

# Георгий Павлов

Преподаватель Ассоциации репетиторов и Профи.ру

Основы описательной статистики,  
виды распределений в Python

Спикер

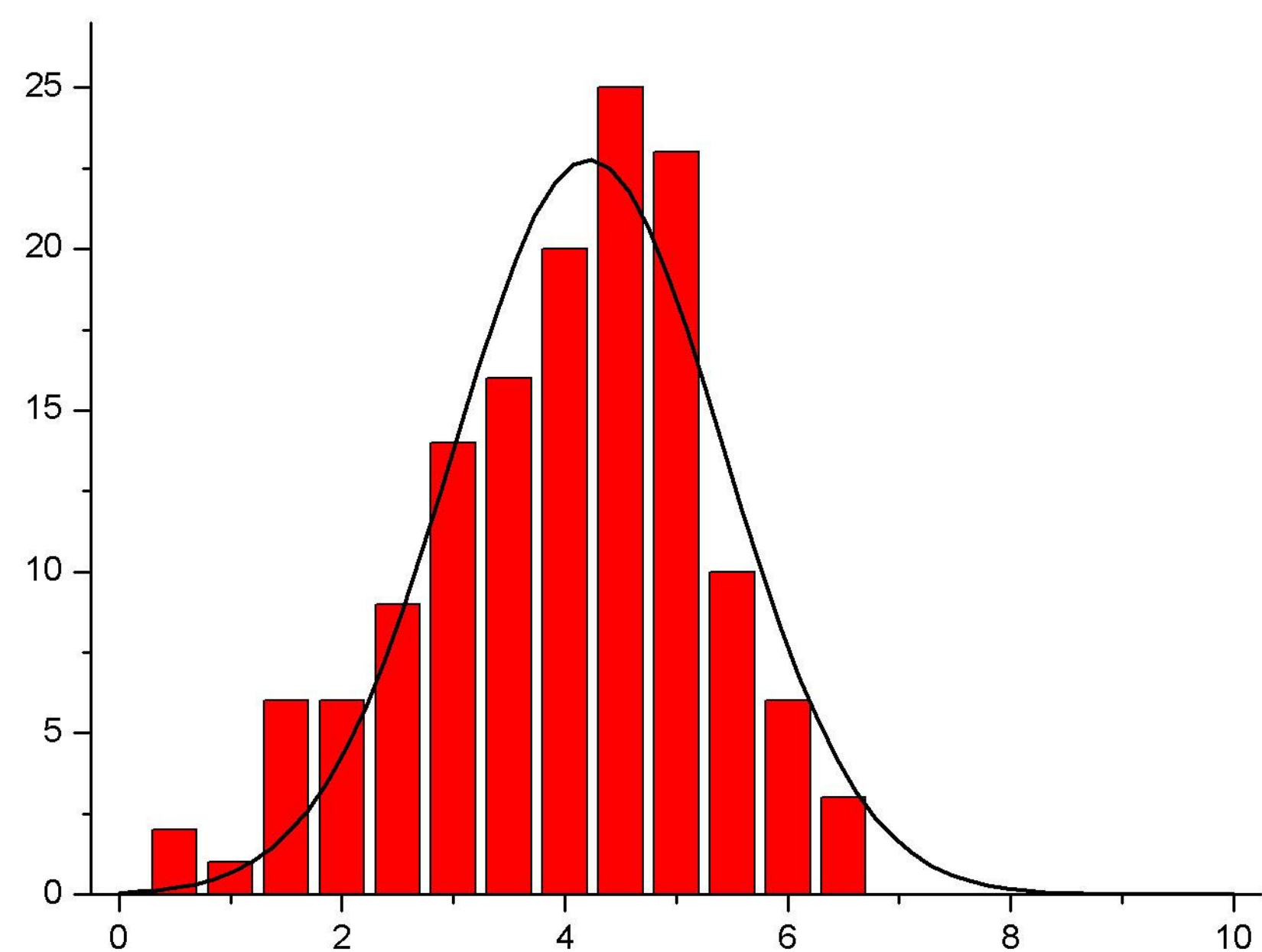


**Георгий Павлов,**  
Преподаватель Ассоциации репетиторов  
и Профи.ру, фрилансер, рисквик.

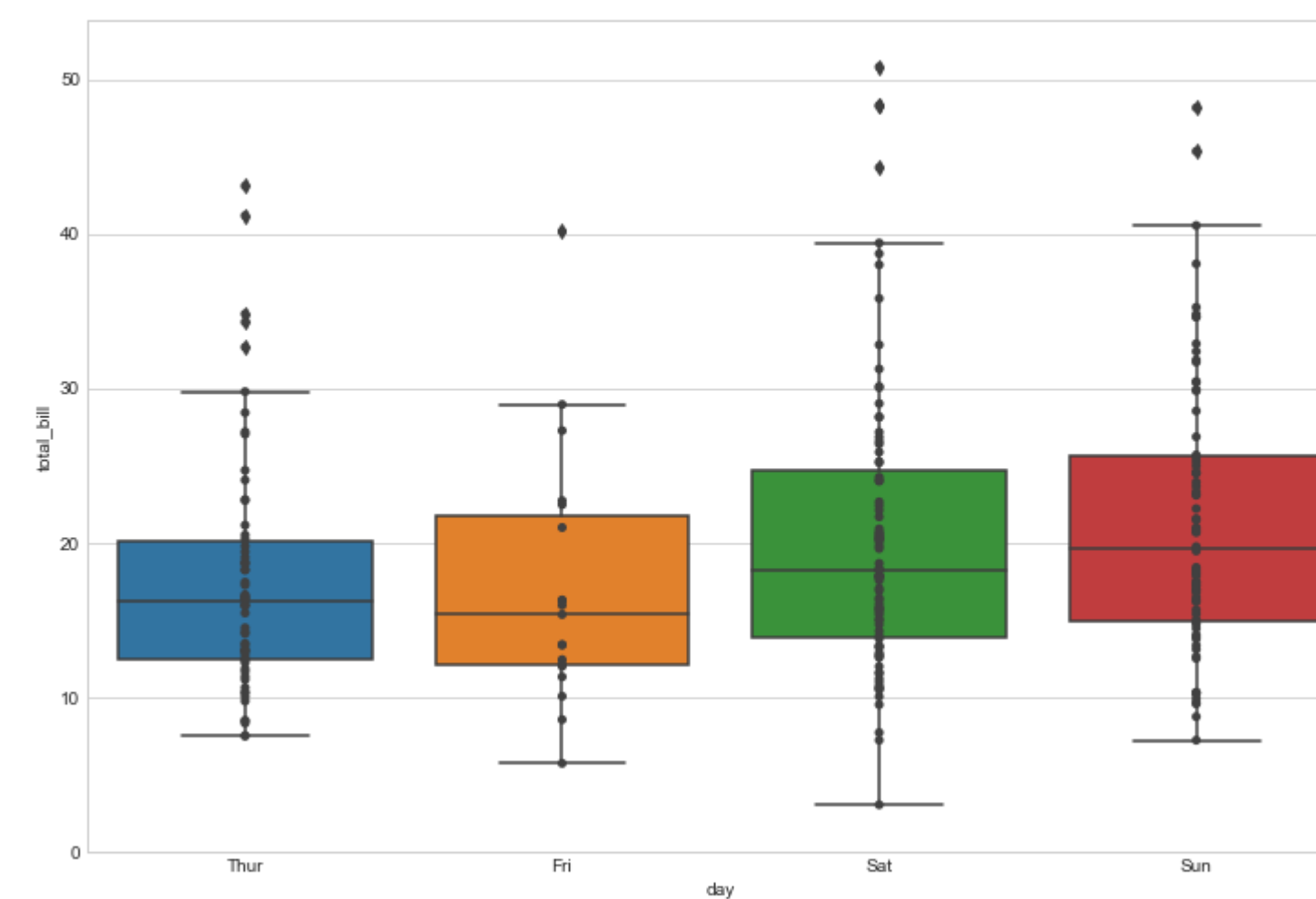
# Темы урока

1. Презентация данных
2. Что такое набор данных или выборка
3. Основные характеристики набора данных
4. Типы распределений

# Презентация данных



Гистограмма

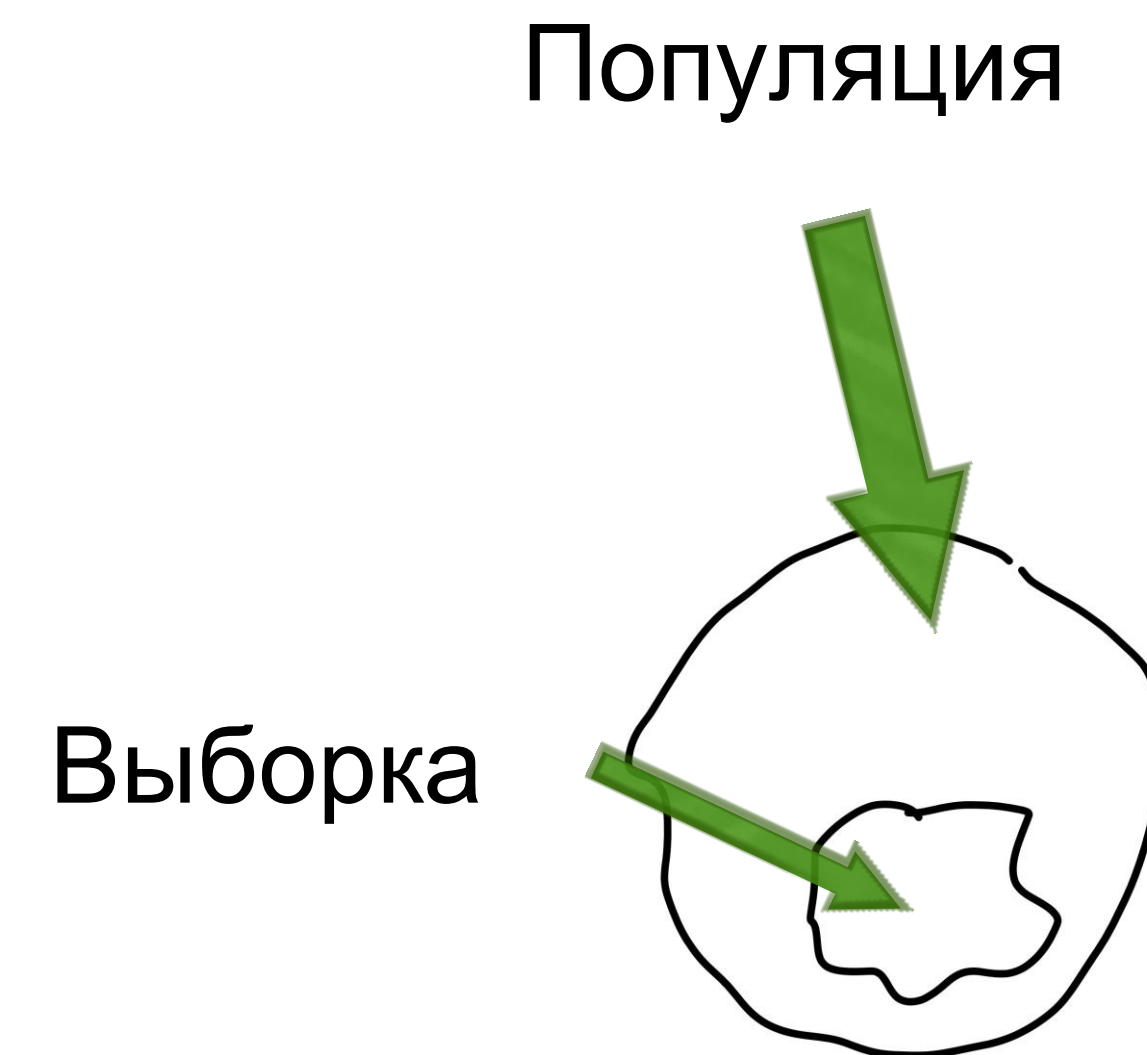


Бокс плот

# Что такое набор данных или выборка

**Популяция - набор всех возможных наблюдений конкретного события**

**Выборка- набор некоторых наблюдений конкретного события**



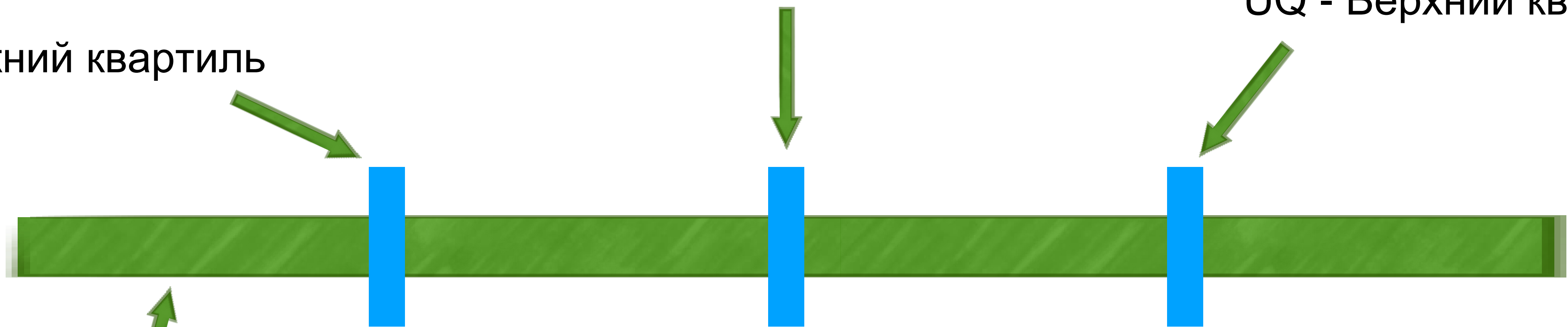
# Основные характеристики данных

М - Медиана (Центральный квартиль)

UQ - Верхний квартиль

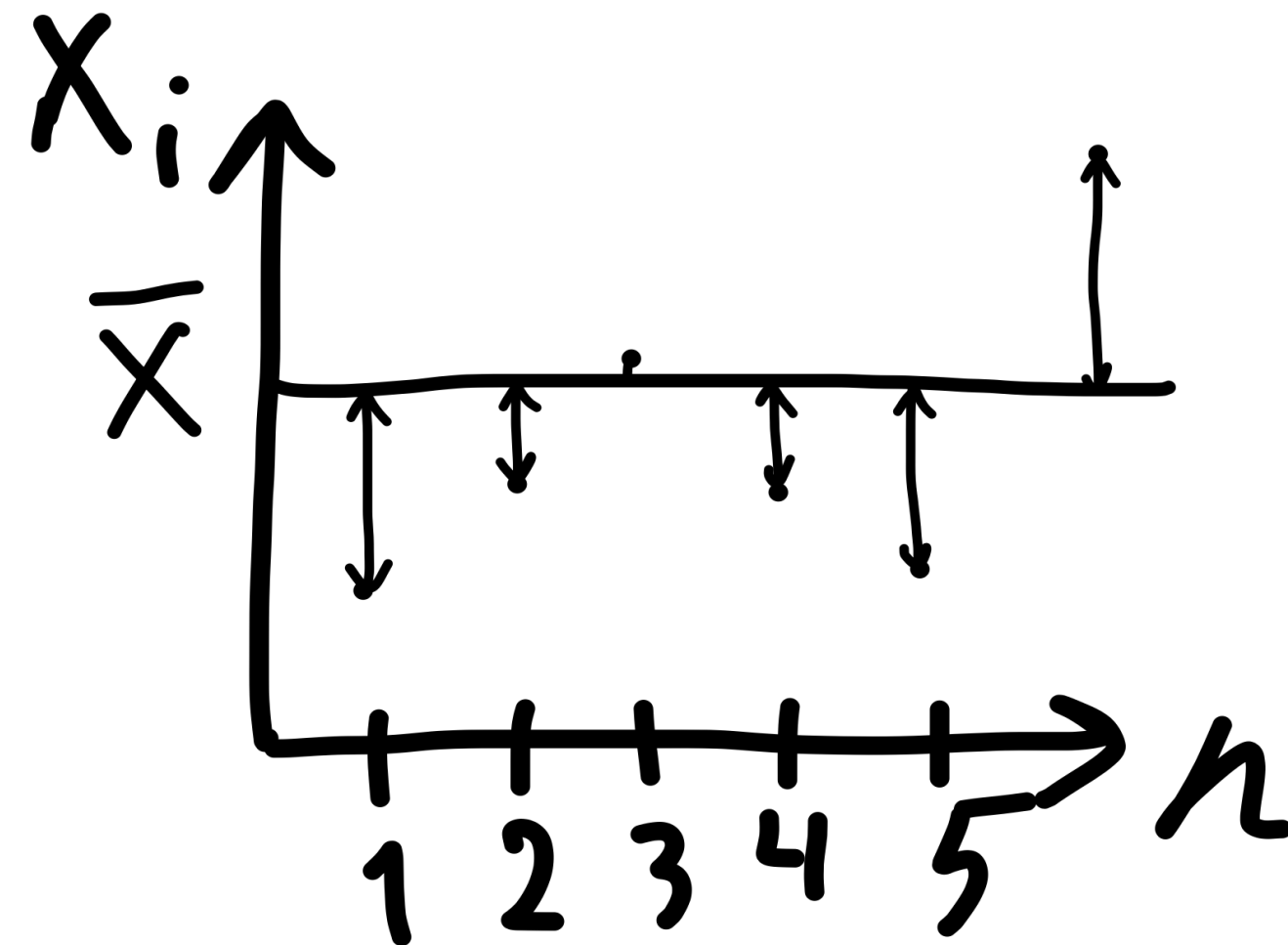
LQ - Нижний квартиль

Отсортированная выборка





# Основные характеристики данных



Стандартная ошибка показывает средний разброс значений выборки относительно Среднего значения выборки.

Расстояние от линии среднего до точки (наблюдения)-это отклонение

Что бы отрицательные и положительные отклонения не ушли в ноль- возводим в квадрат

Суммируем все отклонения и делим на количество- получаем среднее отклонение

# Основные характеристики данных

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$$

Формула среднего

$$V(x) = \frac{1}{n} \sum (x_i - \bar{x})^2$$

Формула дисперсии

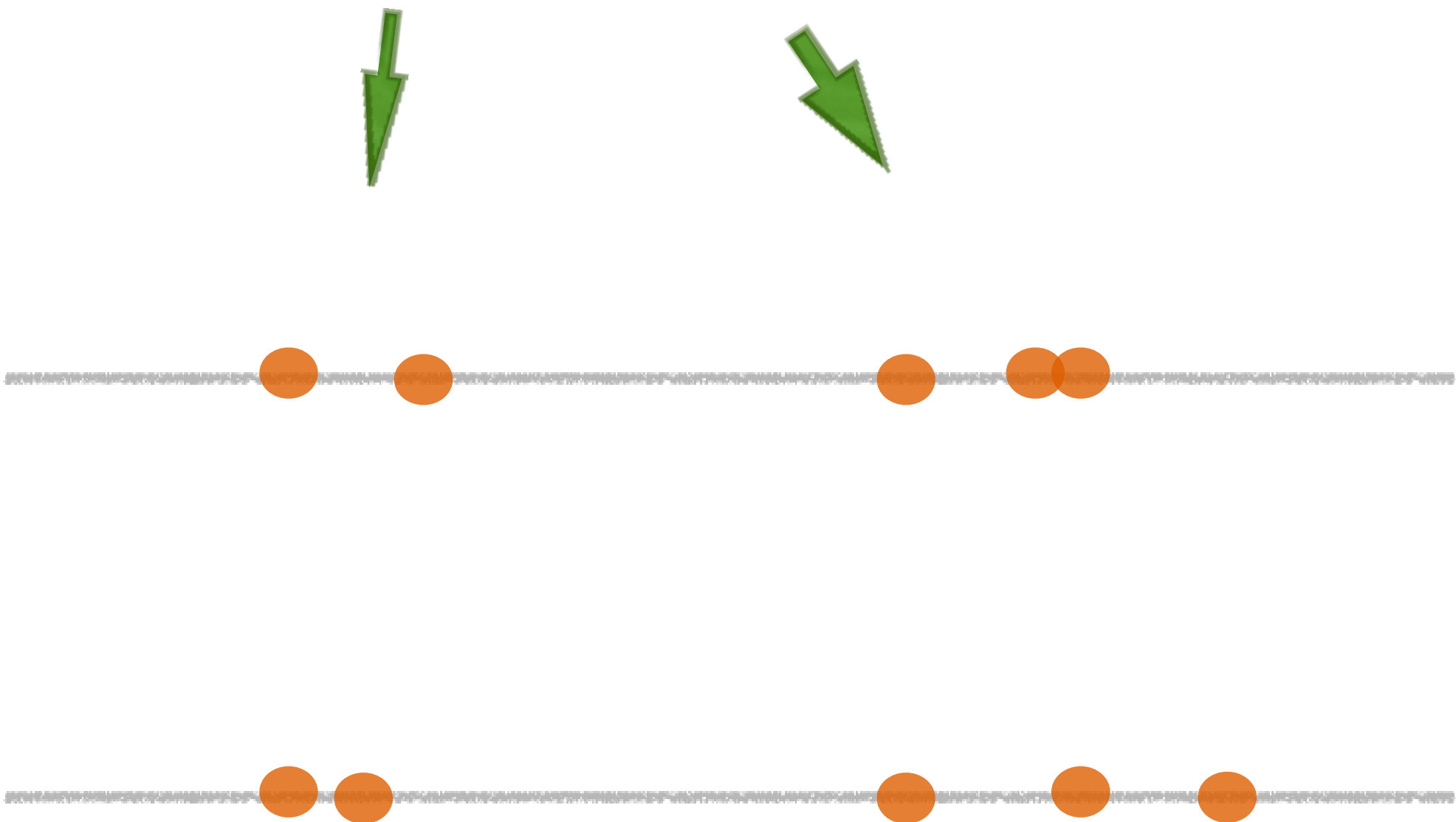
$$\sigma_x = \sqrt{V(x)}$$

Формула стандартного отклонения



# Основные характеристики данных

Скопления точек



|                      |                 |                    |
|----------------------|-----------------|--------------------|
| Если все наблюдения: | Увеличить на 10 | Увеличить в 10 раз |
| Среднее              |                 |                    |
| Ошибка               |                 |                    |

# Основные характеристики данных

Понятие эксцеса

$$E_k = \frac{\mu_4}{\sigma^4} - 3 = \frac{\sum (x_i - \bar{x})^4 f_i}{\sigma^4 \sum f_i} - 3$$

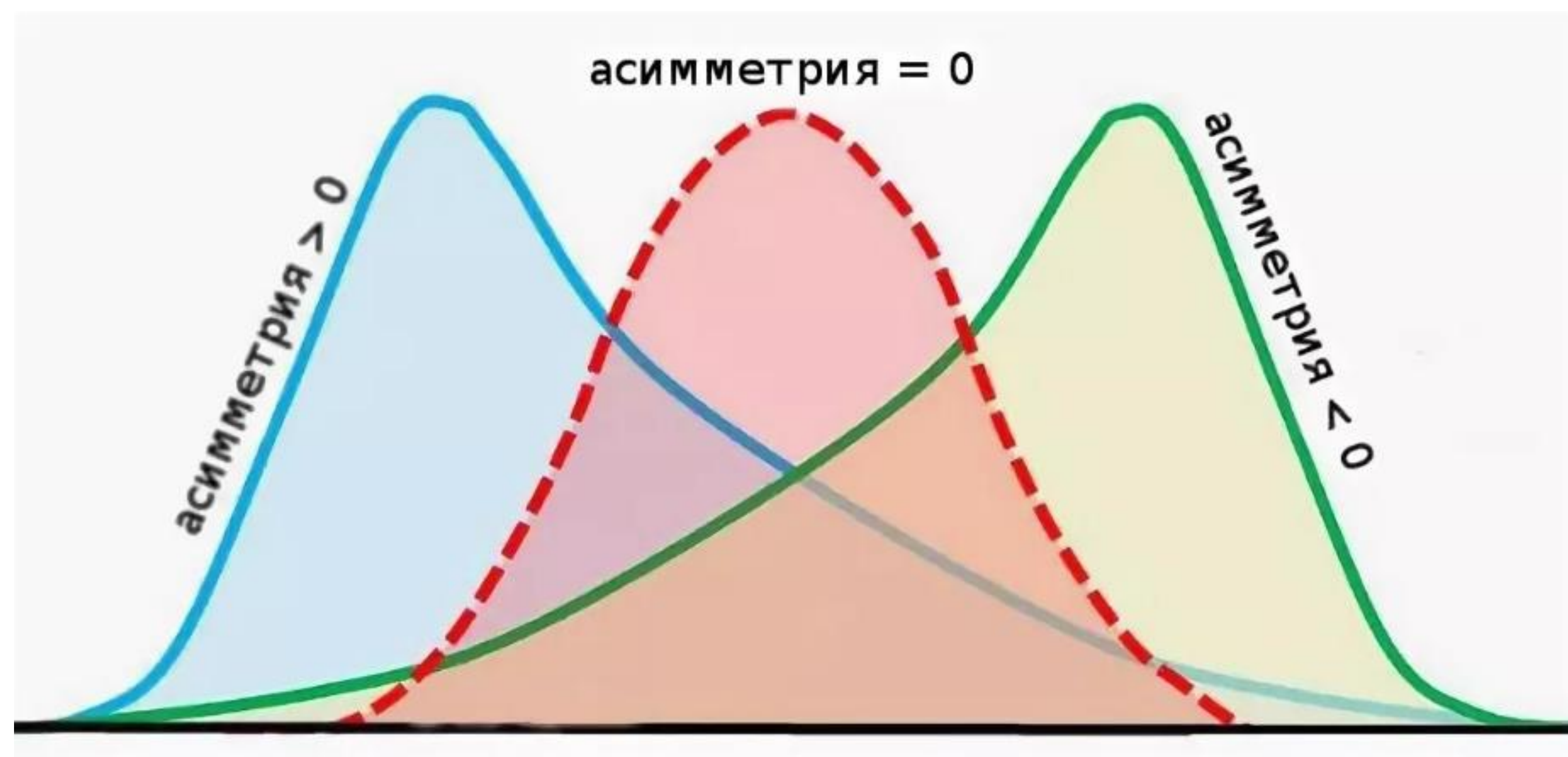


# Основные характеристики данных

Понятие ассиметрии (скошености)

Формула: 
$$\frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2\right)^{3/2}}$$

$$As = \frac{\bar{x} - Mo}{\sigma}$$

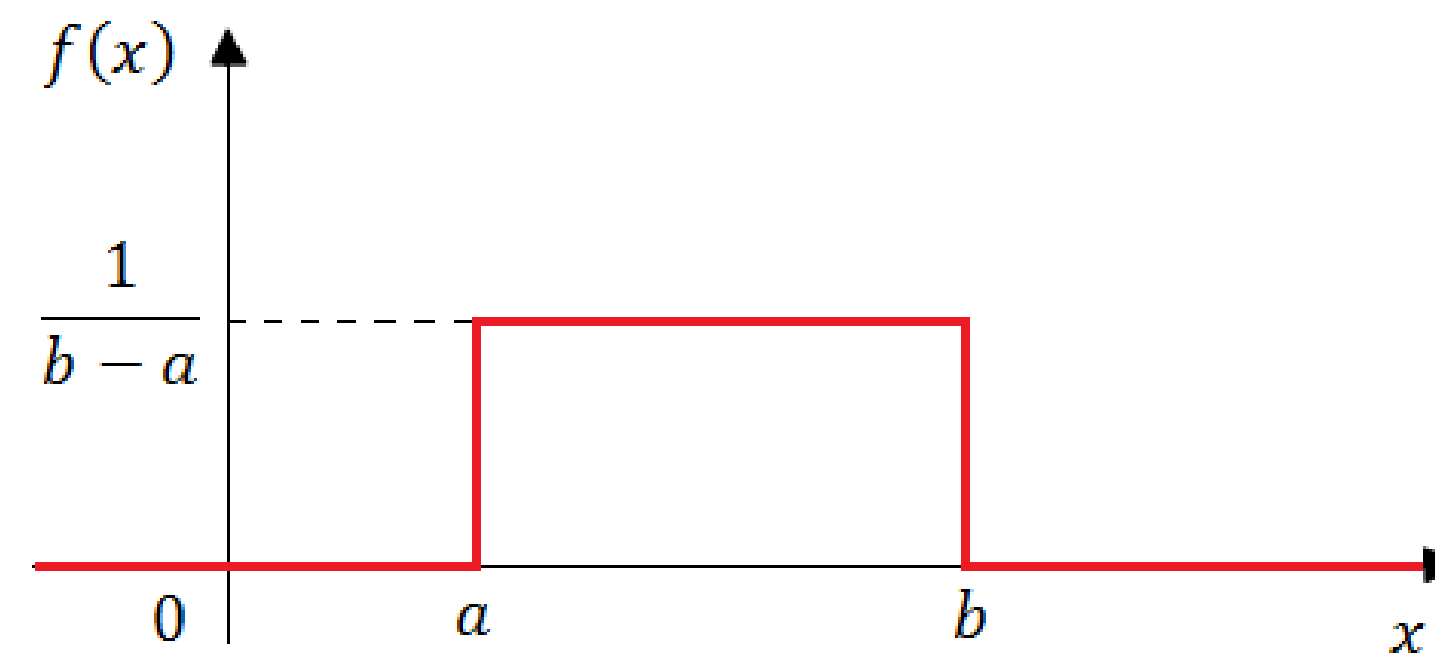


# Равномерное распределение

Функция плотности:  $\frac{1}{b-a}$

Математическое ожидание:  $\frac{a+b}{2}$

Дисперсия:  $\frac{(b-a)^2}{12}$



# Биномиальное распределение

Функция плотности:

$$f(k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

Математическое ожидание:  $N \cdot p$

Дисперсия:  $N \cdot p \cdot (1-p)$



# Нормальное распределение

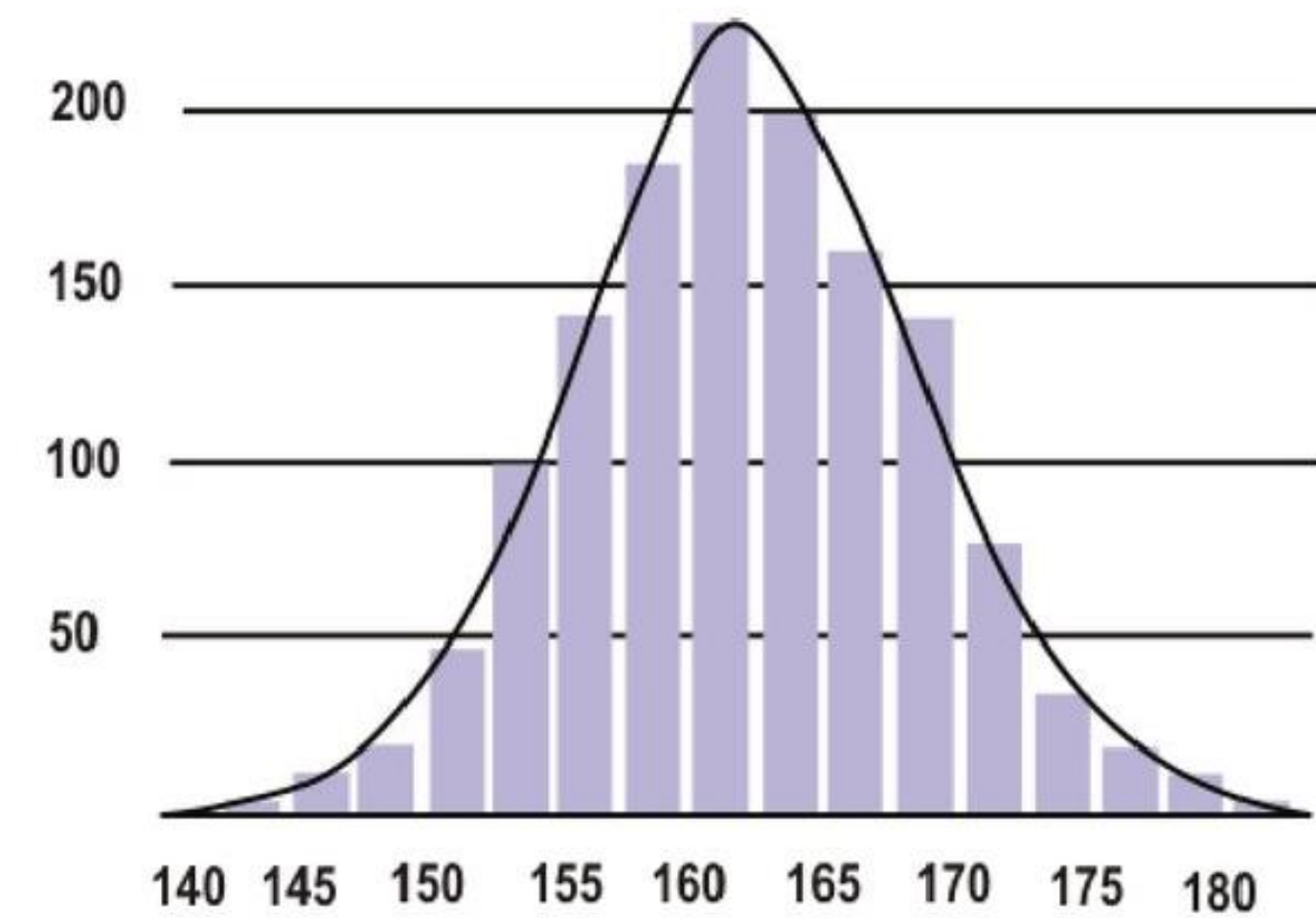
Функция плотности:  $f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{(x - a)^2}{2\sigma^2} \right\}$

Математическое ожидание:

$\mu$

Дисперсия:

$\sigma^2$

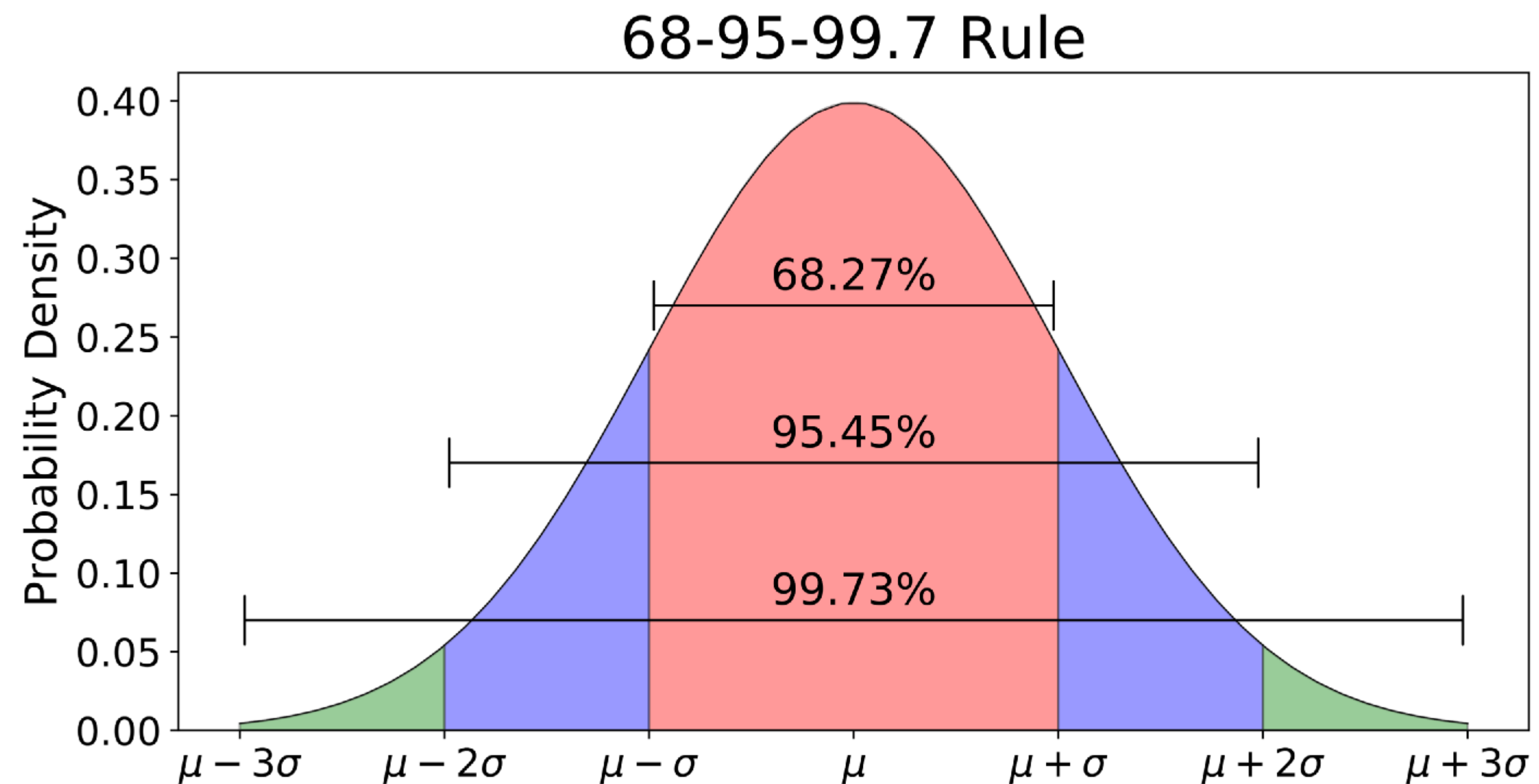


Распределение роста



# Нормальное распределение

Правило трех сигм:

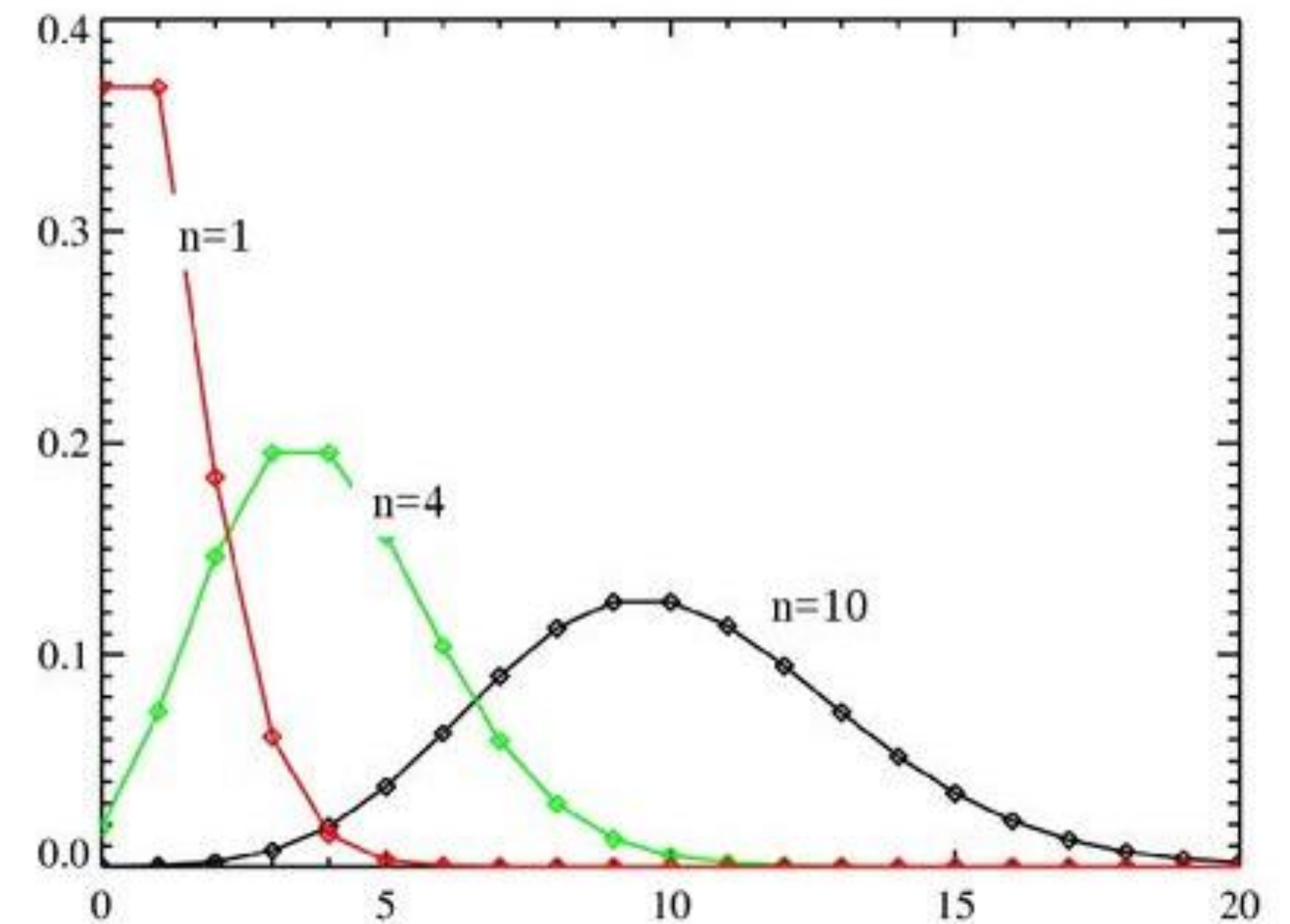


# Распределение Пуассона

Функция плотности:  $p(x) = \frac{x^n}{n!} \exp(-x)$   $0 < x < \infty$ ,  $n$  – целочисленный параметр  
( $n=0, 1, 2, \dots$ )

Математическое ожидание:  $\lambda$

Дисперсия:  $\lambda$

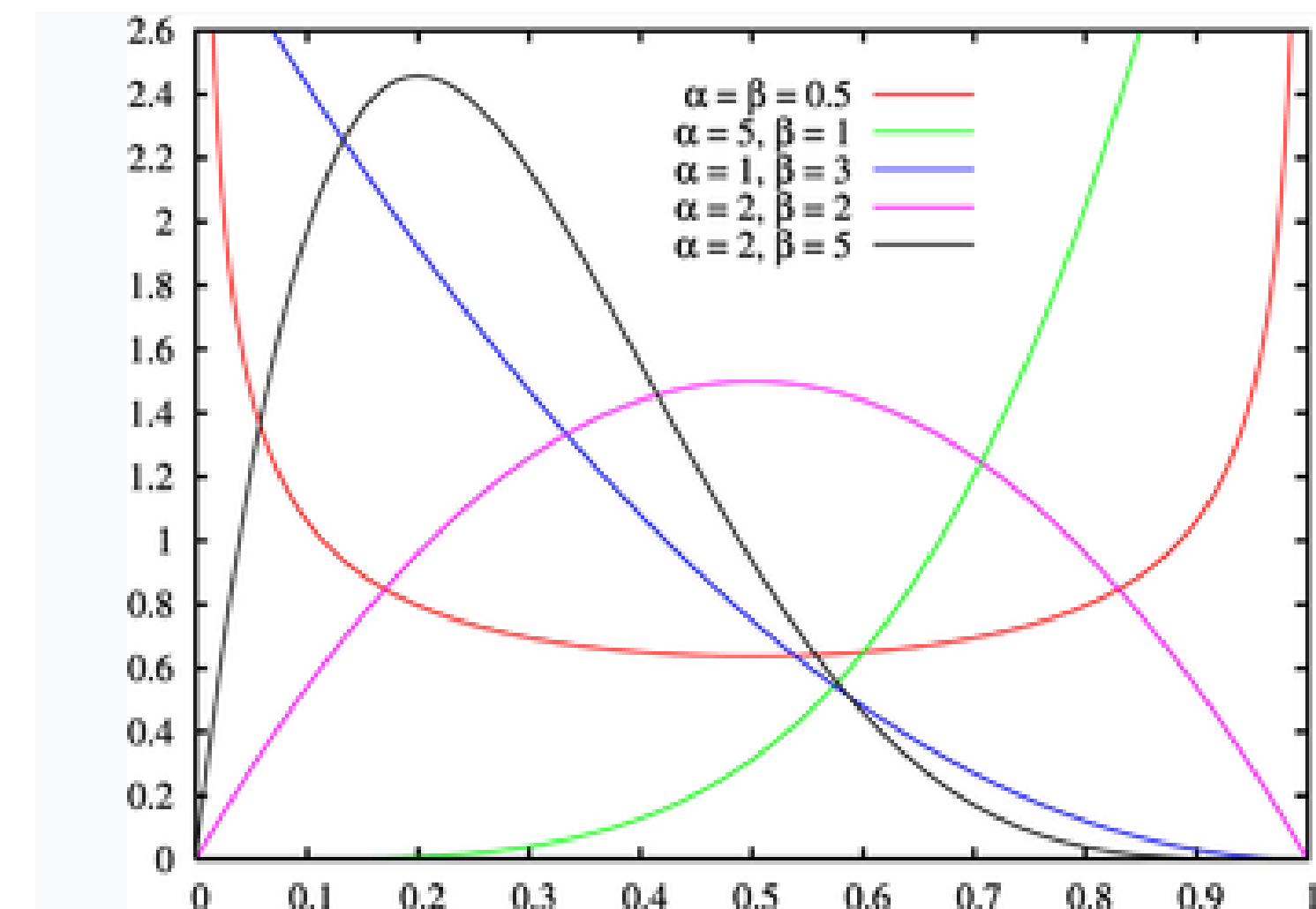


# Бета распределение

Функция плотности:  $\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}$

Математическое ожидание:  $\frac{\alpha}{\alpha + \beta}$

Дисперсия:  $\frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$



# Спасибо за внимание!

**Георгий Павлов,**

Преподаватель Ассоциации репетиторов и Профи.ру

