Лабораторная работа №2

Выполнил Арзуманов Олег

1.1 Оценки математического ожидания, дисперсии, медианы.

Пусть случайная величина ξ имеет распределение, задаваемое плотностью $f_\xi(x) = \theta^2 x e^{-\theta x}$. Для каждого $\theta \in \{0.5, 2, 8\}$:

1. (a) Аналитически вычислить математическое ожидание, дисперсию и математическое ожидание квадрата ξ. Привести в отчет.

(b) Для $k \in \{2^4, 2^5, \dots, 2^{15}\}$ построить выборку из k элементов. Для каждой из них посчитать оценки: математического ожидания, дисперсии и квадрата математического ожидания параметра из варианта. Для каждой из выборок и оценок визуализировать это все на графиках (для каждой оценки — свой график), где по вертикальной оси — оценка, а по горизонтальной — k, плюс, добавьте горизонтальную линию, отвечающую за аналитически полученную оценку.

Решение:

a) Можно заметить, что $f_\xi(x)$ это гамма-распределение с параметрами $\alpha = 2$ (параметр формы) и параметром скорости (rate) $\lambda = \theta$:

$$f(x; \alpha, \lambda) = \begin{cases} \dfrac{\lambda^\alpha \lambda^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}, & x \geq 0 \\ 0, & else \end{cases},$$

где $\Gamma(\alpha)$ - гамма-функция Эйлера, которая определяется как:

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$$

Для целых положительных чисел $n,\ \Gamma(n) = (n-1)!$

Стандартные формулы для гамма(α, λ) дают:

$$E\xi = \frac{\alpha}{\lambda}, \quad Var(\xi) = \frac{\alpha}{\lambda^2}$$

Проверим так ли это и выведем эти результаты через интегралы:

- Расчёт математического ожидания.

$$E\xi = \int_0^\infty x\theta^2 x e^{-\theta x} dx = \theta^2 \int_0^\infty x^2 e^{-\theta x} dx = \theta^2 \times \frac{1}{\theta^3} \int_0^\infty u^2 e^{-u} du = \frac{1}{\theta} \times \Gamma(3) = \frac{2}{\theta}$$

- Расчёт дисперсии $E[\xi^2]$.

Сначала найдём $E[\xi^2]$:

$$E[\xi^2] = \theta^2 \int_0^\infty x^3 e^{-\theta x} dx = \theta^2 \times \frac{1}{\theta^4} \int_0^\infty u^3 e^{-u} du = \frac{1}{\theta^2} \times \Gamma(4) = \frac{6}{\theta^2}$$

$$D\xi = E[\xi^2] - [E\xi]^2 = \frac{6}{\theta^2} - \frac{4}{\theta^2} = \frac{2}{\theta^2}$$

b) Для $\theta = 0.5$:

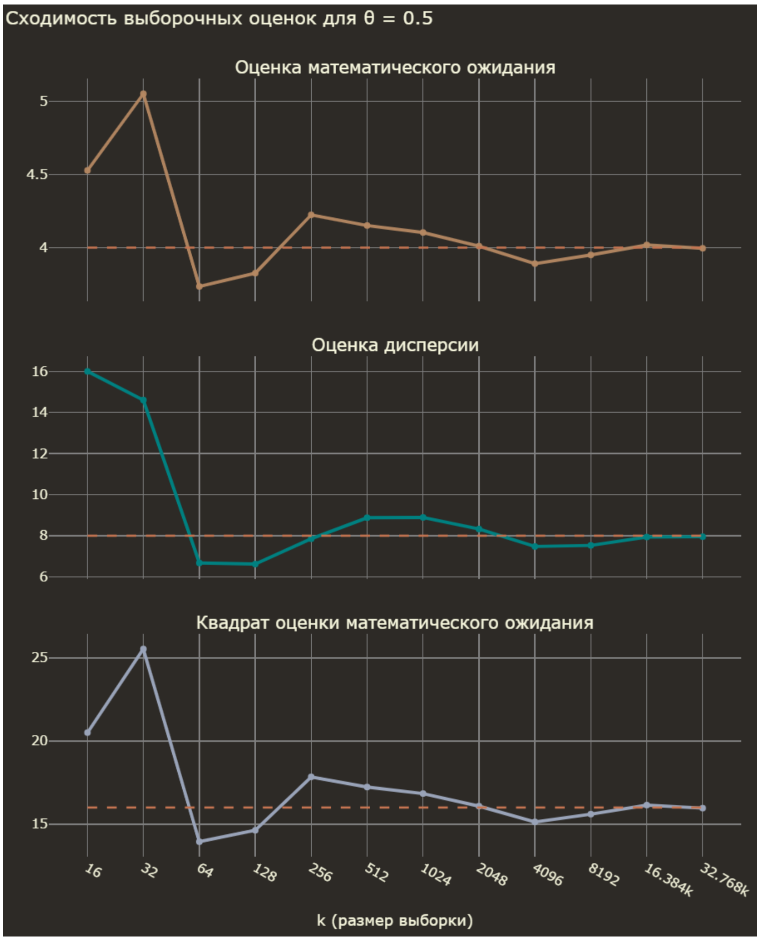

Рисунок 1 – График оценок при $\theta = 0.5$

Для $\theta = 2$:

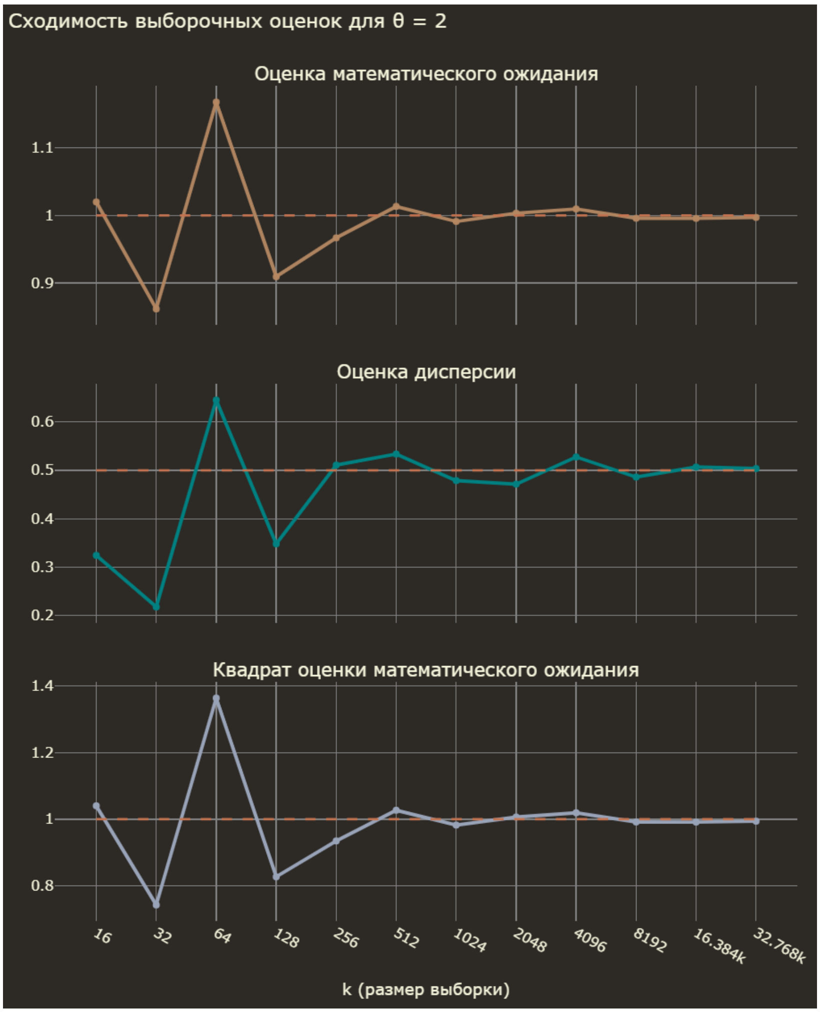

Рисунок 2 – График оценок при $\theta = 2$

Для $\theta = 8$:

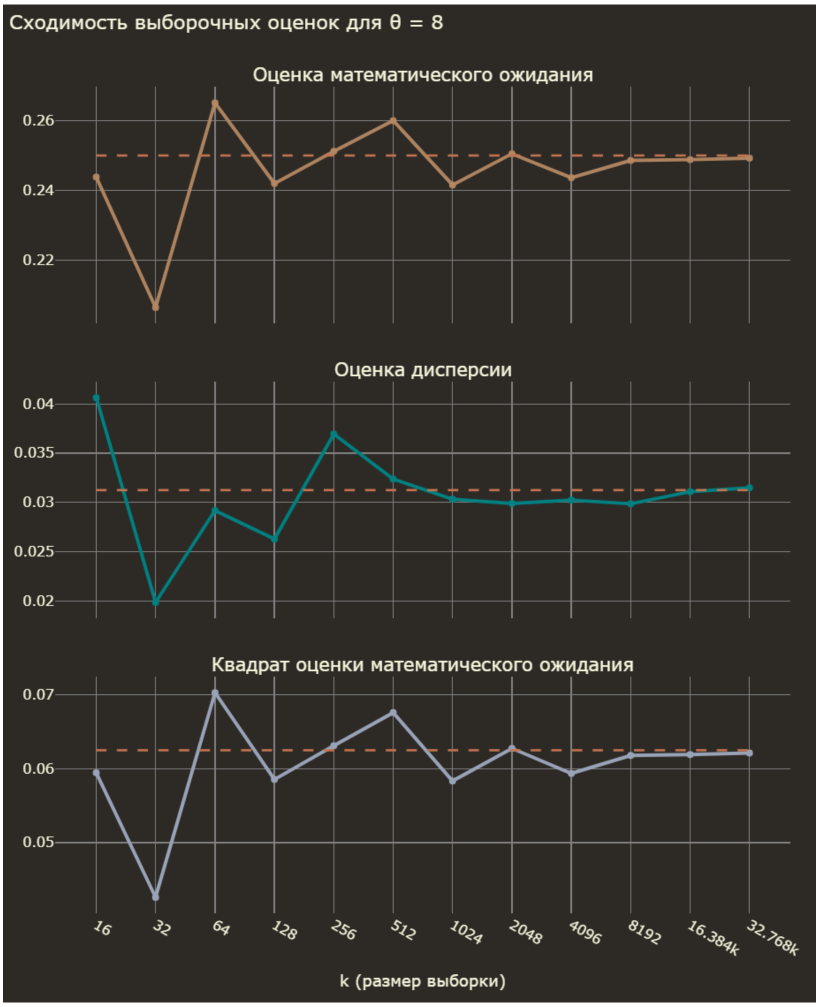

Рисунок 3 – График оценок при $\theta = 8$

Для всех значений θ характерно:

- все наблюдаемые оценки с ростом размера выборки приближаются к их аналитическим значениям;
- при $\theta = 2$ наблюдаемые оценки сходятся к их аналитическим значениям быстрее.

2. Дана плотность распределения случайной величины ξ:

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

Пусть $(\lambda, a) = (2, 2)$

(a) Аналитически вычислите значение моды, математического ожидания и медианы.

(b) Создайте две выборки: одну довольно большого размера (10000 наблюдений, например), а вторую маленькую (например, 20). Постройте оценки моды, математического жидания и медианы.

(c) Постройте для первой выборки на одном графике: гистограмму распределения значений из выборки и три вертикальных линии оценок моды, математического ожидания и медианы. Для второй выборки сделайте то же самое. Постройте ещё график рядом для первой выборки, но с функцией распределения плотности и аналитическими значениями мод, математического жидания и медианы. То же самое, для второй.

(d) Попробуйте поизменять размер выборки и посмотреть на то, например, сходится ли медиана к математическому ожиданию, или нет.

Решение:

a) Мода.

$f_{\xi}^{\lambda,a}(x)$ - сдвинутая экспоненциальная плотность, которая строго убывает на $x \in [a, \infty)$, поэтому максимум плотности (мода) достигается в левой границе при $x = a$.

$$mode = a = 2$$

Математическое ожидание.

$$E\xi = \int_a^\infty x\lambda e^{-\lambda(x-a)}\,dx = \int_0^\infty (t+a)\lambda e^{-\lambda t}\,dt$$

$$= \lambda \int_0^\infty te^{-\lambda t}\,dt\lambda + a\lambda \int_0^\infty e^{-\lambda t}\,dt = \lambda \times \frac{1}{\lambda^2} + a\lambda \times \frac{1}{\lambda} = \frac{1}{\lambda} + a$$

$$E\xi = 2.5$$

Медиана.

$$P(\xi \le m) = 1 - e^{-\lambda(m-a)} = \frac{1}{2} \Longrightarrow e^{-\lambda(m-a)} = \frac{1}{2}$$

$$\Downarrow$$

$$median = \frac{ln\,2}{\lambda} + a \approx 2.347$$

b) Сгенерируем выборки и построим оценки моды, математического жидания и медианы:

```python
import numpy as np
from scipy import stats

gamma = 2
a = 2

np.random.seed(0)

# генерация выборок
sample_large = np.random.exponential(scale=1/gamma, size=10000) + a
sample_small = np.random.exponential(scale=1/gamma, size=20) + a

def estimate_values(sample):
    mean_est = np.mean(sample)
    median_est = np.median(sample)
    mode_est = stats.mode(sample)[0]

    return mean_est, median_est, mode_est

mean_L, med_L, mode_L = estimate_values(sample_large)
mean_S, med_S, mode_S = estimate_values(sample_small)

print("LARGE SAMPLE:")
print("Mode_estimation =", mode_L)
print("Mean_estimation =", mean_L)
print("Median_estimation =", med_L)

print("\nSMALL SAMPLE:")
print("Mode_estimation =", mode_S)
print("Mean_estimation =", mean_S)
print("Median_estimation =", med_S)
```

Вывод:

```
LARGE SAMPLE:
Mode_estimation = 2.000036
Mean_estimation = 2.495367
Median_estimation = 2.3401

SMALL SAMPLE:
Mode_estimation = 2.00078
Mean_estimation = 2.49047
Median_estimation = 2.239
```

c) Построенные графики для двух выборок выглядят следующим образом:

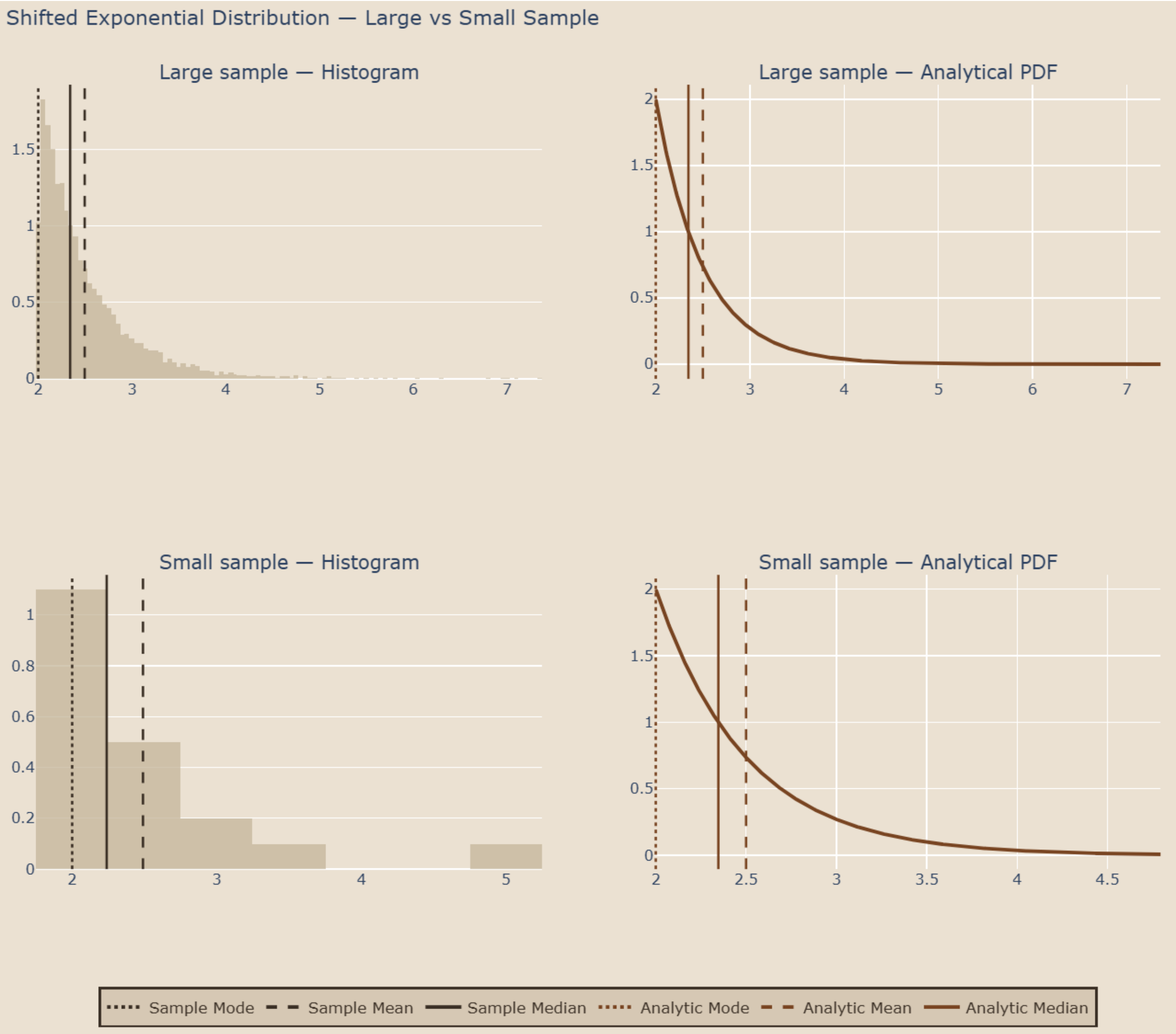

Рисунок 4 – Гистограмма и плотность с оценками для двух выборок

d) Посмотрим сходится ли медиана к математическому ожиданию при изменении размера выборки:

```
sizes = [10, 50, 200, 1000, 5000, 50000, 100000, 500000]
medians = []

for n in sizes:
    sample = np.random.exponential(scale=1/gamma, size=n) + a
    medians.append(np.median(sample))

# график
fig = go.Figure()

fig.add_trace(go.Scatter(
    x=sizes, y=medians,
```

```
    mode="lines+markers",
    name="sample median"
))

fig.add_hline(y=mean_analytic, line_dash="dash",
              annotation_text="analytic mean")

fig.update_layout(
    title="Convergence of sample median",
    xaxis_title="sample size",
    yaxis_title="median"
)

fig.show()
```

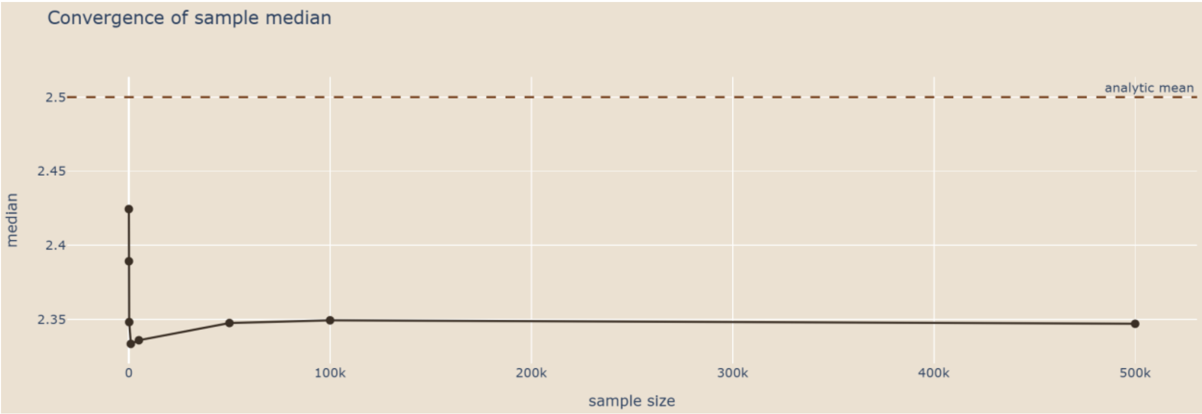

Рисунок 5 – Сходимость медианы

1.2 Моделирование совместного распределения двух СВ

Пусть совместное распределение двух случайных величин задано таблицей

$\xi \backslash \eta$	1	2	3	$\ldots$
-1	$\frac{2}{5}\cdot\frac{1}{2^1}$	$\frac{2}{5}\cdot\frac{1}{2^2}$	$\frac{2}{5}\cdot\frac{1}{2^3}$	$\ldots$
0	$\frac{1}{5}\cdot\frac{1}{2^1}$	$\frac{1}{5}\cdot\frac{1}{2^2}$	$\frac{1}{5}\cdot\frac{1}{2^3}$	$\ldots$
1	$\frac{2}{5}\cdot\frac{1}{2^1}$	$\frac{2}{5}\cdot\frac{1}{2^2}$	$\frac{2}{5}\cdot\frac{1}{2^3}$	$\ldots$

где η принимает все значения из N. Вычислить корреляционую матрицу аналитически и приближенно (на основе моделирования).

Решение:

В таблице видно, что для любого $k \in N$ значения для трёх значений $\xi \in \{-1,0,1\}$ имеют вид:

$$P(\xi = -1,\, \eta = k) = \frac{2}{5} \times \frac{1}{2^k}\,, \quad P(\xi = 0,\, \eta = k) = \frac{1}{5} \times \frac{1}{2^k}\,,$$

$$P(\xi = 1,\, \eta = k) = \frac{2}{5} \times \frac{1}{2^k}$$

$$\Downarrow$$

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

$$\sum_i p_i = 1$$

Так как q_k – это геометрическая прогрессия, то

$$\sum_k q_k = 1$$

По определению, случайные величины ξ и η независимы, если для всех значений:

$$P(\xi = i, \eta = k) = P(\xi = i)\, P(\eta = k)$$

Так как

- $P(\xi = i) = p_i$
- $P(\eta = k) = q_k$

и запись таблицы уже имеет вид:

$$P(\xi = i, \eta = k) = p_i q_k$$

$$\Downarrow$$

Определение независимости выполнено и $cov(\xi, \eta)$ & $corr(\xi, \eta) = 0$. Корреляционная матрица:

$$R = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Вычислим корреляционную матрицу на основе моделирования:

```
import numpy as np

# аналитические значения
corr_analytic = np.array([
    [1, 0],
    [0, 1]
])

print("Аналитическая корреляционная матрица:")
print(corr_analytic)

N = 1000000  # количество испытаний

# моделирование xi
xi_values = np.array([-1, 0, 1])
xi_probabilities = np.array([0.4, 0.2, 0.4])
samples_xi = np.random.choice(xi_values, size=N, p=xi_probabilities)

# моделирование eta
samples_eta = np.random.geometric(0.5, size=N)

# формирование выборки
samples = np.array([samples_xi, samples_eta])

# вычисление выборочной ковариационной матрицы
corr_approximate = np.corrcoef(samples)

print("\n" + f"Приближенная корреляционная матрица (на основе {N} испытаний):")
print(corr_approximate)
```

Вывод:

```
Аналитическая корреляционная матрица:
[[1 0]
 [0 1]]

Приближенная корреляционная матрица (на основе 1000000 испытаний):
[[1.         0.00153823]
 [0.00153823 1.        ]]
```