



ТИНЬКОФФ

Занятие 2

Введение в теорию вероятностей и математическую статистику



Случайные
величины

The diagram features a central yellow circle with the text 'Случайные величины'. Two dashed arrows originate from the top and bottom of this circle, pointing towards two light gray rounded rectangular boxes. The top box is titled 'Дискретные' and contains three white rounded rectangular buttons labeled 'Бернулли', 'Биномиальное', and 'Пуассона'. The bottom box is titled 'Непрерывные' and contains two white rounded rectangular buttons labeled 'Гауссова семейство' and 'Вейбулла'.

Дискретные

Бернулли

Биномиальное

Пуассона

Непрерывные

Гауссова семейство

Вейбулла

Случайные
величины

The diagram features a central yellow circle with the text 'Случайные величины'. Two dashed arrows originate from the top and bottom of this circle, pointing towards two rectangular boxes on the right. The top box is light blue and titled 'Дискретные', containing three white rounded rectangles with the names of discrete distributions: 'Бернулли', 'Биномиальное', and 'Пуассона'. The bottom box is light gray and titled 'Непрерывные', containing two white rounded rectangles with the names of continuous distributions: 'Гауссова семейство' and 'Вейбулла'.

Дискретные

Бернулли

Биномиальное

Пуассона

Непрерывные

Гауссова семейство

Вейбулла

Дискретная случайная величина

Дискретной случайной величиной называется такая случайная величина, значения которой можно закодировать целыми числами (не более чем счетным множеством).



Примеры дискретных с.в.



Число попаданий в мишень
при n выстрелах



Результат подбрасывания монетки
(не обязательно математической)

Случайные
величины

The diagram features a central yellow circle with the text 'Случайные величины'. Two dashed arrows originate from the top and bottom of this circle, pointing towards two distinct categories of random variables. The top category is 'Дискретные' (Discrete), highlighted in a light blue rounded rectangle, and includes three sub-items: 'Бернулли' (Bernoulli) in a blue box, 'Биномиальное' (Binomial) in a white box, and 'Пуассона' (Poisson) in a white box. The bottom category is 'Непрерывные' (Continuous), highlighted in a light gray rounded rectangle, and includes two sub-items: 'Гауссова семейство' (Gaussian family) in a white box and 'Вейбулла' (Weibull) in a white box.

Дискретные

Бернулли

Биномиальное

Пуассона

Непрерывные

Гауссова семейство

Вейбулла

Распределение Бернулли

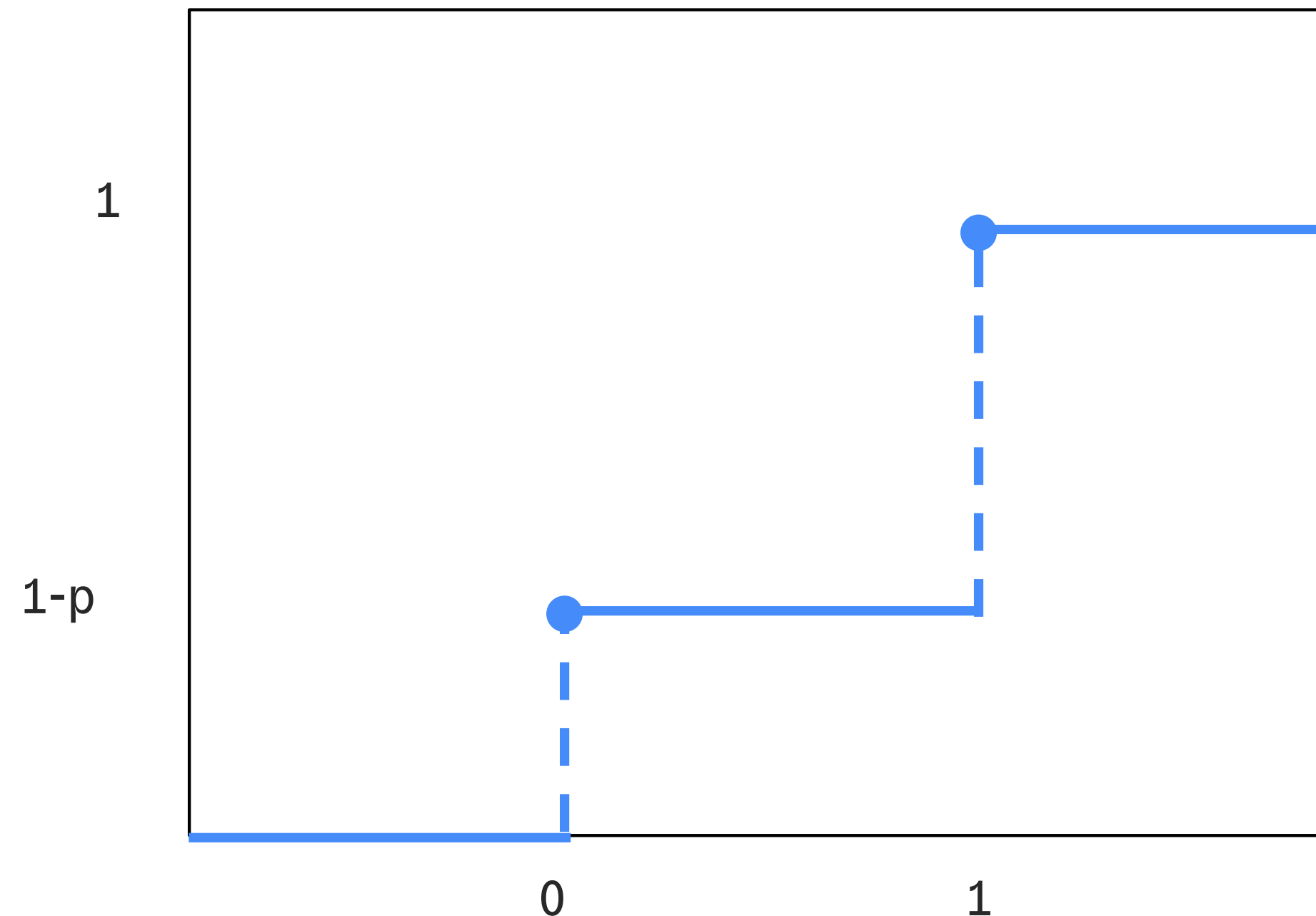
Характеризует случайную величину – результат единоразового эксперимента с бинарным результатом

Обозначение: $\xi \sim Be(p)$, где p – вероятность положительного исхода

Примеры дискретных с.в.

- Возвращение кредита заёмщиком
- Рост актива в заданный момент времени

Функция распределения Бернулли



Случайные величины

The diagram illustrates the classification of random variables. A central yellow circle labeled 'Случайные величины' (Random variables) has two dashed arrows pointing to two categories: 'Дискретные' (Discrete) and 'Непрерывные' (Continuous). The 'Дискретные' category is highlighted in blue and includes 'Бернулли' (Bernoulli), 'Биномиальное' (Binomial), and 'Пуассона' (Poisson). The 'Непрерывные' category is highlighted in light gray and includes 'Гауссова семейство' (Gaussian family) and 'Вейбулла' (Weibull).

Дискретные

Бернулли

Биномиальное

Пуассона

Непрерывные

Гауссова семейство

Вейбулла

Биномиальное распределение

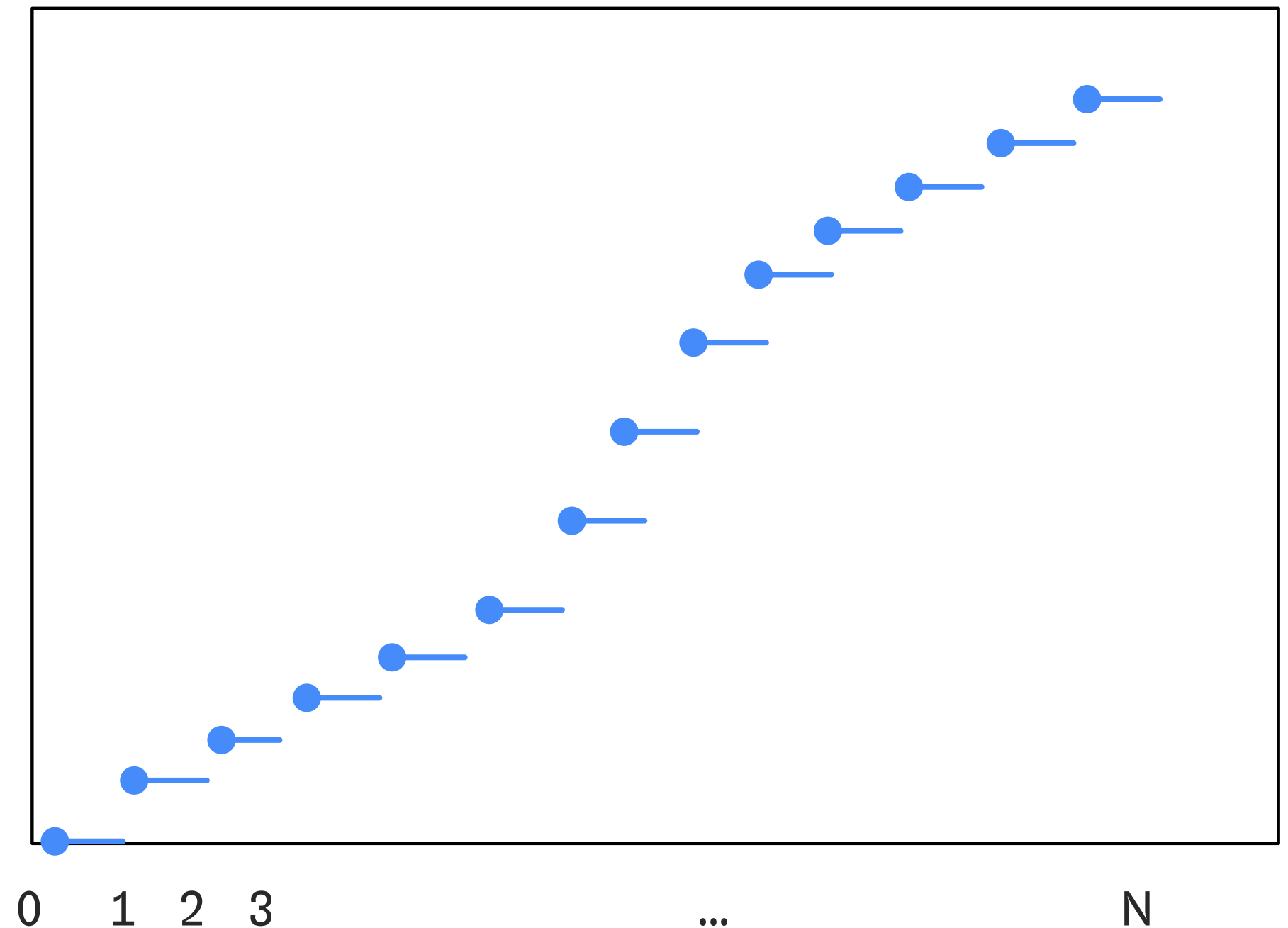
Характеризует случайную величину – количество положительных исходов при проведении n независимых испытаний Бернулли.

Обозначение: $\xi \sim Bi(p, n)$, где p – вероятность положительного исхода, n – число испытаний

Примеры дискретных с.в.

- Кол-во выпадения герба при n бросках монетки
- Кол-во мужчин среди n случайно выбранных людей

Функция распределения биномиального



Случайные
величины

The diagram features a central yellow circle with the text 'Случайные величины'. Two dashed arrows originate from the top and bottom of this circle, pointing towards two distinct categories of random variables. The top category is 'Дискретные' (Discrete), highlighted in a light blue rounded rectangle, and includes three sub-items: 'Бернулли', 'Биномиальное', and 'Пуассона' (the last of which is highlighted in a darker blue box). The bottom category is 'Непрерывные' (Continuous), highlighted in a light gray rounded rectangle, and includes two sub-items: 'Гауссова семейство' and 'Вейбулла'.

Дискретные

Бернулли

Биномиальное

Пуассона

Непрерывные

Гауссова семейство

Вейбулла

Распределение Пуассона

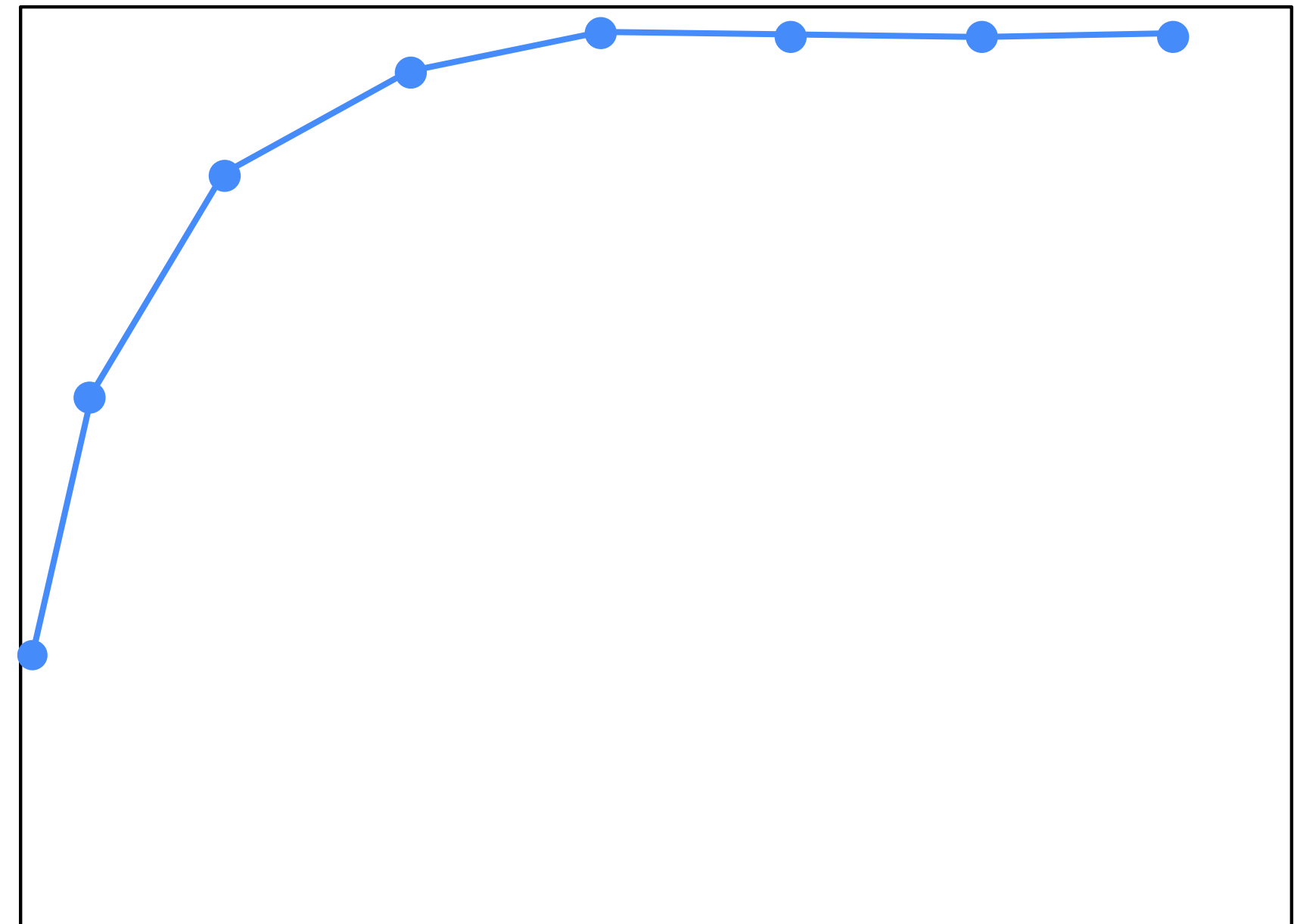
Характеризует случайную величину – схожую с Биномиальным, но применительно к редким событиям при некоторых предположениях о характере процесса

Обозначение: $\xi \sim Pois(\lambda)$, где $\lambda > 0$ – интенсивность события

Примеры дискретных с.в.

- Число частиц участвующих в радиоактивном распаде за время t
- Число запросов в системе за время t

Функция распределения Пуассона



Случайные величины

```
graph LR; A((Случайные величины)) -.-> B[Дискретные]; A -.-> C[Непрерывные]; B --- D[Бернулли]; B --- E[Биномиальное]; B --- F[Пуассона]; C --- G[Гауссова семейство]; C --- H[Вейбулла];
```

The diagram illustrates the classification of random variables. A central yellow circle labeled 'Случайные величины' (Random variables) has two dashed arrows pointing to two main categories: 'Дискретные' (Discrete) and 'Непрерывные' (Continuous). The 'Discrete' category is shown in a light gray box and includes three sub-types: 'Бернулли' (Bernoulli), 'Биномиальное' (Binomial), and 'Пуассона' (Poisson). The 'Continuous' category is shown in a light blue box and includes two sub-types: 'Гауссова семейство' (Gaussian family) and 'Вейбулла' (Weibull).

Дискретные

Бернулли

Биномиальное

Пуассона

Непрерывные

Гауссова семейство

Вейбулла

Непрерывная случайная величина

Случайная величина ξ называется непрерывной, если существует $f_\xi: \mathbb{R} \rightarrow \mathbb{R}_+$ что для неё выполнено

$$\mathbb{F} = P(\xi < x) = \int_{-\infty}^x f_\xi(t) dt, \quad \forall x \in \mathbb{R}$$



Примеры непрерывных с.в.



Размер детали
на заводе



Точные возраст
и вес человека



Измерительная
ошибка прибора

Случайные
величины

The diagram features a central yellow circle with the text 'Случайные величины'. Two dashed arrows originate from the top and bottom of this circle, pointing towards two distinct categories of random variables. The top category is 'Дискретные' (Discrete), highlighted in a light gray box, and includes three sub-items: 'Бернулли' (Bernoulli), 'Биномиальное' (Binomial), and 'Пуассона' (Poisson). The bottom category is 'Непрерывные' (Continuous), highlighted in a light blue box, and includes two sub-items: 'Гауссова семейство' (Gaussian family) and 'Вейбулла' (Weibull). The 'Гауссова семейство' item is distinguished by a blue background, while the others have white backgrounds.

Дискретные

Бернулли

Биномиальное

Пуассона

Непрерывные

Гауссова семейство

Вейбулла

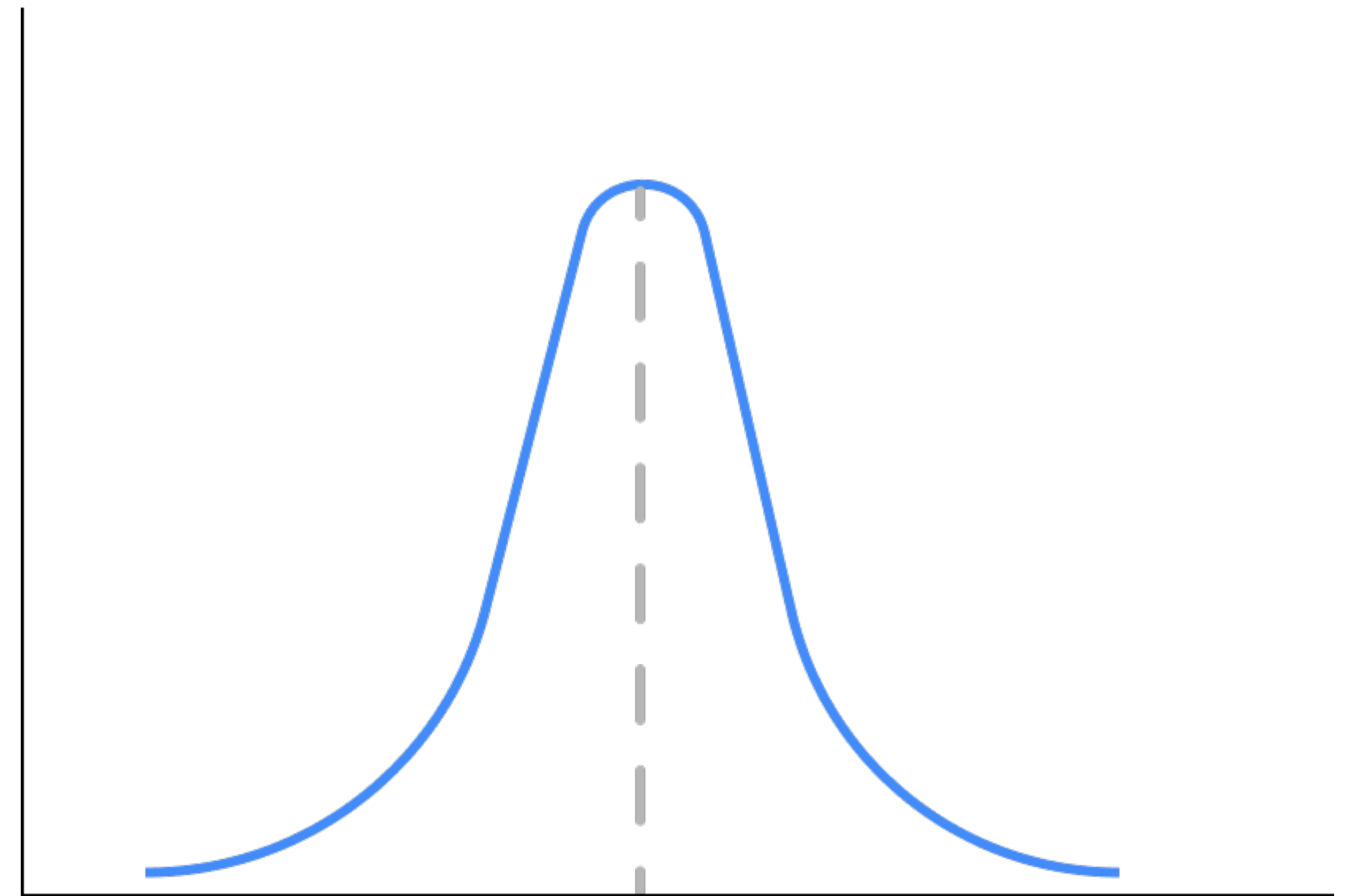
Нормальное распределение

Характеризует случайную величину, зависящую от 2х параметров: среднеквадратичного отклонения и матожидания

Примеры

- Рост человека
- Масса вылавливаемой рыбы одного вида

Плотность вероятности нормального распределения



Случайные величины

A central yellow circle labeled 'Случайные величины' (Random variables) has two dashed arrows pointing to two categories: 'Дискретные' (Discrete) and 'Непрерывные' (Continuous). The 'Discrete' category is in a light gray box and includes 'Бернулли' (Bernoulli), 'Биномиальное' (Binomial), and 'Пуассона' (Poisson). The 'Continuous' category is in a light blue box and includes 'Гауссова семейство' (Gaussian family) and 'Вейбулла' (Weibull). The 'Вейбулла' box is highlighted in a darker blue.

Дискретные

Бернулли

Биномиальное

Пуассона

Непрерывные

Гауссова семейство

Вейбулла

Распределение Вейбулла

Характеризует случайную величину, параметризуемую 3мя величинами, сдвига, формы и масштаба

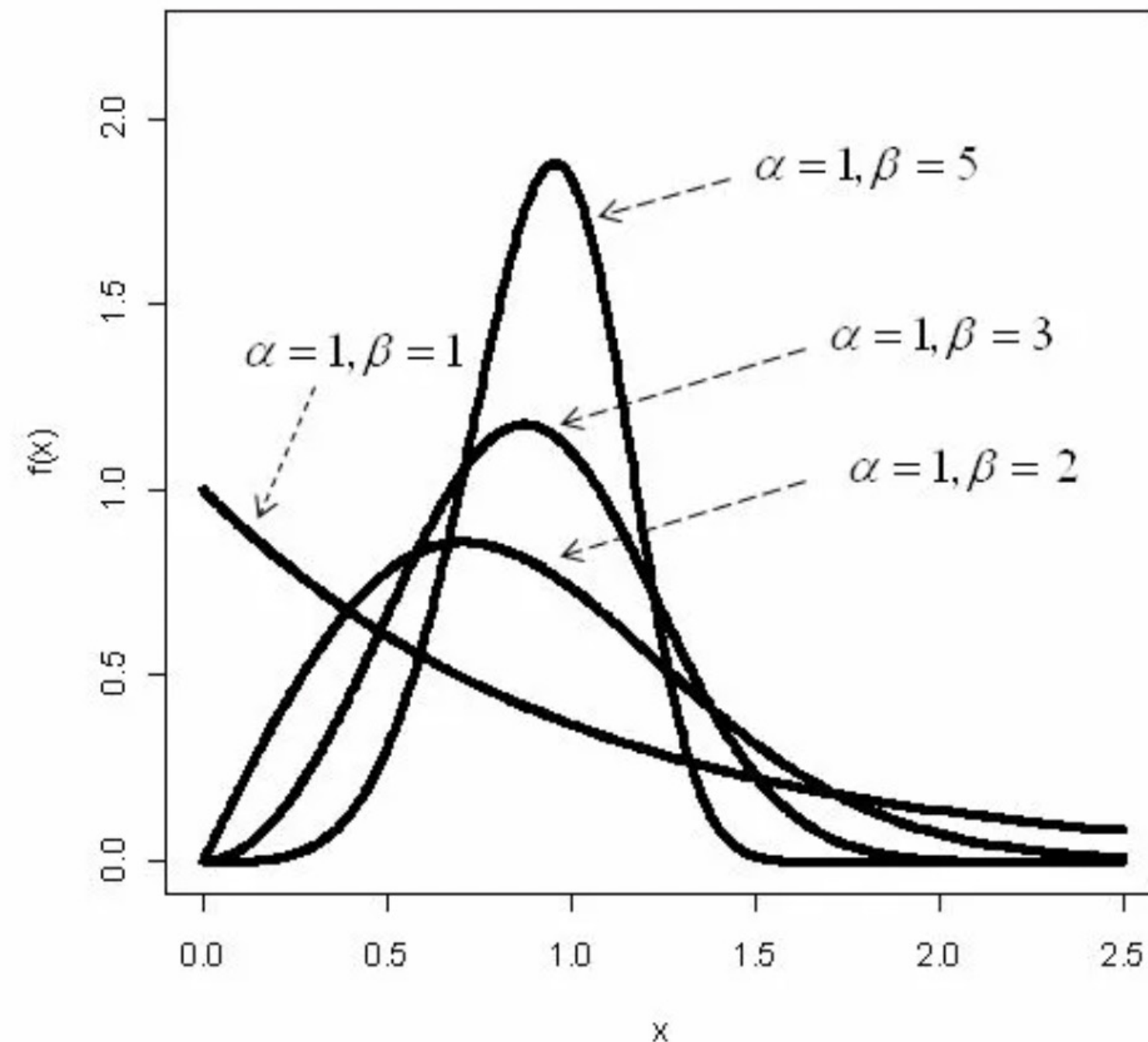
Обозначение:

$$F(x) = 1 - \exp \left\{ - \left(\frac{x-a}{b} \right)^\alpha \right\}, \quad x \geq a$$

Примеры

- Задач оценки рисков
- Анализ выживаемости

Плотность вероятности распределения Вейбулла



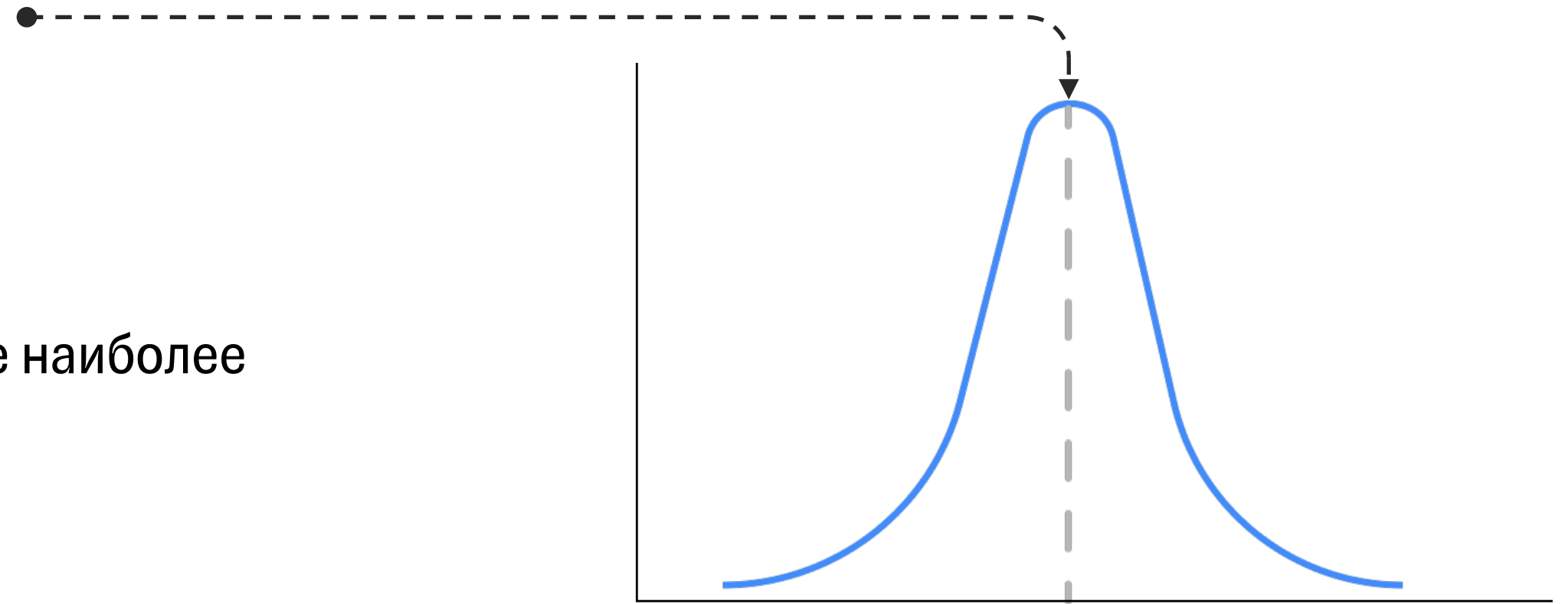
Числовые характеристики случайных величин



- Матожидание – \mathbb{E}
- Дисперсия – \mathbb{D}
- Ковариация – Cov
- Коэффициент корреляции – $corr(\xi, \eta)$

Математическое ожидание

Центр масс случайной величины, в некотором роде наиболее вероятное значение случайной величины



Дискретный случай:

$$\mathbb{E}\xi = \sum_k x_k \cdot \mathbb{P}(\xi = x_k)$$

Непрерывный случай:

$$\mathbb{E}\xi = \int_{-\infty}^{+\infty} x \cdot f_{\xi}(x) dx$$

Математическое ожидание

Свойства

$$a, b \in \mathbb{R} \quad \exists \mathbb{E}\xi, \mathbb{E}\eta \Rightarrow \mathbb{E}(a\xi \pm b\eta) = a\mathbb{E}\xi \pm b\mathbb{E}\eta$$

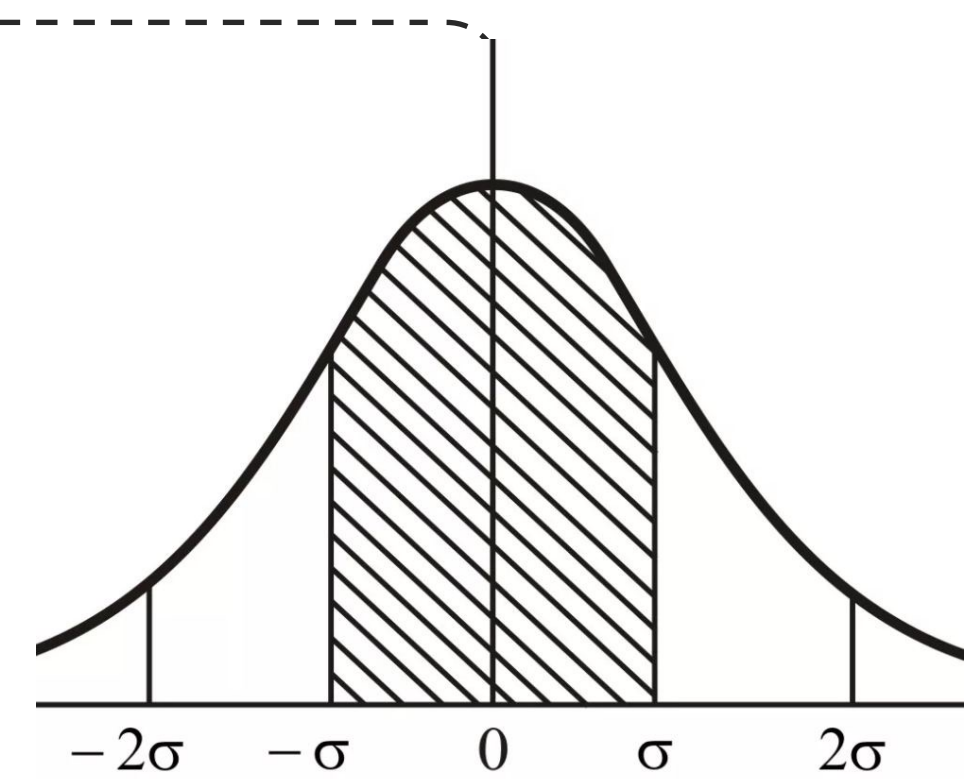
$$\text{Если } \xi \text{ и } \eta \text{ независимы и } \exists \mathbb{E}\xi, \mathbb{E}\eta, \text{ то } \mathbb{E}(\xi\eta) = \mathbb{E}\xi\mathbb{E}\eta$$

Пусть g — \mathcal{B} — измеримая функция, и $\exists \mathbb{E}\xi$, тогда

$$\mathbb{E}g(\xi) = \sum_{i=1}^{+\infty} g(x_i)p_i \text{ для д.с.в. или } \mathbb{E}g(\xi) = \int_{-\infty}^{+\infty} g(x) f_X dx \text{ для н. с. в.}$$

Дисперсия

Математическое ожидание квадрата отклонения случайной величины от ее математического ожидания.



$$\mathbb{D}\xi = \mathbb{E}[(\xi - \mathbb{E}\xi)^2]$$

Ковариация

Ковариация показывает степень линейной зависимости между случайными величинами

$$\text{cov}(\xi, \eta) = \mathbb{E}((\xi - \mathbb{E}\xi)(\eta - \mathbb{E}\eta))$$

Коэффициент корреляции

Инвариантный по масштабу коэффициент линейной зависимости между случайными величинами

$$\text{corr}(\xi, \eta) = \frac{\text{cov}(\xi, \eta)}{\sqrt{\mathbb{D}\xi \mathbb{D}\eta}}$$

Центральная предельная теорема

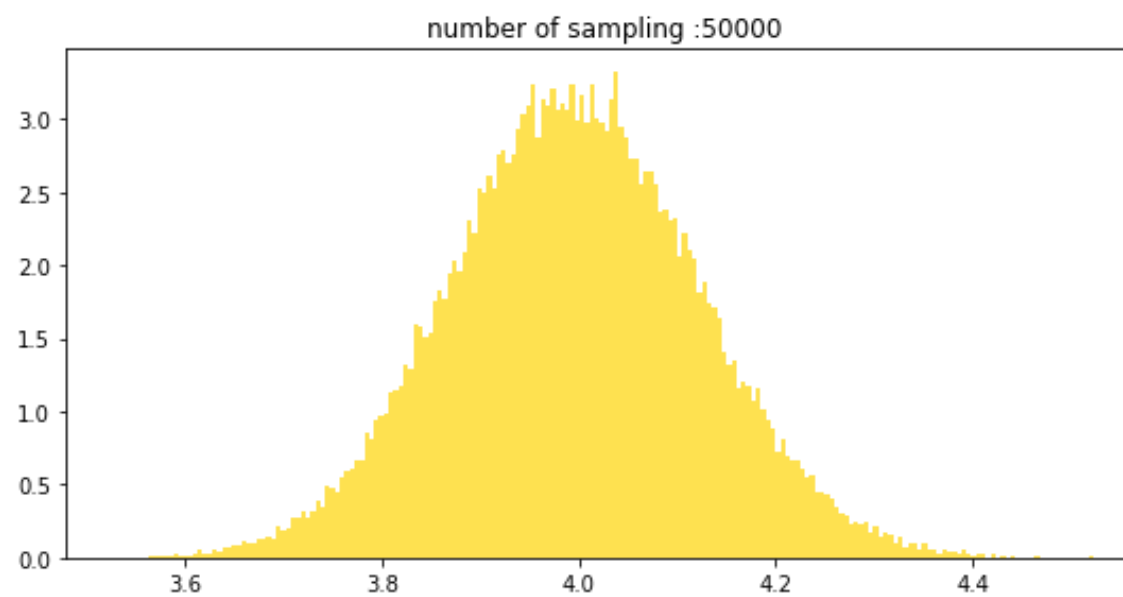
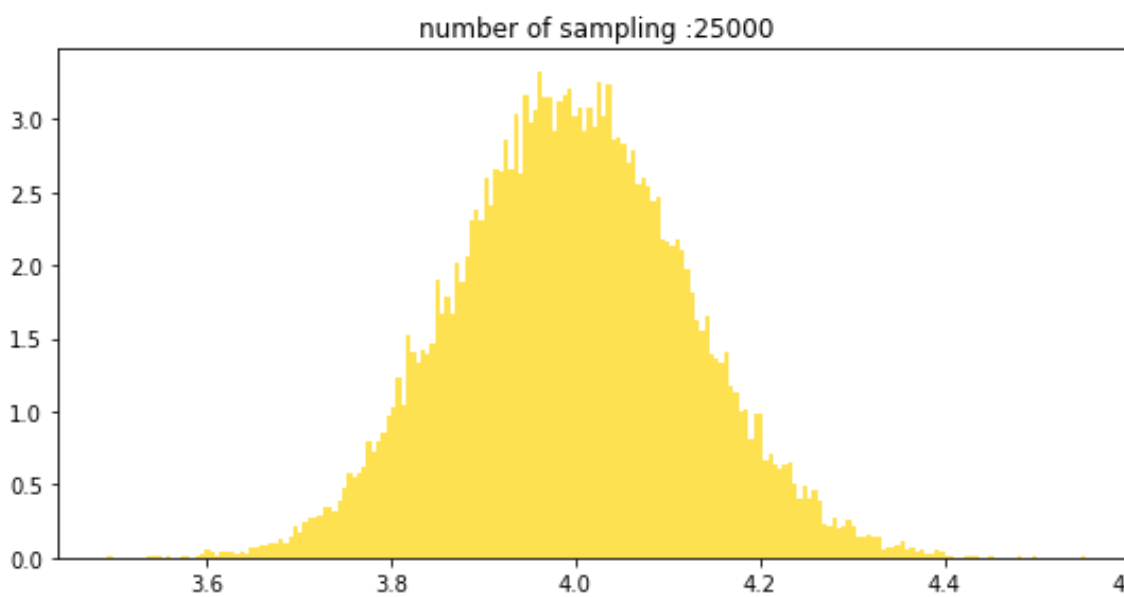
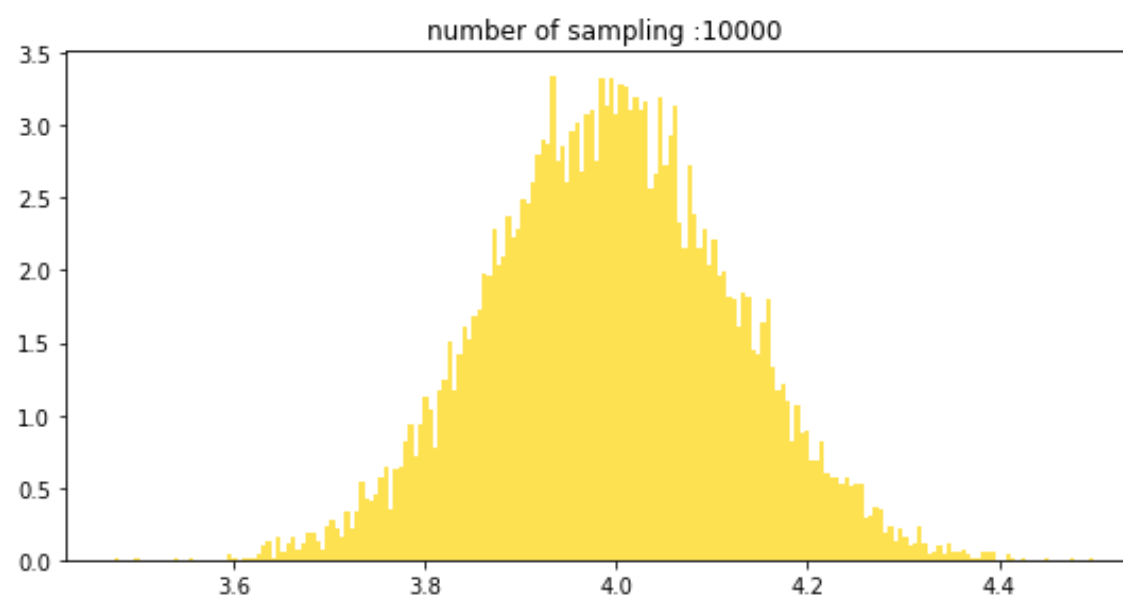
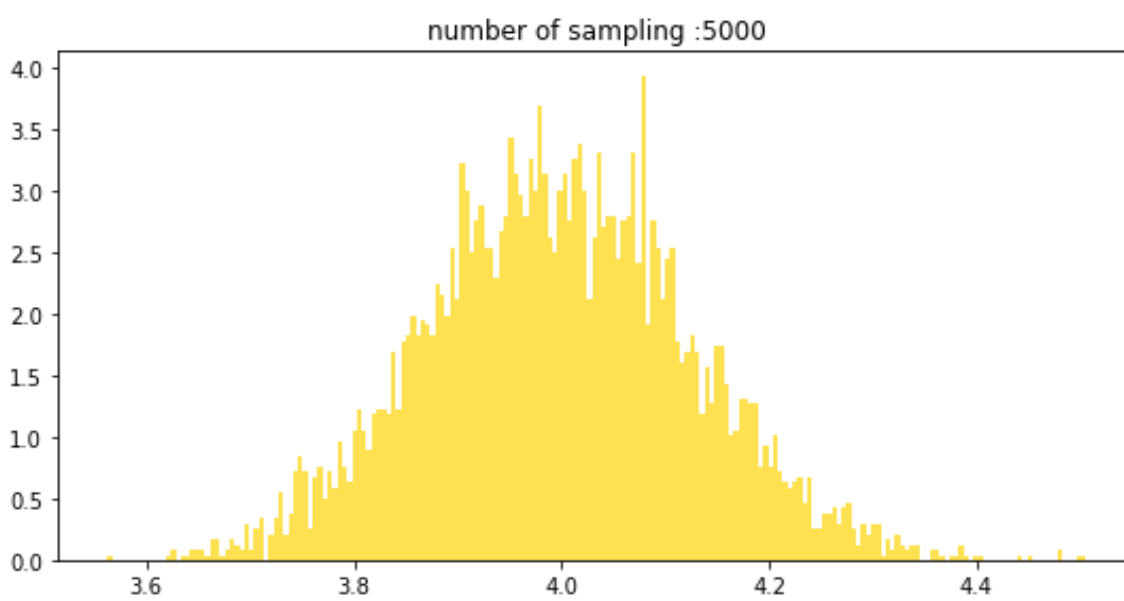
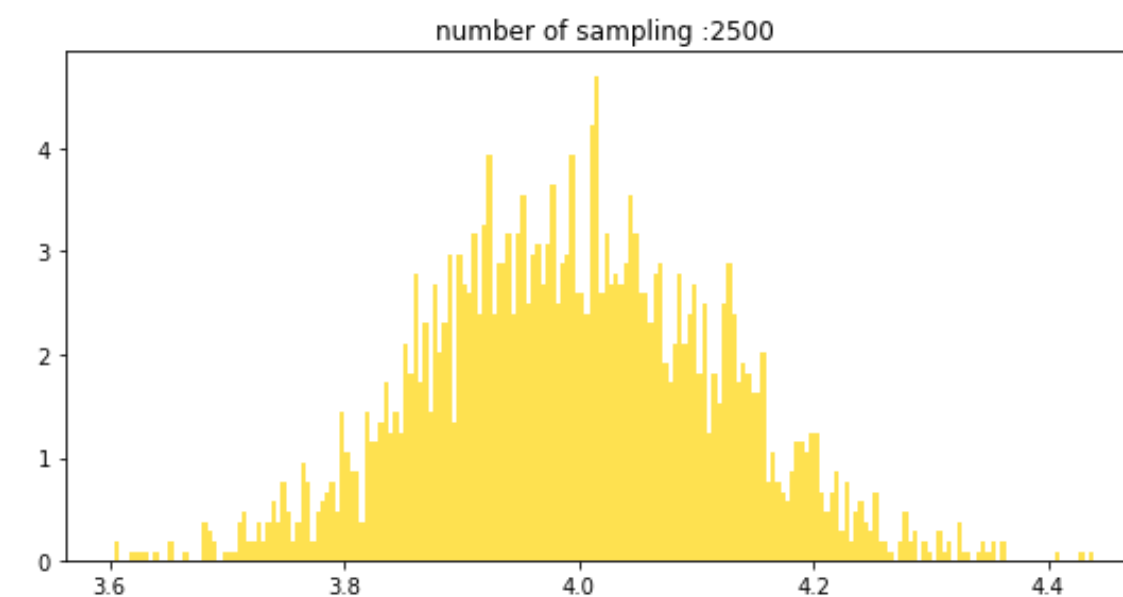
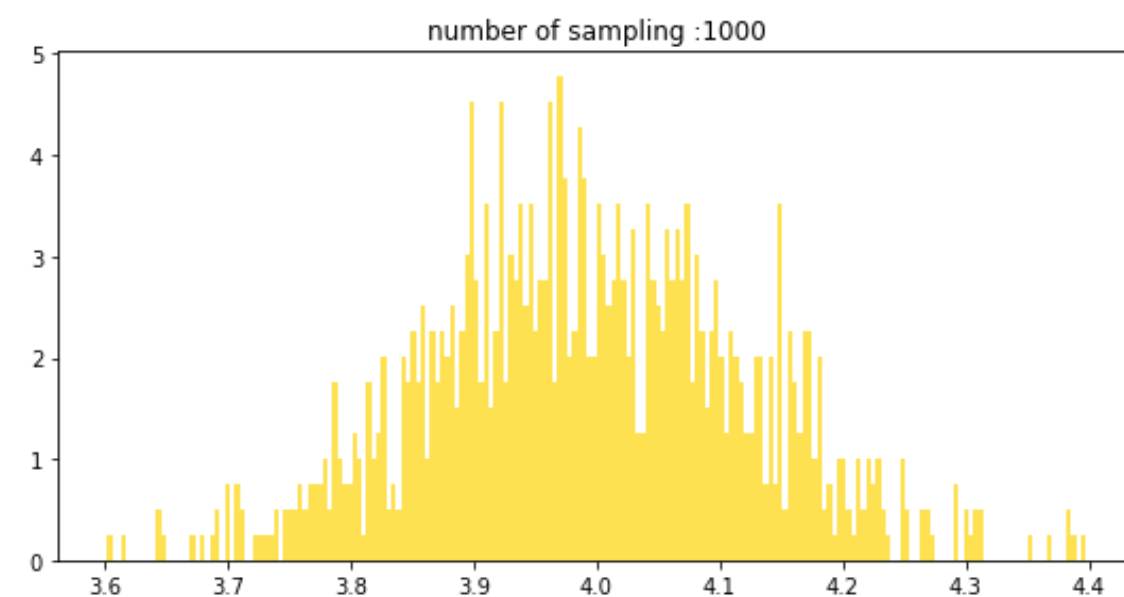


Говорят, что для
последовательности с.в.
 ξ_1, ξ_2, \dots выполняется ЦПТ,
если

$$\frac{S_n - \mathbb{E}S \cdot n}{\sqrt{n \cdot \mathbb{D}S_n}} \rightarrow \mathcal{N}(0,1),$$

по вероятности
при $n \rightarrow \infty$,

где $S_n = \xi_1 + \dots + \xi_n$





Условная вероятность



Независимость событий



Критерий независимости



формула полной вероятности



формула Байеса

Условная вероятность

Пусть задано вероятностное пространство (Ω, F, P) , события $A, B \in F$, $P(B) > 0$.

Условная вероятность события A при событии B :

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

Независимые события

Пусть есть вероятностное пространство (Ω, F, P) . События $A_1, \dots, A_n \in F$

называются независимыми в совокупности, если $\forall k = 2, \dots, n \quad \forall i_1, \dots, i_k: 1 \leq i_1 < i_2 < \dots < i_k \leq n$

выполняется:

$$\mathbb{P} \left(\bigcap_{i=1}^k A_{i_j} \right) = \prod_{j=1}^k \mathbb{P}(A_{i_j})$$

Свойства независимых событий



Замечание

В общем случае из попарной независимости событий A_1, \dots, A_n не следует их независимость в совокупности.

01

Если $A = \emptyset$ или $\mathbb{P}(A) = 0$, то $\forall B: \mathbb{P}(B) > 0$ события A и B независимы.

02

Пусть A и B независимы. Тогда события \bar{A} и B , A и \bar{B} , \bar{A} и \bar{B} также независимы.

03

Пусть $A \subset B$ и $\mathbb{P}(A) > 0, \mathbb{P}(B) < 1$. Тогда A и B зависимы.

04

Если события A и B независимы и $\mathbb{P}(B) > 0$, то $\mathbb{P}(A) = \mathbb{P}(A|B)$

Пример

Рассмотрим правильный тетраэдр, три грани которого окрашены соответственно в красный, синий, зелёный цвета, а четвёртая грань содержит все три цвета.

Событие R (соответственно, G , B) означает, что выпала грань, содержащая красный (соответственно, зелёный, синий) цвета.

$$\mathbb{P}(R) = \mathbb{P}(G) = \mathbb{P}(B) = \frac{1}{2}$$

$$\mathbb{P}(RB) = \mathbb{P}(GB) = \mathbb{P}(RB) = \frac{1}{4}$$

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

т.е. события не являются независимыми
в совокупности

Критерий независимости

Введём обозначение $A_i^{(\delta)} = \begin{cases} A_i, & \delta=1 \\ \overline{A_i}, & \delta=0 \end{cases}$

События A_1, \dots, A_n независимы в совокупности $\Leftrightarrow \forall \delta_1, \dots, \delta_n \in \{0,1\}$ выполнено равенство:

$$\mathbb{P} \left(\bigcap_{j=1}^k A_i^{\delta_i} \right) = \prod_{j=1}^k \mathbb{P}(A_i^{\delta_i})$$

Формула полной вероятности

Пусть даны события $A, B_1, \dots, B_n, \dots$; $\mathbb{P}(B_i) > 0$, причем $B_i B_j = \emptyset (i \neq j) \cup_{i=1}^{\infty} B_i \supset A$

Тогда справедлива следующая формула

$$\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(B_i) \mathbb{P}(A|B_i)$$

Формула(ы) Байеса

Пусть даны события A, H_1, \dots, H_n , $\mathbb{P}(A) > 0$, $\mathbb{P}(H_i) > 0$, причем $H_i H_j = \emptyset (i \neq j)$ и $\bigcup_{i=1}^{\infty} H_i \supset A$

Тогда справедливы формулы Байеса:

$$\mathbb{P}(H_i|A) = \frac{\mathbb{P}(H_i) \mathbb{P}(A|H_i)}{\sum_{j=1}^{\infty} \mathbb{P}(H_j) \mathbb{P}(A|H_j)}, \quad i = 1, n$$

Пример



Тест на рак имеет надёжность 99% (т.е. вероятность как положительной, так и отрицательной ошибки равна 0,01), рак появляется у 1% населения.

Какова вероятность того, что человек болен раком, если у него позитивный результат теста?

Результат теста	Пациент реально болен	
	Да	Нет
Положительный	$0,99 \cdot 0,01$	$0,01 \cdot 0,99$
Отрицательный	$0,01 \cdot 0,01$	$0,99 \cdot 0,99$