1 Лабораторная работа №2

1.1 Оценки математического ожидания, дисперсии, медианы

Выполнил: Диденко Андрей

Задача 1

Пусть ξ имеет плотность

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

Заметим, что это — гамма-распределение.

$$g(x) = \frac{1}{\Gamma(\alpha)\,\beta^\alpha}\, x^{\alpha-1} e^{-x/\beta}, \qquad x \geq 0.$$

Сравнивая с $f_\xi(x)$ видим, что

$$x^{\alpha-1} = x^1 \implies \alpha = 2, \qquad e^{-x/\beta} = e^{-\theta x} \implies \frac{1}{\beta} = \theta \implies \beta = \frac{1}{\theta}.$$

Таким образом $\xi \sim \mathrm{Gamma}(\alpha = 2,\ \beta = 1/\theta)$.

Математическое ожидание и моменты вычисляются как

$$E[\xi] = \int_0^\infty x f_\xi(x) dx, \quad E[\xi^2] = \int_0^\infty x^2 f_\xi(x) dx, \quad D[\xi] = E[\xi^2] - (E[\xi])^2.$$

Аналитические вычисления для каждого θ

$$E[\xi] = \frac{2}{\theta}, \quad E[\xi^2] = \frac{6}{\theta^2}, \quad D[\xi] = \frac{6}{\theta^2} - \left(\frac{2}{\theta}\right)^2 = \frac{2}{\theta^2}.$$

Таким образом, для разных θ:

- $\theta = 0.5$: $E[\xi] = 4$, $D[\xi] = 8$, $E[\xi^2] = 24$.
- $\theta = 2$: $E[\xi] = 1$, $D[\xi] = 0.5$, $E[\xi^2] = 1.5$.
- $\theta = 8$: $E[\xi] = 0.25$, $D[\xi] = 0.03125$, $E[\xi^2] = 0.09375$.

Визуализация

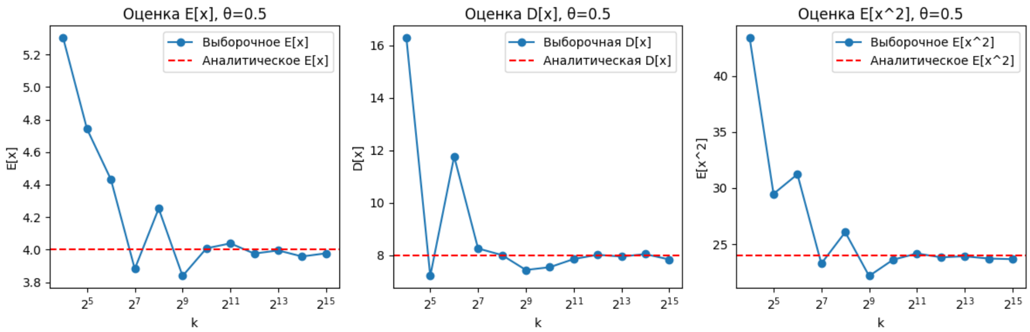

Рис. 1

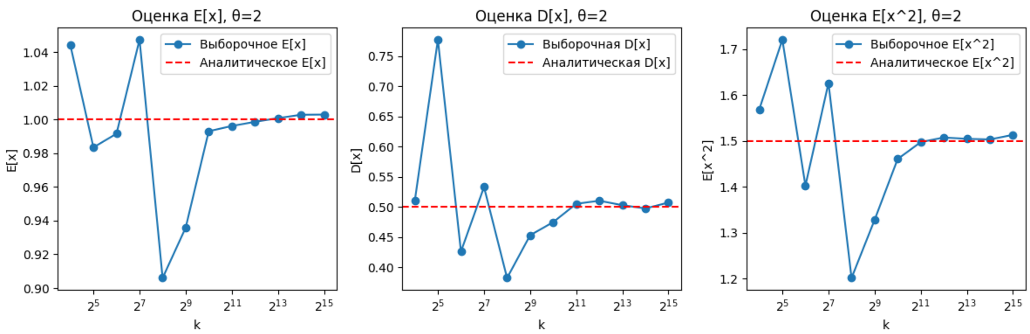

Рис. 2

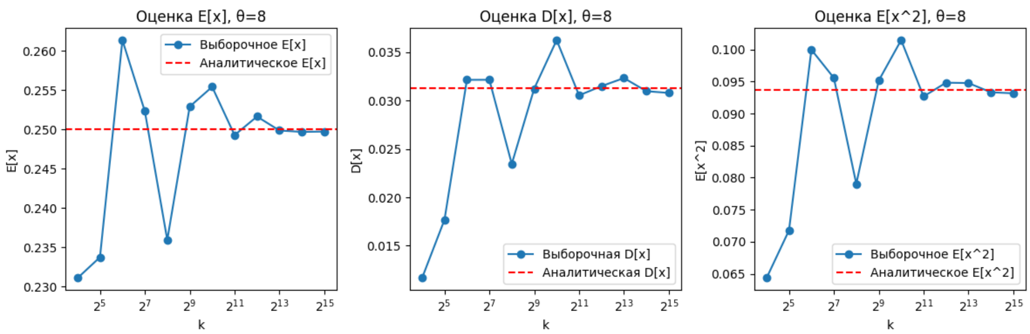

Рис. 3

Полное решение приведено в Jupyter Notebook: `тык`.

Вывод

С ростом объёма выборки k выборочные оценки математического ожидания, дисперсии и среднего квадрата случайной величины становятся всё ближе к их аналитическим значениям. Это подтверждает закон больших чисел и свидетельствует о корректности как аналитических, так и численных результатов моделирования.

Задача 2

Дано распределение случайной величины ξ:

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

(a) Аналитические вычисления: мода, математическое ожидание и медиана

Это сдвинутое экспоненциальное распределение (экспоненциальное распределение с параметром λ и сдвигом на a).

1. Мода.

Мода — точка, в которой плотность достигает максимума. Для $x \geq a$ функция $f_\xi^{(\lambda,a)}(x) = \lambda e^{-\lambda(x-a)}$ монотонно убывает по x (так как $\lambda > 0$). Следовательно максимум достигается при наименьшем допустимом значении аргумента, то есть при $x = a$.

$$\mathrm{mode}(\xi) = a.$$

2. Математическое ожидание.

Вычислим $E[\xi]$:

$$E[\xi] = \int_{-\infty}^{\infty} x f_\xi^{(\lambda,a)}(x)\, dx = \int_a^\infty x\lambda e^{-\lambda(x-a)}\, dx.$$

Выполним замену $y = x - a$ (тогда $x = y + a$, $dy = dx$):

$$E[\xi] = \int_0^\infty (y+a)\lambda e^{-\lambda y}\, dy = a \underbrace{\int_0^\infty \lambda e^{-\lambda y}\, dy}_{=1} + \underbrace{\int_0^\infty y\lambda e^{-\lambda y}\, dy}_{=1/\lambda}.$$

получаем

$$E[\xi] = a + \frac{1}{\lambda}.$$

3. Медиана.
Медиана m определяется условием $F(m) = 0.5$, где F — функция распределения:

$$F(x) = \int_{-\infty}^{x} f_\xi^{(\lambda,a)}(t)\,dt = \begin{cases} 0, & x < a, \\ 1 - e^{-\lambda(x-a)}, & x \ge a. \end{cases}$$

Требуем $F(m) = 1 - e^{-\lambda(m-a)} = 0.5$. Отсюда

$$e^{-\lambda(m-a)} = 0.5 \quad \Longrightarrow \quad -\lambda(m-a) = \ln(0.5) = -\ln 2.$$

Следовательно

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

Численные значения при $\lambda = 2,\ a = 2$. Подставим:

$$\text{мода } = a = 2, \qquad E[\xi] = 2 + \frac{1}{2} = 2.5, \qquad \text{медиана } = 2 + \frac{\ln 2}{2} \approx 2.3466.$$

(b) Создание выборок

Большая выборка (n=10000):

Мода: 2.000,
Среднее: 2.509,
Медиана: 2.353

Маленькая выборка (n=20):

Мода: 2.068,
Среднее: 2.650,
Медиана: 2.489

(c) Визуализация

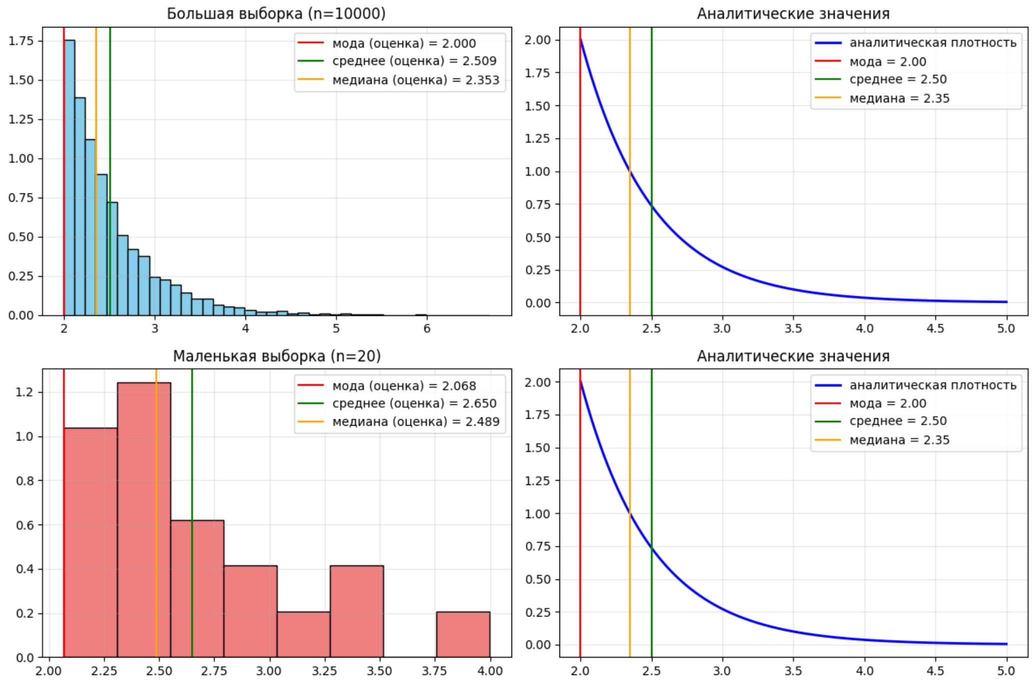

Рис. 4

(d) Анализ медианы и математического ожидания

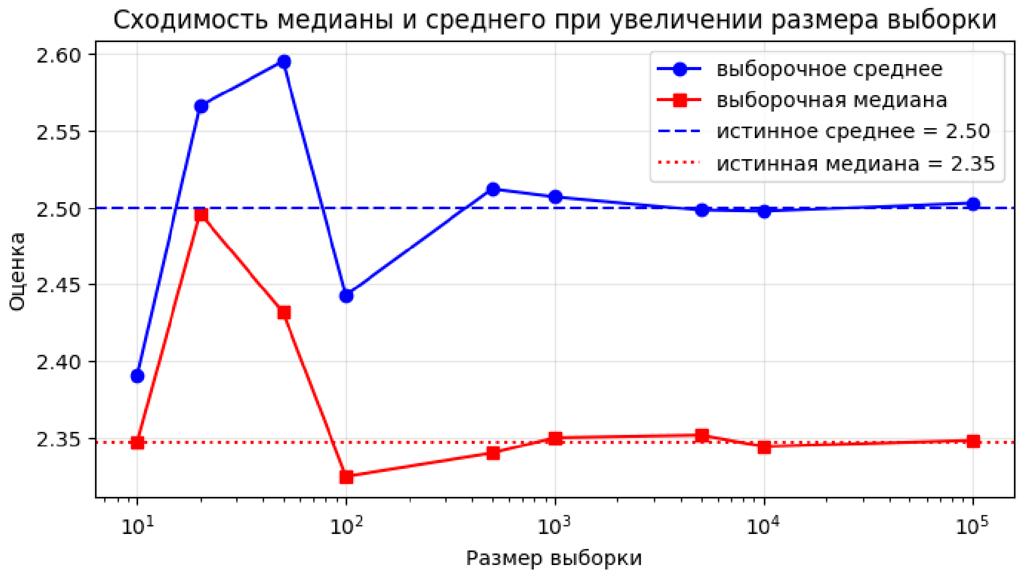

Рис. 5

Полное решение приведено в Jupyter Notebook: `тык`.

Вывод

Из проведённых расчётов видно, что для большой выборки оценки моды, математического ожидания и медианы практически совпадают с аналитическими значениями, а сама гистограмма очень хорошо повторяет форму плотности. Для маленькой выборки всё работает хуже: оценки заметно колеблются, а распределение выглядит «шумным», что ожидаемо при малом количестве данных.

Также при увеличении размера выборки медиана не стремится к математическому ожиданию, поскольку в этом распределении они изначально различаются. Но при этом выборочная медиана всё равно постепенно приближается к своей теоретической медиане.

1.2 Моделирование совместного распределения двух СВ

Пусть случайные величины ξ и η имеют совместное распределение, заданное 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 η принимает значения $1, 2, 3, \dots$ и имеет геометрическое распределение с параметром $p = 0.5$.

Геометрическое распределение

Для дискретной случайной величины η с геометрическим распределением:

$$P(\eta = k) = p(1-p)^{k-1}, \quad k = 1, 2, 3, \dots$$

Для нашей задачи:

$$P(\eta = k) = \frac{1}{2^k}$$

Маргинальные распределения

Для ξ:

$$\sum_{k=1}^{\infty} \frac{1}{2^k} = 1$$

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

Для η: Суммируем совместные вероятности по ξ:

$$P(\eta = 1) = \frac{1}{5} + \frac{1}{10} + \frac{1}{5} = 0.5$$
$$P(\eta = 2) = \frac{1}{10} + \frac{1}{20} + \frac{1}{10} = 0.25$$
$$P(\eta = 3) = \frac{1}{20} + \frac{1}{40} + \frac{1}{20} = 0.125$$
$$\dots$$

Математическое ожидание:

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

$$E[\eta] = \sum_{k=1}^{\infty} k \cdot P(\eta = k) = 1 \cdot 0.5 + 2 \cdot 0.25 + 3 \cdot 0.125 + \cdots = 2$$

Дисперсии:

$$D(\xi) = E[\xi^2] - (E[\xi])^2 = (-1)^2 \cdot \frac{2}{5} + 0 + 1^2 \cdot \frac{2}{5} - 0^2 = \frac{4}{5}$$

$$D(\eta) = E[\eta^2] - (E[\eta])^2 = \sum_{k=1}^{\infty} k^2 P(\eta = k) - 2^2 = 6 - 4 = 2$$

Ковариация:

$$\mathrm{Cov}(\xi, \eta) = E[\xi\eta] - E[\xi]E[\eta]$$

$$\begin{aligned} E[\xi\eta] = (-1) \cdot 1 \cdot \frac{1}{5} + (-1) \cdot 2 \cdot \frac{1}{10} + \ldots \\ + 0 \cdot 1 \cdot \frac{1}{10} + \ldots \\ + 1 \cdot 1 \cdot \frac{1}{5} + 1 \cdot 2 \cdot \frac{1}{10} + \cdots = 0 \end{aligned}$$

$$\mathrm{Cov}(\xi, \eta) = 0$$

Корелляция:

$$\rho_{\xi,\eta} = \frac{\mathrm{Cov}(\xi, \eta)}{\sqrt{D(\xi)}\,\sqrt{D(\eta)}} = \frac{0}{\sqrt{4/5} \cdot \sqrt{2}} = 0.$$

Следовательно, корреляционная матрица:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Моделирование распределения:

n = 10000
Выборочные оценки:

$$\begin{aligned} E[\xi] &\approx 0.008, & D[\xi] &\approx 0.798, \\ E[\eta] &\approx 2.009, & D[\eta] &\approx 2.021, \\ E[\xi\eta] &\approx 0.006, & \mathrm{Cov}(\xi, \eta) &\approx -0.009, \\ \rho_{\xi,\eta} &\approx -0.007 \end{aligned}$$

Матрица корреляции выборки:

$$\begin{bmatrix} 1 & -0.00745872 \\ -0.00745872 & 1 \end{bmatrix}$$

Полное решение приведено в Jupyter Notebook: `тык`.

Вывод

Случайные величины ξ и η независимы. Их ковариация равна нулю, поэтому они некоррелированы. Выборочные оценки подтверждают это – корреляция близка к нулю, а математические ожидания и дисперсии согласуются с аналитическими значениями.