

Санкт-Петербургский Политехнический Университет
Высшая школа прикладной математики и вычислительной физики, ФизМех
01.03.02 Прикладная математика и информатика

Отчет по лабораторной работе № 1
"Гистограммы и плотности распределений"

Выполнил студент гр. 5030102/30201
Преподаватель

Елисеева М. А.
Баженов А. Н.

Санкт-Петербург
2025

Постановка задачи

Задание: сгенерировать выборки объёмом $n = 10, 100, 1000$ для каждого из 5 распределений. Построить на одном графике гистограмму выборки и теоретическую кривую плотности распределения.

Цели и задачи:

- Освоить принцип группировки данных и построения гистограмм.
- Сделать вывод о том, как влияет размер выборки на определение характера распределения величины с помощью гистограммы

Исследуются следующие распределения:

1. Нормальное $N(x; 0, 1)$
2. Коши $C(x; 0, 1)$
3. Лапласа $L(x; 0, 1/\sqrt{2})$
4. Пуассона $P(k; 10)$
5. Равномерное $U(x; -\sqrt{3}, \sqrt{3})$

Теоретическая часть

- Нормальное распределение

$$Normal(x, 0, 1) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$$

- Распределение Коши

$$Cauchy(x, 0, 1) = \frac{1}{\pi} \frac{1}{x^2 + 1}$$

- Распределение Лапласа

$$Laplace\left(x, 0, \frac{1}{\sqrt{2}}\right) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}|x|}$$

- Распределение Пуассона

$$Poisson(k, 10) = \frac{10^k}{k!} e^{-10}$$

- Равномерное распределение

$$Uniform(x, -\sqrt{3}, \sqrt{3}) = \begin{cases} \frac{1}{2\sqrt{3}} & \text{при } |x| \leq \sqrt{3} \\ 0 & \text{при } |x| > \sqrt{3} \end{cases}$$

Реализация

Язык программирования: Python 3.12

Сторонние библиотеки:

`numpy` – генерация случайных выборок

`matplotlib.pyplot` – построение графиков

`scipy.stats` – теоретические плотности распределений

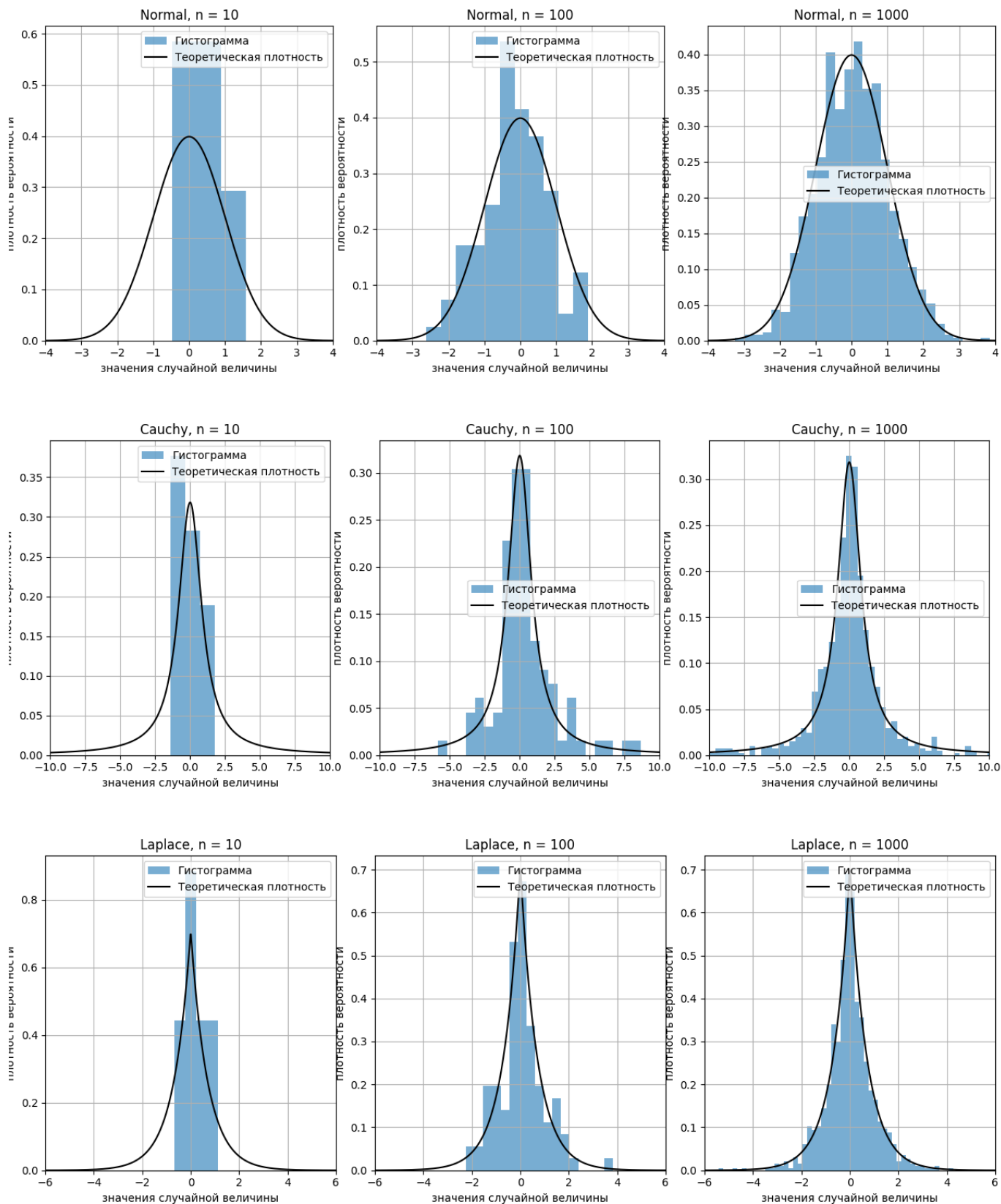
Код: <https://github.com/MaryEliseeva/matstat/blob/main/lab1/lab1.py>

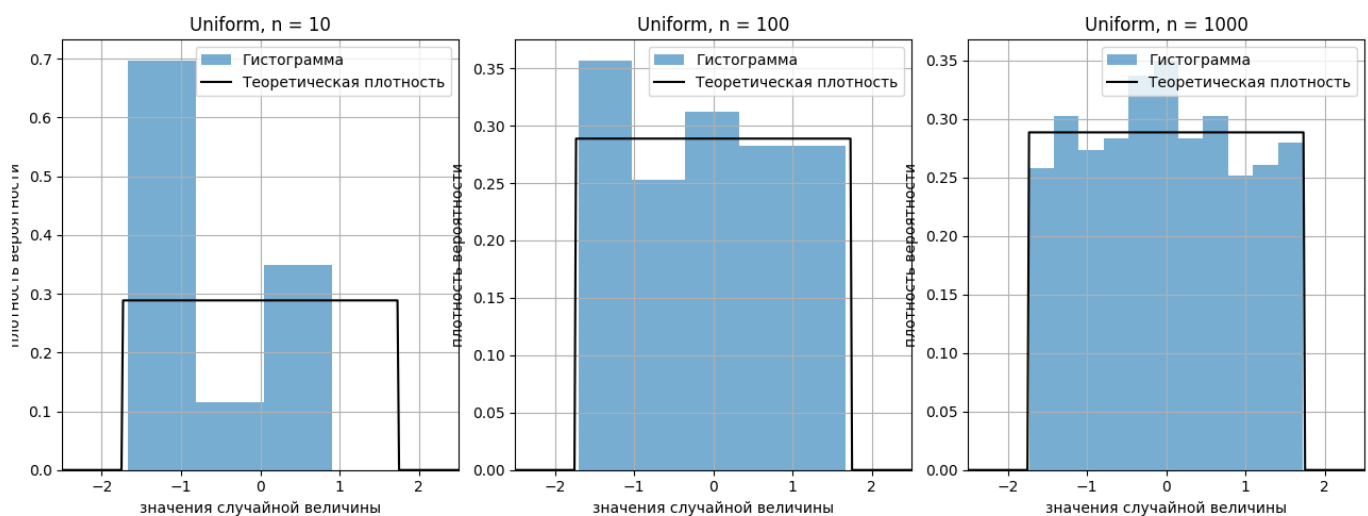
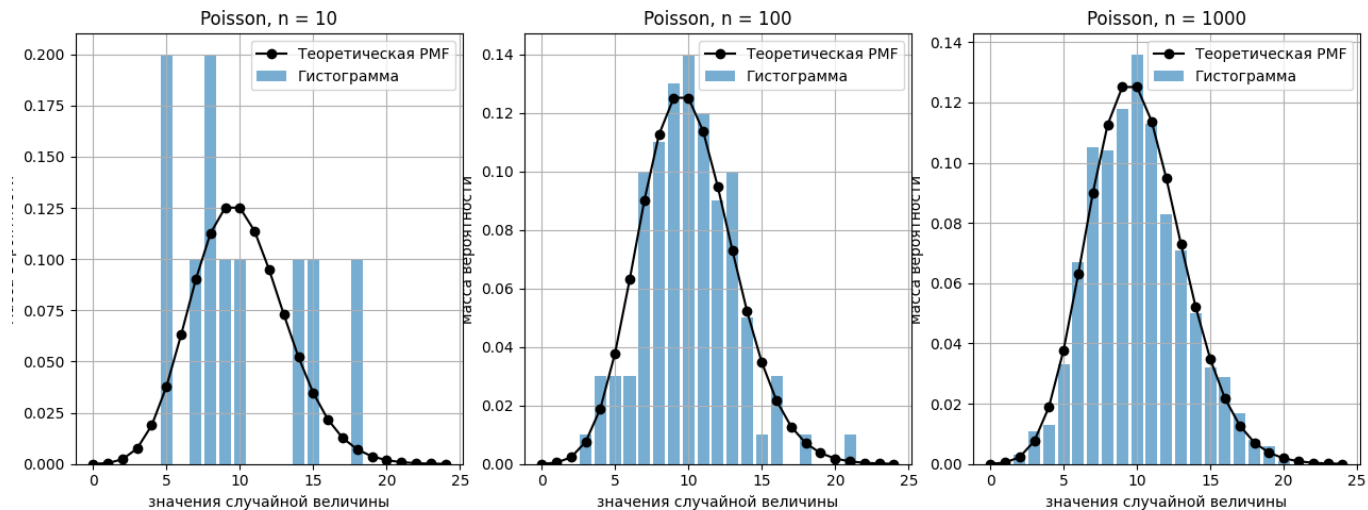
Для непрерывных распределений разбиваем данные на интервалы (бины), строим гистограмму, нормируем так, чтобы площадь под ней была равна 1, и поверх — график теоретической плотности.

Для дискретного распределения Пуассона использовалась столбчатая диаграмма, вместо гистограммы и функция массы вероятности вместо плотности.

Количество интервалов гистограммы выбиралось автоматически из библиотеки `matplotlib`, что обеспечивает баланс между потерей информации и шумом.

Результаты





Выводы

В результате работы были получены графики гистограмм и соответствующих теоретических кривых для каждого распределения при $n=10, 100, 1000$.

При малом объёме выборки гистограммы имеют ступенчатый малоразличимый вид;

По мере увеличения выборки форма распределения начинает проявляться более отчётливо

При большом объёме выборки гистограмма хорошо приближает теоретическую плотность.