

Учебник по Эконометрике

Лекция 3: Основы теории вероятностей

Джон Стачурски

Лекции: Акшай Шенкер

Перевел: Алексей Кедо

1 октября 2020 г.

Вероятность фундаментальна для статистики и эконометрики, но технически сложна:

- множество событий, которым мы хотим присвоить вероятности, может быть очень большим
- нам нужны способы справиться со сложностью

Прежде чем мы начнем:

- множество S исчислимо, если оно конечно или может быть представлено как последовательность
- в противном случае, множество S неисчислимо

Пространство элементарных событий

Пространство элементарных событий можно представить себе как "список" всех возможных результатов в данном случайному эксперименте:

- пространство элементарных событий обычно обозначается как Ω
- пространством элементарных событий может быть только непустое множество
- типичный элемент Ω обозначается как ω

Реализация неопределенности приведет к выбору конкретной $\omega \in \Omega$

Пример. В случайному эксперименте, заключающемся в бросании кости один раз, множество возможных результатов представлено как $\Omega := \{1, \dots, 6\}$

Пример. Обезьяна Бертона Малкиела с завязанными глазами метает дротики в мишень радиусом 1

Введем обычные декартовы координаты с началом в центре мишени

Пусть (h, v) — типичные координаты, измеренные по горизонтали и вертикали соответственно

Естественное пространство элементарных событий — это $\Omega := \{(h, v) \in \mathbb{R}^2 : \|(h, v)\| \leq 1\}$ — также называется **единичным кругом** в \mathbb{R}^2

Неформально, **событие** — это подмножество Ω (мы скоро рассмотрим некоторые предостережения)

Событие A происходит всякий раз, когда отдельный $\omega \in \Omega$, выбранный в случайном эксперименте, оказывается в A

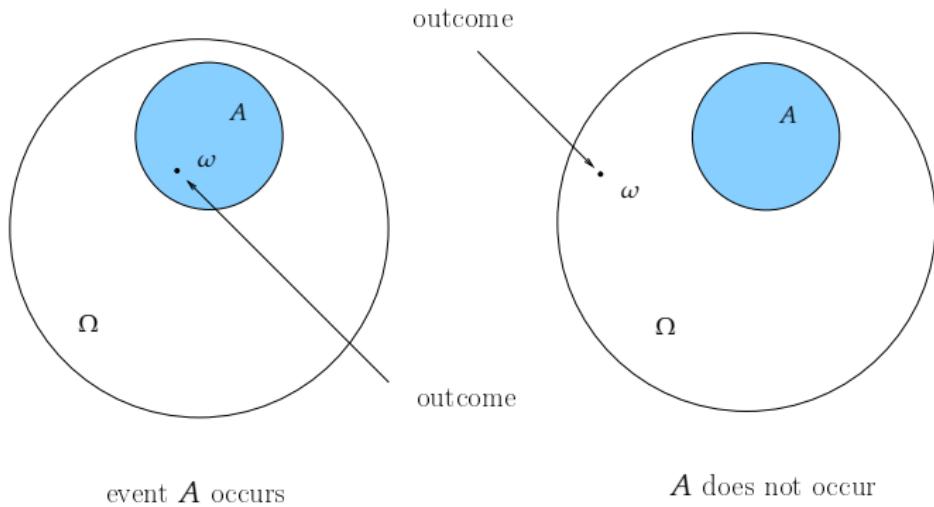


Рис.: Результаты и события

Вероятности и события

Можем ли мы присвоить вероятности каждому $\omega \in \Omega$?

Рассмотрим модель метания дротика, где Ω is \mathbb{R} :

- для $A \subset \Omega$, вероятность того, что дротик попадет в A пропорционален площади A
- вероятность точки $\omega \in \Omega$ будет меньше любой области A , содержащей ω
- для любых $\epsilon > 0$, мы можем найти A , содержащее ω , с площадью меньше ϵ

вероятность попадания в ω меньше ϵ для любых $\epsilon > 0$, значит вероятность попадания в ω должна быть равна нулю!

Итог: когда пространство элементарных событий исчислимо, присваивать вероятности событиям (подмножествам Ω), не каждому $\omega \in \Omega$

Но можем ли мы присваивать вероятности *каждому* попространству Ω ?

В модели дротика:

$$\mathbb{P}(A) = \frac{\lambda(A)}{\pi}$$

где $\lambda(A)$: = площадь множества A

Определение площади A для всех $A \subset \Omega$ проблематично:

- пространство Ω , наш круг для дартс в \mathbb{R}^2 , содержит много подмножеств, что производит странные явления
- Парадокс Банаха — Тарского

Решение: не принимайте множество событий за все подмножества Ω

Возьмите множество событий как определенные "хорошие" подмножества Ω , обозначенные \mathcal{F}

Присваивайте вероятности только подмножествам Ω в \mathcal{F}

Сигма-алгебра

Как мы можем гарантировать, что \mathcal{F} достаточно большой? В разумной вероятностной модели мы в идеале хотим:

- событие "не A " принадлежит \mathcal{F} , если $A \in \mathcal{F}$
- событие " A или B " принадлежит \mathcal{F} , если $A \in \mathcal{F}$ и $B \in \mathcal{F}$

Формально, \mathcal{F} — это **σ -алгебра** множества Ω , если

1. $A \in \mathcal{F} \implies A^c \in \mathcal{F}$,
2. $A_1, A_2, \dots \in \mathcal{F} \implies \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$ и
3. $\Omega \in \mathcal{F}$

1. - 3. подразумевают, что $\emptyset \in \mathcal{F}$, если \mathcal{F} — σ -алгебра

Событие \emptyset называется **невозможным событием**

Событие Ω называется **достоверным событием**

Пример. Множество $\{\Omega, \emptyset\}$ — σ - алгебра, называемая **тривиальной σ -алгеброй**

Борелевская σ -алгебра

σ -алгебра событий меняется от задачи к задаче

В \mathbb{R}^N мы используем Борелевские множества, обозначенные $\mathcal{B}(\mathbb{R}^N)$

- наименьшая σ -алгебра, содержащая все прямоугольники в \mathbb{R}^N

Почему Борелевская σ -алгебра?

- исключает "странные" множества
- включает повседневные полезные множества (включая плоскости и гиперплоскости, круги, сферы, многоугольники, конечные множества и последовательности точек)

Вероятности

Для данного события $B \in \mathcal{F}$, символ $\mathbb{P}(B)$ показывает "вероятность, что событие B случится"

$\mathbb{P}(B)$ показывает вероятность, что когда неопределенность решена и некоторые $\omega \in \Omega$ выбраны "естественно", то утверждение $\omega \in B$ является верным

Нам нужно установить ограничения, чтобы сделать вероятности правильными

Например, мы хотим исключить $\mathbb{P}(B) = -93$ для некоторых B

Пусть Ω — непустое множество и \mathcal{F} — σ -алгебра подпространств Ω . **Вероятность** \mathbb{P} на (Ω, \mathcal{F}) — функция из \mathcal{F} в $[0, 1]$, которая удовлетворяет

1. $\mathbb{P}(\Omega) = 1$ и
2. $\mathbb{P}(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$ для любой непересекающейся последовательности множеств $A_1, A_2, \dots \in \mathcal{F}$

\mathbb{P} также называется **вероятностной мерой**; втроем $(\Omega, \mathcal{F}, \mathbb{P})$ называются **вероятностным пространством**

Аксиома 1.: мы требуем $\mathbb{P}(\Omega) = 1$, так как, по построению, любая возможная ω лежит в множестве Ω .

Аксиома 2. называется **счетной аддитивностью**

В формулировке аксиомы (ii) несовместность попарна: любая различная пара A_i, A_j не имеет общих точек

Счетная аддитивность подразумевает конечную **аддитивность**:

$$\mathbb{P}(A_1 \cup \dots \cup A_k) = \mathbb{P}(A_1) + \dots + \mathbb{P}(A_k) \quad (1)$$

всякий раз, когда A_1, \dots, A_k несовместные

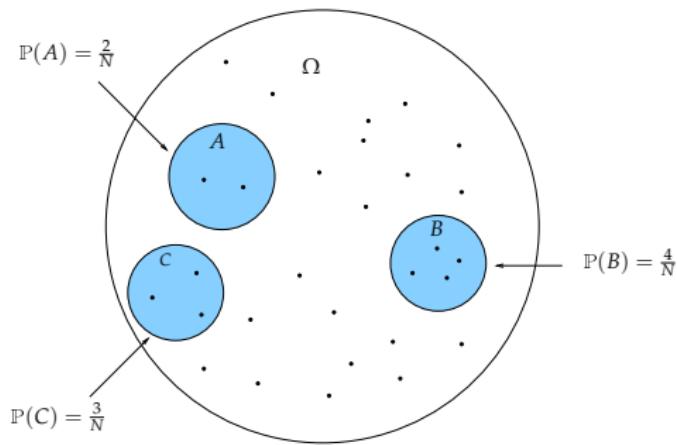


Рис.: Каждая из N точек случается с вероятностью $1/N$

$$\mathbb{P}(A \cup B \cup C) = \frac{9}{N} = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C)$$

Пример. Пусть $\Omega := \{1, \dots, 6\}$ представляют шесть разных граней кубика, как в примере ??

Так как Ω конечна, пусть \mathcal{F} — множество всех подмножеств Ω

Определим вероятность $\mathbb{P}: \mathcal{F} \rightarrow [0, 1]$

$$\mathbb{P}(A) := \frac{|A|}{6} \quad , \text{ где } |A| := \text{количество элементов в множестве } A \quad (2)$$

Легко заметить, что $0 \leq \mathbb{P}(A) \leq 1$ для любых $A \in \mathcal{F}$ и что $\mathbb{P}(\Omega) = 1$

Пример. (прод.) Касательно аддитивности, предположим, что A и B — два несовместных подмножества $\{1, \dots, 6\}$

Тогда $|A \cup B| = |A| + |B|$, значит

$$\mathbb{P}(A \cup B) = \frac{|A \cup B|}{6} = \frac{|A| + |B|}{6} = \frac{|A|}{6} + \frac{|B|}{6} = \mathbb{P}(A) + \mathbb{P}(B)$$

Это доказывает аддитивность для пар множеств. Аналогичный аргумент подтверждает аддитивность для любого конечного набора

Конечная аддитивность в этом случае эквивалентна счетной аддитивности, поскольку общее количество различных событий конечно.

Пример.

Чип памяти состоит из миллиардов переключателей/битов

- Переключатели могут быть выключены или включены (0 или 1)

Генератор случайных чисел обращается к N битам, включая или выключая каждый из них

Получается

- $\Omega := \{(b_1, \dots, b_N) :$
где b_n равняется 0 или 1 для каждого $n\}$
- $\mathbb{P}(A) := 2^{-N}(\#A)$

Упражнение: Покажите, что \mathbb{P} — вероятность

Пример. Рассмотрим снова модель мишени, где Ω — это единичный круг в \mathbb{R}^2

Для пространства событий, возьмем \mathcal{F} как множество Борелевских подмножеств в \mathbb{R}^2 , лежащее в Ω

Для \mathbb{P} мы следуем "равномерному" распределению вероятностей, заданному

То есть, $\mathbb{P}(B) = \lambda(B)/\pi$ для каждого $B \in \mathcal{F}$

Функция λ , которая назначает область для Борелевских множеств, известна как счетно-аддитивная, то есть $\lambda(\bigcup_n A_n) = \sum_{n=1}^{\infty} \lambda(A_n)$ при условии, что эти множества несовместные

Очевидно, $\mathbb{P}(\Omega) = 1$

Мера Лебега

Функция λ , отображающая Борелевские множества на свою "территорию", формально известна как **Мера Лебега**

§?? в ЕТ дает краткое введение в эту концепцию

Свойства вероятностной меры

Факт. (??) Пусть $(\Omega, \mathcal{F}, \mathbb{P})$ — вероятностное пространство и $A, B \in \mathcal{F}$

Если $A \subset B$, то

1. $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$,
2. $\mathbb{P}(A) \leq \mathbb{P}(B)$ (**монотонность**)
3. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$, и
4. $\mathbb{P}(\emptyset) = 0$.

Доказательство. Когда $A \subset B$, имеется $B = (B \setminus A) \cup A$, значит

$$\mathbb{P}(B) = \mathbb{P}(B \setminus A) + \mathbb{P}(A)$$

Все результаты следуют из этого (почему?)

Факт. (??) Если A и B — какие-нибудь (не обязательно несовместные) события, то

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

Докажите в качестве упражнения ?? в ЕТ

Факт подразумевает, что **полуаддитивность**: для любых $A, B \in \mathcal{F}$, имеется

$$\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$$

Условная вероятность и независимость

Условная вероятность A при данном B :

$$\mathbb{P}(A | B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \quad (3)$$

Вероятность A , при данной информации, что B случилось

События A и B называются **независимыми**, если
 $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$

- Если A и B независимы, то

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

Пример. Эксперимент: бросим кубик дважды

$$\Omega := \{(i, j) : i, j \in \{1, \dots, 6\}\} \quad \text{и} \quad \mathbb{P}(E) := \#E/36$$

Теперь рассмотрим события

$$A := \{(i, j) \in \Omega : i \text{ четное}\} \quad \text{и} \quad B := \{(i, j) \in \Omega : j \text{ четное}\}$$

В этом случае мы имеем

$$A \cap B = \{(i, j) \in \Omega : i \text{ и } j \text{ четные}\}$$

Упражнение: убедитесь, что $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$

Следовательно, A и B независимы с вероятностью \mathbb{P}

Закон полной вероятности

Закон полной вероятности утверждает, что:

Факт. (??) Если $A \in \mathcal{F}$ и B_1, \dots, B_M — части Ω с $\mathbb{P}(B_m) > 0$ для всех m , то

$$\mathbb{P}(A) = \sum_{m=1}^M \mathbb{P}(A | B_m) \cdot \mathbb{P}(B_m)$$

Доказательство. Возьмем $A \in \mathcal{F}$ и части B_1, \dots, B_M :

$$\mathbb{P}(A) = \mathbb{P}[A \cap (\bigcup_{m=1}^M B_m)] = \mathbb{P}[\bigcup_{m=1}^M (A \cap B_m)]$$

$$= \sum_{m=1}^M \mathbb{P}(A \cap B_m) = \sum_{m=1}^M \mathbb{P}(A | B_m) \cdot \mathbb{P}(B_m)$$

Теорема Байеса

Теорема Байеса: для любых событий A и B с положительной вероятностью, выполняется

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

Доказательство. Из определения условной вероятности:

$$\mathbb{P}(A | B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \quad \text{и} \quad \mathbb{P}(B | A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$$

Следовательно $\mathbb{P}(A \cap B) = \mathbb{P}(A | B) \mathbb{P}(B) = \mathbb{P}(B | A) \mathbb{P}(A)$

Перегруппировка уравнения (4)

Пример. Банки используют автоматизированные системы для обнаружения мошеннических или незаконных транзакций

Рассмотрим тест, который отвечает на каждую транзакцию с P или N :

- P значит "положительно" (транзакция отмечена как мошенническая)
- N значит "отрицательно" (транзакция отмечена как нормальная)

Пусть F значит мошенническая, предположим

- $\mathbb{P}(P | F) = 0.99$ (тест отмечает 99% мошеннических транзакций),
- $\mathbb{P}(P | F^c) = 0.01$ (процент ложных срабатываний), и
- $\mathbb{P}(F) = 0.001$ (распространенность мошенничества)

Какова вероятность мошенничества при положительном тесте?

Обратите внимание на закон Байеса

$$\mathbb{P}(F | P) = \frac{\mathbb{P}(P | F)\mathbb{P}(F)}{\mathbb{P}(P)}$$

и обратите внимание на закон полной вероятности

$$\mathbb{P}(P) = \mathbb{P}(P | F)\mathbb{P}(F) + \mathbb{P}(P | F^c)\mathbb{P}(F^c)$$

Следовательно

$$\mathbb{P}(F | P) = \frac{0.99 \times 0.001}{0.99 \times 0.001 + 0.01 \times 0.999} = \frac{11}{122} \approx \frac{1}{11}$$

Случайные переменные

Неформально: "значение, которое изменяется случайным образом"

Формально: **случайная переменная** x — функция из Ω в \mathbb{R}

Интерпретация: случайные величины преобразуют исходы в пространстве элементарных событий в числовые исходы.

Главная идея:

- "природа" выбирает ω в Ω
- случайная переменная сообщает результат как $x(\omega) \in \mathbb{R}$

Пример. Предположим, что Ω — множество бесконечных двоичных последовательностей

$$\Omega := \{(b_1, b_2, \dots) : b_n \in \{0, 1\} \text{ для каждого } n\}$$

Мы можем создавать различные отображения случайных переменных $\Omega \rightarrow \mathbb{R}$:

- количество "подбрасываний" до первого "орла":

$$x(\omega) = x(b_1, b_2, \dots) = \min\{n : b_n = 1\}$$

- количество "орлов" в первых 10 "подбрасываниях":

$$x(\omega) = x(b_1, b_2, \dots) = \sum_{n=1}^{10} b_n$$

- количество подбрасываний до первого орла:

$$x(\omega) = x(b_1, b_2, \dots) = \min\{n \in \mathbb{N} : b_n = 1\}$$

- Двоичная случайная величина или случайная величина Бернулли** говорит нам, возникают ли какие-либо орлы в первые 10 подбрасываний:

$$x(\omega) = y(b_1, b_2, \dots) := \min \left\{ \sum_{n=1}^{10} b_n, 1 \right\} \quad (5)$$

Случайная величина Бернулли

Бернулли или **двоичная случайная величина** с.в. x принимает значения $\{0, 1\}$

Теперь мы рассмотрим общий способ создания с.в. Бернулли

Пусть Q — утверждение, например, "а больше 3"

Определение: $\mathbb{1}\{Q\}$ равняется единице, если Q верно, в ином случае нулю.

Определим

$$x(\omega) = \mathbb{1}\{\omega \in A\} \text{ , где } A \in \mathcal{F}$$

С.в. показывает, случается ли событие C

Распространенный вариант обозначений: для произвольного $A \in \mathcal{F}$:

$$\mathbb{1}_A(\omega) := \mathbb{1}\{\omega \in A\} := \begin{cases} 1 & , \text{ если } \omega \in A \\ 0 & , \text{ иначе} \end{cases}$$

Факт. (??) Если A_1, \dots, A_N — подмножества Ω , то

1. $\mathbb{1}_{\bigcap_{n=1}^N A_n} = \prod_{n=1}^N \mathbb{1}_{A_n}$ и
2. $\mathbb{1}_{\bigcup_{n=1}^N A_n} = \sum_{n=1}^N \mathbb{1}_{A_n}$ всякий раз, когда множества несовместные

Смотрите упражнение ?? для доказательства

Здесь равенство означает оценку при любом $\omega \in \Omega$

Условные обозначения

Общие условные обозначения для с.в.:

$\{x \text{ обладает некоторым свойством}\} :=$

$\{\omega \in \Omega : x(\omega) \text{ обладает некоторым свойством}\}$

Пример.

$\{x \leq 2\} := \{\omega \in \Omega : x(\omega) \leq 2\}$

$\therefore \mathbb{P}\{x \leq 2\} := \mathbb{P}\{\omega \in \Omega : x(\omega) \leq 2\}$

Пример. Возьмем случайную величину x и $a \leq b$, мы утверждаем, что

$$\mathbb{P}\{x \leq a\} \leq \mathbb{P}\{x \leq b\}$$

Это выполняется, так как

$$\{x \leq a\} := \{\omega \in \Omega : x(\omega) \leq a\}$$

$$\subset \{\omega \in \Omega : x(\omega) \leq b\} := \{x \leq b\}$$

Теперь применим монотонность: $A \subset B \implies \mathbb{P}(A) \leq \mathbb{P}(B)$

Равенства, неравенства и арифметические операции следует интерпретировать точечно:

- $x \leq y \iff x(\omega) \leq y(\omega)$ для всех $\omega \in \Omega$,
- $x = y \iff x(\omega) = y(\omega)$ для всех $\omega \in \Omega$, и
- $z = \alpha x + \beta y \iff z(\omega) = \alpha x(\omega) + \beta y(\omega)$ для всех $\omega \in \Omega$

Случайные переменные — измеримые функции

Пусть $(\Omega, \mathcal{F}, \mathbb{P})$ — некоторое вероятностное пространство

Пусть B — некоторое подпространство \mathbb{R}

Рассмотрим вероятность

$$\mathbb{P}\{x \in B\} := \mathbb{P}\{\omega \in \Omega : x(\omega) \in B\}$$

где x — некоторая функция из Ω в \mathbb{R}

Нет возможности быть уверенным, что $\{\omega \in \Omega : x(\omega) \in B\}$ является элементом \mathcal{F}

- $\mathbb{P}\{x \in B\}$ может быть не определена

Мы должны ввести ограничения:

- для B мы естественно ограничиваем внимание до $\mathcal{B}(\mathbb{R})$,
Борелевское подмножество \mathbb{R}
- для x мы требуем $\{x \in B\} \in \mathcal{F}$ всюду, где B —
Борелевская множество

Формальное определение случайной величины:

Случайная величина в (Ω, \mathcal{F}) — функция $x: \Omega \rightarrow \mathbb{R}$,
удовлетворяющая

$$\{\omega \in \Omega : x(\omega) \in B\} \in \mathcal{F} \quad \text{для всех } B \in \mathcal{B}(\mathbb{R}) \quad (6)$$

Эти виды функций также называются \mathcal{F} -измеримыми
функциями

Обозначение прообраза: $x^{-1}(B)$ — это все $\omega \in \Omega$, такие что $x(\omega) \in B$

Перепишем (6) как

$$x^{-1}(B) \in \mathcal{F} \quad \text{для всех } B \in \mathcal{B}(\mathbb{R})$$

Таким образом, x "откатывает" Борелевские множества до событий

Измеримые преобразования

Мы хотим обсудить некоторые преобразования x

Например, $y := e^x$. y тоже случайная величина?

Да, при условии, что преобразование удовлетворяет
Борелевской измеримости

Формально, $f: \mathbb{R} \rightarrow \mathbb{R}$ называется **Борелевски измеримой**
или **\mathcal{B} -измеримой**, если

$$f^{-1}(B) \in \mathcal{B}(\mathbb{R}) \quad \text{для всех } B \in \mathcal{B}(\mathbb{R}) \quad (7)$$

Класс \mathcal{B} -измеримых функций огромен: любая непрерывная функция, любая возрастающая функция и т.д.

Предположим, что f \mathcal{B} -измерима и x — случайная величина

Мы имеем $\{y \in B\} \in \mathcal{F}$ для всех $B \in \mathcal{B}(\mathbb{R})$, так как

$$\{y \in B\} = \{f(x) \in B\} = \{x \in f^{-1}(B)\} \quad (8)$$

Таким образом, $y = f(x)$ — случайная величина

Ожидания

Мы хотим определить ожидания для произвольной с.в. x

Грубо говоря, $\mathbb{E}[x] :=$ "сумма" всех возможных значений x , взвешенных по их вероятностям.

"Сумма" в кавычках, потому что может быть бесконечное число слагаемых

Мы используем современный, формальный и строгий подход к определению ожиданий

Для конечных случайных величин дано вероятностное пространство $(\Omega, \mathcal{F}, \mathbb{P})$ и случайная величина x , принимающая только конечное количество различных значений s_1, \dots, s_J , **ожидание** x определяется как

$$\mathbb{E} x = \sum_{j=1}^J s_j \mathbb{P}\{x = s_j\} \quad (9)$$

Пример. Давайте применим это определение к простейшему возможному случаю, в котором случайная величина x удовлетворяет $x(\omega) = \alpha$ для всех $\omega \in \Omega$, где α — некоторое постоянное скалярное значение. В этом случае сумма в (9) имеет только одно слагаемое и

$$\mathbb{E}x = \alpha \mathbb{P}\{x = \alpha\} = \alpha \mathbb{P}\{\omega \in \Omega : x(\omega) = \alpha\} = \alpha \mathbb{P}(\Omega) = \alpha$$

Пример. Чтобы оценить ожидание двоичной случайной величины x , мы применим (9), получаем

$$\mathbb{E} x = 1 \times \mathbb{P}\{x = 1\} + 0 \times \mathbb{P}\{x = 0\} = \mathbb{P}\{x = 1\}$$

Пример. Рассмотрим N подбрасываний честной монеты

Пространство элементарных событий — это $\Omega := \{0, 1\}^N$,
события такие, что $\mathcal{F} :=$ все подмножества Ω , и
 $\mathbb{P}(A) := 2^{-N}|A|$ для всех $A \in \mathcal{F}$

Пусть $x(\omega) = x(b_1, \dots, b_N) = \sum_{n=1}^N b_n$

Прежде всего заметьте, что $0 \leq x \leq N$

По определению \mathbb{P} , для любого k мы имеем

$\mathbb{P}\{x = k\} = 2^{-N}|A_k|$, где

$$A_k := \{x = k\} = \left\{ (b_1, \dots, b_N) \in \Omega : \sum_{n=1}^N b_n = k \right\}$$

Из комбинаторики, $|A_k| = \binom{N}{k}$, где правая сторона называется **биномиальным коэффициентом** для N, k , что удовлетворяет $\sum_{k=0}^N k \binom{N}{k} = N2^{N-1}$ для всех N

Ожидания x :

$$\mathbb{E}x = \sum_{k=0}^N k 2^{-N} |A_k| = 2^{-N} \sum_{k=0}^N k \binom{N}{k} = \frac{N}{2}$$

Для обычного x , приблизительная произвольная случайная величина с конечными случайными величинами:

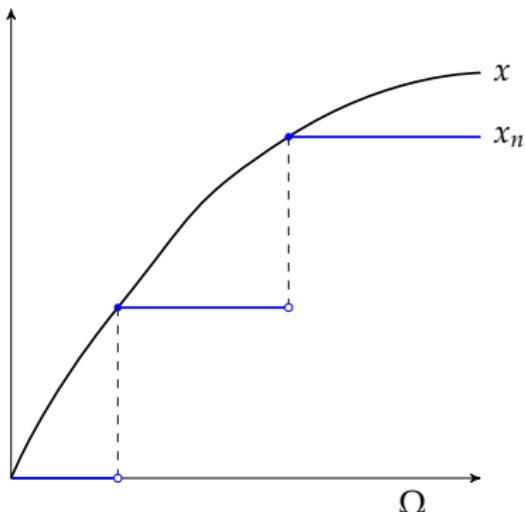


Рис.: Конечное приближение к общей случайной величине

Мы можем улучшать приближение без ограничений – Пусть x_n принимает большее и большее количество различных значений

Процесс дает последовательность конечных случайных величин x_n , сходящихся к x

Определим ожидания x как

$$\mathbb{E} x := \lim_{n \rightarrow \infty} \mathbb{E} x_n$$

$\mathbb{E}x$ также упоминается как **Интеграл Лебега** с.в. x относительно \mathbb{P} , с альтернативными обозначениями
 $\mathbb{E}x = \int x(\omega) \mathbb{P}(d\omega)$

Существует ли последовательность аппроксимации случайной величины? Да, смотрите страницу 94 в ЕТ и **dudley2002real**, утверждение 4.1.5

Если x принимает отрицательные значения, то пишут

$$x = x^+ - x^-$$

$$\text{где } x^+ := \max\{x, 0\} \text{ и } x^- = \min\{x, 0\}$$

Определим ожидания как

$$\mathbb{E}x := \mathbb{E}x^+ - \mathbb{E}x^-$$

Сфокусируемся на **интегрируемых** случайных величинах: всех случайных величинах x , таких что $\mathbb{E}|x| < \infty$

- мы имеем $x^+ \leq |x|$ и $x^- \leq |x|$
- таким образом, $\mathbb{E}x := \mathbb{E}x^+ - \mathbb{E}x^-$ хорошо определены (почему?)

Свойства ожидания

Факт. (??) Возьмем любое вероятностное пространство $(\Omega, \mathcal{F}, \mathbb{P})$, в нем существует однозначно определенная функция \mathbb{E} , что отображает каждую интегрируемую случайную величину x на $(\Omega, \mathcal{F}, \mathbb{P})$ в качестве значения

$$\mathbb{E}x = \int x(\omega) \mathbb{P}(d\omega) \quad (10)$$

в \mathbb{R} , называемой **математическим ожиданием** x . Функция имеет следующие свойства:

1. $\mathbb{E}\alpha = \alpha$ для всех $\alpha \in \mathbb{R}$
2. $\mathbb{E}1_A = \mathbb{P}(A)$ для всех $A \in \mathcal{F}$
3. $x \leq y \implies \mathbb{E}x \leq \mathbb{E}y$
4. $\mathbb{E}[\alpha x + \beta y] = \alpha \mathbb{E}x + \beta \mathbb{E}y$ для всех интегрируемых x, y и констант α, β

Чтобы напомнить себе об основной вероятностной мере \mathbb{P} , мы можем написать $\mathbb{E}_{\mathbb{P}} x$ вместо $\mathbb{E} x$

Обратите внимание на выражение $\mathbb{E} \alpha$ понимаемое как ожидание постоянной случайной величины, равной α

- следует из 4. при $x = \mathbb{1}_{\Omega}$ и $\beta = 0$

Упражнение: проверьте 3. для $x(\omega) := \mathbb{1}\{\omega \in A\}$ и $y(\omega) := \mathbb{1}\{\omega \in B\}$

Подсказка: что означает $x \leq y$ относительно A и B ?

Для дальнейших подробности и ссылок на доказательства вышеуказанного факта,смотрите страницу 96 в ЕТ

Мы сейчас докажем, что если x — конечная случайная величина с диапазоном $\{s_j\}_{j=1}^J$ и h — любая \mathcal{B} -измеримая функция, то

$$\mathbb{E} h(x) = \sum_{j=1}^J h(s_j) \mathbb{P}\{x = s_j\} \quad (11)$$

Сначала заметьте, что $\sum_{j=1}^J \mathbb{1}\{x = s_j\} = 1$, и значит мы можем записать $h(x)$ как

$$h(x) = h(x) \sum_{j=1}^J \mathbb{1}\{x = s_j\} = \sum_{j=1}^J h(s_j) \mathbb{1}\{x = s_j\}$$

Используем линейность ожиданий:

$$\mathbb{E} h(x) = \sum_{j=1}^J h(s_j) \mathbb{E} \mathbb{1}\{x = s_j\}$$

Применение части 2. факта ?? приводит к (11)

Неравенство Чебышёва: Факт. (??) Для любой неотрицательной случайной величины x и любой $\delta > 0$, мы имеем

$$\mathbb{P}\{x \geq \delta\} \leq \frac{\mathbb{E}x}{\delta} \quad (12)$$

Распространенная разновидность неравенства Чебышева имеет вид

$$\mathbb{P}\{|x| \geq \delta\} \leq \frac{\mathbb{E}x^2}{\delta^2} \quad (13)$$

Смотрите упражнение ?? для доказательства

Моменты

Пусть x — случайная величина и $k \in \mathbb{N}$. Если x^k интегрируема, то

- $\mathbb{E}[x^k]$ называется **k -ым моментом** x
- $\mathbb{E}[(x - \mathbb{E}x)^k]$ называется **k -ым центральным моментом** x

Если $\mathbb{E}[|x|^k] = \infty$, то говорят, что k -ый момент не существует. Для некоторых случайных величин даже первый момент не существует

В ином случае, каждый момент существует

Факт. (??) Если k -ый момент x существует, то также существует и j -ый для всех $j \leq k$

Доказательство: Упражнение ??

Неравенство Коши — Буняковского для случайных величин:

Факт. (??) Если x и y — случайные величины с конечным вторым моментом, то

$$|\mathbb{E}[xy]| \leq \sqrt{\mathbb{E}[x^2]\mathbb{E}[y^2]} \quad (14)$$

Вторым центральным моментом x называется **дисперсия** x :

$$\text{var } x := \mathbb{E}[(x - \mathbb{E}x)^2]$$

Стандартное отклонение x :

$$\sigma_x := \sqrt{\text{var } x}$$

Ковариация случайной величины x и y :

$$\text{cov}[x, y] := \mathbb{E}[(x - \mathbb{E}x)(y - \mathbb{E}y)]$$

Факт. (??) Если x и y имеют конечные вторые моменты, то

1. $\text{var } x$ и $\text{cov}[x, y]$ конечны
2. $\text{var } x = \mathbb{E}[x^2] - [\mathbb{E} x]^2$, и
3. $\text{cov}[x, y] = \mathbb{E}[xy] - \mathbb{E}[x]\mathbb{E}[y]$

Часть 1. следует из 2.–3., неравенства Коши — Буняковского и факта ??

Части 2.–3. следуют из линейности \mathbb{E} и нескольких простых манипуляций

Факт. (??) Если x_1, \dots, x_N — случайные величины и $\alpha_0, \alpha_1, \dots, \alpha_N$ — постоянные скаляры, то

$$\text{var} \left[\alpha_0 + \sum_{n=1}^N \alpha_n x_n \right] = \sum_{n=1}^N \alpha_n^2 \text{var}[x_n] + 2 \sum_{n < m} \alpha_n \alpha_m \text{cov}[x_n, x_m]$$

Некоторые простые выводы:

1. $\text{var}[\alpha + \beta x] = \beta^2 \text{var}[x]$ и
2. $\text{var}[\alpha x + \beta y] = \alpha^2 \text{var}[x] + \beta^2 \text{var}[y] + 2\alpha\beta \text{cov}[x, y]$.

Корреляция x и y :

$$\text{corr}[x, y] := \frac{\text{cov}[x, y]}{\sigma_x \sigma_y}$$

Если $\text{corr}[x, y] = 0$, то x и y некорреляированы

Положительная корреляция означает, что $\text{corr}[x, y]$ положительна, и отрицательная корреляция значит, что $\text{corr}[x, y]$ отрицательна

Факт. (??) Возьмем любые две случайные величины x, y и положительные константы α, β , мы имеем

$$-1 \leq \text{corr}[x, y] \leq 1 \quad \text{и} \quad \text{corr}[\alpha x, \beta y] = \text{corr}[x, y]$$

первая часть следует из факта ??; вторая — из алгебры

Лучшие линейные предсказатели

Рассмотрим задачу предсказания значения случайной величины y , учитывая знание значения второй случайной величины x

Мы ищем функцию f , такую что $f(x)$ в среднем близка к y

Для измерения последнего мы будем использовать **среднеквадратичную ошибку**, которая в данном случае составляет

$$\mathbb{E} [(y - f(x))^2]$$

В §??, чтобы получить минимизатор среднеквадратичного отклонения по всем функциям x , мы выбираем

$$f(x) = \mathbb{E}[y | x]$$

Здесь мы рассмотрим поиск хорошего предсказателя y среди класса "линейных" функций

$$\mathcal{H}_\ell := \{ \text{ все функции вида } \ell(x) = \alpha + \beta x \}$$

Рассмотрим:

$$\min_{\ell \in \mathcal{H}_\ell} \mathbb{E} [(y - \ell(x))^2] = \min_{\alpha, \beta \in \mathbb{R}} \mathbb{E} [(y - \alpha - \beta x)^2] \quad (15)$$

Если α и β решают (15), то функция

$$\ell^*(x) := \alpha^* + \beta^* x \quad (16)$$

называется **лучшим линейным предсказателем** y при
данном x

Пример. (??) Взаимосвязь между доходностью данного актива R_a и рыночной доходностью R_m называется коэффициентом **бета**

Измеряет подверженность системному риску в отличие от нерыночного риска.

Бета R_a часто определяется как коэффициент β^* в лучшем линейном предсказании (16), когда x — это рыночная доходность и $y = R_a$

Чтобы решить (15), раскроем квадрат с правой стороны и используем линейность \mathbb{E} , чтобы написать целевую функцию как

$$\psi(\alpha, \beta) := \mathbb{E}[y^2] - 2\alpha\mathbb{E}[y] - 2\beta\mathbb{E}[xy] + 2\alpha\beta\mathbb{E}[x] + \alpha^2 + \beta^2\mathbb{E}[x^2]$$

Вычисление производных и решение условий первого порядка:

$$\beta^* := \frac{\text{cov}[x, y]}{\text{var}[x]} \quad \text{and} \quad \alpha^* := \mathbb{E}[y] - \beta^*\mathbb{E}[x] \quad (17)$$

Смотрите упр. ??

Распределения

Возьмем случайную величину x на вероятностное пространство $(\Omega, \mathcal{F}, \mathbb{P})$

Вероятность, x примет значение из Борелевского множества B

$$\mathbb{P}\{x \in B\}$$

На практике, удобнее представлять вероятность как *распределение* в \mathbb{R}

Специфицируем Ω в \mathbb{R} и возьмем множество событий в \mathbb{R} как $\mathcal{B}(\mathbb{R})$

- Вероятностная мера, определенная в $\mathcal{B}(\mathbb{R})$, называется **законом** или **распределением**

Формально, распределение P — отображение из $\mathcal{B}(\mathbb{R})$ в $[0, 1]$, такое что

1. $P(\mathbb{R}) = 1$ и
2. $P(\bigcup_{n=1}^{\infty} B_n) = \sum_{n=1}^{\infty} P(B_n)$ для любой несовместной последовательности $\{B_n\}$

Если существует Борелевское множество S с $P(S) = 1$, то мы говорим, что S является **носителем распределения** P

Распределения характеризуются **функцией распределения**, или функция распределения, которой может быть любая функция $F: \mathbb{R} \rightarrow [0, 1]$, удовлетворяющая следующим условиям:

1. монотонность: $s \leq s'$ подразумевает, что $F(s) \leq F(s')$,
2. непрерывность справа: $F(s_n) \downarrow F(s)$ всюду, где $s_n \downarrow s$, и
3. $\lim_{s \rightarrow -\infty} F(s) = 0$ и $\lim_{s \rightarrow \infty} F(s) = 1$.

Функция распределения и распределение в \mathbb{R} можно поставить во взаимно однозначное соответствие

Распределение P полностью характеризуется значениями функции

$$F(s) := P((-\infty, s]) \quad (s \in \mathbb{R}) \quad (18)$$

Факт. (??) Следующие утверждения верны:

1. Если P — любое распределение в \mathbb{R} , то функция F в (18) — функция распределения.
2. Возьмем любую функция распределения F в \mathbb{R} , существует ровно одно распределение P , удовлетворяющее (18)

Для полного доказательства, смотрите
williams1991probability, лемма 1.6, или **dudley2002real**,
теорема 9.1.1.

Здесь ограничимся демонстрацией, что функция F в (18)
удовлетворяет части 1. определения функция распределения

- заметим, что $s \leq s'$ подразумевает, что $(-\infty, s] \subset (-\infty, s']$
- вспомним, что $P(A) \leq P(B)$, если $A \subset B$
- тогда мы имеем $P((-\infty, s]) \leq P((-\infty, s'])$ и $F(s) \leq F(s')$,
как и было заявлено

Пример. Одномерные нормальные распределения или распределения Гаусса относятся к распределениям классов, обозначенным как функция распределения вида

$$F(s) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^s \exp\left\{-\frac{(t-\mu)^2}{2\sigma^2}\right\} dt \quad (s \in \mathbb{R})$$

где $\mu \in \mathbb{R}$ и $\sigma > 0$

Мы представляем распределение, связанное с (μ, σ) с помощью $N(\mu, \sigma^2)$

Распределение $N(0, 1)$ называется **стандартным нормальным распределением**

Мы используем символ Φ для его функции распределения

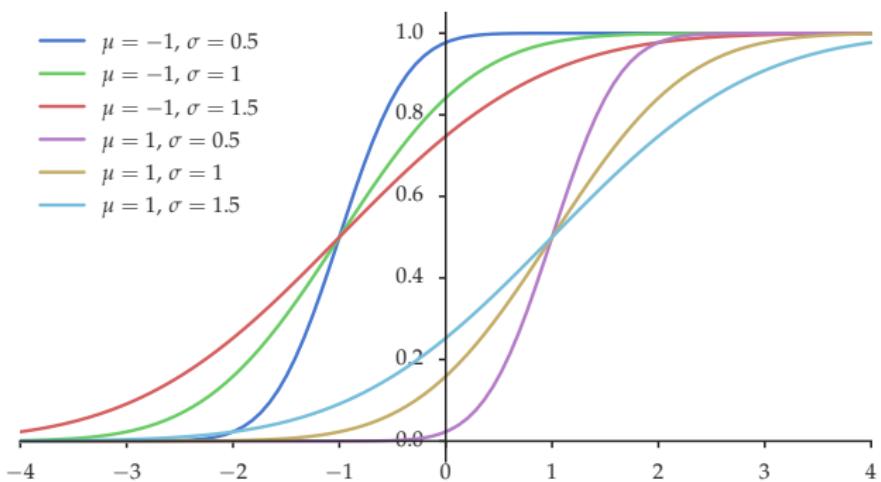


Рис.: Функция распределения для нормальных распределений

Пример. Распределение Парето — одномерное распределение с функцией распределения вида

$$F(s) = \begin{cases} 0 & , \text{ если } s < s_0 \\ 1 - \left(\frac{s_0}{s}\right)^\alpha & , \text{ если } s_0 \leq s \end{cases} \quad (s \in \mathbb{R}, s_0, \alpha > 0)$$

Распределения Парето часто используются для моделирования явлений с тяжелым правым хвостом, таких как распределение богатства или дохода.

Пример. Класс **функции распределения бета** дан с помощью

$$F(s) = \begin{cases} 0 & , \text{ если } s \leq 0 \\ \frac{1}{B(\alpha, \beta)} \int_0^s u^{\alpha-1} (1-u)^{\beta-1} du & , \text{ если } 0 < s < 1 \\ 1 & , \text{ если } 1 \leq s \end{cases}$$

где $\alpha, \beta > 0$.

В этом примере $B(\alpha, \beta)$ — **функция бета**

$$B(\alpha, \beta) := \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)} \quad \text{where} \quad \Gamma(a) := \int_0^\infty u^{a-1} e^{-u} du$$

Функция Γ называется **функция гамма**.

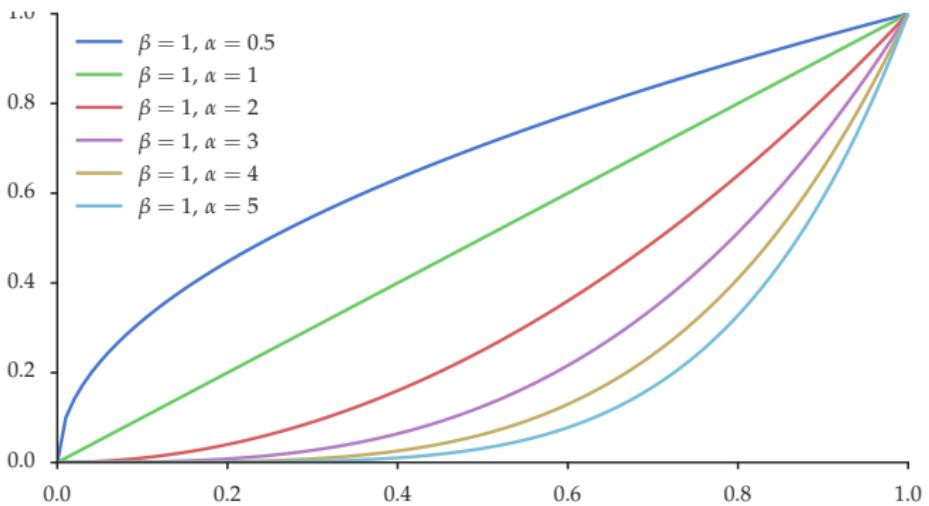


Рис.: Функции распределения бета

Пример. класс **распределений Коши** определяется как

$$F(s) = \frac{1}{\pi} \arctan\left(\frac{s - \tau}{\gamma}\right) + \frac{1}{2} \quad (s \in \mathbb{R})$$

параметры $\tau \in \mathbb{R}$ и $\gamma > 0$ — параметры местоположения и масштаба соответственно

Если $\tau = 0$ и $\gamma = 1$, то F называется **стандартным распределением Коши**

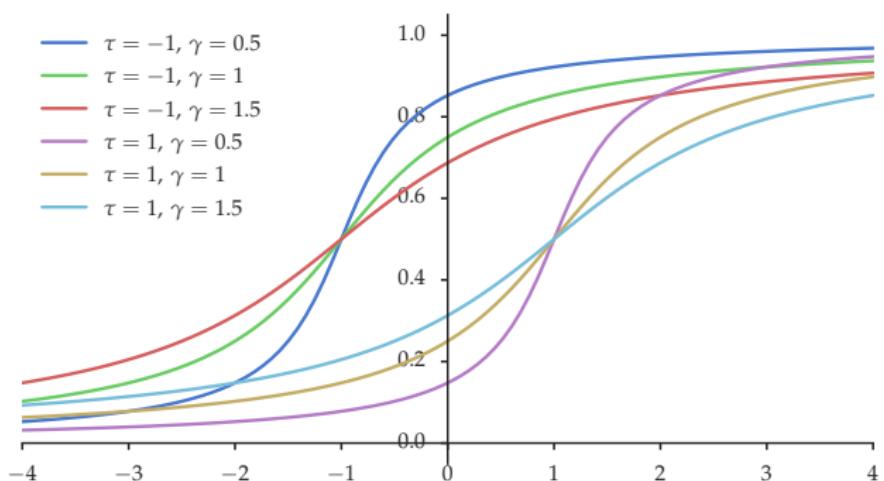


Рис.: Функции распределения Коши

Пример. Возьмем $a < b$, **равномерная функция распределения** на промежутке $[a, b]$ — это

$$F(s) = \begin{cases} 0 & , \text{ если } s \leq a \\ \frac{s-a}{b-a} & , \text{ если } a < s < b \\ 1 & , \text{ если } b \leq s \end{cases}$$

мы обозначаем это распределение как $U[a, b]$

Плотности и функции вероятности

Два удобных частных случая

- дискретный: функция распределения просто скачет (ступенчатая функция)
- абсолютно непрерывный случай: функция непрерывна, без скачков

Дискретный случай

Распределение P называется **дискретным**, если оно имеет носитель распределения в счетном множестве; that is, если существует счетное множество $\{s_j\}_{j \geq 1}$ с $P(\{s_j\}_{j \geq 1}) = 1$

для такого P пусть

$$p_j := P\{s_j\} := P(\{s_j\}) = \text{вероятность в одной точке } s_j$$

Функция вероятности — любая неотрицательная последовательность (конечная или бесконечная), сумма которой равна единице.

Упражнение: покажите, что $\{p_j\}_{j \geq 1}$ является **функцией вероятности**

Мы можем показать связь функции распределения с P как:

$$F(s) = \sum_{j \geq 1} \mathbb{1}\{s_j \leq s\} p_j \quad (19)$$

так как

$$F_x(s) := \mathbb{P}\{x \leq s\} = \mathbb{P} \bigcup_{j \text{ s.t. } s_j \leq s} \{x = s_j\}$$

$$= \sum_{j \text{ s.t. } s_j \leq s} \mathbb{P}\{x = s_j\} = \sum_{j=1}^J \mathbb{1}\{s_j \leq s\} p_j$$

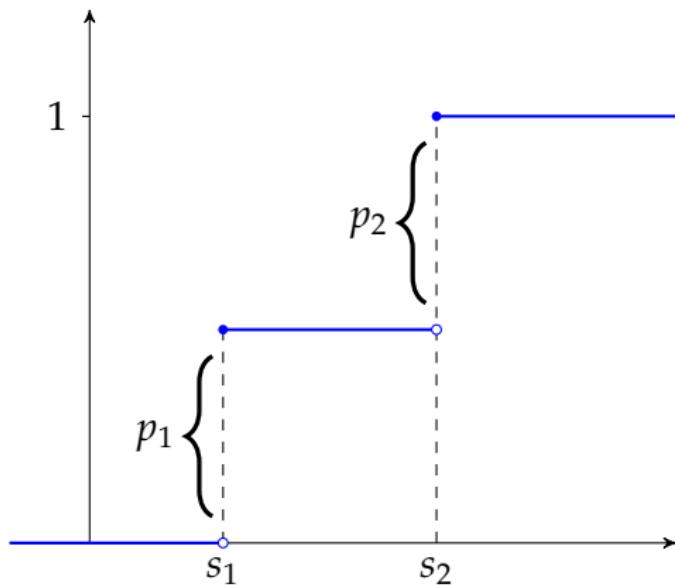


Рис.: Дискретная функция распределения

Пример. Возьмем $N \in \mathbb{N}$ и $\pi \in (0, 1)$, последовательность $\{p_0, \dots, p_N\}$, определяемая как

$$p_j = \binom{N}{j} \pi^j (1 - \pi)^{N-j}$$

называется **биномиальной функцией вероятности**

Значение p_j — вероятность, j успехов в N независимых испытаниях, вероятность успеха каждого случая равна π

Абсолютно непрерывный случай

Функция плотности — неотрицательная функция p в \mathbb{R} , которая интегрируется в 1

Распределение P определяется функцией плотности p , если p — функция плотности и

$$P(B) = \int_B p(s) \, ds \quad \text{для всех } B \in \mathcal{B}(\mathbb{R})$$

Заметим, что

$$\int_B p(s) \, ds := \int_{-\infty}^{\infty} \mathbb{1}_B(s) p(s) \, ds$$

Точное необходимое и достаточное условие существования функции плотности — абсолютная непрерывность

Распределение P в Борелевских подмножествах \mathbb{R} называется **абсолютно непрерывным**, если $P(B) = 0$ всюду, где мера Лебега B равна нулю (смотрите §??)

- Любое счетное подмножество \mathbb{R} имеет меру Лебега равную нулю

Факт. (??) Если P абсолютно непрерывная, то $P(C) = 0$ всюду, где C счетно

Если распределение абсолютно непрерывно:

- вероятность в каждой точке равна нулю
- соответствующая функция распределения не содержит скачков
- теорема Ньютона — Лейбница говорит, что $F(s)$ дифференцируема во всех точках непрерывности p , и:

$$F'(s) = p(s) \quad \text{для всех } s \in \mathbb{R}, \text{ таких что } p \text{ непрерывна в } s$$

Пример. Нормальные функции распределения дифференцируемы для всех μ, σ , с функцией плотности

$$p(s) = F'(s) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(s-\mu)^2}{2\sigma^2}\right\}$$

Мы используем символ ϕ для стандартного нормального распределения

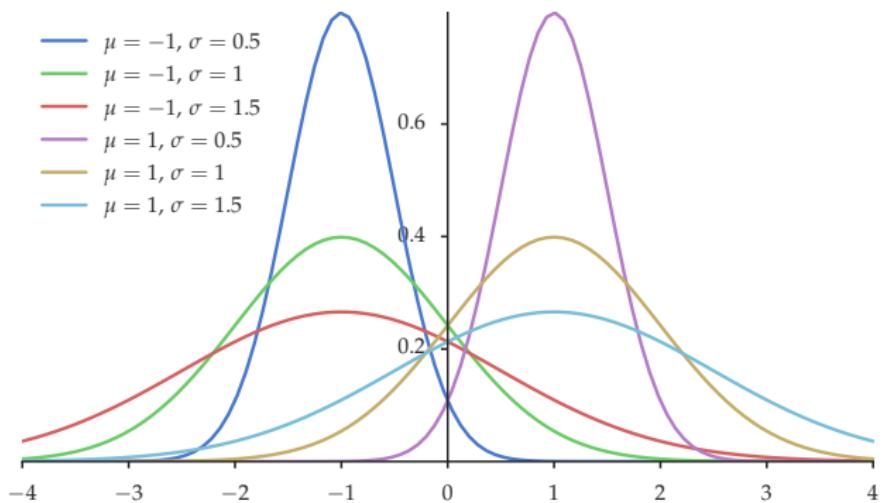


Рис.: Функции плотности нормального распределения

Пример. Функция распределения Коши имеет функцию плотности

$$p(s) = \frac{1}{\pi\gamma} \left[1 + \left(\frac{s - \tau}{\gamma} \right)^2 \right]^{-1} \quad (s \in \mathbb{R}, \gamma > 0, \tau \in \mathbb{R})$$

Функции плотности Коши более остроконечны около своих мод и имеют большую массу в хвосте, чем нормальная функция плотности

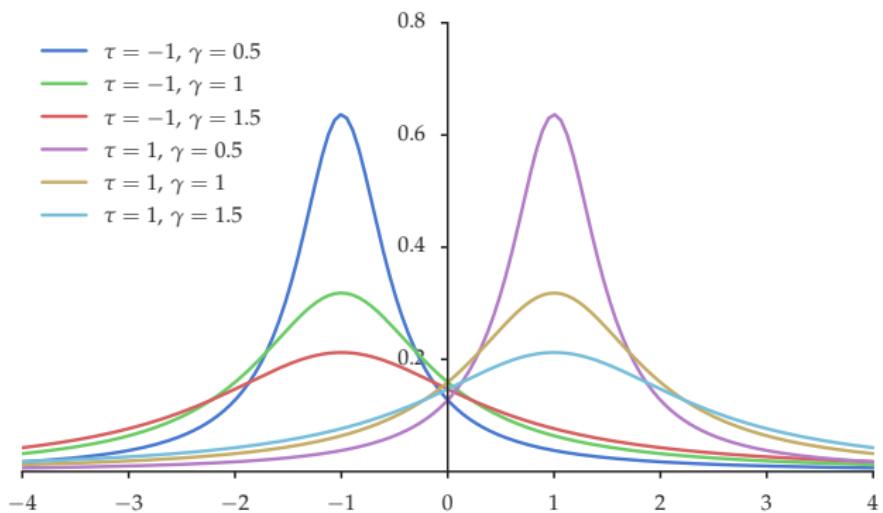


Рис.: Функции плотности Коши

Пример. Бета имеет функцию плотности, определяемую как

$$p(s) = \frac{s^{\alpha-1}(1-s)^{\beta-1}}{B(\alpha, \beta)} \quad (\alpha, \beta > 0)$$

Когда $0 < s < 1$ и 0 в ином случае

Пример. $U[a, b]$ распределение представлено функцией плотности

$$p(s) = \frac{1}{b-a} \mathbb{1}\{a \leq s \leq b\} \quad (s \in \mathbb{R}, a, b \in \mathbb{R}, a < b)$$

Пример. Гамма-распределение с параметром формы α и параметром масштаба β — распределение с функцией плотности

$$p(s) = \frac{s^{\alpha-1} e^{-s/\beta}}{\beta^\alpha \Gamma(\alpha)} \quad (\alpha, \beta > 0)$$

Когда $0 < s < 1$ и 0 в ином случае

Пример. **Хи-квадрат распределение с k степенями свободы** — распределение с функцией плотности

$$p(s) := \frac{1}{2^{k/2}\Gamma(k/2)}s^{k/2-1}e^{-s/2} \quad (s > 0, k \in \mathbb{N})$$

Это распределение представлено символом $\chi^2(k)$

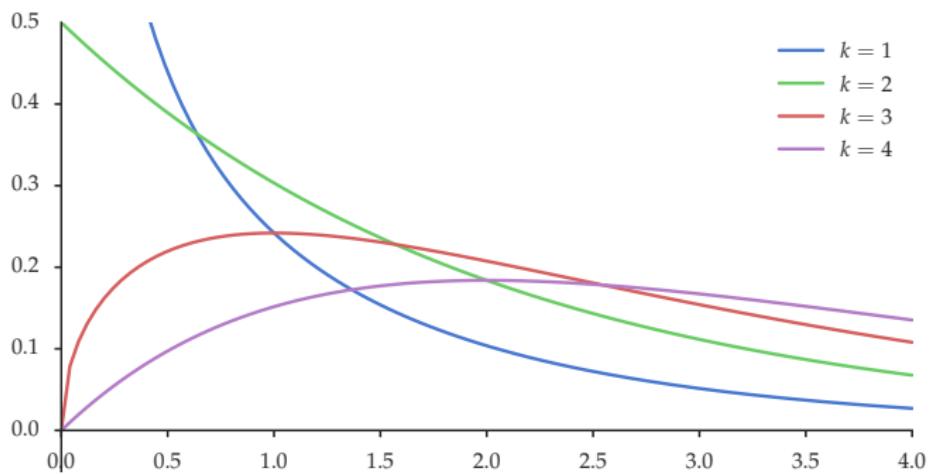


Рис.: Функции плотности хи-квадрата

Пример. Распределение Стьюдента t с k степенями свободы, или, проще, t -распределение с k степенями свободы, — распределение в \mathbb{R} с функцией плотности

$$p(s) := \frac{\Gamma(\frac{k+1}{2})}{(k\pi)^{1/2}\Gamma(\frac{k}{2})} \left(1 + \frac{s^2}{k}\right)^{-(k+1)/2} \quad (s \in \mathbb{R}, k > 0)$$

Пример. *F-распределение* с параметрами k_1, k_2 – распределение с функцией плотности

$$p(s) := \frac{\sqrt{(k_1 s)^{k_1} k_2^{k_2} / [k_1 s + k_2]^{k_1+k_2}}}{s B(k_1/2, k_2/2)} \quad (s \geq 0, k_1, k_2 > 0)$$

F-распределение возникает при проверке ряда гипотез, как обсуждается ниже.

Интегрирование

Рассмотрим обычный интервал $\int_a^b h(s) \, ds$ функции h на некотором интервале $[a, b]$

Предположим, мы хотим сделать этот интервал взвешенным, придав больше массы различным частям $[a, b]$:

$$\int_a^b h(s)p(s) \, ds$$

Например:

- h — функция благосостояния и p плотность агентов
- p — плотность, указывающая на вероятности исходов, h функция прибыли

Предположим, что P не имеет функции плотности, но мы все еще хотим взвесить интеграл с помощью P

- мы хотим определить $\int h(s)P(\mathrm{d}s)$

Возьмем распределение P в \mathbb{R} и рассмотрим $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P)$ как вероятностное пространство

$(\mathbb{R}, \mathcal{B}(\mathbb{R}), P)$ имеет собственный оператор ожидания \mathbb{E}_P

Предположим, что h — случайная величина в $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P)$, тогда

$$\mathbb{E}_P h := \int h(s)P(\mathrm{d}s) := \text{ожидания } h \text{ при } P$$

Факт. Пусть $h: \mathbb{R} \rightarrow \mathbb{R}$ — \mathcal{B} -измерима и P — распределение в \mathbb{R}

Если P дискретна, с функцией вероятности $\{p_j\}_{j \geq 1}$ и носителем распределения $\{s_j\}_{j \geq 1}$, то

$$\int h(s)P(\mathrm{d}s) = \sum_{j \geq 1} h(s_j)p_j$$

Если P абсолютно непрерывно с функцией плотности p , то

$$\int h(s)P(\mathrm{d}s) = \int_{-\infty}^{\infty} h(s)p(s) \mathrm{d}s$$

Распределения и случайные переменные

Каждая случайная переменная определяет распределение on \mathbb{R}

Пусть x — случайная величина на некотором вероятностном пространстве $(\Omega, \mathcal{F}, \mathbb{P})$:

- вероятность $\mathbb{P}\{x \in B\}$ определена для каждого $B \in \mathcal{B}(\mathbb{R})$ (смотрите (6))
- множество функций P определено как

$$P(B) = \mathbb{P}\{x \in B\} \quad (B \in \mathcal{B}(\mathbb{R})) \quad (20)$$

является **распределением x**

Функция распределения соответствующая распределению P случайной переменной x удовлетворяет

$$F(s) = \mathbb{P}\{x \leq s\} \quad (s \in \mathbb{R}) \quad (21)$$

мы пишем $\mathcal{L}(x) = F$, чтобы подчеркнуть, что F означает распределение x

Факт. (??) Если $\mathcal{L}(x) = F$, то $\mathbb{P}\{a < x \leq b\} = F(b) - F(a)$ для любых $a \leq b$

Доказательство выполните в качестве упражнения (или смотрите страницу 111 в ЕТ)

Для каждой функции распределения F , существует вероятностное пространство $(\Omega, \mathcal{F}, \mathbb{P})$ и случайная величина $x: \Omega \rightarrow \mathbb{R}$, такая что $\mathcal{L}(x) = F$; §?? показывает построение

Если $\mathcal{L}(x) = P$ и P имеет функцию плотности p , мы говорим, что x имеет функцию плотности p

Если распределение x дискретно, мы будем называть x **дискретной случайной величиной**

Факт. Если x имеет функцию плотности, то $\mathbb{P}\{x = s\} = 0$ для всех $s \in \mathbb{R}$, и для любых $a < b$,

$$\mathbb{P}\{a < x < b\} = \mathbb{P}\{a < x \leq b\}$$

$$= \mathbb{P}\{a \leq x < b\} = \mathbb{P}\{a \leq x \leq b\}$$

Распределения преобразований

Факт. (??) Если $\mathcal{L}(x) = F$ и $y := \psi(x)$, где $\psi: \mathbb{R} \rightarrow \mathbb{R}$ строго возрастает, то $\mathcal{L}(y) = G$, где $G(s) := F(\psi^{-1}(s))$.

Доказательство. При таких гипотезах ψ^{-1} существует и (медленно) возрастает. Следовательно

$$\mathbb{P}\{y \leq s\} = \mathbb{P}\{\psi(x) \leq s\} = \mathbb{P}\{x \leq \psi^{-1}(s)\} = F(\psi^{-1}(s))$$

Заметьте, как монотонность используется во втором равенстве

Пример. Если $\mathcal{L}(x) = F$ и $y := \exp(x)$, то функция распределения y — это $G(s) := F(\ln(s))$

Факт. (??) Если x имеет плотность p в \mathbb{R} и $y := \psi(x)$, где ψ — диффеоморфизм в \mathbb{R} , то распределение y абсолютно непрерывное с функцией плотности

$$q(s) = p(\psi^{-1}(s)) \left| \frac{d\psi^{-1}(s)}{ds} \right| \quad (s \in \mathbb{R})$$

термин **дiffeоморфизм** значит, что ψ — биекция в \mathbb{R} и оба ψ и его обратное дифференцируемы

Пример. Если x имеет функцию плотности p в \mathbb{R} , и μ и σ — константы с $\sigma > 0$, то функция плотности $y := \mu + \sigma x$ — это

$$q(s) = p\left(\frac{s - \mu}{\sigma}\right) \frac{1}{\sigma} \quad (s \in \mathbb{R})$$

Когда x стандартное нормальное: $y = \mu + \sigma x$ is $N(\mu, \sigma^2)$

Почему?

- Возьмем p функцию плотности стандартного нормального распределения ϕ
- Вспомним

$$p(s) = F'(s) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(s - \mu)^2}{2\sigma^2}\right\}$$

Пусть x — случайная величина в вероятностном пространстве $(\Omega, \mathcal{F}, \mathbb{P})$

Распределение x кодирует всю информацию для расчета ожидания x или любого \mathcal{B} -измеримого преобразования $h(x)$

Во-первых, пусть x конечно. Предположим, что

- $\mathcal{L}(x) = P$
- функция $h: \mathbb{R} \rightarrow \mathbb{R}$ — любая \mathcal{B} -измеримая функция
- P помещает все вероятности в конечное множество $\{s_j\}_{j=1}^J$

Используем $\mathbb{P}\{x = s_j\} = P\{s_j\}$ и определение ожиданий:

$$\mathbb{E}h(x) = \sum_{j=1}^J h(s_j)\mathbb{P}\{x = s_j\} = \sum_{j=1}^J h(s_j)P\{s_j\} = \sum_{j=1}^J h(s_j)p_j$$

Ожидания $h(x)$ в $(\Omega, \mathcal{F}, \mathbb{P})$ равны ожиданиям h в $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P)$

Верно также и для бесконечного случая:

Факт. Пусть x — случайная величина в некотором вероятностном пространстве $(\Omega, \mathcal{F}, \mathbb{P})$, пусть $\mathcal{L}(x) = P$ и h — \mathcal{B} -измеримая функция, такая что $h(x)$ интегрируема. Ожидания $\mathbb{E}h(x)$ полностью определены h и P . В частности,

$$\mathbb{E}h(x) = \int h(s)P(ds)$$

где $\int h(s)P(ds)$ — ожидания h в $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P)$

Пример. Пусть x — случайная величина, чье распределение P является равномерным распределением в $[a, b]$

Применить определение функции плотности равномерного распределения

$$\mathbb{E}x = \int sP(ds) = \int sp(s) ds = \int_{-\infty}^{\infty} \frac{s}{b-a} \mathbb{1}\{a \leq s \leq b\} ds$$

Решение интеграла дает $\mathbb{E}x = \mu := (a + b)/2$. Дисперсия равна

$$\text{var}[x] = \int (s - \mu)^2 P(ds)$$

$$= \int_a^b \left(s - \frac{a+b}{2}\right)^2 \frac{1}{b-a} ds = \frac{1}{12}(b-a)^2$$

Пример. Предположим, что $\mathcal{L}(x) = \text{N}(\mu, \sigma)$

Если $\sigma > 0$, среднее значение может быть вычислено с помощью

$$\mathbb{E}x = \int_{-\infty}^{\infty} s \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(s-\mu)^2}{2\sigma^2}\right\} ds = \mu$$

Дисперсия определяется как:

$$\text{var}[x] = \int_{-\infty}^{\infty} (s - \mu)^2 \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(s-\mu)^2}{2\sigma^2}\right\} ds = \sigma^2$$

Моменты распределений

Любые две случайные величины с одинаковыми распределениями имеют одинаковые моменты

Следовательно моменты лучше всего рассматривать как свойство распределения, не как случайную величину

Таким образом, мы определяем

- среднее P как $\mu = \int sP(ds)$,
- k -ый момент P как $\int s^k P(ds)$,
- дисперсию P как $\int (s - \mu)^2 P(ds)$,

и так далее

Функция квантилей

Пусть F — строго возрастающая функция распределения в \mathbb{R}

Возьмем $\tau \in (0, 1)$, **τ -ая квантиль** F — это $\xi \in \mathbb{R}$, который является решением $F(\xi) = \tau$

Согласно нашим предположениям о F , такое ξ существует и однозначно определено

0.5-ая квантиль называется **медианой** F

Функция квантилей:

$F^{-1}(\tau) :=$ единственный ξ , такой что $F(\xi) = \tau$ ($0 < \tau < 1$)

Пример. Функция квантилей связанная со стандартным распределением Коши — это $F^{-1}(\tau) = \tan[\pi(\tau - 1/2)]$

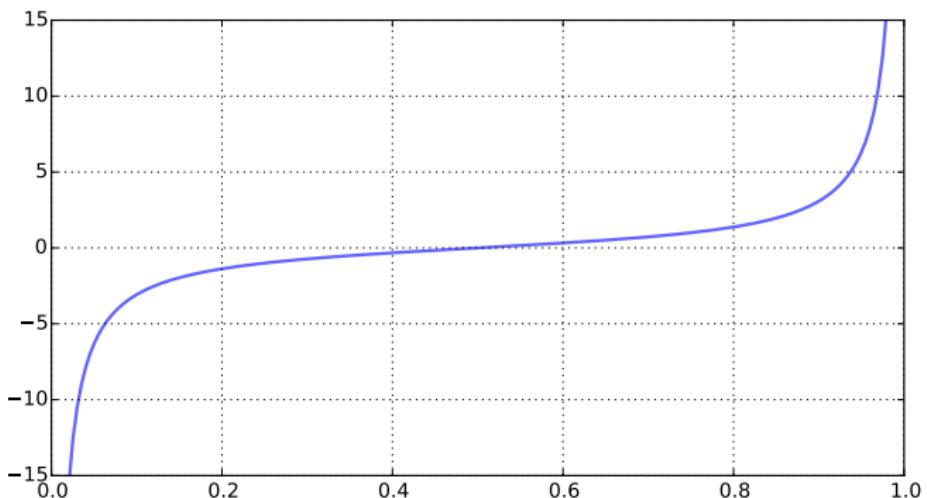


Рис.: Функция квантилей распределения Коши (горизонтальная ось — это $\tau \in (0, 1)$)

Когда F не строго возрастающая, F^{-1} не определено

Мы можем задать:

$$F^{-1}(\tau) := \inf\{s \in \mathbb{R} : F(s) \geq \tau\} \quad (0 < \tau < 1) \quad (22)$$

Функция плотности p симметрична, если $p(s) = p(-s)$ для всех $s \in \mathbb{R}$

Распространенный сценарий проверки гипотез

Факт. (??) Пусть x — случайная величина с функцией плотности p . Если p симметрична, то функция распределения G $y := |x|$ определяется как

$$G(s) := \mathbb{P}\{y \leq s\} = \begin{cases} 2F(s) - 1 & , \text{ если } s \geq 0 \\ 0 & \text{в ином случае} \end{cases}$$

Докажите в качестве упражнения ??

Факт эквивалентен $F(s) = 1 - F(-s)$

Возьмем случайную величину x с $\mathcal{L}(x) = F$ и заданной константой $\alpha \in (0, 1)$

Рассмотрим c , являющийся решением $\mathbb{P}\{-c \leq x \leq c\} = 1 - \alpha$

Факт. Если $\mathcal{L}(x) = F$, x имеет симметричную функцию плотности и F — строго возрастающая, то

$$c = F^{-1}(1 - \alpha/2) \implies \mathbb{P}\{-c \leq x \leq c\} = 1 - \alpha \quad (23)$$

Когда F стандартная нормальная функция распределения Φ , с обычно обозначается как $z_{\alpha/2}$:

$$z_{\alpha/2} := \Phi^{-1}(1 - \alpha/2) \quad (24)$$

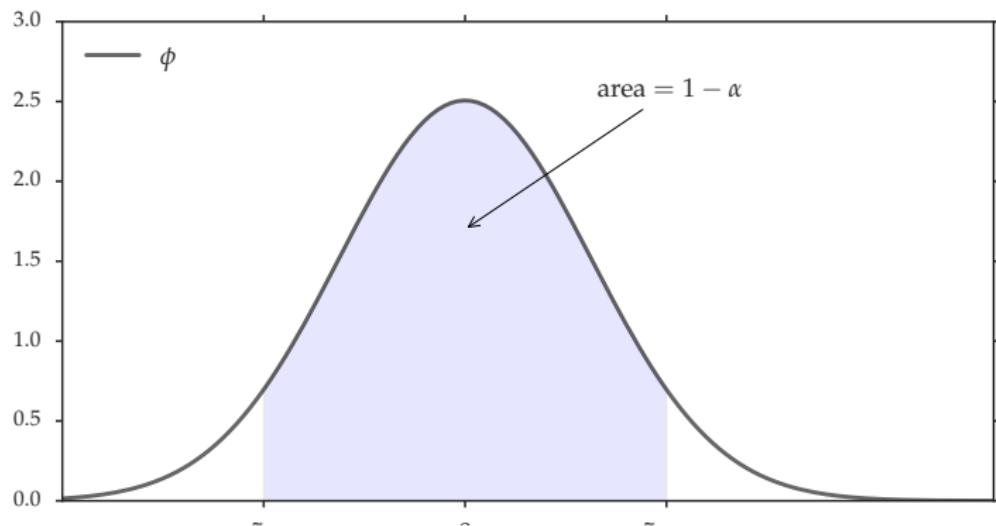


Рис.: Критические значения для стандартной нормальной плотности