

# Лабораторная работа №2

## 1.1. Оценки математического ожидания, дисперсии, медианы

**Задача 1. Формулы для мат. ожидания и дисперсии, моделирование.** Пусть случайная величина  $\xi$  имеет распределение с плотностью

$$f_{\xi}(x) = \theta^2 x e^{-\theta x}, \quad x \geq 0, \theta > 0.$$

По условию рассматриваются значения  $\theta \in \{0.5, 2, 8\}$ .

### (а) Аналитические значения $\mathbb{E}\xi$ , $D(\xi)$ , $\mathbb{E}\xi^2$

Заметим, что

$$f_{\xi}(x) = \theta^2 x e^{-\theta x}, \quad x \geq 0$$

— это гамма-распределение с параметрами формы  $k = 2$  и интенсивностью  $\theta$ :

$$\xi \sim \Gamma(k = 2, \text{rate} = \theta).$$

Для гамма-распределения  $\Gamma(k, \theta)$  (где  $\theta$  — параметр интенсивности) известны формулы:

$$\mathbb{E}\xi = \frac{k}{\theta}, \quad D(\xi) = \frac{k}{\theta^2}, \quad \mathbb{E}\xi^2 = \frac{k(k+1)}{\theta^2}.$$

Подставляя  $k = 2$ , получаем:

$$\mathbb{E}\xi = \frac{2}{\theta}, \quad D(\xi) = \frac{2}{\theta^2}, \quad \mathbb{E}\xi^2 = \frac{2 \cdot 3}{\theta^2} = \frac{6}{\theta^2}.$$

Для конкретных значений параметра:

$\theta$	$\mathbb{E}\xi$	$D(\xi)$	$\mathbb{E}\xi^2$
0.5	$\frac{2}{0.5} = 4$	$\frac{2}{0.5^2} = 8$	$\frac{6}{0.5^2} = 24$
2	$\frac{2}{2} = 1$	$\frac{2}{2^2} = \frac{1}{2}$	$\frac{6}{2^2} = \frac{3}{2}$
8	$\frac{2}{8} = \frac{1}{4}$	$\frac{2}{8^2} = \frac{1}{32}$	$\frac{6}{8^2} = \frac{3}{32}$

### (b) Моделирование выборок и построение графиков

Пусть для заданного  $\theta$  строится выборка

$$X_1, \dots, X_k \sim \xi \quad (\text{независимые и одинаково распределённые св}), \quad k \in \{2^4, 2^5, \dots, 2^{15}\}.$$

Для каждой выборки размера  $k$ :

- Оценка математического ожидания

$$\hat{\mu}_k = \frac{1}{k} \sum_{i=1}^k X_i.$$

- Оценка дисперсии (классическая несмещённая)

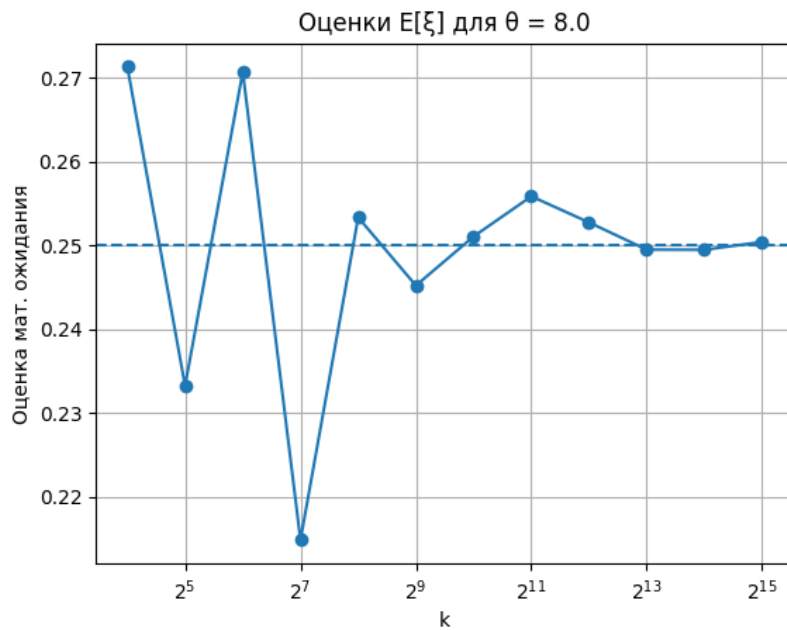
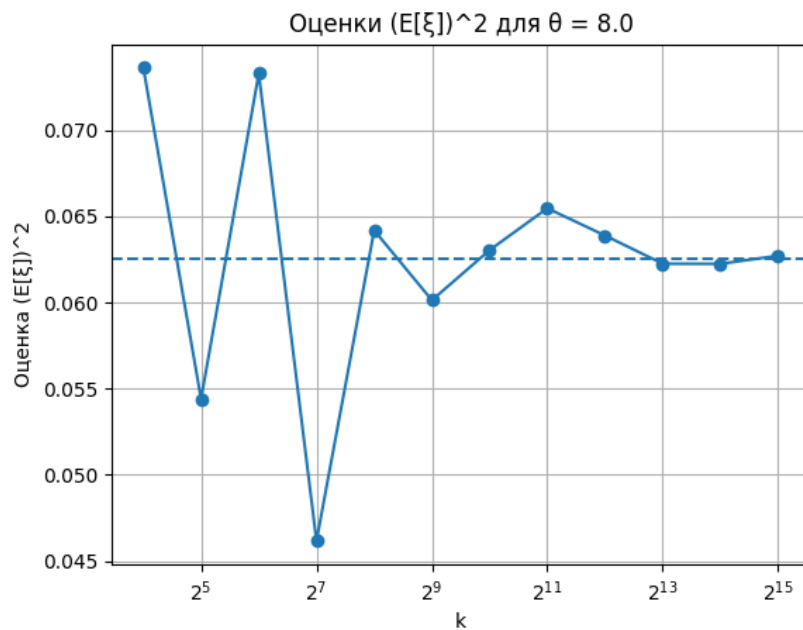
$$\hat{\sigma}_k^2 = \frac{1}{k-1} \sum_{i=1}^k (X_i - \hat{\mu}_k)^2.$$

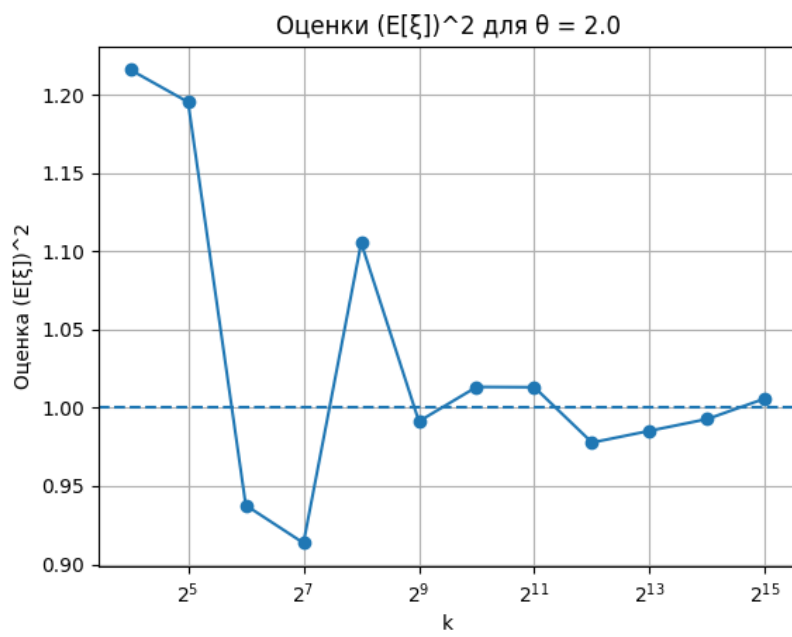
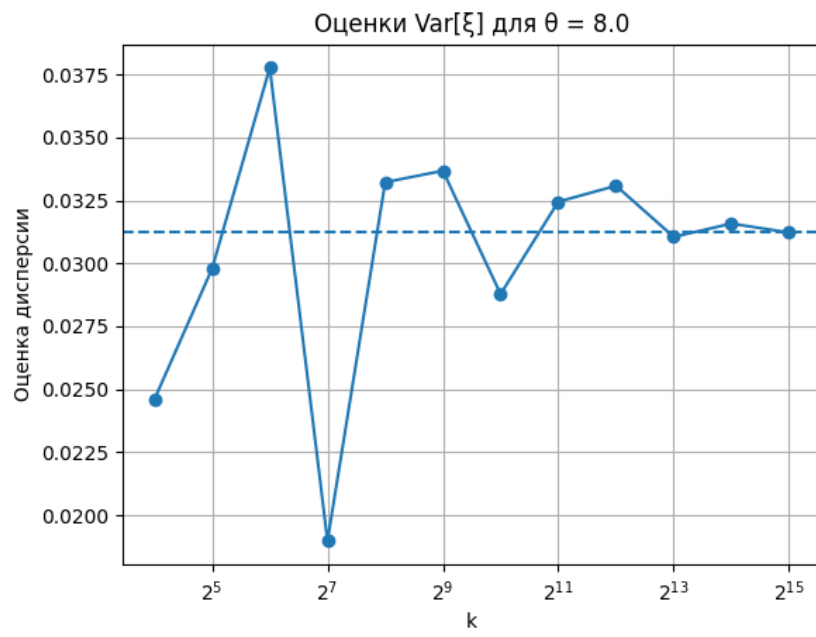
- Оценка квадрата математического ожидания

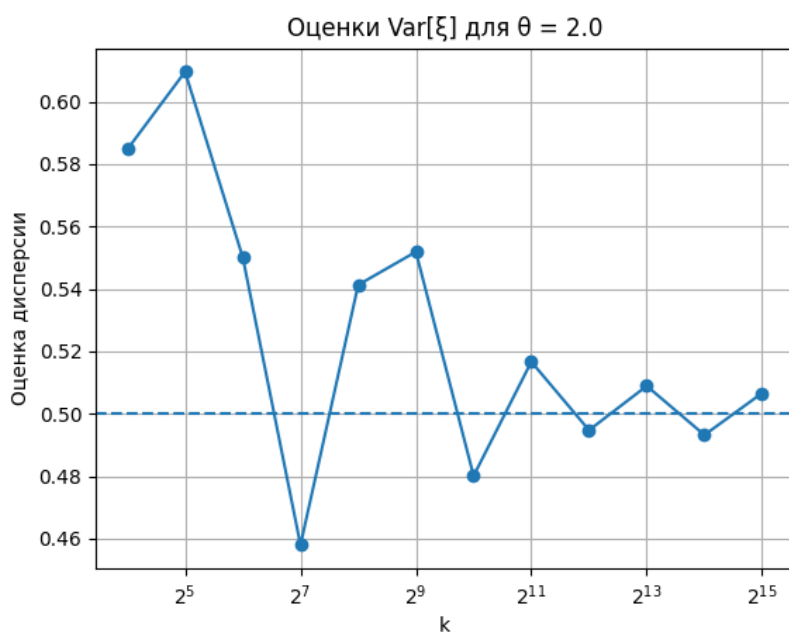
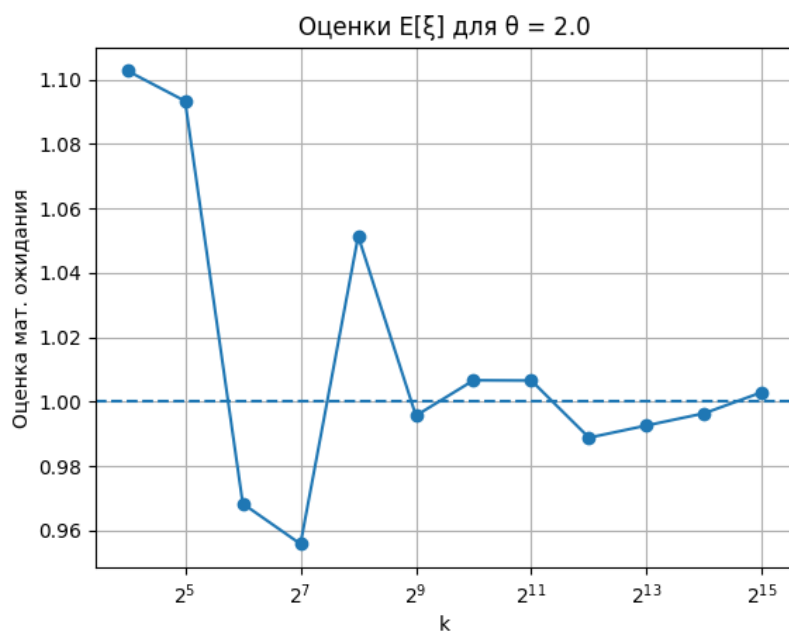
$$\widehat{(\mathbb{E}\xi)^2}_k = (\hat{\mu}_k)^2.$$

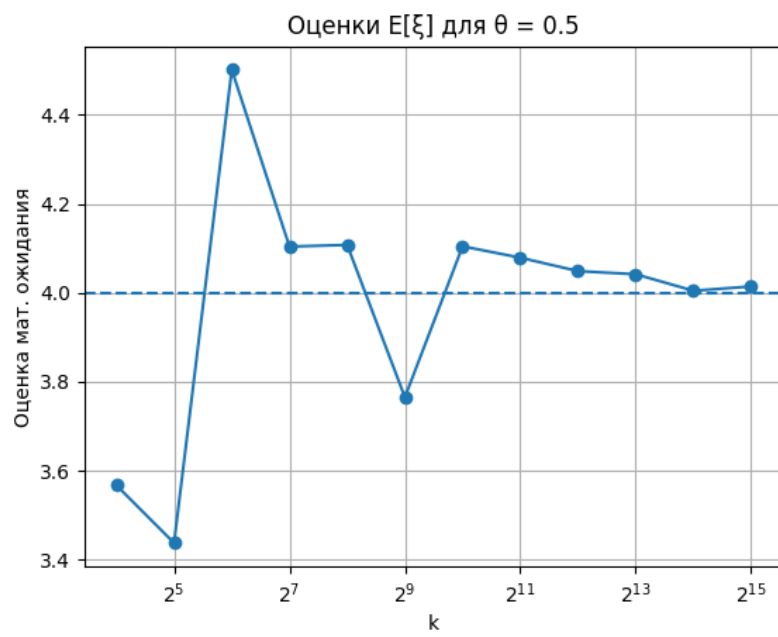
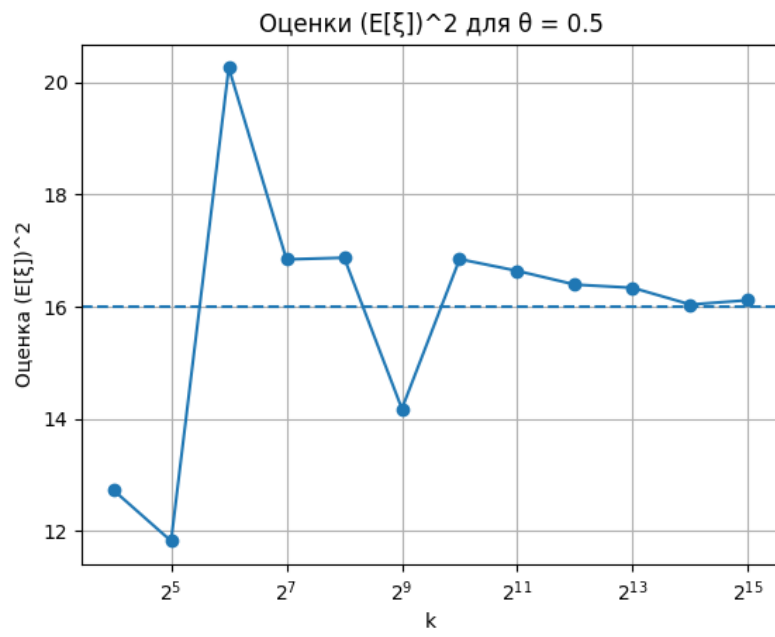
Далее, для каждого  $\theta$  и каждой оценки строится график:

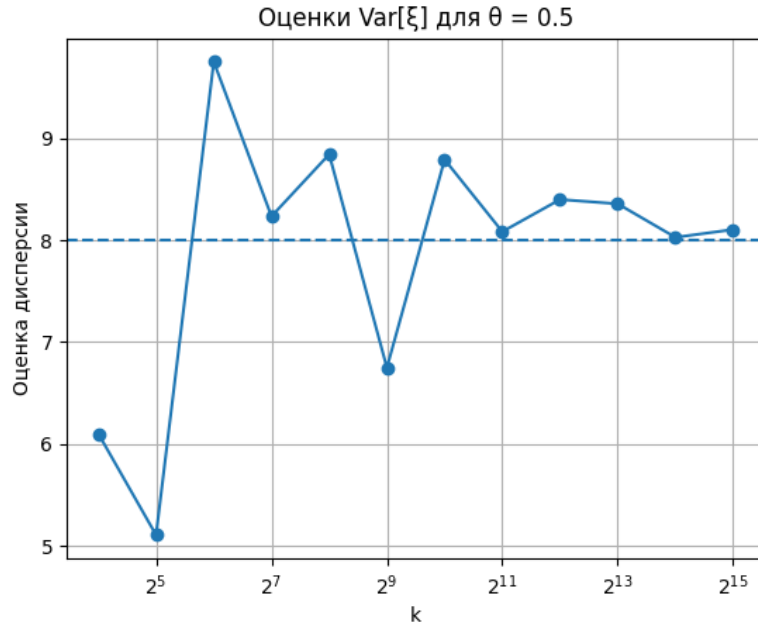
- по горизонтали — размер выборки  $k$ ;
- по вертикали — значение оценки при этом  $k$ ;
- добавляется горизонтальная прямоугольная линия на уровне истинного аналитического значения  $(\mathbb{E}\xi, D(\xi)$  или  $(\mathbb{E}\xi)^2$ ).











**Задача 2. Мода, мат. ожидание и медиана сдвинутого экспоненциального распределения.** Дана плотность

$$f_{\lambda,a}(x) = \begin{cases} \lambda e^{-\lambda(x-a)}, & x \geq a, \\ 0, & x < a, \end{cases}$$

где по условию  $(\lambda, a) = (2, 2)$ .

Заметим, что

$$\xi \stackrel{d}{=} a + Y, \quad Y \sim \text{Exp}(\lambda),$$

то есть  $\xi$  — сдвинутая экспоненциальная случайная величина.

**(а) Аналитические значения моды, мат. ожидания и медианы**

- **Мода.** Плотность для  $x \geq a$  убывает по  $x$ , поэтому максимум достигается в точке сдвига:

$$\text{mode}(\xi) = a = 2.$$

- **Математическое ожидание.** Для несдвинутой экспоненты  $Y \sim \text{Exp}(\lambda)$  имеем  $\mathbb{E}Y = 1/\lambda$ . Тогда

$$\mathbb{E}\xi = \mathbb{E}(a + Y) = a + \frac{1}{\lambda} = 2 + \frac{1}{2} = \frac{5}{2} = 2.5.$$

- **Медиана.** Пусть  $m$  — медиана  $\xi$ , то есть

$$\mathbb{P}(\xi \leq m) = \frac{1}{2}.$$

Для  $x \geq a$  функция распределения  $\xi$ :

$$F_{\xi}(x) = \mathbb{P}(\xi \leq x) = 1 - e^{-\lambda(x-a)}.$$

Тогда медиана  $m$  удовлетворяет

$$1 - e^{-\lambda(m-a)} = \frac{1}{2} \quad \Rightarrow \quad e^{-\lambda(m-a)} = \frac{1}{2} \quad \Rightarrow \quad m - a = \frac{\ln 2}{\lambda}.$$

Значит,

$$m = a + \frac{\ln 2}{\lambda} = 2 + \frac{\ln 2}{2}.$$

Итого:

$$\text{mode}(\xi) = 2, \quad \mathbb{E}\xi = \frac{5}{2}, \quad \text{med}(\xi) = 2 + \frac{\ln 2}{2}.$$

### (b)–(c) Моделирование выборок, оценки и графики

Пусть  $\xi$  моделируется как

$$\xi = a + Y, \quad Y \sim \text{Exp}(\lambda = 2).$$

Создаём две выборки:

- “большая” выборка размера  $n_{\text{big}} \approx 10000$ ;
- “маленькая” выборка размера  $n_{\text{small}} \approx 20$ .

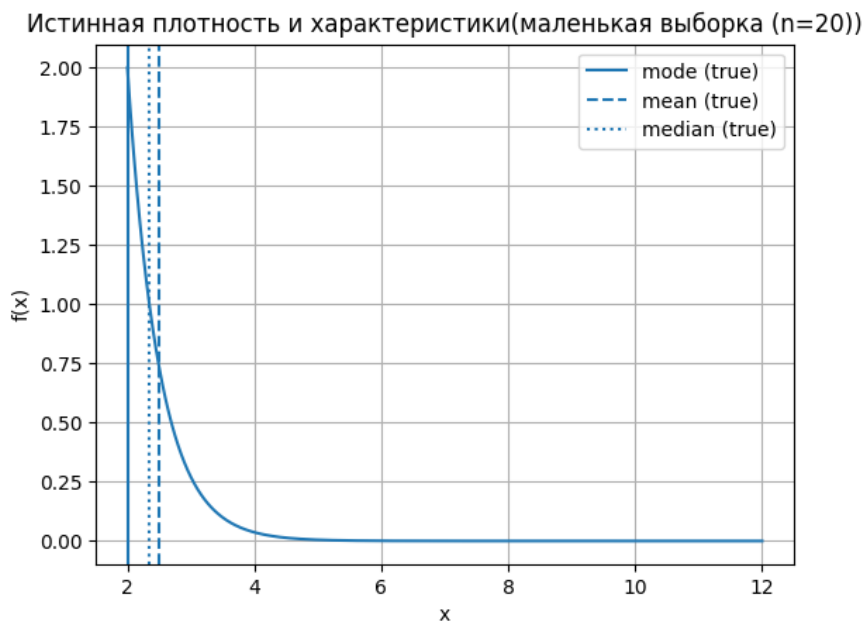
Для каждой выборки:

- оцениваем **математическое ожидание**:  $\hat{\mu} = \frac{1}{n} \sum \xi_i$ ;
- оцениваем **медиану** как выборочную медиану;
- оцениваем **моду**:
  - строим гистограмму;
  - за оценку моды берём центр столбика, в котором наблюдается максимальная высота.

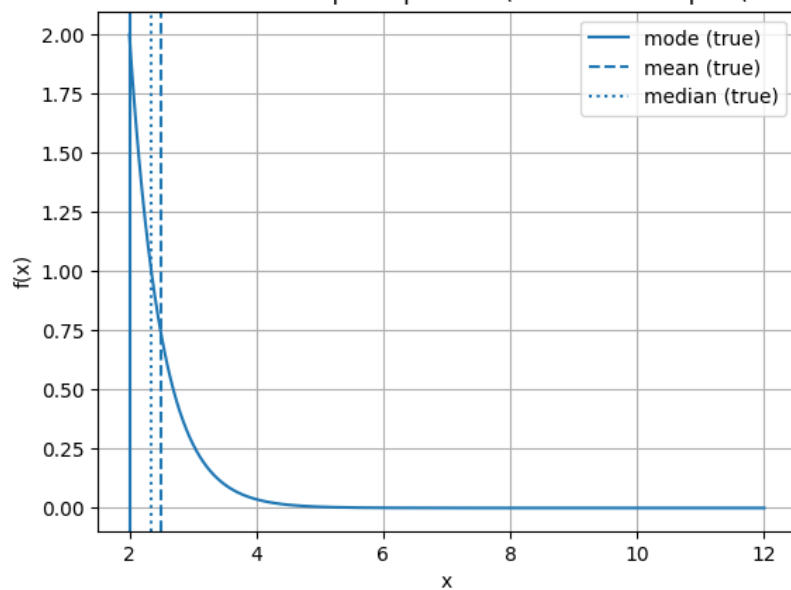
Для каждой выборки строятся графики:

- 1) гистограмма выборки + три вертикальные линии, соответствующие оценкам моды, мат. ожидания и медианы;
- 2) плотность (аналитическая кривая  $f_{\lambda,a}(x)$ ) + три вертикальные *аналитические* линии:

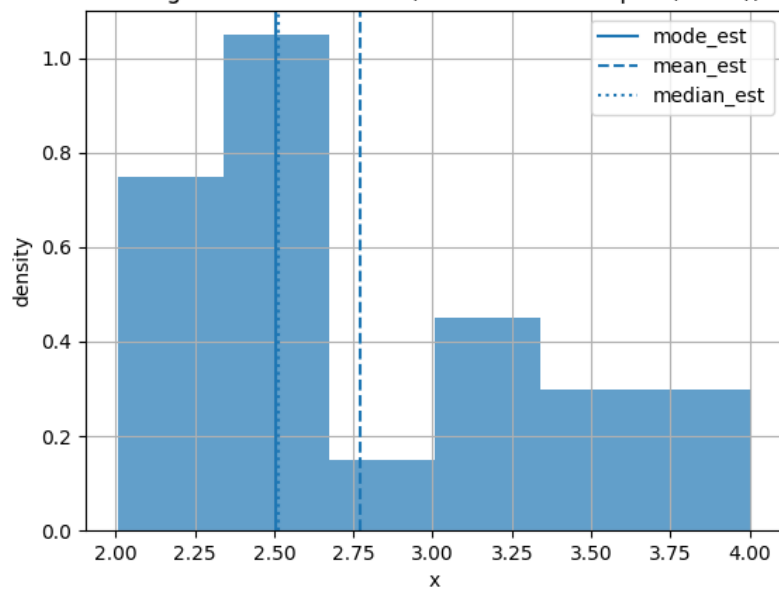
$$x = \text{mode}(\xi), \quad x = \mathbb{E}\xi, \quad x = \text{med}(\xi).$$

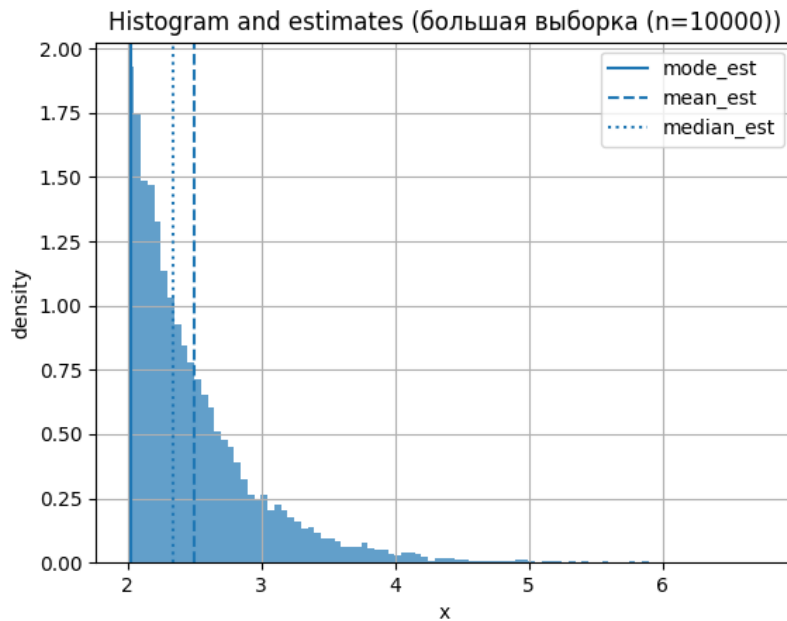


Истинная плотность и характеристики(большая выборка (n=10000))



Histogram and estimates (маленькая выборка (n=20))





#### (d) Поведение медианы при увеличении размера выборки

Аналитически для рассматриваемого распределения:

$$\mathbb{E}\xi = 2 + \frac{1}{\lambda} = 2 + \frac{1}{2} = \frac{5}{2},$$

$$\text{med}(\xi) = 2 + \frac{\ln 2}{\lambda} = 2 + \frac{\ln 2}{2}.$$

Разность:

$$\mathbb{E}\xi - \text{med}(\xi) = \frac{1}{\lambda} - \frac{\ln 2}{\lambda} = \frac{1 - \ln 2}{\lambda} > 0.$$

То есть истинные значения мат. ожидания и медианы *не совпадают*.

При увеличении размера выборки  $n$ :

- выборочное среднее  $\hat{\mu}_n$  сходится (по вероятности) к  $\mathbb{E}\xi$ ;
- выборочная медиана  $\widehat{\text{med}}_n$  сходится к  $\text{med}(\xi)$ .

Поэтому выборочная медиана *не будет сходиться к математическому ожиданию*; она будет стремиться к своему собственному пределу — истинной медиане распределения.

Это можно проиллюстрировать, изменяя размер выборки в коде и наблюдая, что:

$$\hat{\mu}_n \rightarrow \frac{5}{2}, \quad \widehat{\text{med}}_n \rightarrow 2 + \frac{\ln 2}{2}.$$

## 1.2. Моделирование совместного распределения двух СВ

По условию совместное распределение  $(\xi, \eta)$  задано таблицей:

$\xi \backslash \eta$	1	2	3	...
−1	$\frac{2}{5} \cdot \frac{1}{2^1}$	$\frac{2}{5} \cdot \frac{1}{2^2}$	$\frac{2}{5} \cdot \frac{1}{2^3}$	...
0	$\frac{1}{5} \cdot \frac{1}{2^1}$	$\frac{1}{5} \cdot \frac{1}{2^2}$	$\frac{1}{5} \cdot \frac{1}{2^3}$	...
1	$\frac{2}{5} \cdot \frac{1}{2^1}$	$\frac{2}{5} \cdot \frac{1}{2^2}$	$\frac{2}{5} \cdot \frac{1}{2^3}$	...

где  $\eta$  принимает все значения из  $\mathbb{N} = \{1, 2, 3, \dots\}$ .

То есть

$$\mathbb{P}(\xi = x, \eta = k) = p_x \cdot \frac{1}{2^k}, \quad x \in \{-1, 0, 1\}, \quad k \in \mathbb{N},$$

где

$$p_{-1} = \frac{2}{5}, \quad p_0 = \frac{1}{5}, \quad p_1 = \frac{2}{5}.$$

**Задача.** Вычислить корреляционную матрицу аналитически и приближённо (на основе моделирования).

### Аналитическое решение

#### Маргинальные распределения.

Для  $\xi$ :

$$\mathbb{P}(\xi = -1) = \frac{2}{5}, \quad \mathbb{P}(\xi = 0) = \frac{1}{5}, \quad \mathbb{P}(\xi = 1) = \frac{2}{5}.$$

Для  $\eta$ :

$$\mathbb{P}(\eta = k) = \sum_{x \in \{-1, 0, 1\}} \mathbb{P}(\xi = x, \eta = k) = \left( \frac{2}{5} + \frac{1}{5} + \frac{2}{5} \right) \frac{1}{2^k} = \frac{1}{2^k}, \quad k \geq 1.$$

То есть  $\eta$  имеет геометрическое распределение с параметром  $p = \frac{1}{2}$  (поддержка  $\{1, 2, 3, \dots\}$ ):

$$\mathbb{P}(\eta = k) = p(1-p)^{k-1} = \frac{1}{2} \left( \frac{1}{2} \right)^{k-1} = \frac{1}{2^k}.$$

#### Математические ожидания и дисперсии.

Для  $\xi$ :

$$\mathbb{E}\xi = (-1) \cdot \frac{2}{5} + 0 \cdot \frac{1}{5} + 1 \cdot \frac{2}{5} = 0,$$

$$\mathbb{E}\xi^2 = (-1)^2 \cdot \frac{2}{5} + 0^2 \cdot \frac{1}{5} + 1^2 \cdot \frac{2}{5} = \frac{2}{5} + 0 + \frac{2}{5} = \frac{4}{5},$$

$$D(\xi) = \mathbb{E}\xi^2 - (\mathbb{E}\xi)^2 = \frac{4}{5} - 0 = \frac{4}{5}.$$

Для  $\eta \sim \text{Geom}(p = \frac{1}{2})$  (с поддержкой  $\{1, 2, \dots\}$ ) известны:

$$\mathbb{E}\eta = \frac{1}{p} = 2, \quad D(\eta) = \frac{1-p}{p^2} = \frac{\frac{1}{2}}{(\frac{1}{2})^2} = 2.$$

#### Независимость и ковариация.

Совместное распределение факторизуется:

$$\mathbb{P}(\xi = x, \eta = k) = p_x \cdot \frac{1}{2^k} = \mathbb{P}(\xi = x) \cdot \mathbb{P}(\eta = k),$$

то есть  $\xi$  и  $\eta$  **независимы**.

Следовательно,

$$\text{cov}(\xi, \eta) = \mathbb{E}(\xi\eta) - \mathbb{E}\xi \mathbb{E}\eta = \mathbb{E}\xi \mathbb{E}\eta - \mathbb{E}\xi \mathbb{E}\eta = 0.$$

#### Ковариационная и корреляционная матрицы.

Ковариационная матрица вектора  $(\xi, \eta)$ :

$$\Sigma_{\xi, \eta} = \begin{pmatrix} D(\xi) & \text{cov}(\xi, \eta) \\ \text{cov}(\xi, \eta) & D(\eta) \end{pmatrix} = \begin{pmatrix} \frac{4}{5} & 0 \\ 0 & 2 \end{pmatrix}.$$

Корреляционный коэффициент:

$$\rho_{\xi, \eta} = \frac{\text{cov}(\xi, \eta)}{\sqrt{D(\xi)D(\eta)}} = 0.$$

Поэтому корреляционная матрица:

$$R_{\xi, \eta} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

### Приближённое вычисление (моделирование)

Чтобы получить приближённую (выборочную) корреляционную матрицу, можно смоделировать  $n$  независимых реализаций пары  $(\xi, \eta)$ .

#### Алгоритм моделирования:

- Сначала моделируем  $\xi_i$  независимо с распределением

$$\mathbb{P}(\xi = -1) = \frac{2}{5}, \quad \mathbb{P}(\xi = 0) = \frac{1}{5}, \quad \mathbb{P}(\xi = 1) = \frac{2}{5}.$$

- Затем моделируем  $\eta_i$  независимо от  $\xi_i$  по геометрическому распределению  $\text{Geom}(p = \frac{1}{2})$  на  $\{1, 2, \dots\}$ .
- Пара  $(\xi_i, \eta_i)$  имеет нужное совместное распределение.

После генерации выборок  $\{(\xi_i, \eta_i)\}_{i=1}^n$  можно вычислить выборочный корреляционный коэффициент

$$r_{\xi, \eta} = \frac{\sum_{i=1}^n (\xi_i - \bar{\xi})(\eta_i - \bar{\eta})}{\sqrt{\sum_{i=1}^n (\xi_i - \bar{\xi})^2} \sqrt{\sum_{i=1}^n (\eta_i - \bar{\eta})^2}},$$

где  $\bar{\xi}, \bar{\eta}$  — выборочные средние. При достаточно большом  $n$  он должен быть близок к нулю, а выборочные дисперсии близки к  $\frac{4}{5}$  и 2 соответственно.

Приближённая корреляционная матрица по результатам моделирования:

$$\hat{R}_{\xi, \eta} \approx \begin{pmatrix} 1 & 0.00436 \\ 0.00436 & 1 \end{pmatrix}.$$