

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

Гусев Г. А.
Баженков А. Н.

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 $P(k; 5)$
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 3.14

Сторонние библиотеки:

`numpy` – генерация случайных выборок

`matplotlib.pyplot` – построение графиков

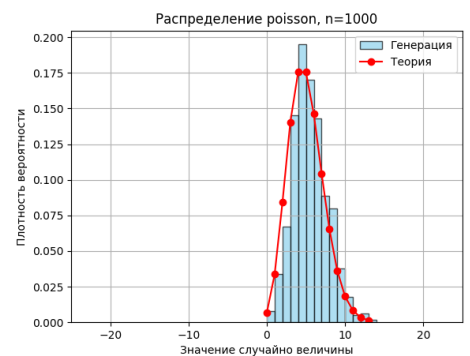
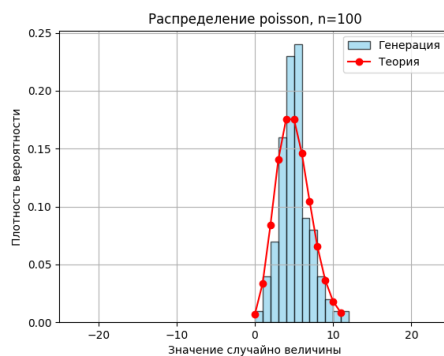
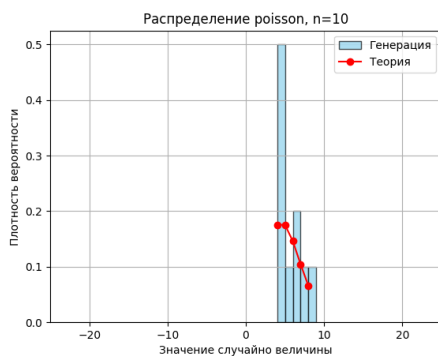
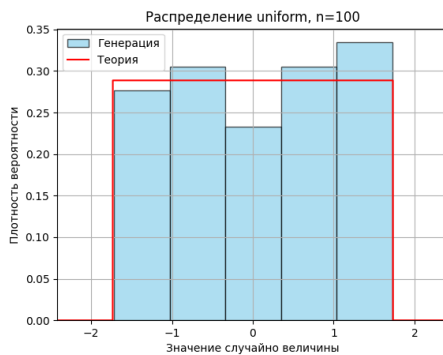
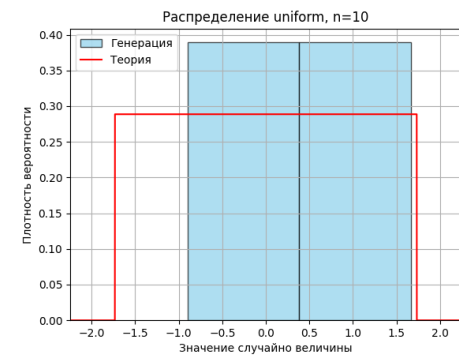
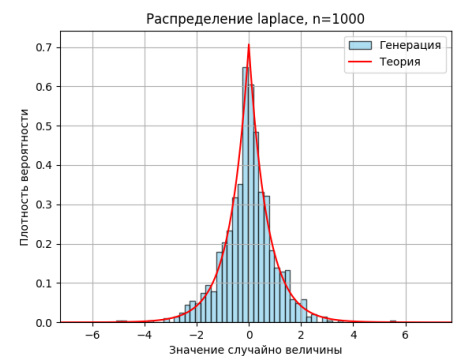
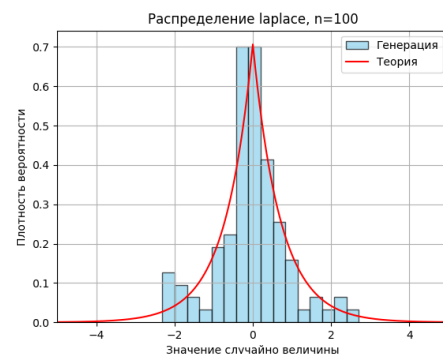
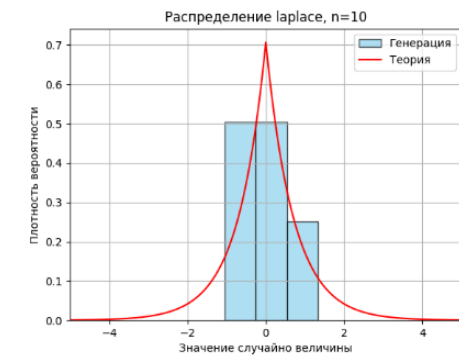
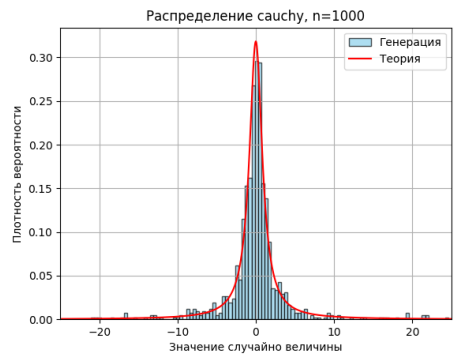
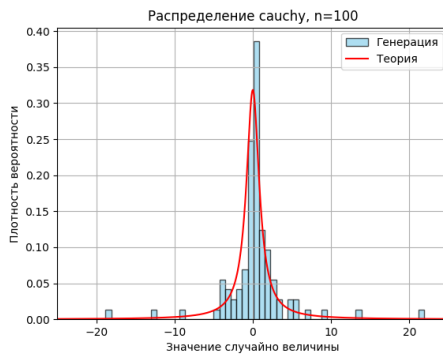
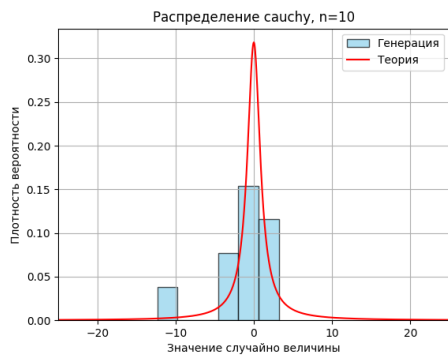
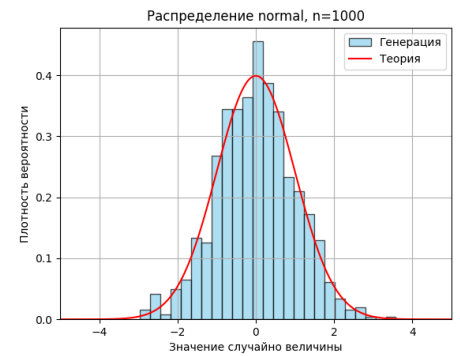
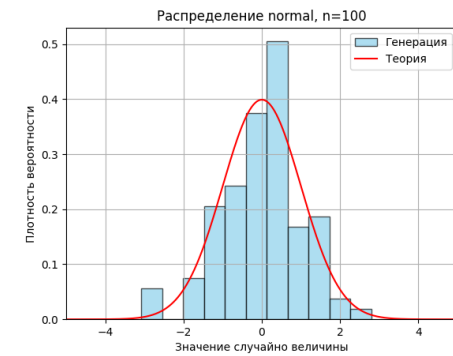
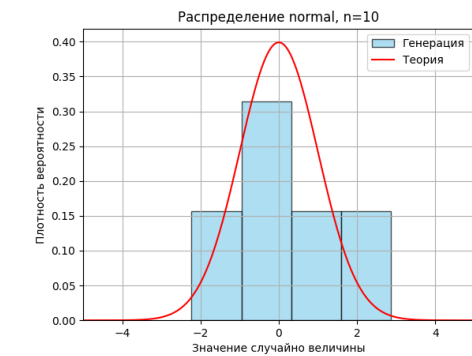
`scipy.stats` – теоретические плотности распределений

Код: https://github.com/Gusen14ka/MatStat_Lab1.git

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

Количество интервалов гистограммы выбиралось правилом Фридмана–Диакониса, что обеспечивает баланс между потерей информации и шумом.

Результаты



Обсуждение

В ходе выполнения лабораторной работы были построены гистограммы выборок объёмом ($n = 10, 100, 1000$) для пяти заданных распределений и выполнено их сопоставление с соответствующими теоретическими плотностями (или функцией вероятности для дискретного распределения).

Анализ полученных графиков показывает, что объём выборки существенно влияет на вид гистограммы и на возможность корректного определения характера распределения случайной величины.

При малом объёме выборки ($n = 10$) гистограммы носят выраженный случайный характер: наблюдаются значительные колебания частот, форма распределения и его основные особенности выражены слабо, что затрудняет идентификацию типа распределения.

При увеличении объёма выборки до $n = 100$ начинают отчётливо проявляться основные свойства распределений: симметрия или асимметрия, наличие тяжёлых хвостов, форма плотности.

Гистограммы становятся более сглаженными и лучше согласуются с теоретическими кривыми.

При большом объёме выборки ($n = 1000$) эмпирические распределения в большинстве случаев хорошо аппроксимируют соответствующие теоретические плотности. Наблюдается наглядная иллюстрация сходимости выборочного распределения к теоретическому при увеличении числа наблюдений.

Для нормального, лапласовского и равномерного распределений подтверждается ожидаемое поведение: с ростом объёма выборки гистограмма всё точнее повторяет форму теоретической плотности. Для распределения Пуассона, являющегося дискретным, во всех случаях сохраняется дискретный характер распределения, что проявляется в ступенчатой структуре столбцов.

Таким образом, влияние объёма выборки на качество аппроксимации теоретического распределения эмпирическими данными подтверждено на практике.

Выводы

В результате выполнения лабораторной работы были построены гистограммы выборок объемом ($n = 10, 100, 1000$) и соответствующие им теоретические кривые для каждого из исследуемых распределений.

Установлено, что:

- при малом объеме выборки гистограммы имеют нерегулярный и малоинформативный характер;
- при увеличении числа наблюдений форма распределения становится более выраженной;
- при большом объеме выборки гистограмма хорошо приближает теоретическую плотность (или функцию вероятности).

Поставленные цели работы достигнуты: продемонстрировано влияние объема выборки на качество графического представления распределения случайной величины и на возможность корректного определения его характера.