ФКН

Департамент больших данных и информационного поиска

Москва 2025

Лекция 2
Случайные величины и их характеристики

Машинное обучение в цифровом продукте

Полякова И.Ю.

Случайная величина – величина, которая в результате опыта принимает то или иное значение, причем заранее неизвестно, какое именно

(Сами СВ: X, Y, Z или θ, β, φ; значения СВ: $x_1, x_2, \dots, y_1, y_2$)

- Число очков, появляющееся при бросании игральной кости
- Время безотказной работы прибора
- Число выстрелов до первого попадания
- Средняя оценка за экзамен по курсу «Теория вероятностей и математическая статистика»
- …

СВ также по определению является функцией,
которая отображает

Закон распределения случайной величины – «любое» правило (таблица, график, функция), позволяющее находить вероятности произвольных событий, в частности, указывающее вероятности отдельных значений случайной величины или множества этих значений

Случайные величины

Дискретные

Непрерывные

Дискретная случайная величина

$X = \{x_1, x_2, \dots, x_n\}$ - «можно» пересчитать

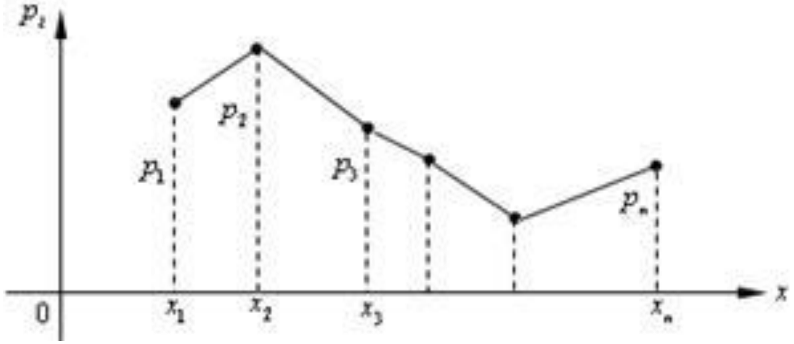
Формула закона распределения: $p_i = P\{X = x_i\}, \ i = 1, \dots, n$

Таблица распределения:

x_i	x_1	x_2	x_3	……..	x_n
p_i	p_1	p_2	p_3	……..	p_n

$$\sum_{i=1}^{i=n} p_i = 1.$$

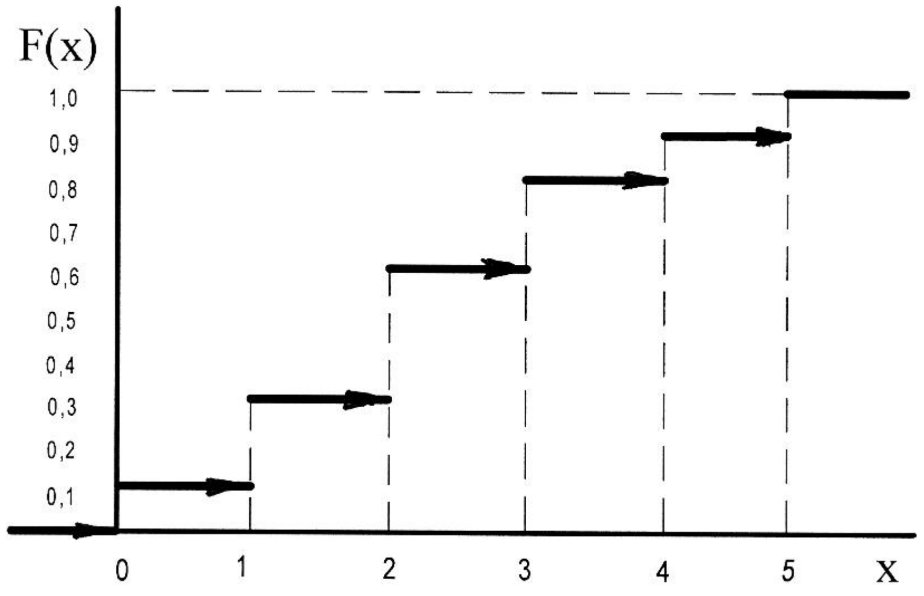
Многоугольник распределения:



Дискретная случайная величина

Функция распределения: F(x) = P(X < x)

«Вероятность того, что случайная величина примет значение меньше x»



Дискретная случайная величина

- **Суммой/ разностью/ произведением** для д.с.в. X, принимающей значения x_i с вероятностями $p_i = \{\mathrm{X} = x_i\}, i = 1, \dots, n$ и для д.с.в. Y, принимающей значения y_i с вероятностями $p_j = \{\mathrm{Y} = y_j\}, j = 1, \dots, m$ называется д.с.в

$$Z = X + Y \ (Z = X - Y; Z = X * Y),$$

- Принимающая значения

$$z_{ij} = x_i + y_j \ (z_{ij} = x_i - y_j \ ; z_{ij} = x_i * y_j \ )$$

- С вероятностями

$$p_{ij} = P\{X = x_i, Y = y_j\}$$

**в случае совпадения некоторых сумм/разностей/произведений соответствующие вероятности складываются*

Дискретная случайная величина

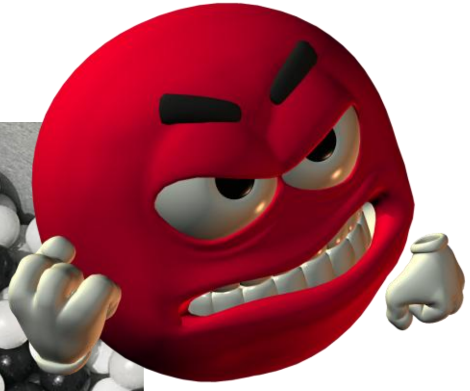
- **Произведением д.с.в на число** называется д.с.в cX, принимающая значения cx_i с вероятностями $p_i = P\{X = x_i\}$

- Две д.с.в называются **независимыми**, если события $\{X = x_i\}$ и $\{Y = y_j\}$ независимы для любых i и j, т.е
$$P\{X = x_i; Y = y_j\} = P\{X = x_i\} * P\{Y = y_j\}$$

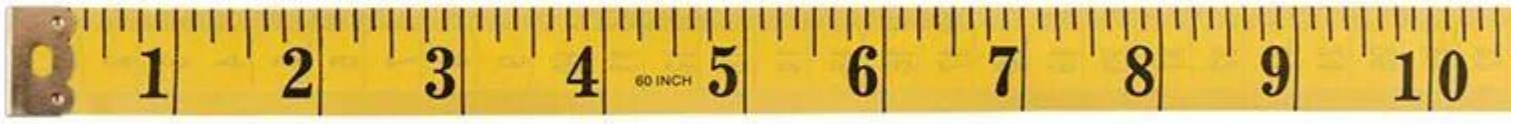
- **Взаимно независимые с.в** – такие с.в, что закон распределения любой из них не зависит от того, какие возможные значения приняли остальные величины

Задача

В урне 8 шаров, из которых 5 белых, остальные – черные. Вынимают наудачу 3 шара. Найдите закон и функцию распределения числа белых шаров в выборке



Непрерывная случайная величина



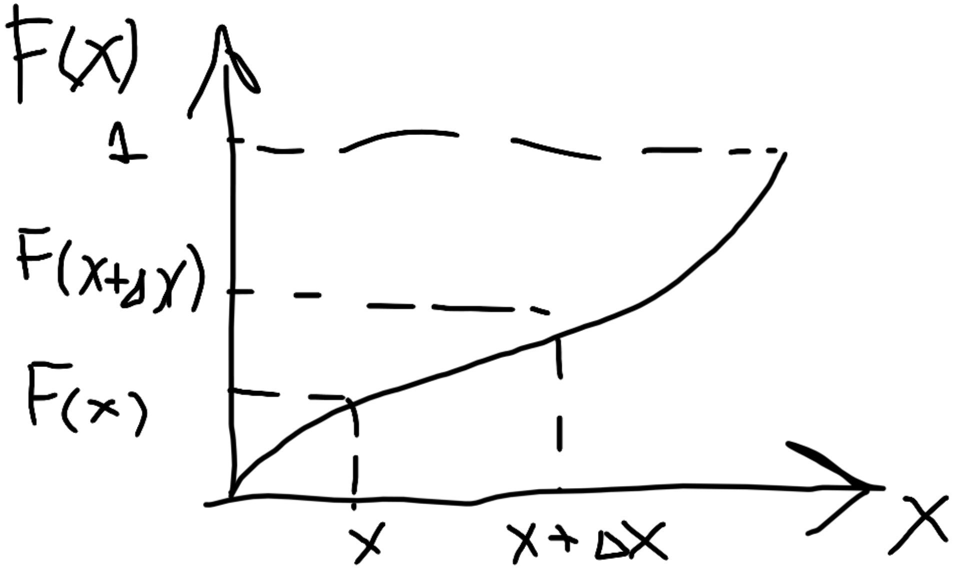
Непрерывная случайная величина

- {X = x} – событие с нулевой вероятностью, но при этом возможное!
- Будем рассматривать вероятность события {X < x}

Функция распределения: F(x) = P(X < x)

Свойства:

1. $0 \leq F(x) \leq 1$
2. $F(x_2) \geq F(x_1)$
3. $F(-\infty) = 0$
4. $F(+\infty) = 1$
5. $P\{a \leq X \leq b\} = F(b) - F(a)$



Непрерывная случайная величина

- **Случайную величину X** называют **непрерывной**, если ее функция распределения непрерывна в любой точке и дифференцируема всюду, кроме, может быть, отдельных точек

Непрерывная случайная величина

- **Плотность вероятности:** $\boldsymbol{f(x) = F'(X)}$

*Но более корректно определять немного наоборот (см. след слайд)

Плотность вероятности

$$\boldsymbol{F(x) = \int_{-\infty}^{x} f(x)dx},$$ при условии интегрируемости $f(x)$

Свойства:

1. $\forall\, x \in R^D : f(x) \geq 0$
2. $\int_{-\infty}^{+\infty} f(x)dx = 1$ (условие нормировки вероятностей)
3. $P(a \leq x \leq b) = \int_{a}^{b} f(x)dx = F(b) - F(a)$

Характеристики СВ

Мат.ожидание – среднее значение X

Для дискретной СВ: $\boldsymbol{E}(\boldsymbol{X}) = \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{n}} \boldsymbol{x_i p_i}$

Для непрерывной СВ: $\boldsymbol{E}(\boldsymbol{X}) = \int_{-\infty}^{+\infty} \boldsymbol{x f}(\boldsymbol{x}) \boldsymbol{d x}$

Аналог в статистике: выборочное среднее

Характеристики СВ

Свойства мат.ожидания:

1. $E(const) = const$
2. $E(const * X) = const * E(X)$
3. $E(X + Y) = E(X) + E(Y)$
4. $E\big(X - E(X)\big) = 0$
5. $E(XY) = E(X) * E(Y)$ – для независимых СВ

Мат. ожидание НЕ является СВ!

Характеристики СВ

Дисперсия – “мера разброса” X / мат. ожидание квадрата отклонения X от мат. ожидания X

$$\boldsymbol{D(X) = E[X - E(X)]^{\ 2}}$$

или

«Рабочая» формула: $\boldsymbol{D(X) = E\left(X^2\right) - [E(X)]^2}$

Аналог в статистике: выборочная дисперсия

$$D(X) = E\left[(X - E(X))^2\right]$$ − по определению

$$D(X) = E\left[X^2 - 2 \cdot X \cdot E(X) + [E(X)]^2\right] =$$

$$= E(X^2) - 2E(X) \cdot E(X) + [E(X)]^2 =$$

$$= E(X^2) - [E(X)]^2$$

$E(E(X)) = E(X)$ т.к $E(X)$ не СВ, а значит $E(X) = const$

Характеристики СВ

Свойства дисперсии:

1. $D(const) = 0$
2. $D(const * X) = const^2 * D(X)$
3. $D(X \pm Y) = D(X) + D(Y) -$ для независимых СВ
4. $D(X + const) = D(X)$
5. $D(XY) = E(X^2) * E(Y^2) \ - (E(X))^2 * (E(Y))^2$

– для независимых СВ

Среднеквадратичное отклонение (СКО): $\sigma = \sqrt{D(X)}$

СКО более “приятно” интерпретировать, чем дисперсию

Характеристики СВ

Квантиль уровня $\boldsymbol{\alpha}$ $(\boldsymbol{X_\alpha})$, $\alpha \in (0,1)$

$$X_\alpha : P(X \le X_\alpha) \ge \alpha, P(X \ge X_\alpha) \ge 1 - \alpha$$

Медиана – квантиль уровня 0.5

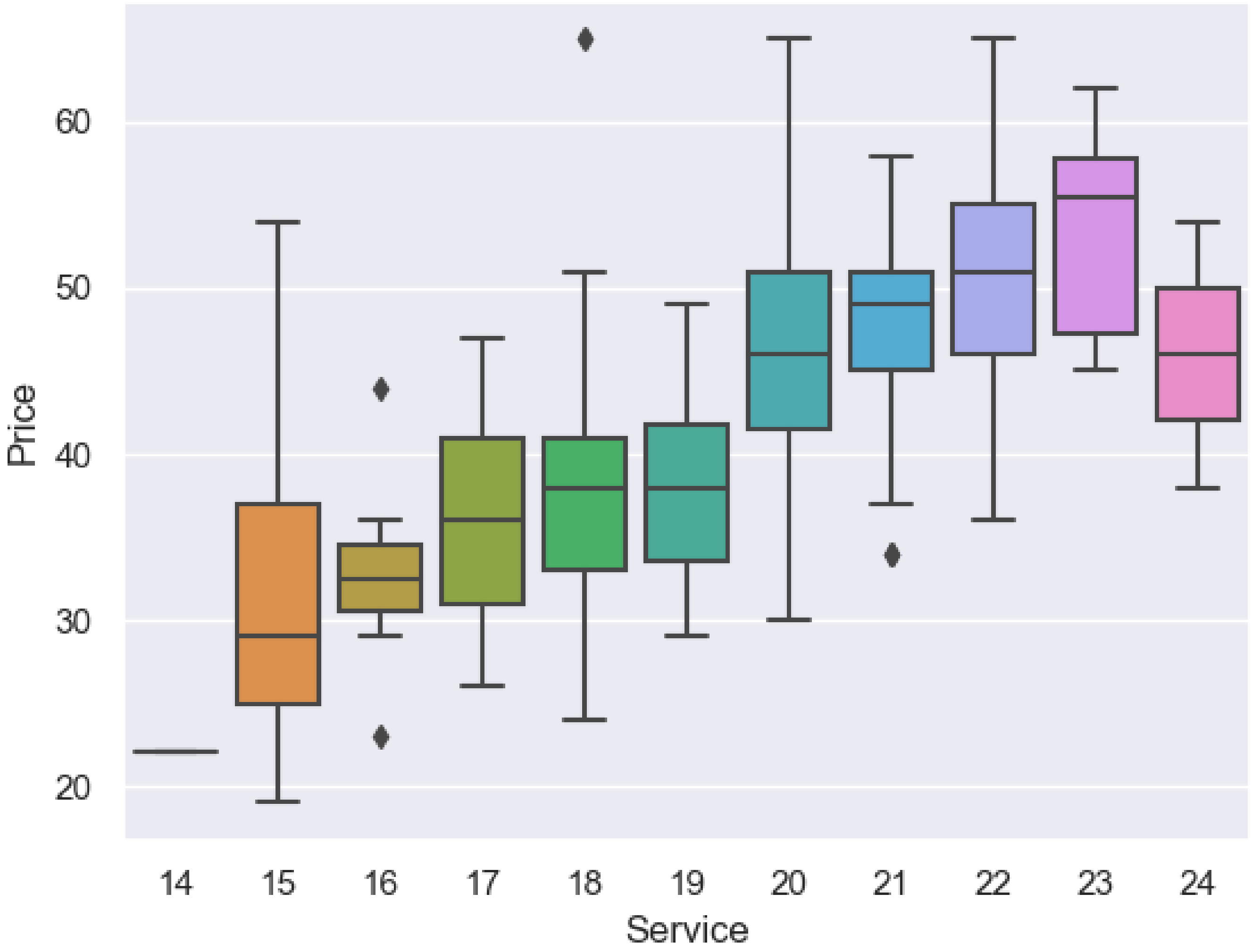
$$med\ X : P(X \le med\ X) \ge 0.5, P(X \ge med\ X) \ge 0.5$$

В статистике: медиана – это такое значение, больше которого 50% выборки, и меньше которого 50% выборки. Иначе говоря, центральный элемент или среднее двух центральных (если в выборке четное кол-во элементов)

Интерквартильный размах – разность между квантилем 0.75 и квантилем 0.25

$$IQR = X_{0.75} - X_{0.25}$$

Boxplot: интерпретация



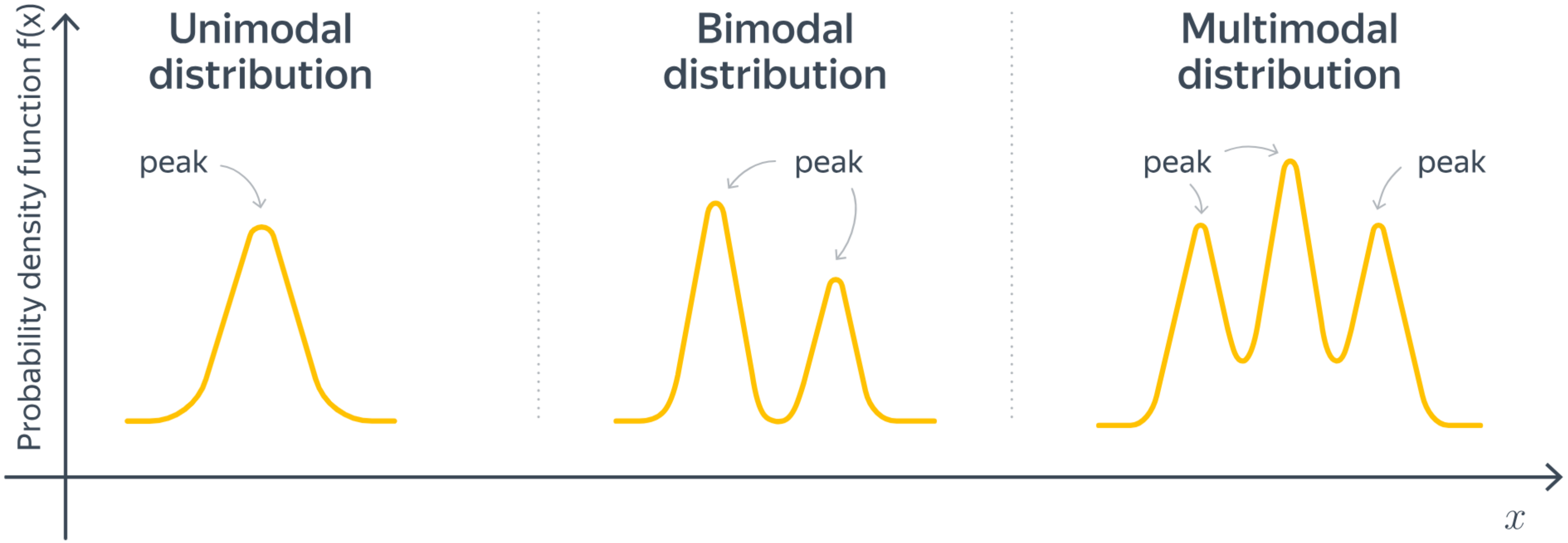
Характеристики СВ

Мода – точка максимума плотности вероятности

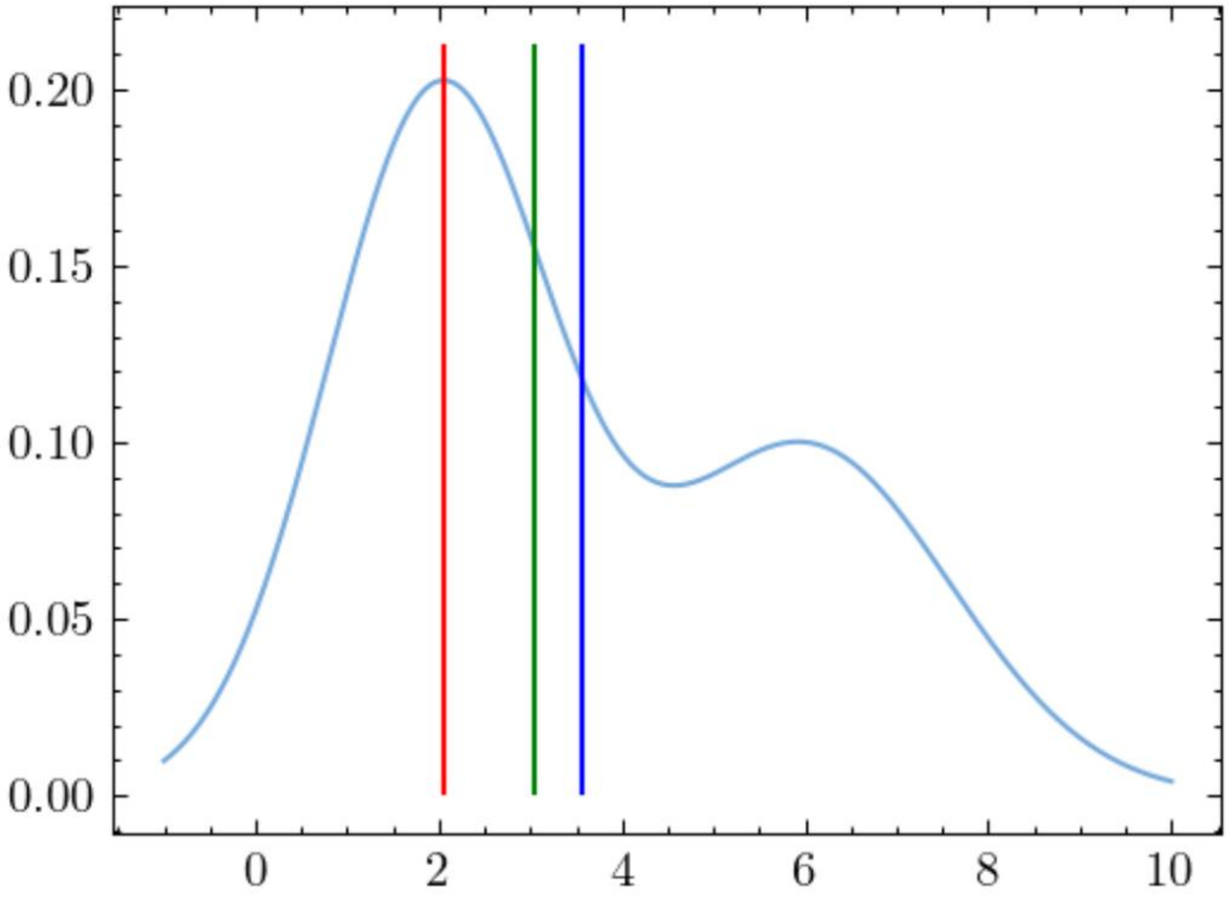
$$\text{mode}\, X = \underset{x}{\operatorname{argmax}}\, f(x)$$

В статистике: самое часто встречающееся значение в выборке

Характеристики СВ



Источник: https://education.yandex.ru/handbook/ml/article/veroyatnostnye-raspredeleniya

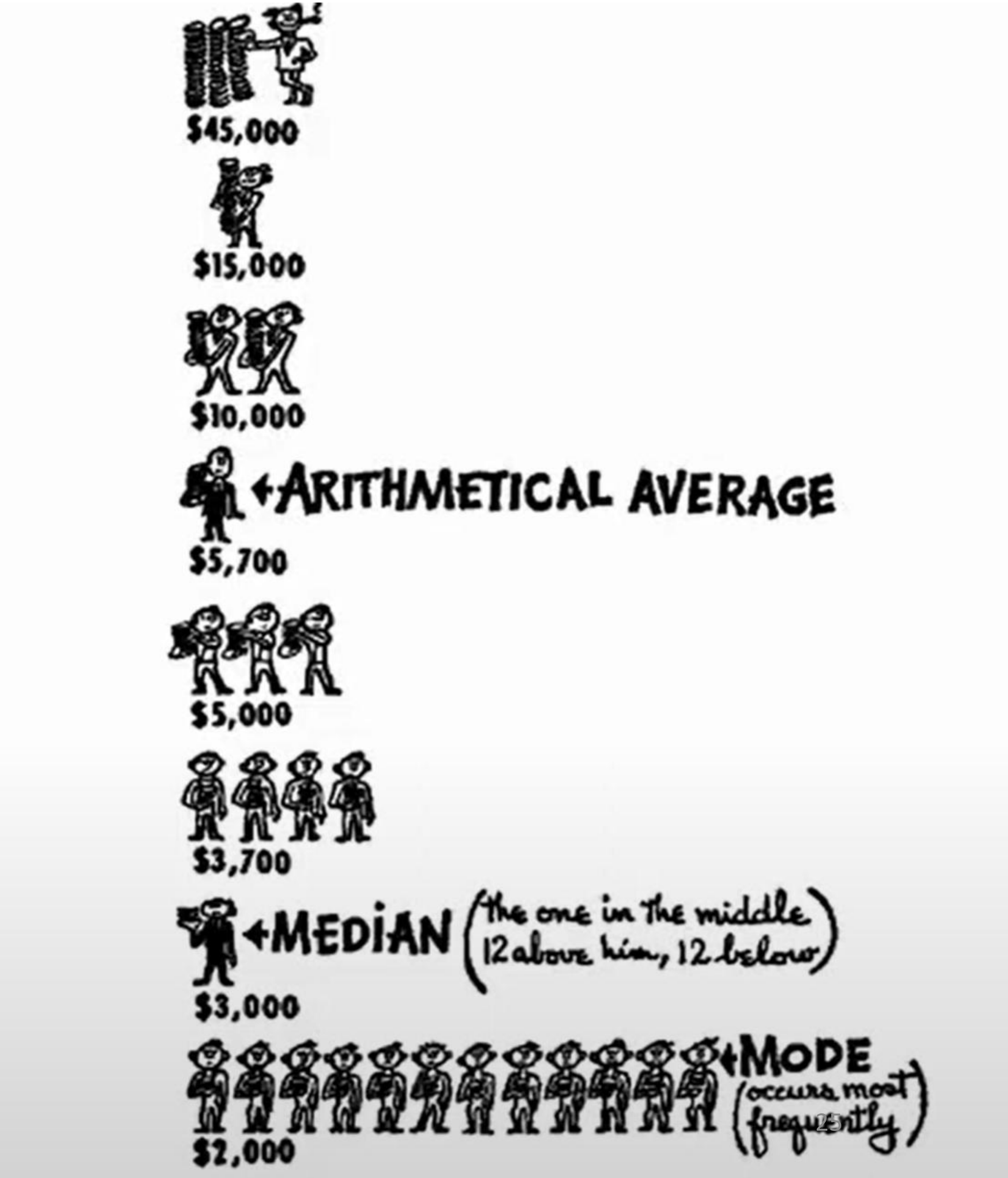


Среднее, медиана, мода?

Оценки центральной тенденции

Huff, 1954

Какова “средняя” зарплата населения?



Источник: *Дарелл Хафф.* Как лгать при помощи статистики = How to Lie with Statistics. — М.: Альпина Паблишер, 2015. — 163 с. — ISBN 978-5-9614-5212-9

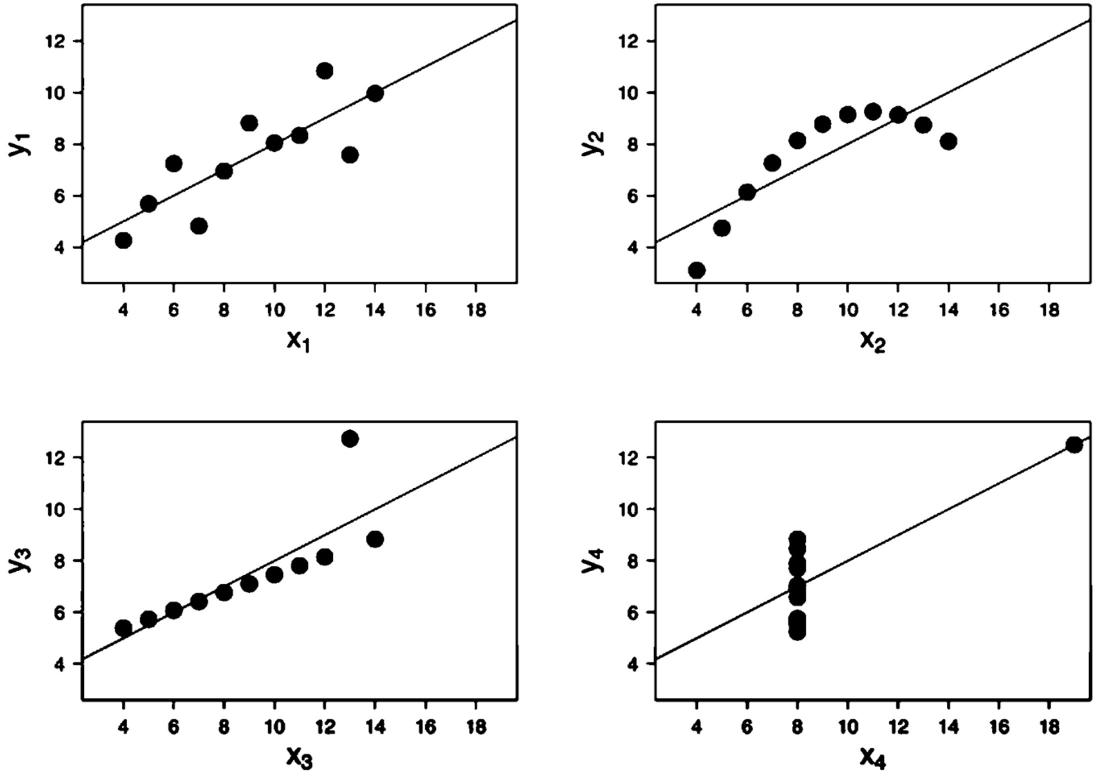
И еще об ограничениях характеристик

Есть четыре выборки с очень схожими статистиками:

№	1	2	3	4
$\bar{x}$	9	9	9	9
S_x	11	11	11	11
$\bar{y}$	7.5	7.5	7.5	7.5
S_y	4.127	4.127	4.128	4.128
r_{xy}	0.816	0.816	0.816	0.816

И еще об ограничениях характеристик

Есть четыре выборки с очень схожими статистиками:



Характеристики СВ

Моменты k-го порядка:

$$v_1 = E\left(X^k\right)$$

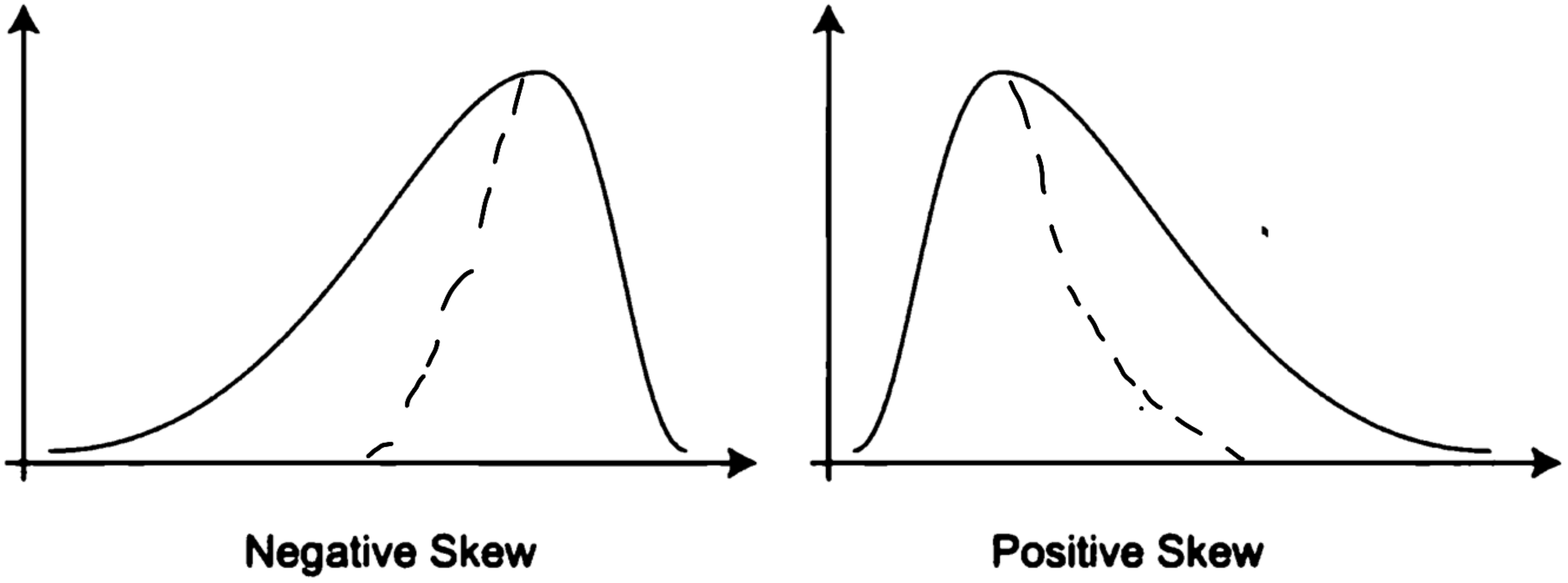
Центральные моменты k-го порядка:

$$m_k = E[\left(X - E(X)\right)^k]$$

Характеристики СВ

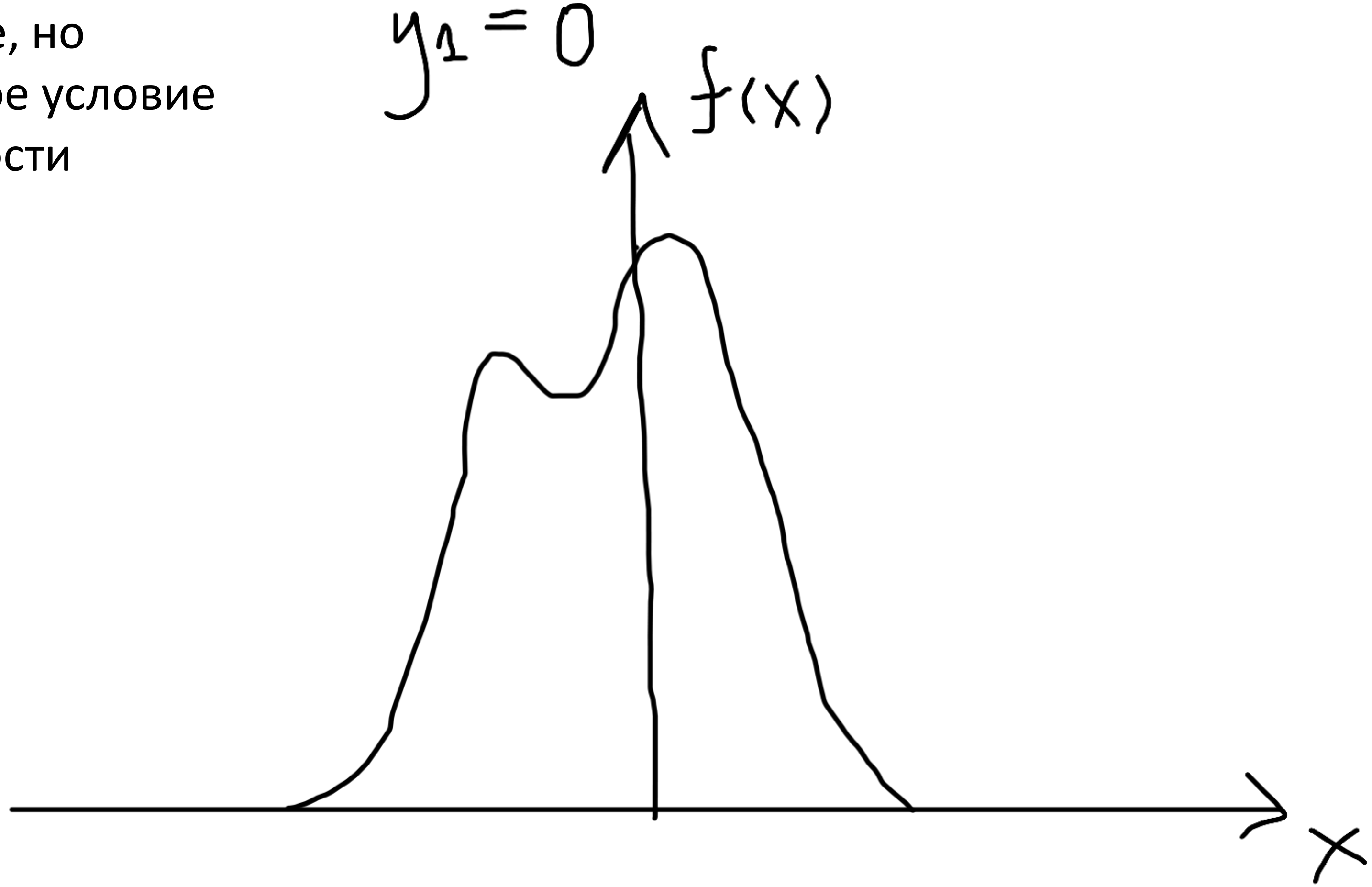
Коэффициент ассиметрии (skewness):

$$\gamma_1 = \mathbb{E}\left(\frac{X - \mathbb{E}X}{\sqrt{\mathbb{D}X}}\right)^3$$



Характеристики СВ

Необходимое, но недостаточное условие симметричности



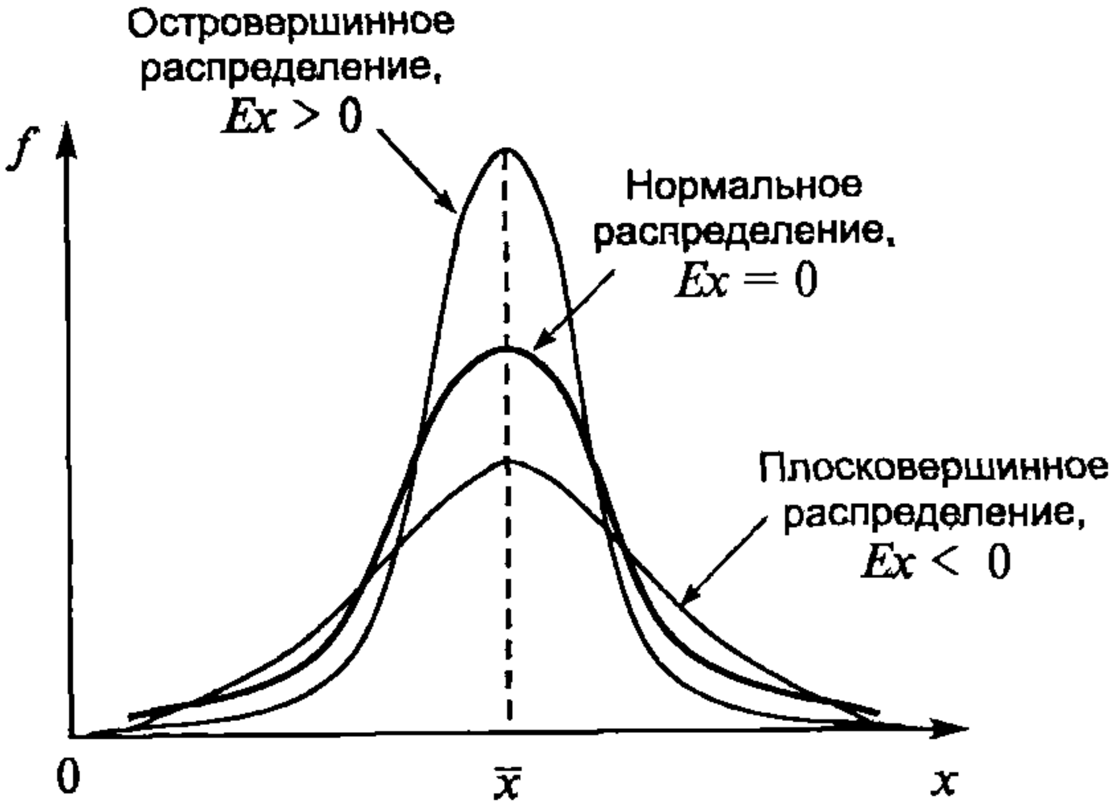
Характеристики СВ

Коэффициент эксцесса (excess):

Без вычитания тройки - kurtosis

$$\gamma_2 = \frac{E(X-E(X))^4}{(D(X))^2} - 3$$

K=4



Более общий термин для таких характеристик – моменты k-го порядка, либо центральные моменты k-го порядка

Предельные теоремы теории вероятностей

- **Закон больших чисел**: при большом количестве испытаний среднее арифметическое наблюдений стремится к математическому ожиданию (теоретическому среднему)

Или

- Среднее арифметическое этих СВ сходится по вероятности к среднему арифметическому их мат. ожидания (ЗБЧ в форме Чебышева, 1866)

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n \to \infty]{P} \frac{1}{n}\sum_{i=1}^{n} M(X_i)$$

Предельные теоремы теории вероятностей

- **Сходимость по вероятности**

Говорят, что последовательность случайных величин $X_1, X_2, \dots, X_n, \dots$ **сходится по вероятности** к числу a, если для любого сколь угодно малого положительного числа $\epsilon > 0$ выполняется условие:

$$\lim_{n\to\infty} P(| X_n - a | \geq \epsilon) = 0$$

Предельные теоремы теории вероятностей

- **Центральная предельная теорема:** при большом количестве испытаний распределение выборочного среднего становится близким к нормальному распределению, независимо от того, какое было исходное распределение данных

Или

- Пусть СВ $X_1, \ldots, X_n$ независимы, одинаково распределены, имеют конечное

мат. ожидание $M(X_i) = \alpha$ и дисперсию $D(X_i) = \sigma^2, i = \overline{1, n}$. Тогда функция распределения центрированной и нормированной суммы этих случайных величин стремится при $n \to \infty$ к функции распределения стандартной нормальной величины

Или

- При **достаточно большом n** сумма $S_n = X_1 + \cdots + X_n$ **приближенно**

распределена по нормальному закону: $S_n \sim N(na, \sqrt{n\sigma})$

* Распределение средних зачастую становится близко к нормальному уже даже при небольших n (n>30)

Пример задачи

Независимые СВ X_i распределены равномерно на отрезке [0,1].
Найти закон распределения Y и вероятность того, что 55< Y< 70

$$Y = \sum_{i=1}^{100} X_i$$

Подсказка: нужно смоделировать нормальное распределение с параметрами с пред. слайда, для ответа на второй вопрос воспользоваться формулой Ньютона-Лейбница и рассчитать нужные значения из функции Лапласа

Ограничения предельных теорем

Для закона больших чисел

- **Отсутствие конечного мат. ожидания:** если у распределения не существует конечного среднего значения, ЗБЧ не работает

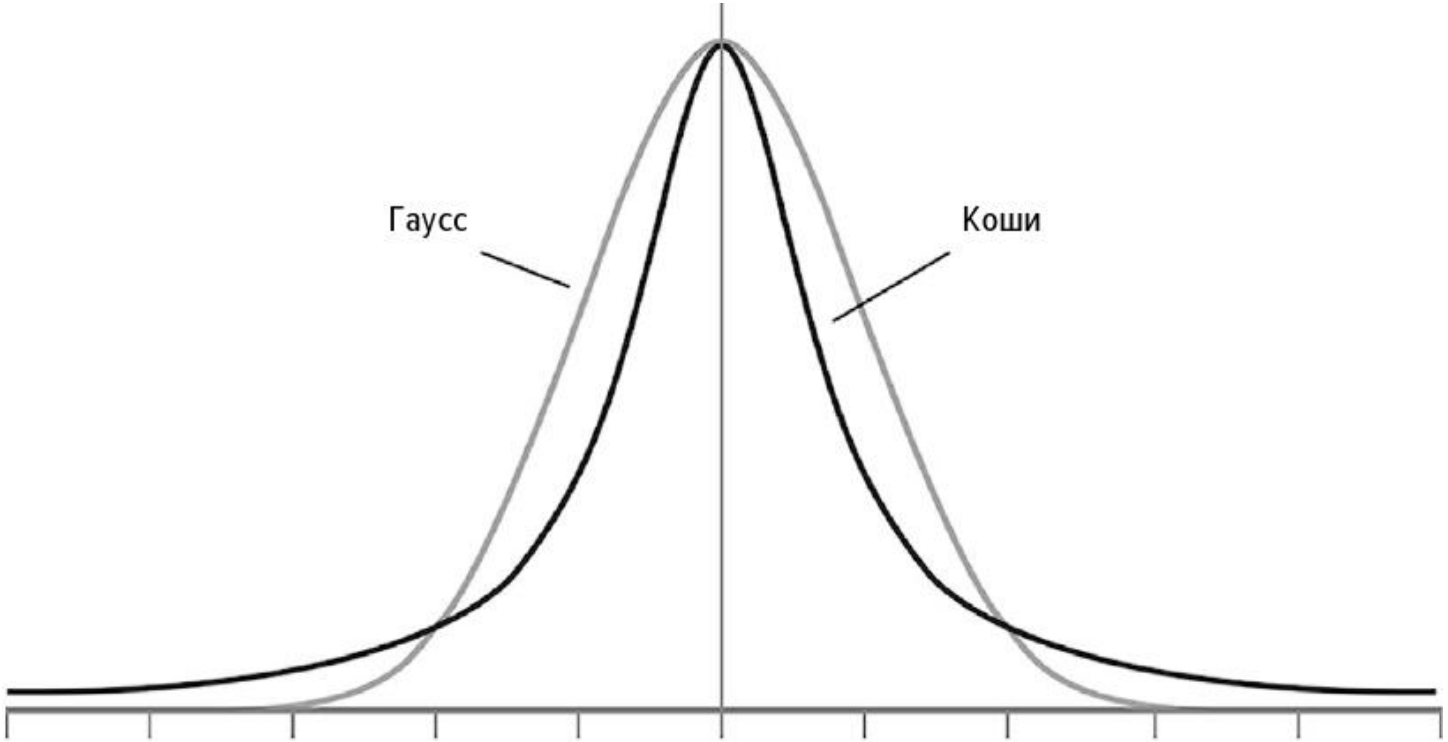
Пример: распределение Коши

$$f(x) = \frac{1}{\pi(1+x^2)}$$

$$E(X) = \int x f(x)\,dx = $$

$$= \int \frac{x}{\pi(1+x^2)}\,dx = +\infty$$

Ограничения предельных теорем



Из-за «тяжести хвостов» первый момент (мат. ожидание) не вычисляется

Ограничения предельных теорем

Для закона больших чисел

- **Зависимые наблюдения:** классический ЗБЧ требует независимости наблюдений

Пример: финансовые данные, временные ряды…

Ограничения предельных теорем

Для закона больших чисел

- **«Неидентичная распределенность»:** данные происходят из разных распределений

Пример: смешение разных групп потребителей для оценки спроса

Ограничения предельных теорем

Для закона больших чисел

- **«Тяжелые хвосты» распределений:** сходимость по вероятности будет достигаться очень медленно, на практике потребуются огромные выборки (хоть моменты и вычисляются)

Пример: распределение Парето с $\alpha < 2$

Ограничения предельных теорем

Для центральной предельной теоремы

- Быстрая сходимость для близких к нормальным/равномерным распределениям
- Намного медленнее для асимметричных распределений, распределений с «тяжелыми хвостами»
- Не выполняется в случае бесконечной дисперсии
- Не выполняется в случае сильной зависимости СВ
- Может плохо работать с «экстремальными» значениями: например, некоторыми квантилями, где α близко к 0 или 1

Основные распределения:

Лучше, чем в теормине
в хэндбуке от Яндекса
пока нигде не писали :)

А здесь конспект
студентов 2025/2026

https://education.yandex.ru/handbook/ml/article/veroyatnostnye-raspredeleniya

Дополнительно

- Конспект курса Б.Б. Демешева: https://hse.liferooter.dev/probability.pdf
- Д. Письменный, Конспект лекций по теории вероятностей, математической статистике и случайным процессам, 2004 (5 глава)
- Про ЦПТ и ЗБЧ подробнее и с док-вами: http://math-info.hse.ru/a/2013-14/ps-aa/statlecture6.pdf
- Про применение метода Монте-Карло для вычисления разных интегралов: https://habr.com/ru/articles/835870/

ВЫСШАЯ • ШКОЛА • ЭКОНОМИКИ •