

**Санкт-Петербургский Политехнический университет имени Петра
Великого**

Кафедра прикладной математики и информатики

ОТЧЁТ

по лабораторной работе №1

ГИСТОГРАММЫ И ПЛОТНОСТИ РАСПРЕДЕЛЕНИЙ

Выполнил:
студент группы 5030102/30003
Крутянский Роман Игоревич

Преподаватель: Баженов Александр
Николаевич

Санкт-Петербург
2025

1. Постановка задачи

Целью данной лабораторной работы является освоение основных методов визуализации распределений случайных величин с помощью гистограмм, а также исследование влияния размера выборки на точность восстановления теоретической плотности распределения.

1. Сгенерировать выборки объёмом $n = 10, 100, 1000$ для каждого из пяти распределений:
2. Построить гистограммы выборок с наложением теоретических кривых плотности,
3. Проанализировать влияние количества интервалов (бинов) на качество визуализации,
4. Сделать выводы о сходимости эмпирического распределения к теоретическому при росте n .

Исследуются следующие распределения:

№	Распределение	Обозначение	Параметры
1	Нормальное	$N(x; 0, 1)$	$\mu = 0, \sigma = 1$
2	Коши	$C(x; 0, 1)$	$x_0 = 0, \gamma = 1$
3	Лапласа	$L\left(x; 0, \frac{1}{\sqrt{2}}\right)$	$\mu = 0, b = \frac{1}{\sqrt{2}}$
4	Пуассона	$P(k; 10)$	$\lambda = 10$
5	Равномерное	$U\left(x; -\sqrt{3}, \sqrt{3}\right)$	$a = -\sqrt{3}, b = \sqrt{3}$

2. Теоретическая часть

2.1. Гистограмма

Гистограмма — это графическое представление распределения количественных данных, где диапазон значений разбивается на интервалы (бины), и для каждого интервала отображается частота попадания значений выборки в этот интервал.

Пусть имеется выборка x_1, x_2, \dots, x_n объёма n . Диапазон значений $[x_{\min}, x_{\max}]$ разбивается на k интервалов шириной h :

$$h = \frac{x_{\max} - x_{\min}}{k}$$

Частота попадания в i -й интервал:

$$n_i = \{j : x_j \in [a_i, a_{i+1})\}$$

Нормированная гистограмма (оценка плотности):

$$\hat{f}(x) = \frac{n_i}{n * h}, \quad x \in [a_i, a_{i+1})$$

3. Реализация

3.1. Инструментарий

Работа выполнена на языке программирования Python 3.10 с использованием следующих библиотек:

Библиотека	Назначение
numpy	Генерация выборок, численные расчёты
matplotlib	Построение графиков и гистограмм
scipy.stats	Теоретические распределения и их функции

3.2. Алгоритм работы

Программа реализует следующий алгоритм:

1. Для каждого из 5 распределений генерируются 3 выборки объёмом $n = 10, 100, 1000$
2. Для каждой выборки строится нормированная гистограмма с автоматическим выбором количества бинов
3. На тот же график наносится теоретическая кривая плотности распределения
4. Все графики сохраняются в картинку в формате png

4. Результаты

4.1. Визуализация распределений

На рисунках Рис. 1–Рис. 5 представлены гистограммы для всех исследуемых распределений при различных объемах выборки.

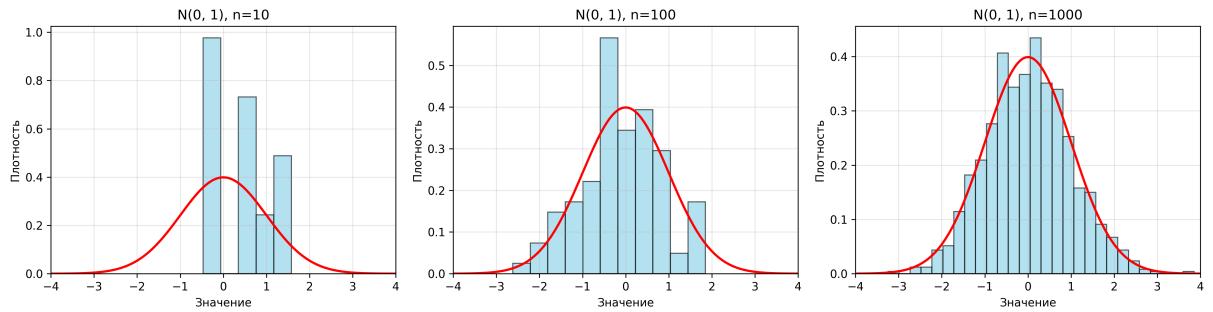


Рис. 1. Гистограммы нормального распределения $N(0, 1)$ при $n = 10, 100, 1000$

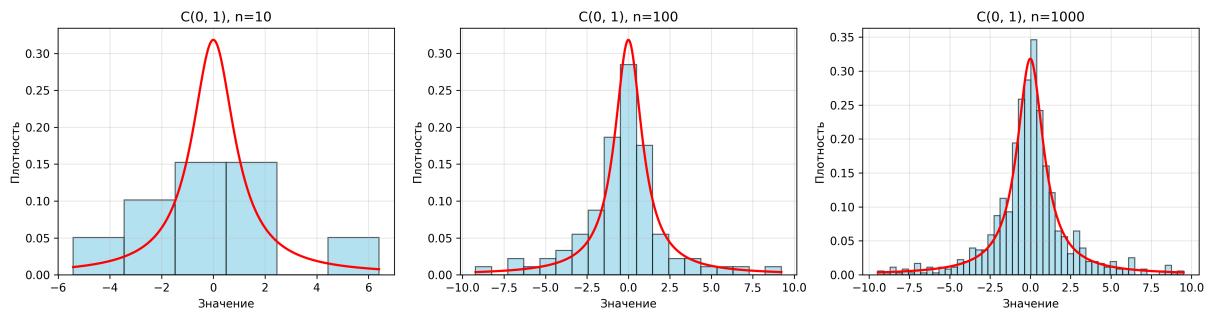


Рис. 2. Гистограммы распределения Коши $C(0, 1)$ при $n = 10, 100, 1000$

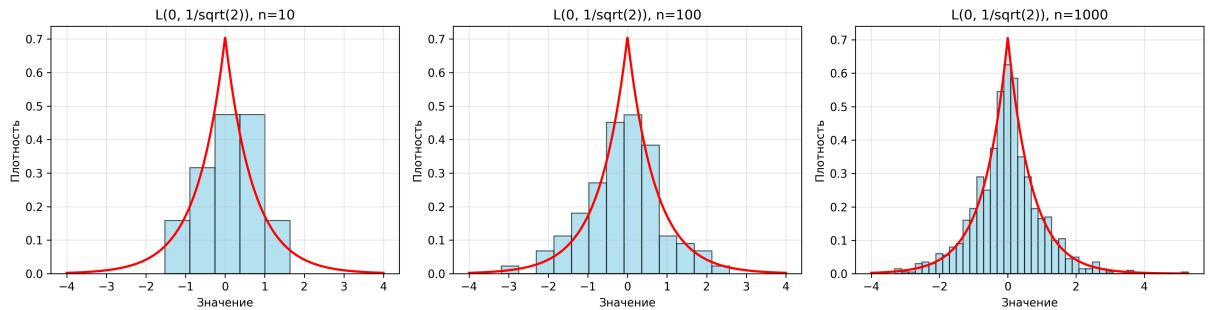


Рис. 3. Гистограммы распределения Лапласа $L\left(0, \frac{1}{\sqrt{2}}\right)$ при $n = 10, 100, 1000$

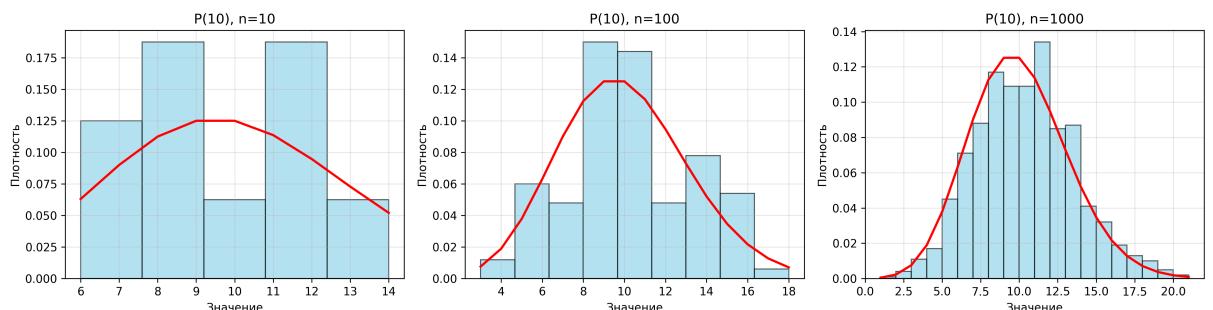


Рис. 4. Гистограммы распределения Пуассона $P(10)$ при $n = 10, 100, 1000$

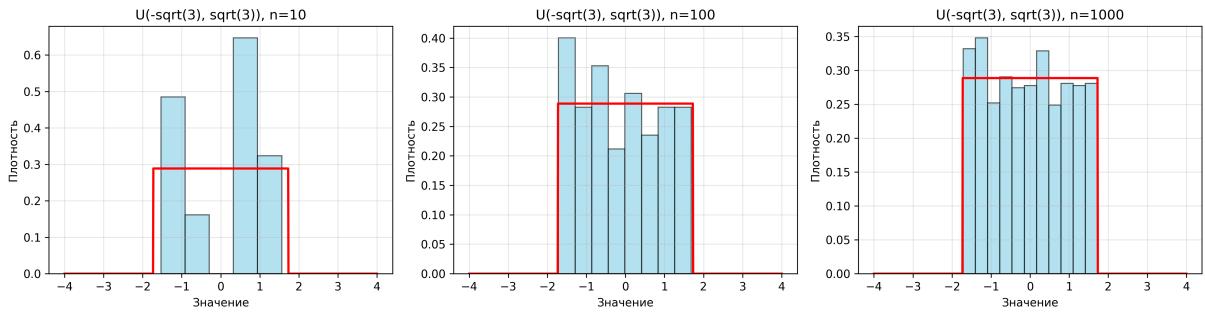


Рис. 5. Гистограммы равномерного распределения $U(-\sqrt{3}, \sqrt{3})$ при $n = 10, 100, 1000$

4.2. Сводный график

На рисунке Рис. 6 представлены все распределения на одном графике для сравнения.

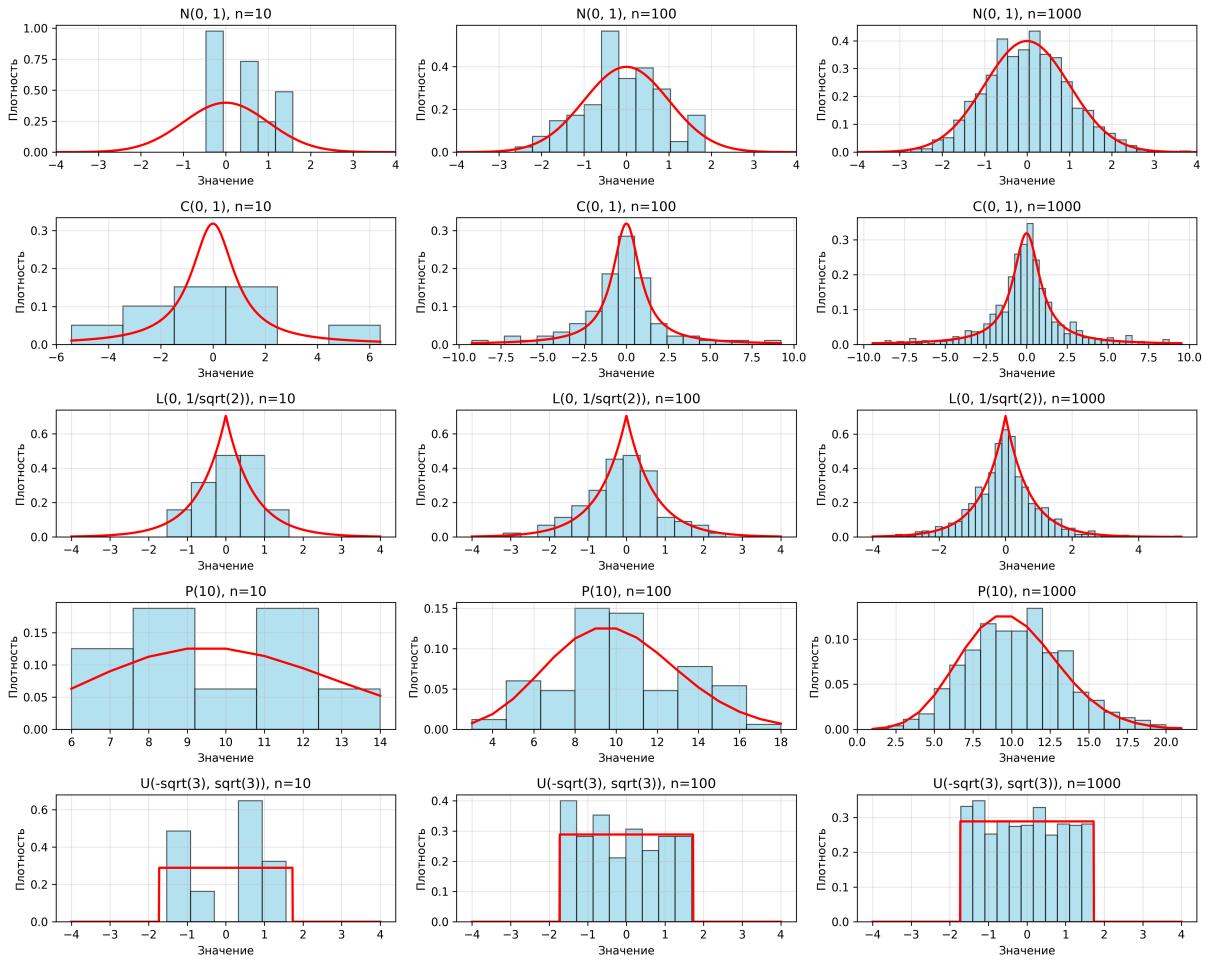


Рис. 6. Сравнение гистограмм всех распределений при $n = 10, 100, 1000$

5. Обсуждение

5.1. Влияние размера выборки

Анализ полученных результатов позволяет сделать следующие наблюдения:

Размер выборки	Характеристики гистограммы
$n = 10$	Сильный шум, форма распределения неочевидна, большие отклонения от теоретической кривой
$n = 100$	Появляется узнаваемая форма, но сохраняются заметные флуктуации
$n = 1000$	Хорошее соответствие теоретической плотности, гладкая форма гистограммы

5.2. Особенности распределений

Нормальное распределение. При $n \geq 100$ гистограмма хорошо аппроксимирует колоколообразную кривую. Симметрия относительно $x = 0$ становится очевидной.

Распределение Коши. Характеризуется тяжёлыми хвостами, что приводит к появлению выбросов даже при больших n . Гистограмма имеет более широкое основание по сравнению с нормальным распределением.

Распределение Лапласа. Отличается более острым пиком в центре ($x = 0$) и экспоненциальным убыванием хвостов.

Распределение Пуассона. Дискретная природа распределения хорошо видна на гистограмме. При $\lambda = 10$ распределение приближается к нормальному (центральная предельная теорема).

Равномерное распределение. Гистограмма стремится к прямоугольной форме. При малых n наблюдаются значительные отклонения от равномерности.

5.3. Проблема выбора бинов

Экспериментально подтверждено, что:

- Слишком малое количество бинов приводит к потере информации о форме распределения

- Слишком большое количество бинов создаёт «шумный» график с пустыми интервалами
- Автоматический выбор (`bins='auto'` в `matplotlib`) даёт разумный компромисс для большинства случаев

6. Выводы

В ходе выполнения лабораторной работы были достигнуты следующие результаты:

1. Освоен принцип группировки данных и построения нормированных гистограмм
2. Продемонстрирована сходимость эмпирического распределения к теоретическому при росте объёма выборки
3. Получены навыки работы с непрерывными и дискретными распределениями в Python

Наиболее существенное влияние размера выборки наблюдается для распределений с тяжёлыми хвостами (Коши) и дискретных распределений (Пуассона).