперь перейти к пределу при $n \rightarrow \infty$, то левая часть этого сквозного неравенства стремится к единице, и вероятность, стоящая в средней части, «намертво» впечатывается в единицу. В этом и состоит высший смысл Закона Больших Чисел: при стремлении n к бесконечности и для любого $\varepsilon > 0$ неравенство $\left|\frac{S_n - nM\xi_1}{n}\right| < \varepsilon$ становится достоверным событием.

Поэтому ЗБЧ есть строгое математическое выражение свойства статистической устойчивости (см. опред 4.4). Заметим, что $M\xi_1$ и $D\xi_1$ являются константами, а $\varepsilon > 0$ хотя и любое, но тоже постоянное число.

Идея как-то оторвать упомянутую вероятность от единицы заключается в следующем: а не слишком ли на большую степень числа n мы поделили в выражении $\frac{S_n - nM\xi_1}{n}$?

Нельзя ли поделить на что-нибудь, растущее к бесконечности медленнее чем n так, чтобы последовательность событий $\left\{\left|\frac{S_n - nM\xi_1}{n^\alpha}\right| < \varepsilon\right\}$ при некотором $0 < \alpha < 1$ перестала бы сходиться к достоверному (и, само собой разумеется, не сходилась бы к невозможному событию)?

Оказывается, если выбрать $\alpha = 1/2$, то при $n \rightarrow \infty$ последовательность случайных величин $\eta_n = \frac{S_n - nM\xi_1}{\sqrt{n}}$ слабо сходится к случайной величине η , имеющей нормальное (!) распределение.

Теорема 26.3 (ЦПТ в классической форме). *Если случайные величины*

$$\xi_1, \xi_2, \dots, \xi_n, \dots$$

независимы, одинаково распределены и имеют конечные математическое ожидание $M\xi = a < \infty$ и дисперсию $D\xi = \sigma^2 < \infty$ и $\sigma \neq 0$, тогда последовательность случайных величин

$$\eta_n = \frac{\xi_1 + \dots + \xi_n - na}{\sigma\sqrt{n}}$$

слабо сходится к случайной величине, имеющей стандартное нормальное распределение, т.е.

$$\lim_{n \rightarrow \infty} \mathbf{P} \left(\frac{\xi_1 + \dots + \xi_n - na}{\sigma \sqrt{n}} \leq x \right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-u^2/2} du,$$

или кратко

$$\lim_{n \rightarrow \infty} \mathbf{P} (\eta_n \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-u^2/2} du,$$

Доказательство. Для упрощения выкладок вместо $\xi_1, \xi_2, \dots, \xi_n, \dots$ введём стандартные случайные величины $\bar{\xi}_1, \bar{\xi}_2, \dots, \bar{\xi}_n, \dots$ по формулам

$$\bar{\xi}_i = \frac{\xi_i - a}{\sigma}$$

Легко видеть, что случайные величины $\bar{\xi}_i$ являются независимыми, одинаково распределёнными, имеют нулевое математическое ожидание и единичную дисперсию и следовательно имеют одинаковые характеристические функции. Обозначим их характеристические функции тем же символом как и для первой случайной величины $\theta_{\bar{\xi}_i}(t) = \theta_{\xi_1}(t)$. Используя теор. 24.12, разложим её в ряд Тейлора, содержащий три первых члена плюс остаточный член:

$$\theta_{\bar{\xi}_i}(t) = 1 - \frac{1}{2}t^2 + o(t^2),$$

где мы учли, что $\mathbf{M}\bar{\xi}_1 = 0$ и $\mathbf{M}\bar{\xi}_1^2 = \mathbf{D}\bar{\xi}_1 + (\mathbf{M}\bar{\xi}_1)^2 = 1$.

Случайная величина η_n может быть записана в виде

$$\eta_n = \frac{\bar{\xi}_1 + \dots + \bar{\xi}_n}{\sqrt{n}} = \frac{\bar{\xi}_1}{\sqrt{n}} + \dots + \frac{\bar{\xi}_n}{\sqrt{n}}$$

Запишем и преобразуем её характеристическую функцию, получим

$$\begin{aligned} \theta_{\eta_n}(t) &= \theta_{\frac{\bar{\xi}_1}{\sqrt{n}} + \dots + \frac{\bar{\xi}_n}{\sqrt{n}}}(t) \stackrel{24.8}{=} \theta_{\frac{\bar{\xi}_1}{\sqrt{n}}}(t) \cdot \dots \cdot \theta_{\frac{\bar{\xi}_n}{\sqrt{n}}}(t) \stackrel{24.6}{=} \\ &= \theta_{\bar{\xi}_1} \left(\frac{t}{\sqrt{n}} \right) \cdot \dots \cdot \theta_{\bar{\xi}_n} \left(\frac{t}{\sqrt{n}} \right) = \left[\theta_{\bar{\xi}_1} \left(\frac{t}{\sqrt{n}} \right) \right]^n = \left[1 - \frac{t^2}{2n} + o \left(\frac{t^2}{n} \right) \right]^n \end{aligned}$$

Т.к. по второму замечательному пределу имеем

$$\lim_{n \rightarrow \infty} \left[1 - \frac{t^2}{2n} + o \left(\frac{t^2}{n} \right) \right]^n = e^{-t^2/2}$$

то это означает, что последовательность характеристических функций $\{\theta_{\eta_n}(t)\}_{n=1}^{\infty}$ поточечно сходится к характеристической функции $\theta_{\eta}(t) = e^{-t^2/2}$.

По пункту 25.5 эта функция является характеристической функцией стандартного нормального распределения с плотностью вероятности $f_{\eta}(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$.

По теор. 24.13 (Леви о непрерывном соответствии), последовательность случайных величин $\{\eta_n\}_{n=1}^{\infty}$ слабо сходится к случайной величине η , а по опред. 21.9 слабой сходимости имеем

$$\lim_{n \rightarrow \infty} P(\eta_n \leq x) \stackrel{11.3}{=} \lim_{n \rightarrow \infty} F_{\eta_n}(x) \stackrel{21.9}{=} F_{\eta}(x) \stackrel{12.1}{=} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-u^2/2} du.$$

Что и требовалось доказать. □

Следствие 26.4. Если случайные величины $\xi_1, \xi_2, \dots, \xi_n, \dots$ независимы, одинаково распределены и имеют конечную ненулевую дисперсию и $\eta_n = \frac{\xi_1 + \xi_2 + \dots + \xi_n - na}{\sigma\sqrt{n}}$, то следующие утверждения эквивалентны ЦПТ в классической форме.

1) Для любых $x < y$ имеет место равенство:

$$\lim_{n \rightarrow \infty} P(x < \eta_n < y) = \frac{1}{\sqrt{2\pi}} \int_x^y e^{-u^2/2} du.$$

2) Для любых $x < y$ имеет место равенство:

$$\lim_{n \rightarrow \infty} P(x \leq \eta_n \leq y) = \frac{1}{\sqrt{2\pi}} \int_x^y e^{-u^2/2} du.$$

Теорема 26.5 (ЦПТ в форме Ляпунова). Пусть случайные величины

$$\xi_1, \xi_2, \dots, \xi_n, \dots$$

независимы и имеют конечные абсолютные начальные моменты 3-го порядка ($M|\xi_n|^3 < \infty$). Введём следующие обозначения:

$$A_n = \sum_{k=1}^n M\xi_k, \quad B_n^2 = \sum_{k=1}^n D\xi_k, \quad C_n^3 = \sum_{k=1}^n M|\xi_k - M\xi_k|^3.$$

Если

$$\lim_{n \rightarrow \infty} \frac{C_n}{B_n} = 0,$$

тогда последовательность случайных величин

$$\frac{\xi_1 + \xi_2 + \dots + \xi_n - A_n}{B_n}$$

слабо сходится к случайной величине, имеющей стандартное нормальное распределение, т.е.

$$\lim_{n \rightarrow \infty} \mathbf{P} \left(\frac{\xi_1 + \xi_2 + \dots + \xi_n - A_n}{B_n} \leq x \right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-u^2/2} du.$$

В качестве следствия из ЦПТ докажем предельную теорему Муавра-Лапласа. Подобно ЗБЧ в форме Бернулли предельная теорема Муавра-Лапласа является утверждением только для схемы Бернулли. Напомним её формулировку.

Теорема 26.6 (Интегральная теорема Муавра-Лапласа). *Если случайная величина η_n имеет биномиальное распределение $\mathbf{P}(\eta_n = k) = C_n^k p^k q^{n-k}$, тогда для любых вещественных k_1 и k_2 имеет место равенство*

$$\lim_{n \rightarrow \infty} \mathbf{P}(k_1 \leq \eta_n \leq k_2) = \frac{1}{\sqrt{2\pi}} \int_x^y e^{-u^2/2} du,$$

где $x = \frac{k_1 - np}{\sqrt{npq}}$ и $y = \frac{k_2 - np}{\sqrt{npq}}$.

Доказательство. Приведём обозначения теоремы Муавра-Лапласа в соответствие с обозначениями ЦПТ. Случайная величина ξ есть число появлений события A в результате n испытаний в схеме Бернулли с вероятностью $\mathbf{P}(A) = p$.

Если обозначить через ξ_i число (равное 0 или 1) появлений события A в результате i -го испытания, где $i = 1, \dots, n$, то случайную величину η_n можно представить в виде суммы $\eta_n = \xi_1 + \dots + \xi_n$ независимых, одинаково распределённых случайных величин, имеющих одинаковые конечные математические ожидания $\mathbf{M}_{\xi_i} = p$ и дисперсии $\mathbf{D}_{\xi_i} = p(1 - p) = pq$.

Тогда по пункту 18.9.2) имеем $\mathbf{M}_{\eta_n} = np$ и $\mathbf{D}_{\eta_n} = npq$.

Заметим, что неравенство $k_1 \leq \eta_n \leq k_2$ эквивалентно неравенству

$$\frac{k_1 - np}{\sqrt{npq}} \leq \frac{\eta_n - np}{\sqrt{npq}} \leq \frac{k_2 - np}{\sqrt{npq}}.$$

По следствию 26.4.2) получим требуемый результат. □

Пример 26.7 (Ср. пример 21.7). Монету подбрасывают 10 000 раз. Найти вероятность того, что относительная частота выпадения герба отличается от классической вероятности $1/2$ не менее чем на $0,01$.

Решение. Другими словами, требуется найти $P\left(\left|\frac{n(A)}{n} - \frac{1}{2}\right| \geq 0,01\right)$, где $n = 10^4$, $n(A) = \sum_{k=1}^n \xi_k$ – число выпадений герба, а ξ_k являются независимыми случайными величинами, имеющие одно и то же распределение Бернулли с $p = \frac{1}{2}$, и равные единице, если выпал герб, и нулю – в противном случае.

Умножим обе части неравенства под знаком вероятности на $\sqrt{n} = 100$ и разделим на корень из дисперсии $\sqrt{D\xi_1} = \sqrt{pq} = \frac{1}{2}$.

$$\begin{aligned} P\left(\left|\frac{n(A)}{n} - p\right| \geq 0,01\right) &= 1 - P\left(\left|\frac{n(A) - np}{n}\right| < 0,01\right) = \\ &= 1 - P\left(\frac{\sqrt{n}}{\sqrt{D\xi_1}} \left|\frac{n(A) - np}{n}\right| < 0,01 \frac{\sqrt{n}}{\sqrt{D\xi_1}}\right) = \\ &= 1 - P\left(\left|\frac{n(A) - np}{\sqrt{npq}}\right| < 2\right) = \frac{1}{\sqrt{2\pi}} \int_{-2}^2 e^{-u^2/2} du = \\ &= 1 - 2\Phi(2) \simeq 1 - 2 \cdot 0,47725 = 0,0455, \end{aligned}$$

где значение $\Phi(2)$ интеграла вероятности взято из таблицы. Заметим, что применение ЦПТ и предельной теоремы Муавра-Лапласа доставляет лучшую оценку чем ЗБЧ в форме Бернулли (см. ответ в примере 21.7.).

27. Сферическое, ξ^2 -распределение и распределение Стьюдента

27.1. Сферическое распределение

Определение 27.1. Пусть случайные величины ξ_1, \dots, ξ_n независимы и одинаково распределены по нормальному закону распределения

$$f_{\xi_i}(x_i) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x_i^2/2\sigma^2},$$

то есть $a_i = 0$ и $\sigma_i = \sigma$ для всех $i = 1, \dots, n$. Тогда распределение случайного вектора $\xi = (\xi_1, \dots, \xi_n)$ называется нормальным сферическим распределением или проще сферическим распределением.

Лемма 27.2. Плотность вероятности сферического распределения есть

$$f_{\xi}(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2} \sigma^n} e^{-\frac{1}{2\sigma^2}(x_1^2 + \dots + x_n^2)}$$

Доказательство. Это утверждение очевидно следует из 27.1 и 15.6.2). \square

Замечание 27.3. Так как поверхности равной вероятности, то есть $f_{\xi}(x_1, \dots, x_n) = \text{const}$, являются концентрическими $(n-1)$ -мерными сферами $x_1^2 + \dots + x_n^2 = R^2$ с центром в начале координат, то сферическое распределение инвариантно относительно любого ортогонального преобразования $A : \mathbb{R}^n \rightarrow \mathbb{R}^n$. То есть, если $\eta = A \cdot \xi$, то $f_{\eta}(x_1, \dots, x_n) = f_{\xi}(x_1, \dots, x_n)$, где через A обозначена ортогональная матрица.

Положим в сферическом распределении $\sigma = 1$, получим плотность вероятности

$$f_{\xi}(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}(x_1^2 + \dots + x_n^2)}$$

и назовём её плотностью вероятности стандартного сферического распределения.

27.2. χ - и χ^2 -распределения

Определение 27.4. Пусть случайные величины ξ_1, \dots, ξ_n независимы и одинаково распределены по стандартному нормальному закону распределения. Случайные величины

$$\chi = \sqrt{\xi_1^2 + \dots + \xi_n^2}$$

и

$$\chi^2 = \xi_1^2 + \dots + \xi_n^2$$

называются соответственно χ - и χ^2 -распределениями с n степенями свободы.

Наша ближайшая задача найти их плотности вероятности $f_{\chi}(x)$ и $f_{\chi^2}(x)$.

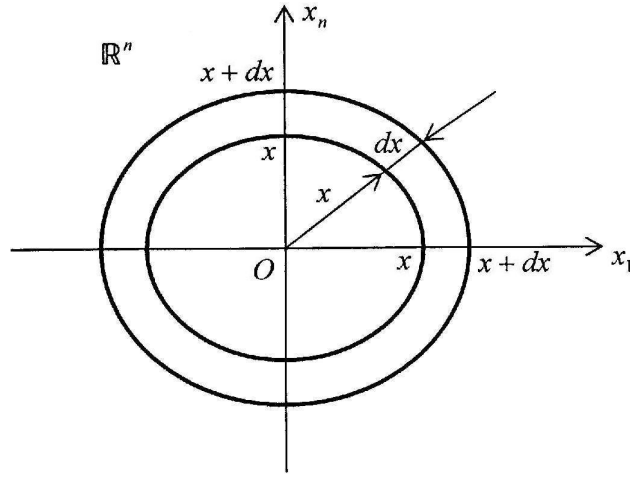


Рис. 26: К вычислению плотности χ -распределения

Теорема 27.5. Плотность вероятности χ -распределения задаётся по формуле

$$f_{\chi}(x) = \frac{x^{n-1}}{2^{\frac{n}{2}-1} \Gamma(\frac{n}{2})} e^{-\frac{1}{2}x^2}, x \geq 0$$

где $\Gamma(x)$ — гамма-функция.

Доказательство. Вероятность события $\{x < \chi < x + dx\}$ можно вычислить, интегрируя плотность вероятности $f_{\xi}(x_1, \dots, x_n)$ стандартного сферического распределения по n -мерному сферическому слою радиуса x и толщины dx (см. рис. 26).

$$P(x < \chi < x + dx) = \int \dots \int_{x < \sqrt{x_1^2 + \dots + x_n^2} < x + dx} f_{\xi}(x_1, \dots, x_n) dx_1 \dots dx_n.$$

Откуда получаем

$$\begin{aligned} f_{\xi}(x) &= \frac{P(x < \chi < x + dx)}{dx} = \frac{\int \dots \int_{\sqrt{x_1^2 + \dots + x_n^2} = x} f_{\xi}(x_1, \dots, x_n) dx_1 \dots dx_n}{\sqrt{x_1^2 + \dots + x_n^2} = x} = \\ &= \frac{1}{(2\pi)^{n/2}} \int \dots \int_{\sqrt{x_1^2 + \dots + x_n^2} = x} e^{-\frac{1}{2}(x_1^2 + \dots + x_n^2)} dx_1 \dots dx_n = \\ &= \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}x^2} \int \dots \int_{\sqrt{x_1^2 + \dots + x_n^2} = x} dx_1 \dots dx_n = C x^{n-1} e^{-\frac{1}{2}x^2} \end{aligned}$$

где последнее равенство справедливо потому, что площадь $(n-1)$ -мерной сферы радиуса x пропорциональна x^{n-1} . Чтобы найти константу C , воспользуемся свойством плотности $\int_0^{\infty} f_{\chi}(x)dx = 1$, откуда получаем

$$C \int_0^{\infty} x^{n-1} e^{-\frac{1}{2}x^2} dx = 2^{\frac{n}{2}-1} \Gamma\left(\frac{n}{2}\right) C = 1$$

откуда и следует требуемый результат. \square

Теорема 27.6. Плотность $f_{\chi^2}(x)$ случайной величины χ^2 есть

$$f_{\chi^2}(x) = \frac{x^{\frac{n}{2}-1}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} e^{-\frac{x}{2}}, \quad x \geq 0$$

Доказательство. Ясно, что случайные величины χ и χ^2 связаны функцией $g(x) = x^2$. Поэтому по теор. 14.2, находим, что

$$f_{\chi^2}(x) = |(g^{-1}(x))'| \cdot f_{\chi}(g^{-1}(x)) = \frac{1}{2\sqrt{x}} f_{\chi}(\sqrt{x}) = \frac{x^{\frac{n}{2}-1}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} e^{-\frac{x}{2}}, \quad x \geq 0$$

\square

Замечание 27.7. Заметим, что χ^2 -распределение с двумя степенями свободы ($n = 2$) является показательным распределением 12.7 с параметром $\lambda = 1/2$ и с плотностью вероятности $\frac{1}{2}e^{-\frac{x}{2}}, x \geq 0$. А χ -распределение с тремя степенями свободы ($n = 3$) называется распределением Максвелла и описывает в кинетической теории газов распределение модуля скорости частиц в состоянии термодинамического равновесия.

27.3. О распределении Стьюдента

Замечание 27.8. Этот пункт имеет информативный характер. Распределение Стьюдента²⁷ часто возникает в прикладных задачах математической статистики. Пусть случайные величины $\xi_0, \xi_1, \dots, \xi_n$ независимы и одинаково распределены по стандартному нормальному закону распределения

²⁷Стьюдент — псевдоним английского химика и статистика Вильяма Госсета (William Sealy Gosset, 1876 — 1937). По окончании университета был (с 1899) младшим управляющим пивоваренной компании в Дублине. В то время в совете директоров компании уже понимали важность методов теории вероятностей и статистики для массового производства и направили Госсета на стажировку в Лондонский университет для изучения статистики. Вскоре Госсет получил теоретические результаты в статистике, которые оказались важными для пивоварения и которые он хотел опубликовать. В компании не возражали против его публикаций, однако совет директоров хотел, чтобы конкуренты как можно дольше не знали о применении статистики на пивоварне. Это объясняет появление псевдонима Стьюдент.

$$f_{\xi_i}(x_i) = \frac{1}{\sqrt{2\pi}} e^{-x_i^2/2}$$

для всех $i = 0, 1, \dots, n$. Тогда случайная величина

$$\tau_n = \frac{\xi_0}{\sqrt{\frac{1}{n}(\xi_1^2 + \dots + \xi_n^2)}}$$

называется отношением Стьюдента, её распределение называется распределением Стьюдента с n степенями свободы. Формула плотности вероятности τ_n имеет вид

$$f_{\tau_n}(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad -\infty < x < \infty$$

Замечание 27.9. Легко проверить, что последовательность отношений Стьюдента $\{\tau_n\}_{n=1}^{\infty}$ слабо сходится к стандартному нормальному распределению, то есть

$$\lim_{n \rightarrow \infty} f_{\tau_n}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

28. Цепи Маркова

Непосредственным обобщением схемы независимых испытаний (см. опред. 10.3) является схема цепей Маркова.

Определение 28.1. Пусть эксперимент состоит в том, что

1. проводят последовательность испытаний с номерами $s = 1, 2, \dots$;
2. в каждом испытании появляется одно из k несовместных событий

$$A_1, A_2, \dots, A_k,$$

при этом то, что в s -ом испытании произошло событие A_i , будем обозначать A_i^s ;

3. вероятность появления события A_j^{s+1} (то есть j -го события в $(s+1)$ -ом испытании) зависит только от того, какое событие произошло в предыдущем s -ом испытании и не зависит от того, какие события произошли в

испытаниях с номерами $s - 1, s - 2, \dots$. Тогда такая последовательность событий называется (простой) цепью Маркова.

Пример. Согласно предложенной Н. Бором²⁸ модели атома водорода электрон может находиться только на одной из допустимых орбит. Обозначим через A_i событие, состоящее в том, что электрон находится на i -ой орбите. Эти события Бор назвал стационарными состояниями атома. Вероятность перехода электрона с i -ой на j -ю орбиту зависит только от i и j , потому что разность $j - i$ зависит от количества испускаемой или поглощаемой атомом энергии и не зависит от того, на каких орбитах находился электрон в прошлом. Этот пример доставляет цепь Маркова теоретически с бесконечным числом состояний атома.

Теория цепей Маркова довольно обширна, поэтому мы ограничимся изучением так называемых однородных цепей Маркова.

Рассмотрим условную вероятность $P(A_j^{s+1}|A_i^s)$, т.е. вероятность появления события A_j^{s+1} при условии, что событие A_i^s произошло.

Определение 28.2. Цепь Маркова называется однородной, если условная вероятность $P(A_j^{s+1}|A_i^s)$ не зависит от номера s .

При этом вероятность $P(A_j^{s+1}|A_i^s)$ называется вероятностью перехода от события A_i^s к событию A_j^{s+1} и обозначается

$$p_{ij} = P(A_j^{s+1}|A_i^s)$$

Замечание 28.3. Ясно, что

- 1) $0 \leq p_{ij} \leq 1$ и
- 2) $p_{i1} + p_{i2} + \dots + p_{ik} = 1$ для всех $i = 1, 2, \dots, k$.

Определение 28.4. Полная информация о вероятностях перехода в цепи Маркова содержится в таблице

$$\pi_1 = \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1k} \\ p_{21} & p_{22} & \dots & p_{2k} \\ \dots & \dots & \dots & \dots \\ p_{k1} & p_{k2} & \dots & p_{kk} \end{pmatrix}$$

которая называется матрицей перехода.

²⁸Нильс Хенрик Давид Бор (Niels Henrik David Bohr, 1865 — 1962), датский физик, один из создателей квантовой физики. Создал первую квантовую модель атома, участвовал в разработке основ квантовой механики, теории атомного ядра, ядерных реакций и взаимодействия элементарных частиц со средой.

Замечание 28.5. Главной задачей теории цепей Маркова является нахождение вероятности перехода от события A_i^s к событию A_j^{s+n} , произошедшему через n испытаний. Обозначим эту вероятность

$$p_{ij}(n) = P(A_j^{s+n} | A_i^s),$$

а искомую матрицу перехода через n испытаний через

$$\pi_n = \begin{pmatrix} p_{11}(n) & p_{12}(n) & \dots & p_{1k}(n) \\ p_{21}(n) & p_{22}(n) & \dots & p_{2k}(n) \\ \dots & \dots & \dots & \dots \\ p_{k1}(n) & p_{k2}(n) & \dots & p_{kk}(n) \end{pmatrix}$$

Теорема 28.6. Если $0 < m < n$, то $\pi_n = \pi_m \cdot \pi_{n-m}$.

Доказательство. Рассмотрим какое-нибудь промежуточное испытание с номером $s + m$, то есть $s < s + m < s + n$. Пусть в этом испытании появится какое-то событие A_r^{s+m} , где $1 \leq r \leq k$. В наших обозначениях вероятность перехода от события A_i^s к событию A_r^{s+m} равна

$$P(A_r^{s+m} | A_i^s) = p_{ir}(m),$$

а вероятность перехода от события A_r^{s+m} к событию A_j^{s+n} равна

$$P(A_j^{s+n} | A_r^{s+m}) = p_{jr}(n - m)$$

По формуле полной вероятности 7.2 имеем

$$p_{ij}(n) = \sum_{r=1}^k p_{ir}(m) \cdot p_{jr}(n - m).$$

Последняя формула в точности совпадает с формулой произведения матриц, поэтому получаем $\pi_n = \pi_m \cdot \pi_{n-m}$. \square

Теорема 28.7. Для любого $n \geq 1$, имеет место формула $\pi_n = \pi_1^n$.

Доказательство. Доказательство проведём методом математической индукции.

1. Проверяем тождество при $n = 1$: $\pi_1 = \pi_1^1 = \pi_1$.
2. Пусть при $n = q$ выполнено тождество $\pi_q = \pi_1^q$.
3. По теор. 28.6 имеем $\pi_{q+1} = \pi_1 \cdot \pi_q = \pi_1 \cdot \pi_1^q = \pi_1^{q+1}$.

Что и требовалось доказать. \square

Список литературы

- [1] Гнеденко Б.В. Курс теории вероятностей. –М.:Наука, 1988.
- [2] Чистяков В.П. Курс теории вероятностей. –М.:Наука, 1996.
- [3] Боровков А.А. Теория вероятностей. – М.:Эдиториал УРСС, 1999.
- [4] Севастьянов Б.А. Курс теории вероятностей и математической статистики. –М.:Наука, 1982.
- [5] Двайт Г.Б. Таблицы интегралов и другие математические формулы. 9. –СПб.:Лань, 2005.
- [6] Корн Г., Корн Т. Справочник по математике для научных работников и инженеров. М.: Наука, 1968.
- [7] Задачи по теории вероятностей (Сост. А.А. Дубков, А.Т. Гаврилин) Метод. разработка для студентов РФ ф-та ННГУ. –Н.Новгород: ННГУ, 1999.
- [8] Сборник задач по теории вероятностей, математической статистике и теории случайных функций (под редакцией А.А. Свешникова). М.: Наука, 1970.
- [9] Коваленко И.Н., Филиппова А.А. Теории вероятностей и математической статистики. –М.: Высшая школа, 1992.
- [10] Вентцель Е.С. Теория Вероятностей. –М.: Высшая школа, 1998.
- [11] Гихман И.И., Скороход А.В., Ядренко М.И. Теории вероятностей и математической статистики. –Киев: Высшая школа, 1988.
- [12] Агапов Г.И. Задачник по теории вероятностей –М.: Высшая школа, 1996.
- [13] Чернова Н.И. Теория вероятностей: Учеб. пособие/Новосиб. гос ун-т. Новосибирск, 2007. 160 с.