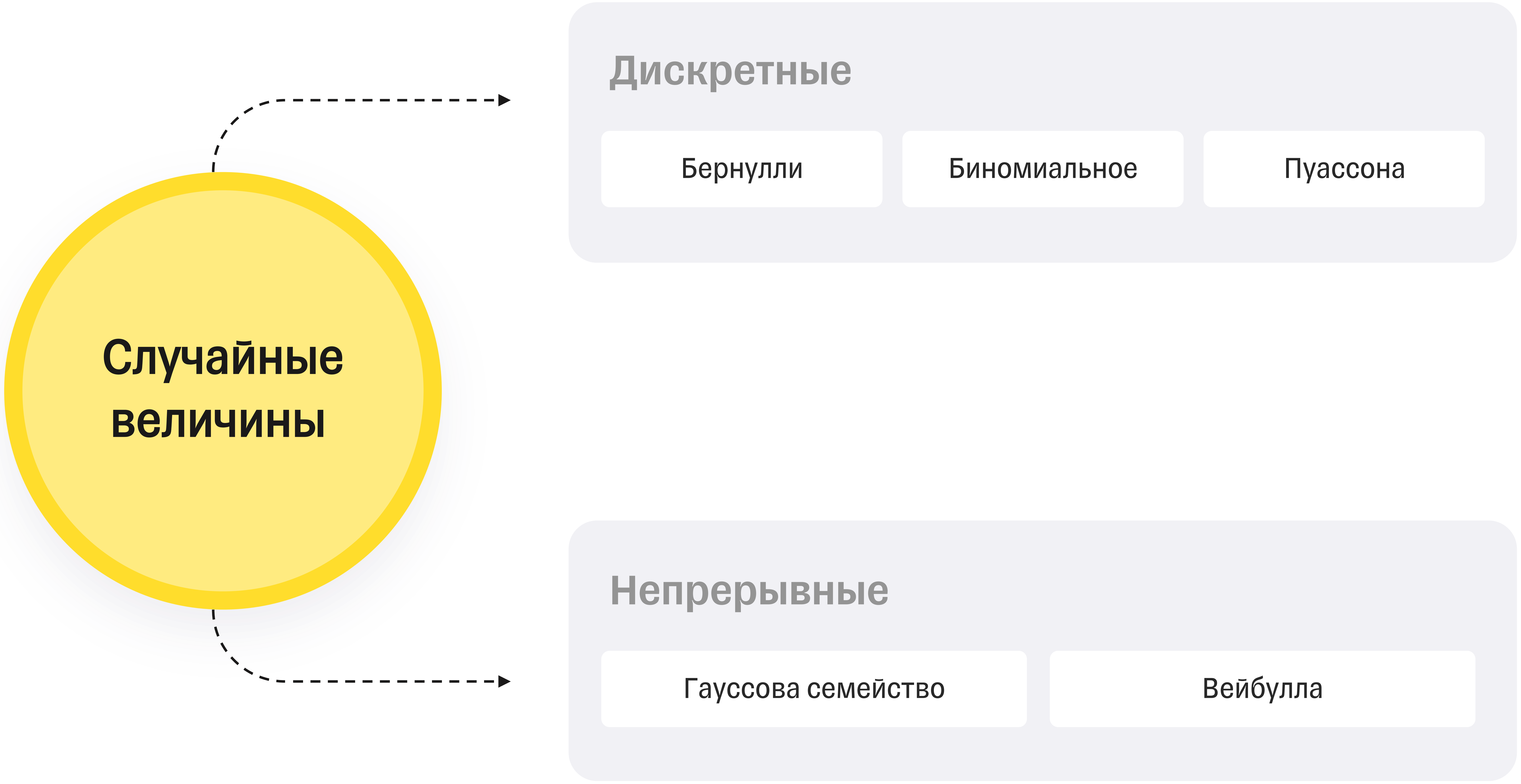
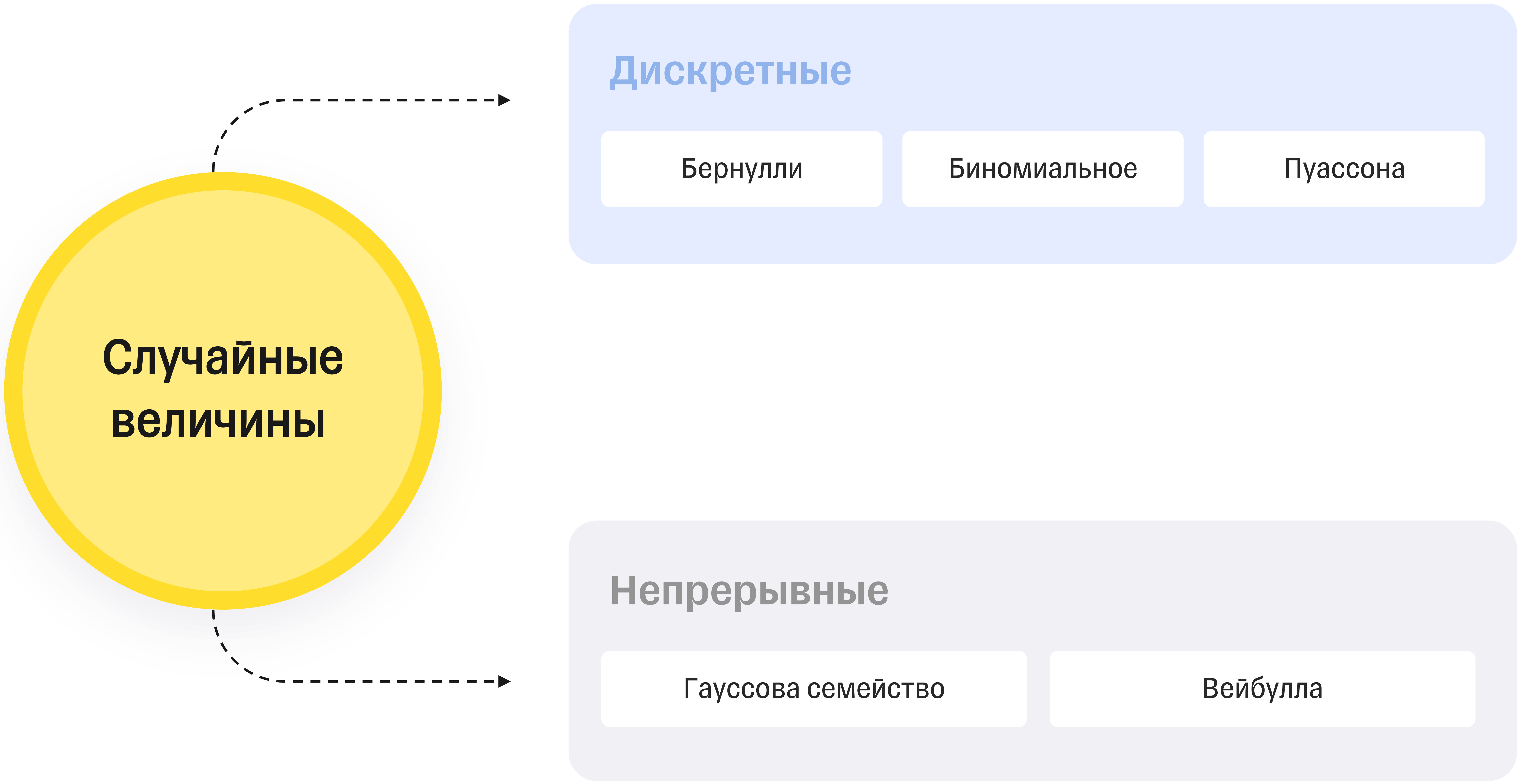
ТИНЬКОФФ

Занятие 2

Введение в теорию вероятностей и математическую статистику


Случайные величины
Дискретные
Бернулли
Биномиальное
Пуассона
Непрерывные
Гауссова семейство
Вейбулла

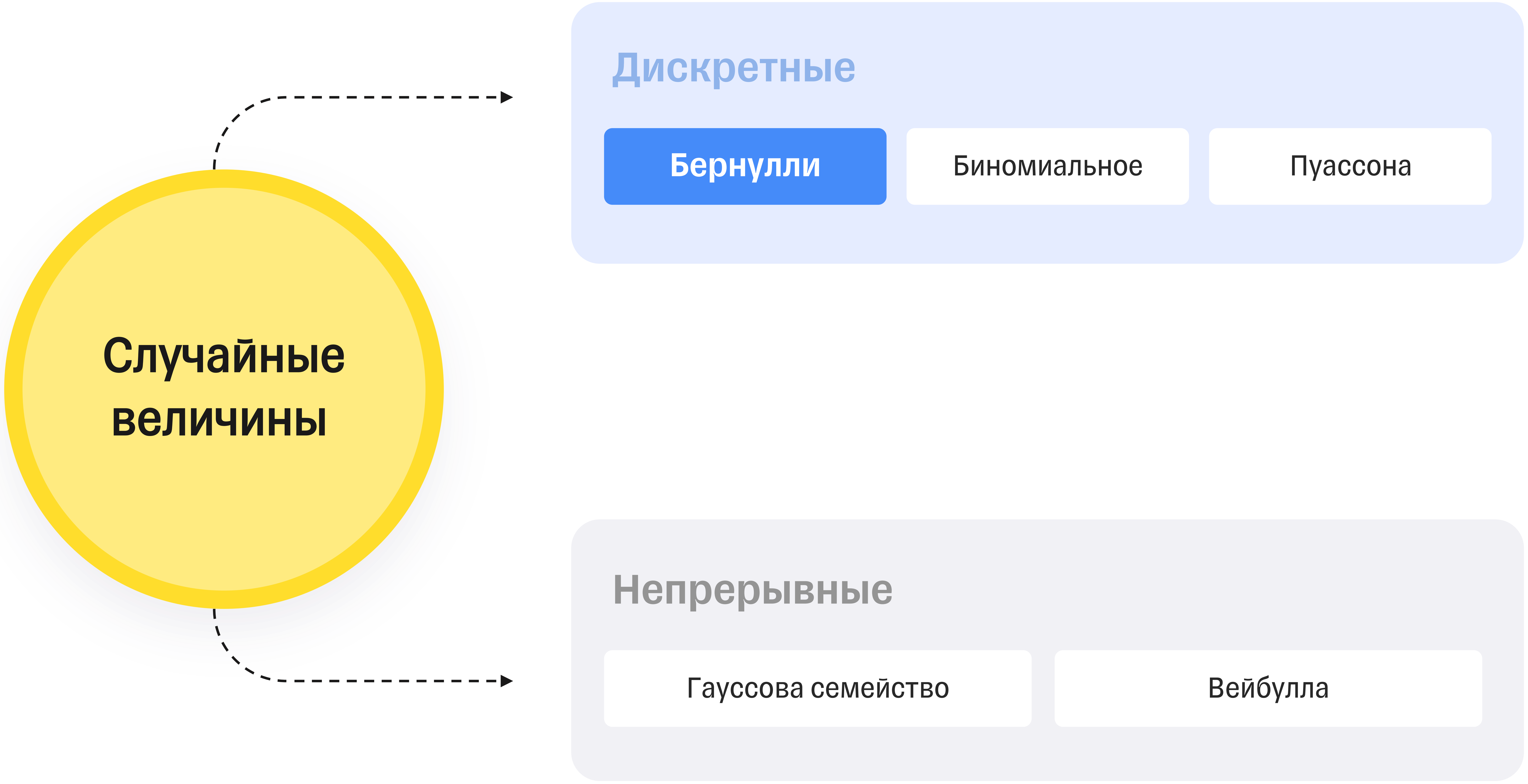

Случайные величины
Дискретные
Бернулли
Биномиальное
Пуассона
Непрерывные
Гауссова семейство
Вейбулла

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
Дискретные
Бернулли
Биномиальное
Пуассона
Непрерывные
Гауссова семейство
Вейбулла

Распределение Бернулли

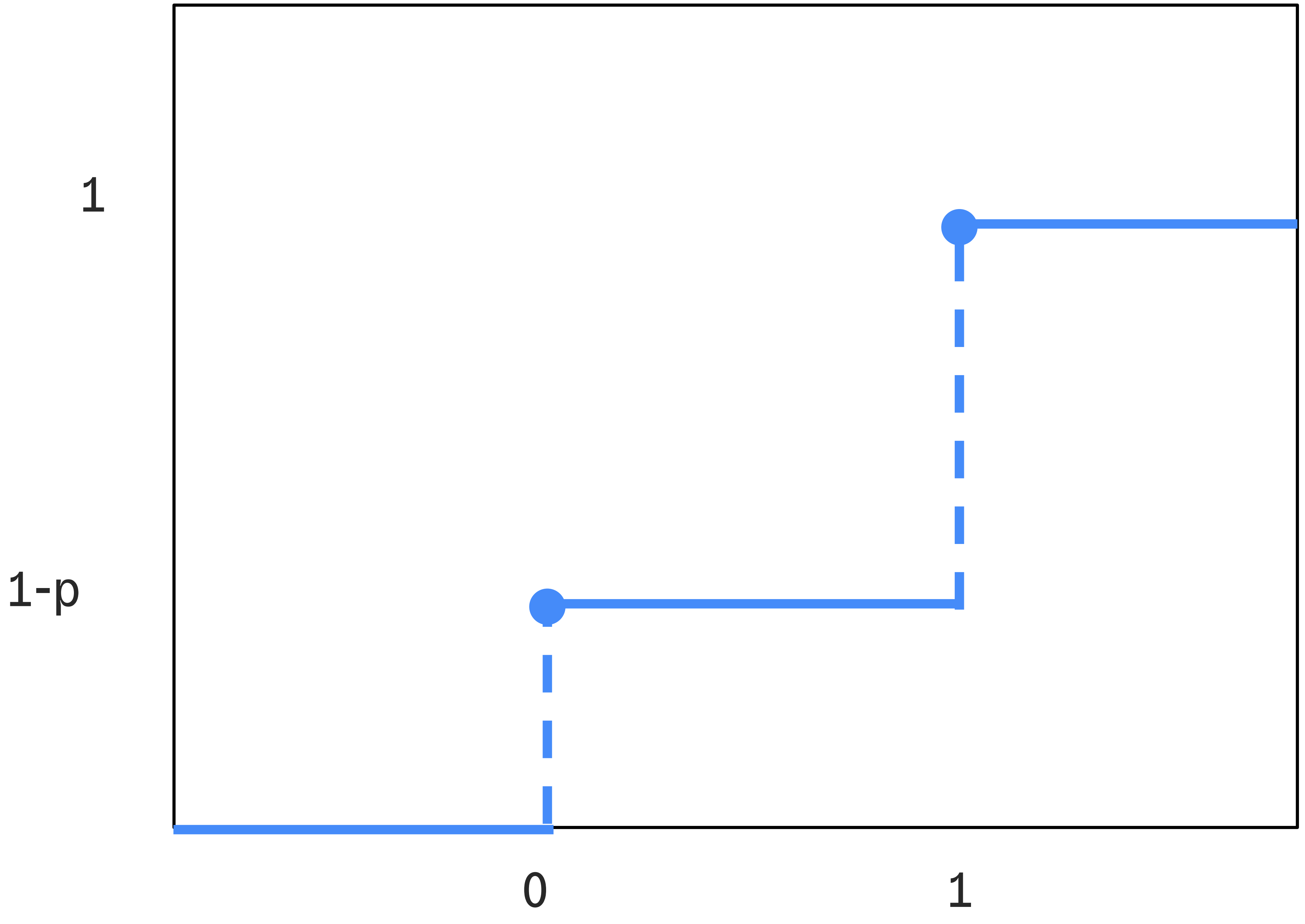
Характеризует случайную величину – результат единоразвого эксперимента с бинарным результатом

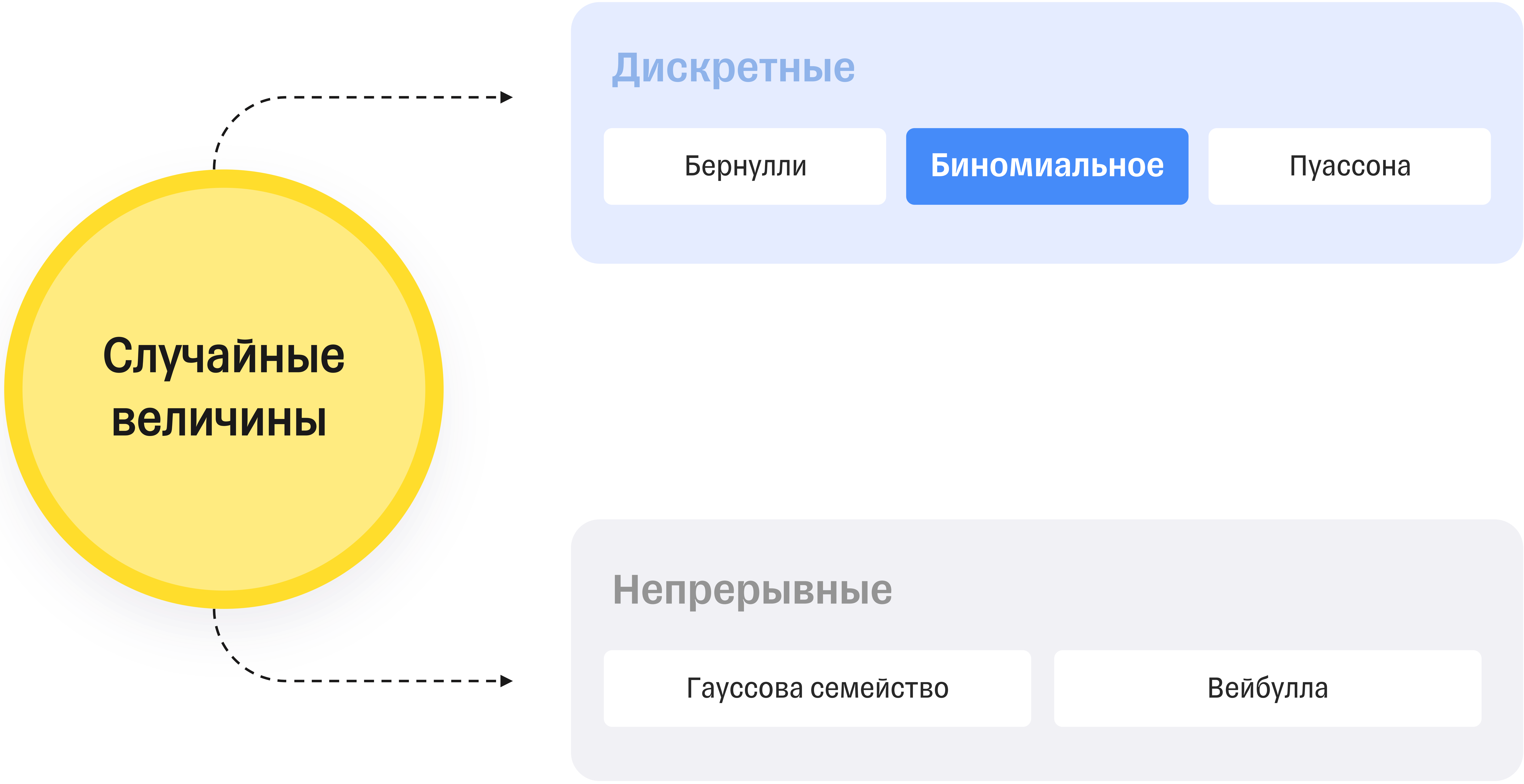
Обозначение: $\xi \sim Be(p)$, где p – вероятность положительного исхода

Примеры дискретных с.в.

- Возвращение кредита заёмщиком
- Рост актива в заданный момент времени

Функция распределения Бернулли




Случайные величины
Дискретные
Бернулли
Биномиальное
Пуассона
Непрерывные
Гауссова семейство
Вейбулла

Биномиальное распределение

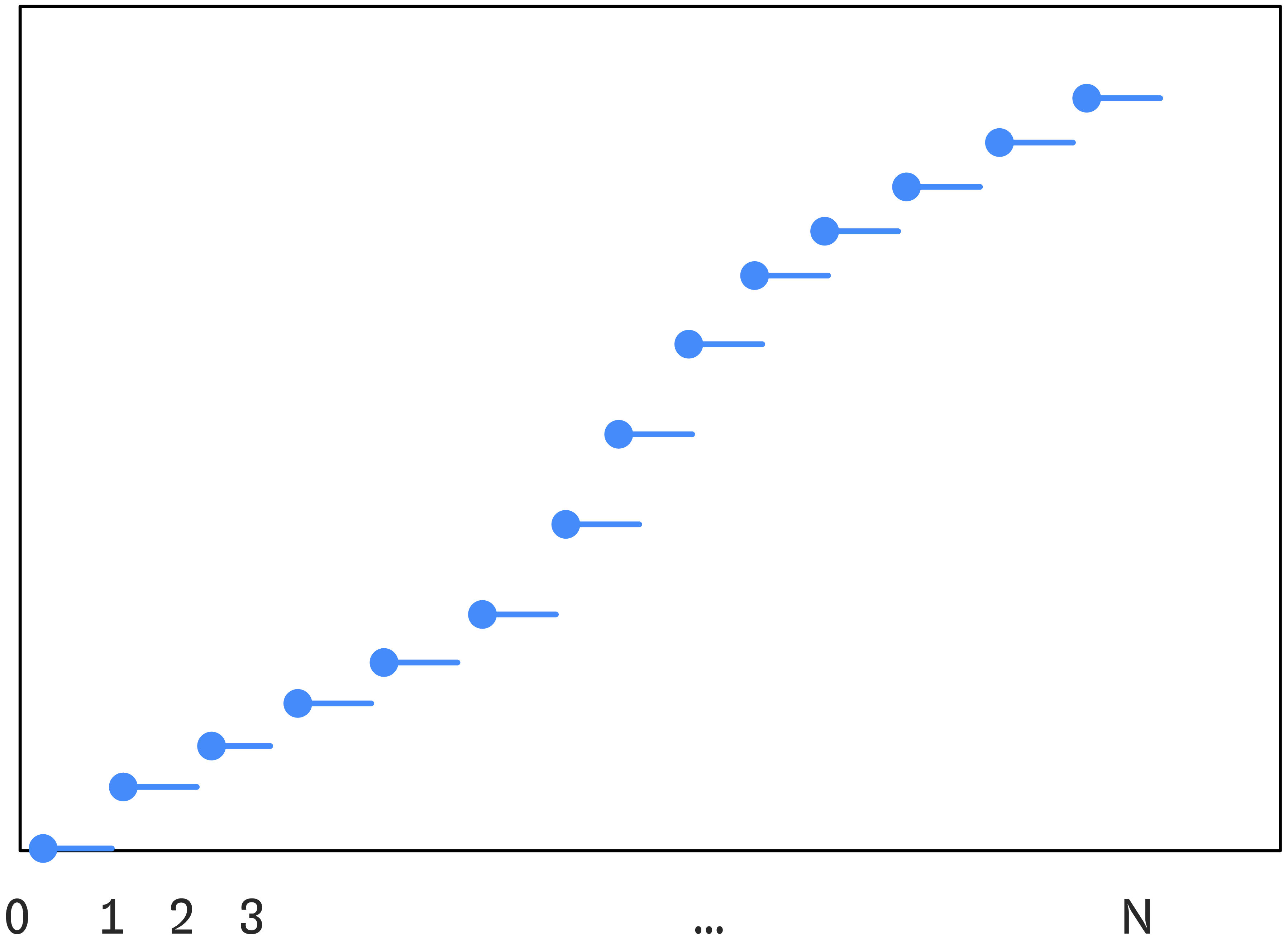
Характеризует случайную величину – количество положительных исходов при проведении n независимых испытаний Бернулли.

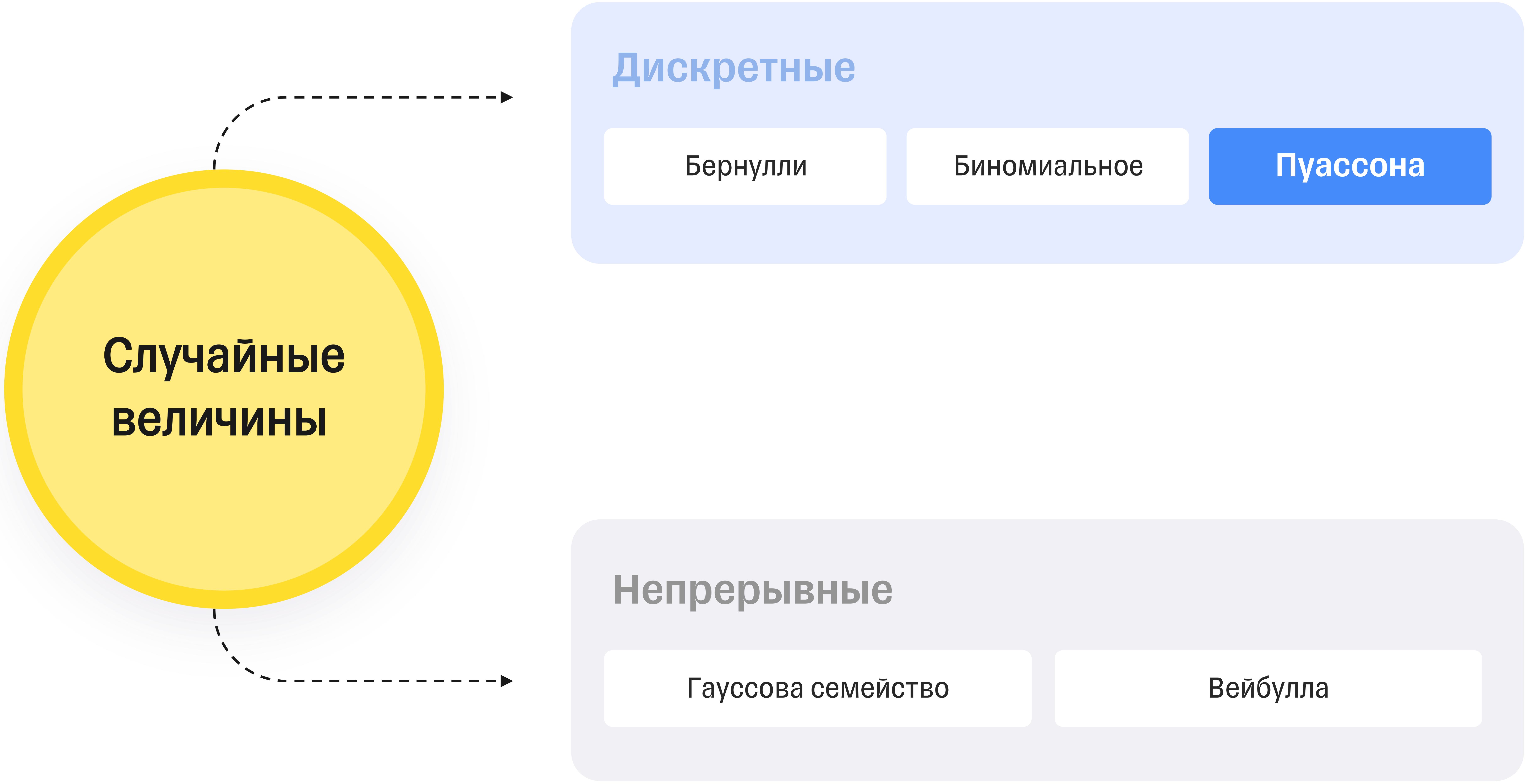
Обозначение: $\xi \sim Bi(p, n)$, где p – вероятность положительного исхода, n – число испытаний

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
Дискретные
Бернулли
Биномиальное
Пуассона
Непрерывные
Гауссова семейство
Вейбулла

Распределение Пуассона

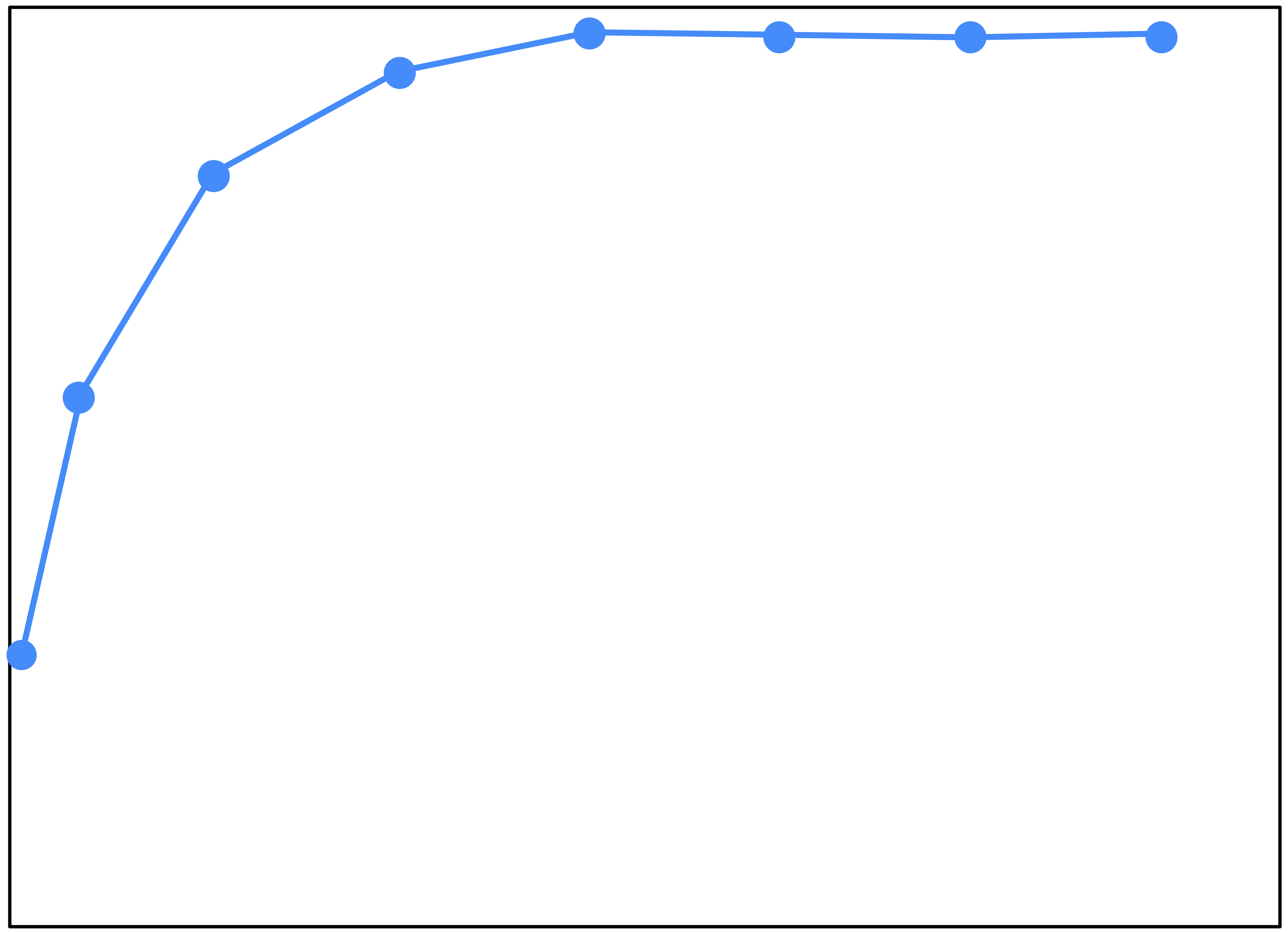
Характеризует случайную величину – схожую с Биномиальным, но применительно к редким событиям при некоторых предположениях о характере процесса

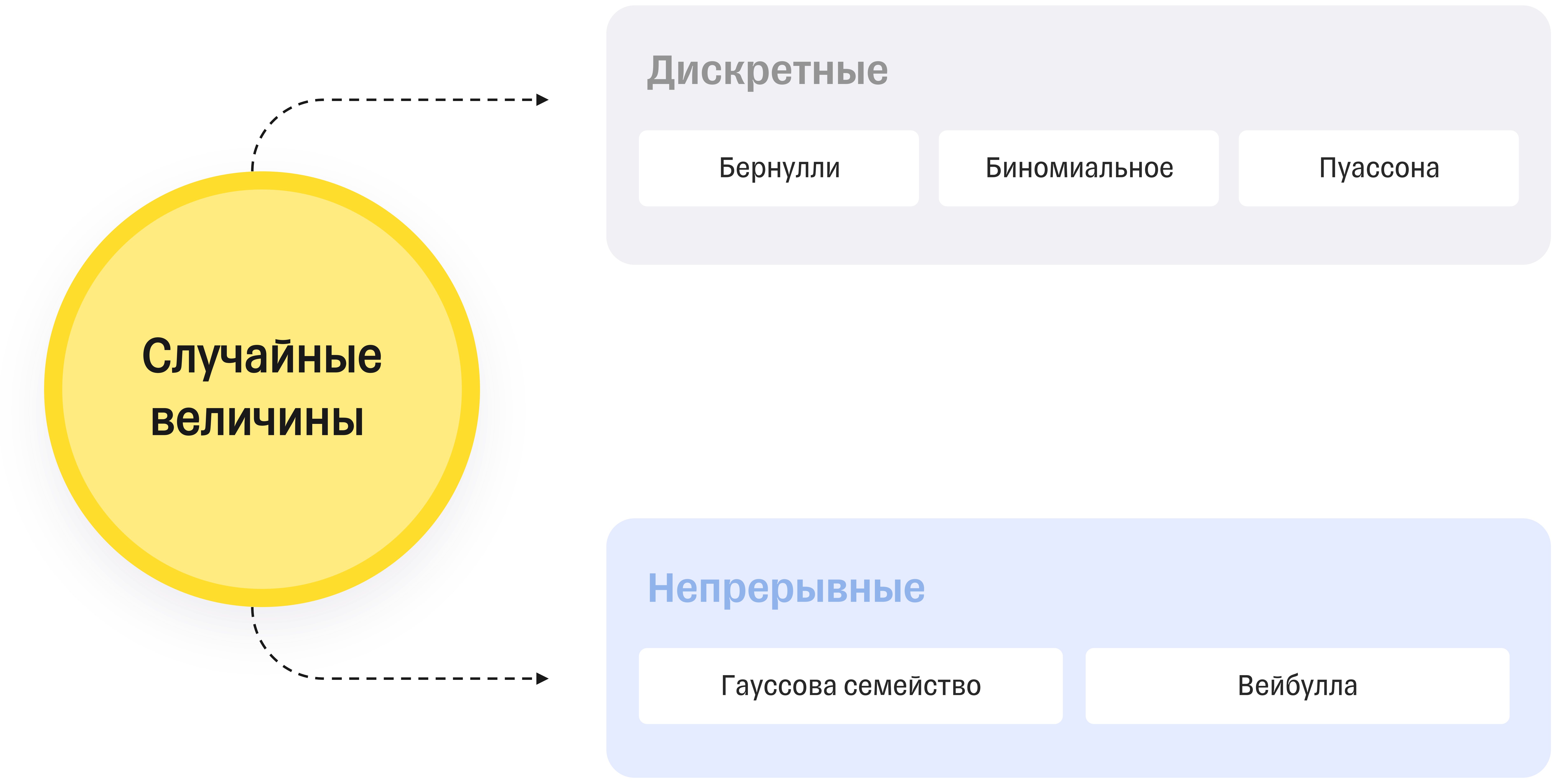
Обозначение: $\xi \sim Pois(\lambda)$, где $\lambda > 0$ – интенсивность события

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
Дискретные
Бернулли
Биномиальное
Пуассона
Непрерывные
Гауссова семейство
Вейбулла

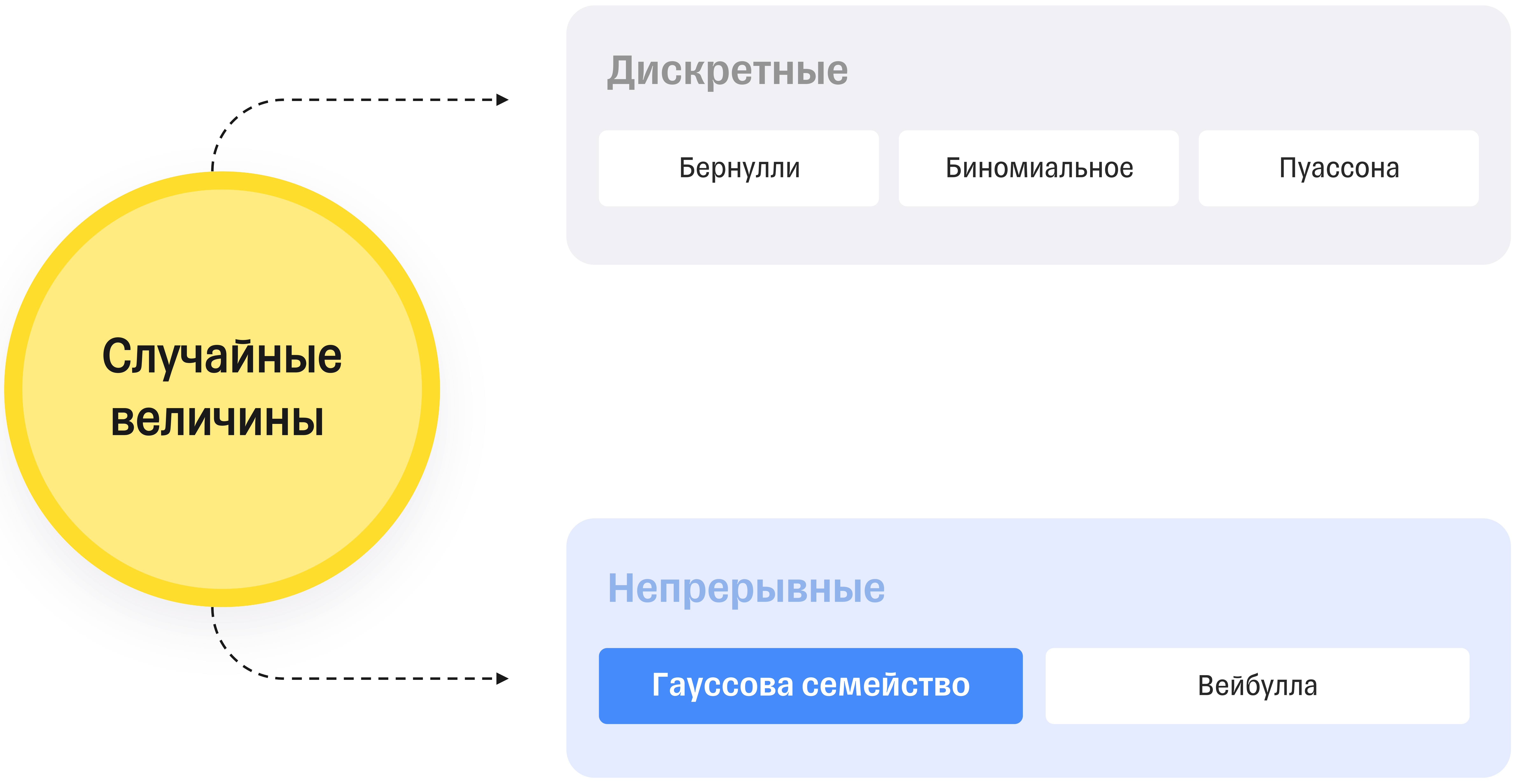
Непрерывная случайная величина

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

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

Примеры непрерывных с.в.

- Размер детали на заводе
- Точные возраст и вес человека
- Измерительная ошибка прибора


Случайные величины
Дискретные
Бернулли
Биномиальное
Пуассона
Непрерывные
Гауссова семейство
Вейбулла

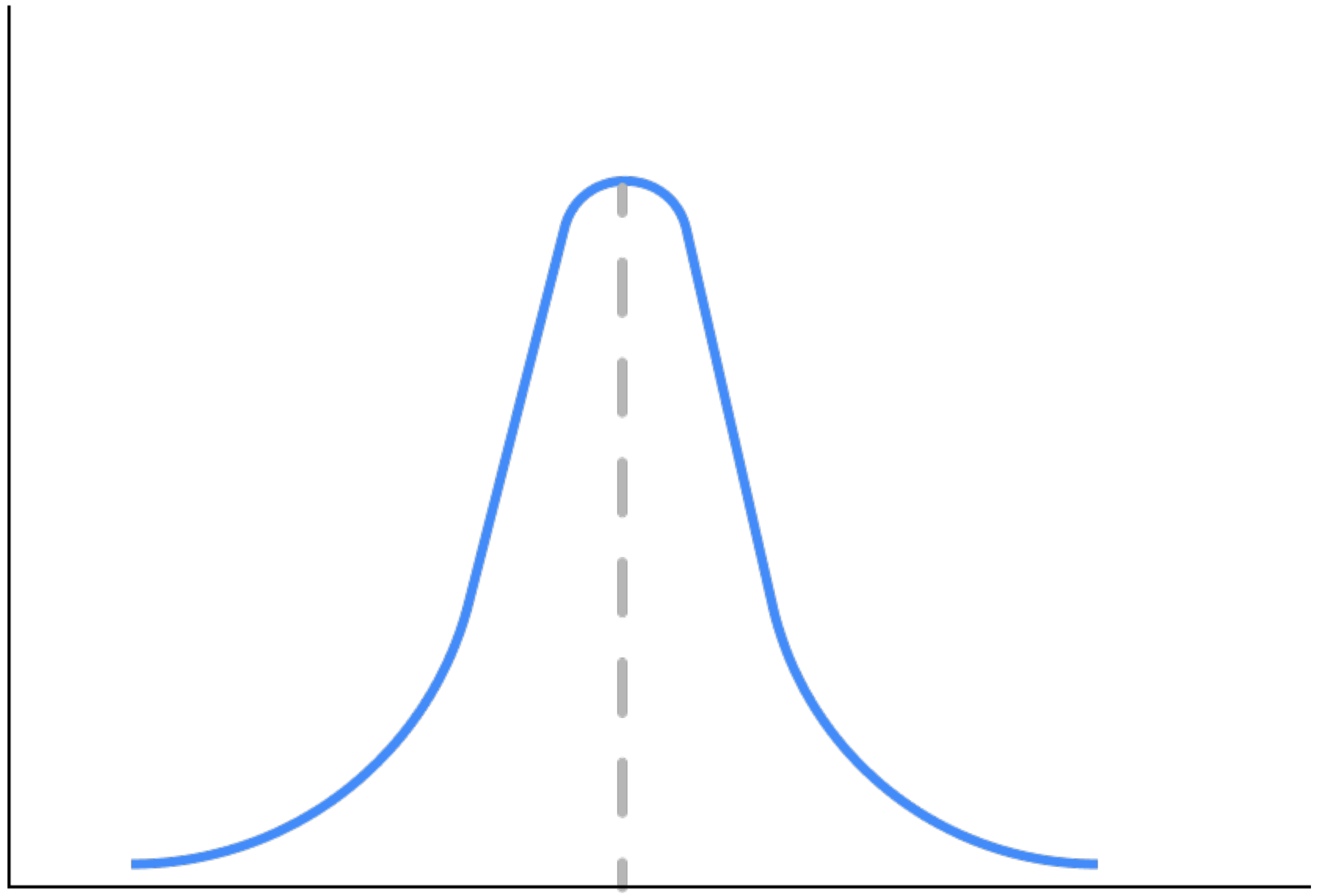
Нормальное распределение

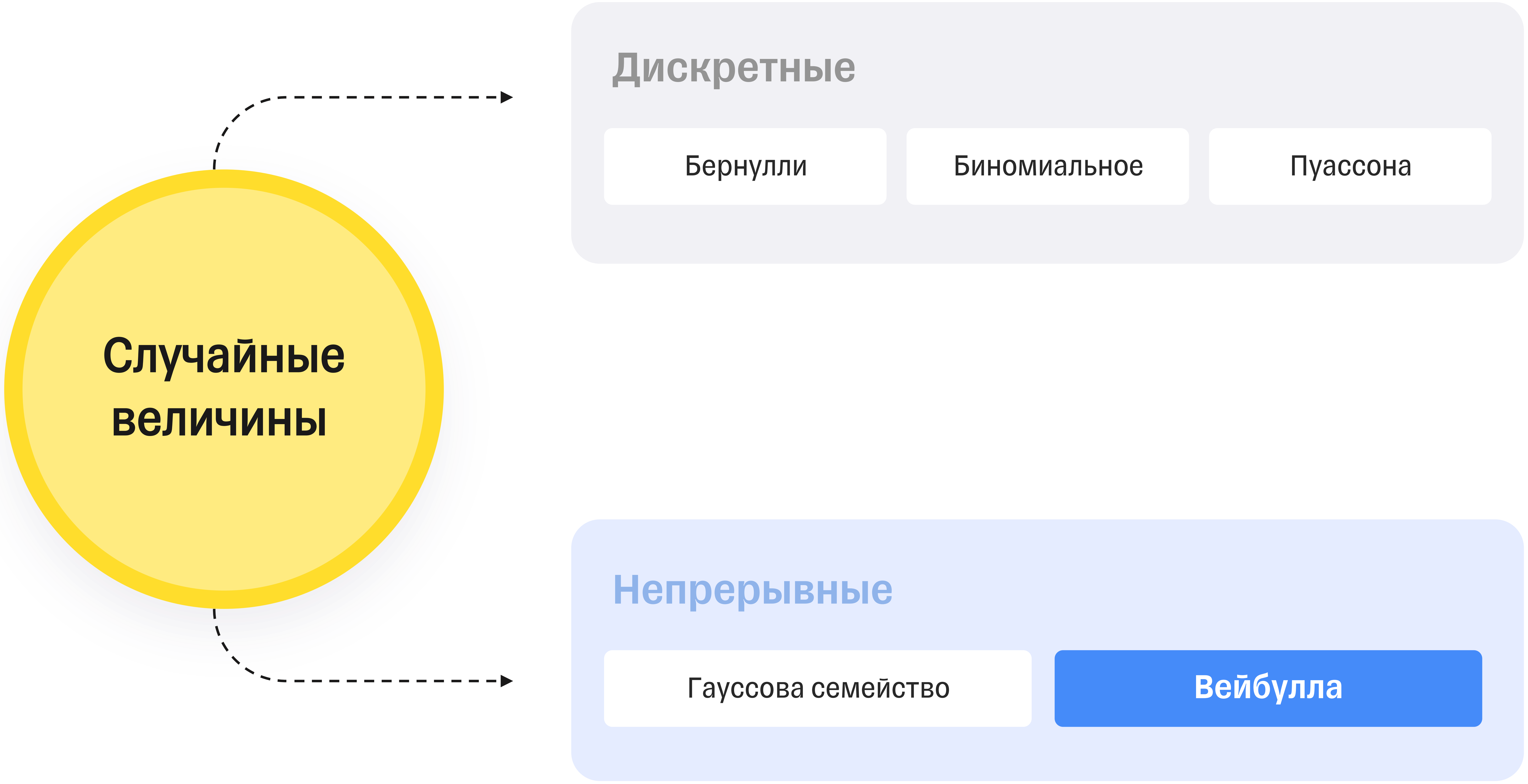
Характеризует случайную величину, зависящую от 2х параметров: среднеквадратичного отклонения и матожидания

Примеры

- Рост человека
- Масса вылавливаемой рыбы одного вида

Плотность вероятности нормального распределения




Случайные величины
Дискретные
Бернулли
Биномиальное
Пуассона
Непрерывные
Гауссова семейство
Вейбулла

Распределение Вейбулла

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

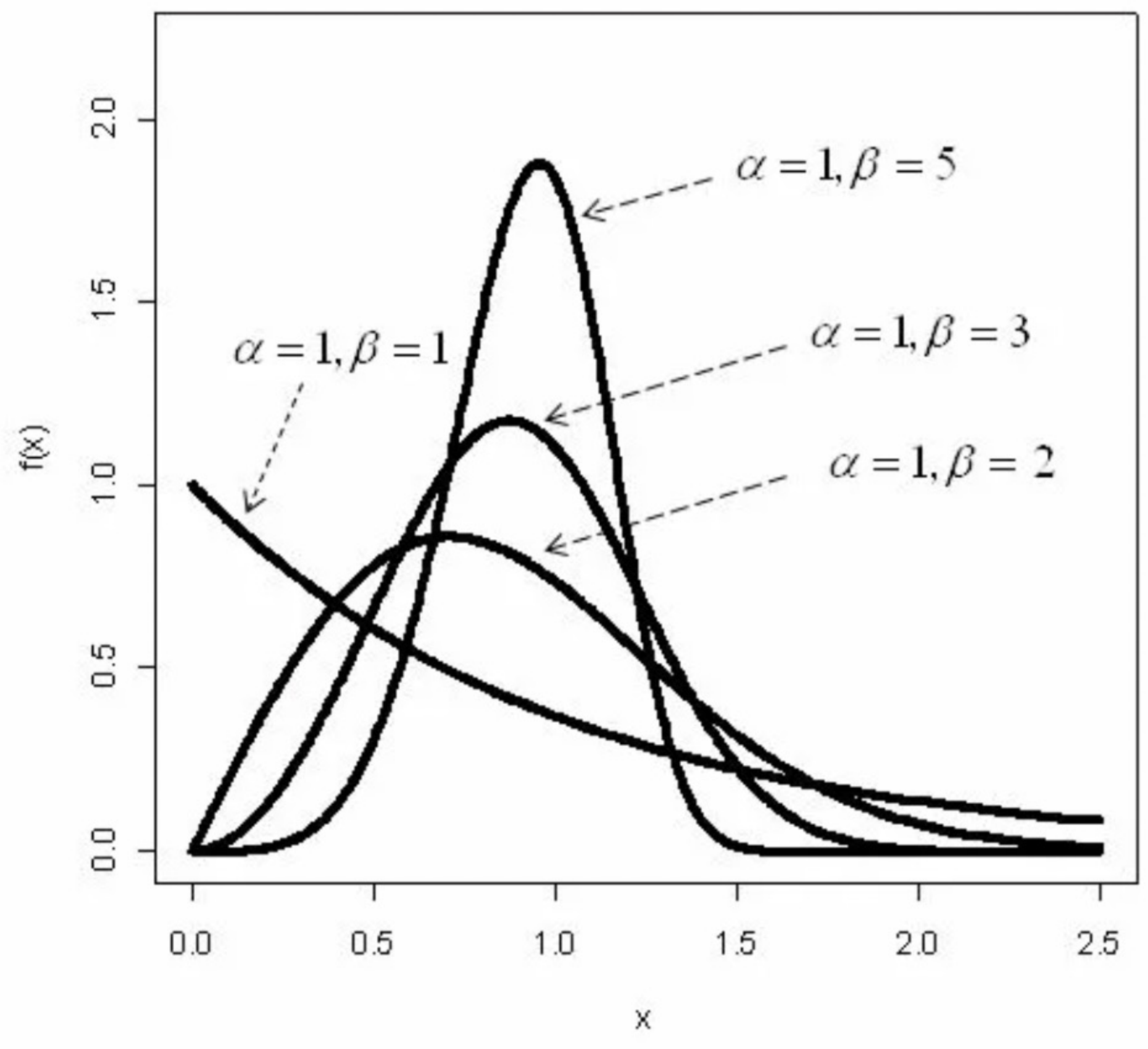
Обозначение:

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

Примеры

- Задач оценки рисков
- Анализ выживаемости

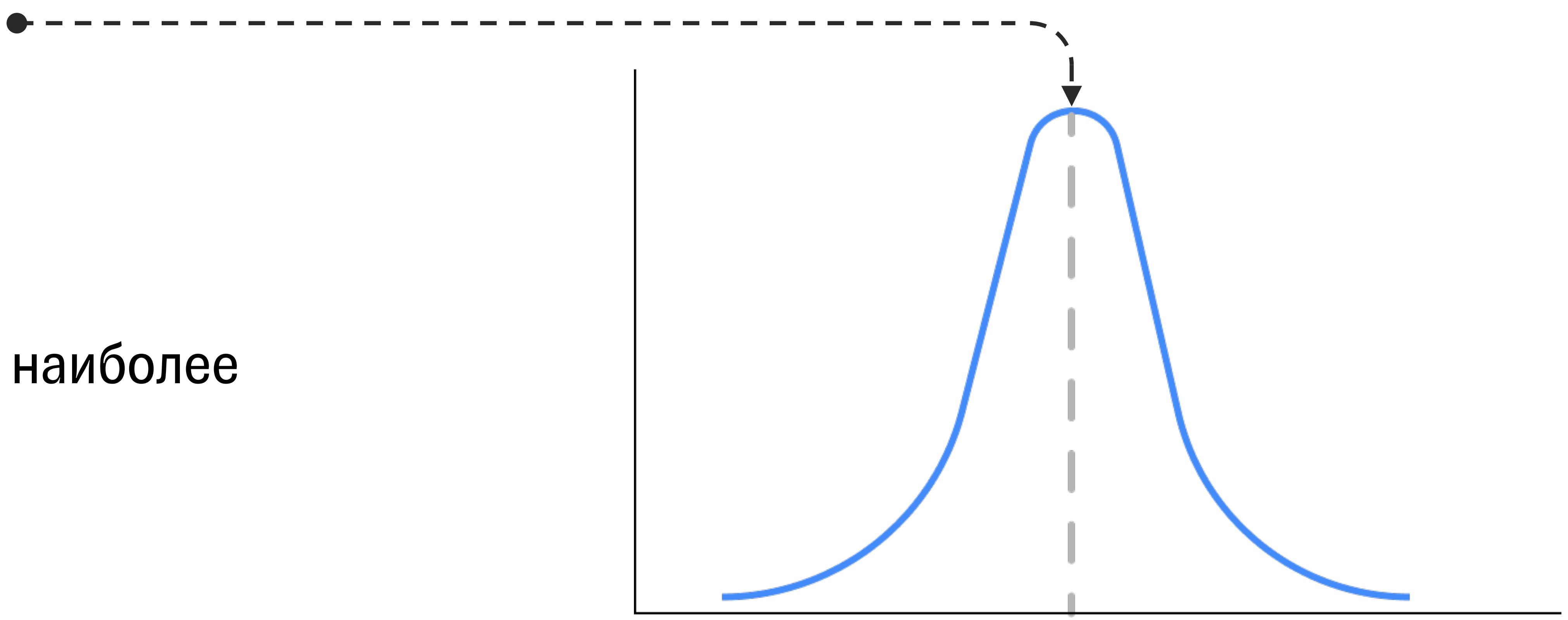
Плотность вероятности распределения Вейбулла



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 случайно величины

Дискретный случай:

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

Непрерывный случай:

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

Математическое ожидание

Свойства

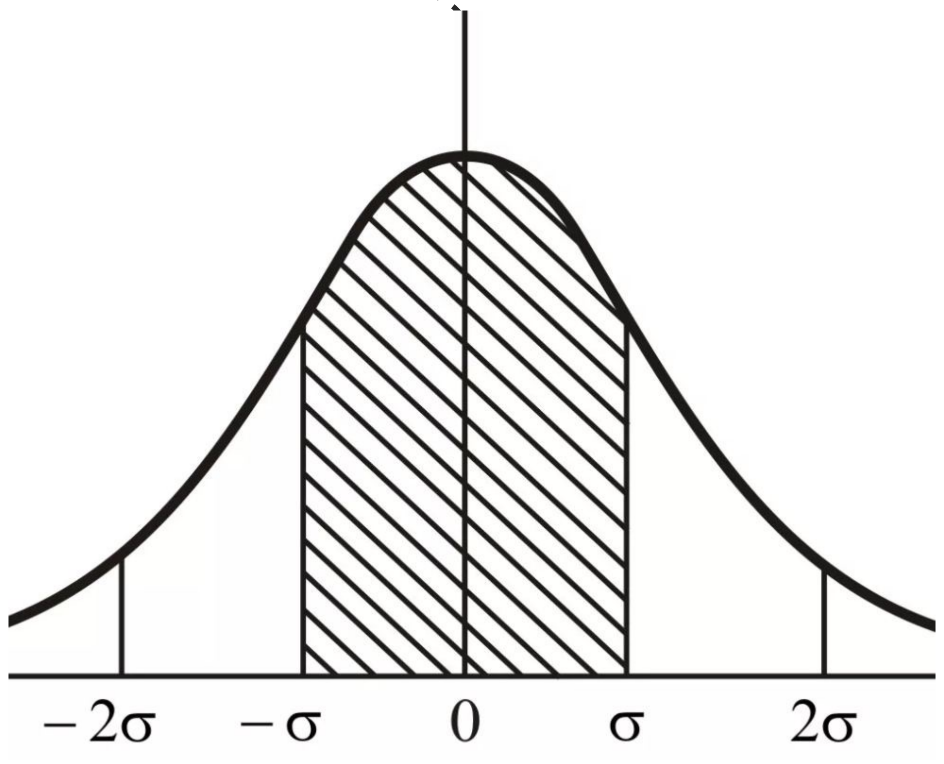
$a, b \in \mathbb{R}\ \exists \mathbb{E}\xi, \mathbb{E}\eta\ \Rightarrow\ \mathbb{E}(a\xi \pm b\eta) = a\mathbb{E}\xi \pm b\mathbb{E}\eta$

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

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

$\mathbb{E}g(\xi) = \sum_{i=1}^{+inf} g(xi)pi$ для д.с.в. или $\mathbb{E}g(\xi) = \int_{-inf}^{+inf} g(x)\, f_X\, dx$ 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

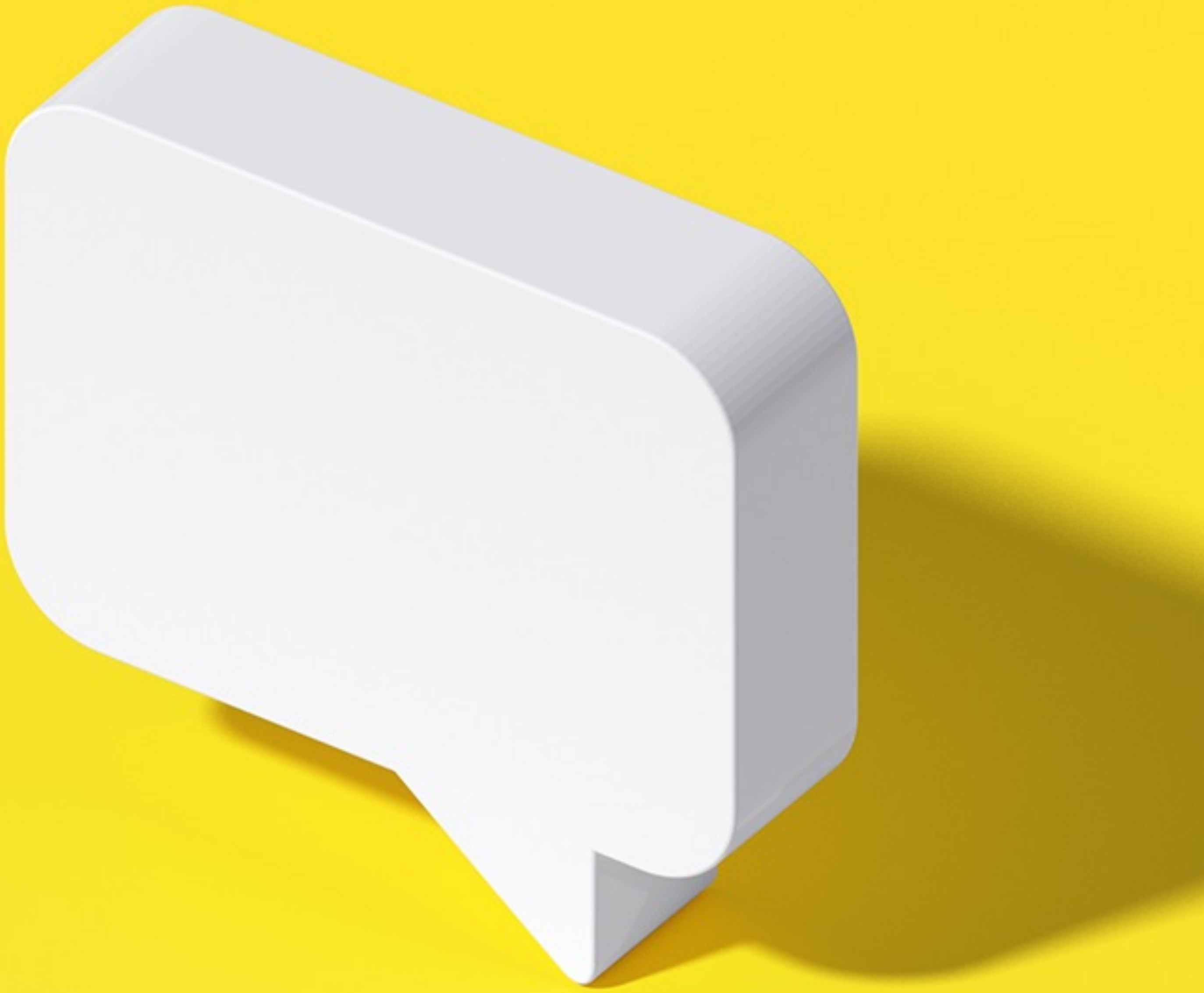
$$cov(\xi, \eta) = \mathbb{E}((\xi - \mathbb{E}\xi)(\eta - \mathbb{E}\eta))$$

Коэффициент корреляции

Инвариантный по масштабу коэффициент линейной зависимости между случайными величинами

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

Центральная предельная теорема

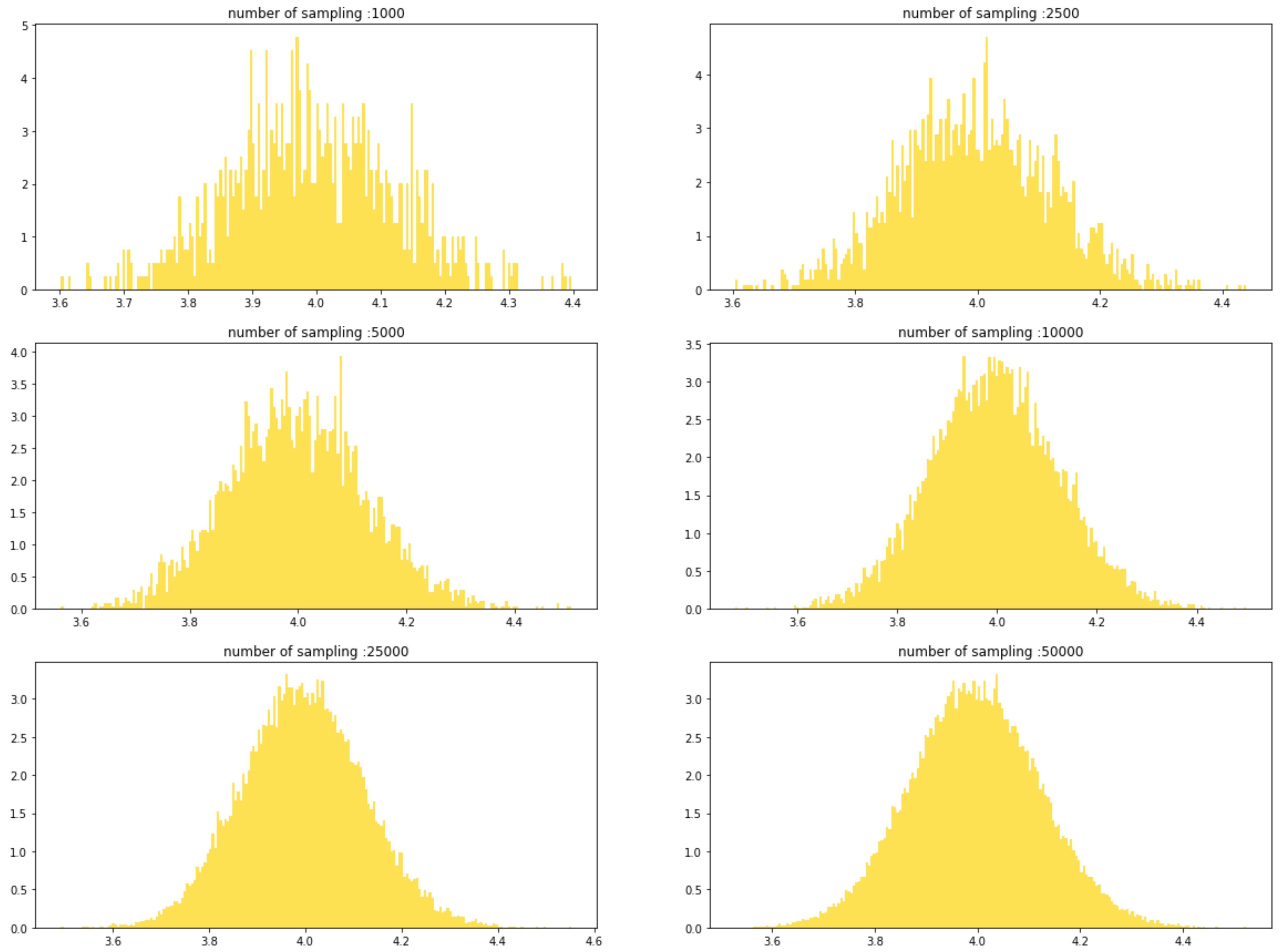


Говорят, что для последовательности с.в. $\xi_1, \xi_2, \ldots$ выполняется ЦПТ, если

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

по вероятности при $n \longrightarrow \infty$,

где $S_n = \xi_1 + \cdots + \xi_n$



- 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 $A_1, \dots, An \in F$ называются независимыми в совокупности, если $\forall k = 2, \dots, n\ \ \forall i_1, \dots, i_k\colon\ 1 \le i_1 < i_2 < \cdots < i_k \le n$ выполняется:

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

Свойства независимых событий

01

Если $A = \emptyset$ или $\mathbb{P}(A) = 0$, то $\forall B\colon\ B\colon\ \mathbb{P}(B) > 0$ 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 $\mathbb{P}(A) = \mathbb{P}(A|B)$

Замечание

В общем случае из попарной независимости событий $A_1, \ldots, An$ не следует их независимость в совокупности.

Пример

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

Событие $\boldsymbol{R}$ (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 $A, B_1, \ldots, Bn, \ldots$; $\mathbb{P}(B_i) > 0$, причем $B_i B_j = \emptyset (i \neq j)$ $\bigcup_{i=1}^{\infty} B_i \supset A$

Тогда справедлива следующая формула

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

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:

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

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	$0{,}01 \cdot 0{,}01$	$0{,}99 \cdot 0{,}99$