Санкт-Петербургский политехнический университет
Петра Великого

Институт прикладной математики и механики
Кафедра «Прикладная математика»

Отчёт
по лабораторной работе №1
по дисциплине
«Математическая статистика»

Выполнил студент:
Самутичев Евгений Романович
группа: 3630102/70201

Проверил:
к.ф.-м.н., доцент
Баженов Александр Николаевич

Санкт-Петербург
2020 г.

Содержание

1 Постановка задачи

Для каждого из 5 распределений:

1. Нормального $N(x, 0, 1)$
2. Коши $C(x, 0, 1)$
3. Лапласа $L(x, 0, \frac{1}{\sqrt{2}})$
4. Пуассона $P(k, 10)$
5. Равномерного $U(x, -\sqrt{3}, \sqrt{3})$

сгенерировать массив случайных данных (выборку) размера: 10, 50, 1000 и построить графики плотности вероятности (функции вероятности для распределения Пуассона как дискретного).

2 Теория

2.1 Распределения

Пусть задано вероятностное пространство $(\Omega, \mathcal{F}, \mathbf{P})$, на котором определена *случайная величина* $\xi : \Omega \to \mathbb{R}$ т.е. функция $\xi(\omega)$ такая что $\xi^{-1}(B) \in \mathcal{F}, \forall B \in \mathcal{B}(\mathbb{R})$. Она индуцирует вероятностную меру на $\mathbb{R}$ как $\mathbf{P}_\xi(B) = \mathbf{P}(\xi^{-1}(B))$ которая и носит название *распределения вероятностей* случайной величины [1].

Функция $F_\xi(x) = \mathbf{P}_\xi(-\infty, x], x \in \mathbb{R}$ называется *функцией распределения* случайной величины ξ. Случайная величина может быть:

1. *дискретной*, если распределение представимо в виде $\mathbf{P}_\xi(B) = \sum\limits_{k:x_k \in B} p(x_k)$, где $p(x_k) = \mathbf{P}_\xi\{x_k\}$ для конечного $\{x_1, ..., x_n\}$ или счетного $\{x_1, ..., x_k, ...\}$ подмножества вещественных чисел. В этом случае функция $p(x_k)$ называется таблицей распределения.

2. *непрерывной*, если $F(x)$ непрерывна

3. *абсолютно непрерывной*, если существует такая неотрицательная функция $f_\xi(x)$ называемая *плотностью вероятности*, что $F(x) = \int\limits_{-\infty}^{x} f(y)dy$

В работе рассматриваются следующие распределения:

1. *Нормальное* $N(x, 0, 1)$ - абсолютно непрерывное, задается плотностью

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

2. *Коши* $C(x, 0, 1)$ - абсолютно непрерывное, задается плотностью

$$f_C(x) = \frac{1}{\pi(x^2 + 1)}$$

3. *Лапласа* $L(x, 0, \frac{1}{\sqrt{2}})$ - абсолютно непрерывное, задается плотностью

$$f_L(x) = \frac{1}{2\sqrt{2}} e^{-\frac{1}{\sqrt{2}}|x|}$$

4. *Пуассона* $P(k, 10)$ - дискретное, задается на $\{1, 2, ..., k, ...\}$ как

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

5. *Равномерное* $U(x, -\sqrt{3}, \sqrt{3})$ - абсолютно непрерывное, задается плотностью

$$f_U(x) = \begin{cases} \frac{1}{2\sqrt{3}} & \text{если } x \in [-\sqrt{3}, \sqrt{3}] \\ 0 & \text{иначе} \end{cases}$$

2.2 Гистограмма

Все приведенные распределения характеризуются таблицей (для дискретных) или плотностью (для абсолютно непрерывных). Эмпирическим аналогом таблицы или плотности является *гистограмма* [2]. Гистрограмма строится по группированным данным. Предполагаемую область значений случайной величины ξ делят на некоторое количество интервалов:

Пусть $A_1, ..., A_k$ - интервалы на прямой. Обозначим $\nu_j, j \in \{1, ..., k\}$ - число элементов выборки, попавших в интервал A_j. Размер выборки в этих обозначениях равен $n = \sum\limits_{j=1}^{k} \nu_j$. На каждом из интервалов строят прямоугольник, площадь которого пропорциональна ν_j, общая площадь всех прямоугольников должна равняться единице (нормировка гистограммы), поэтому высота каждого определяется как $f_j = \frac{\nu_j}{nl_j}$. Полученная фигура из объединения прямоугольников и называется гистограммой.

3 Реализация

Работа выполнена с использованием языка **Python** в интегрированной среде разработки **PyCharm**, были задействованы библиотеки:

- **NumPy** - работа с массивами данных
- **SciPy** - модуль **stats** для генерации данных по распределениям
- **Matplotlib** - отрисовка гистограмм и графиков плотности/функции вероятности

Для случайной генерации было выбрано зерно 102 (для повторяемости в дальнейших экспериментах), исходный код работы приведен в приложении. Число интервалов гистограммы выбрано как округление к большему к целому $\sqrt{n}$, где n - размер выборки.

4 Результаты

4.1 Гистограммы и графики

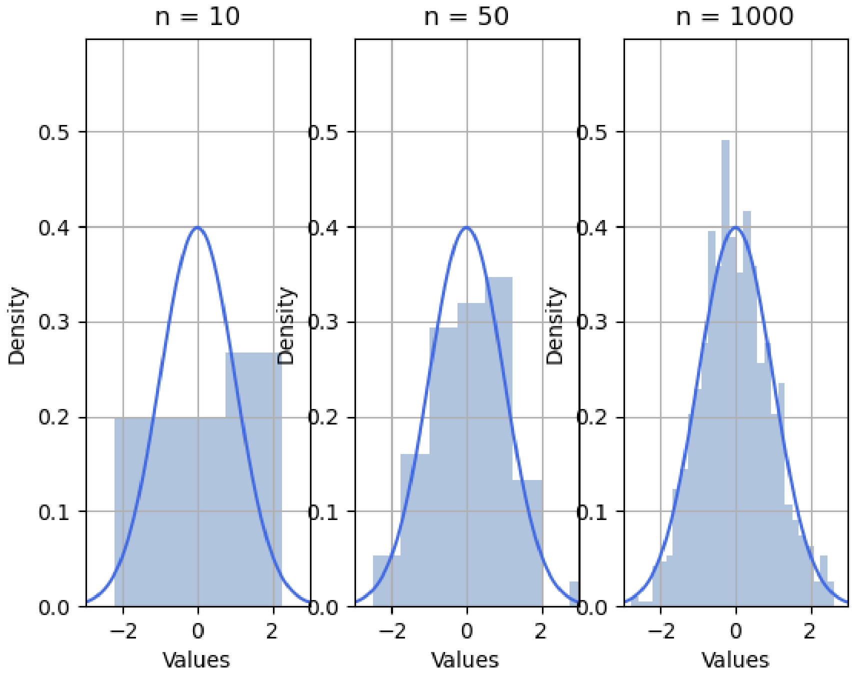

Рис. 1: Нормальное распределение

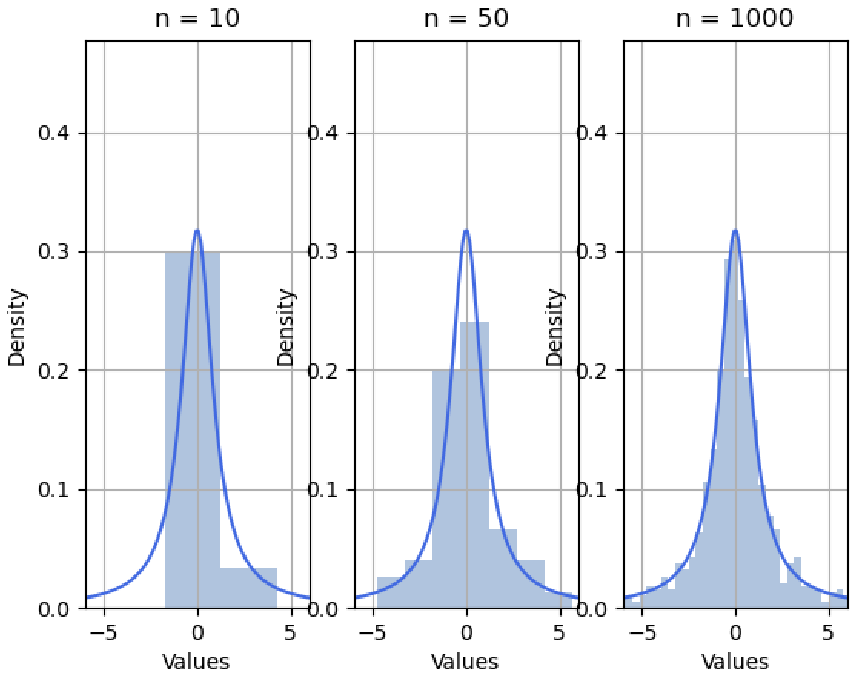

Рис. 2: Распределение Коши

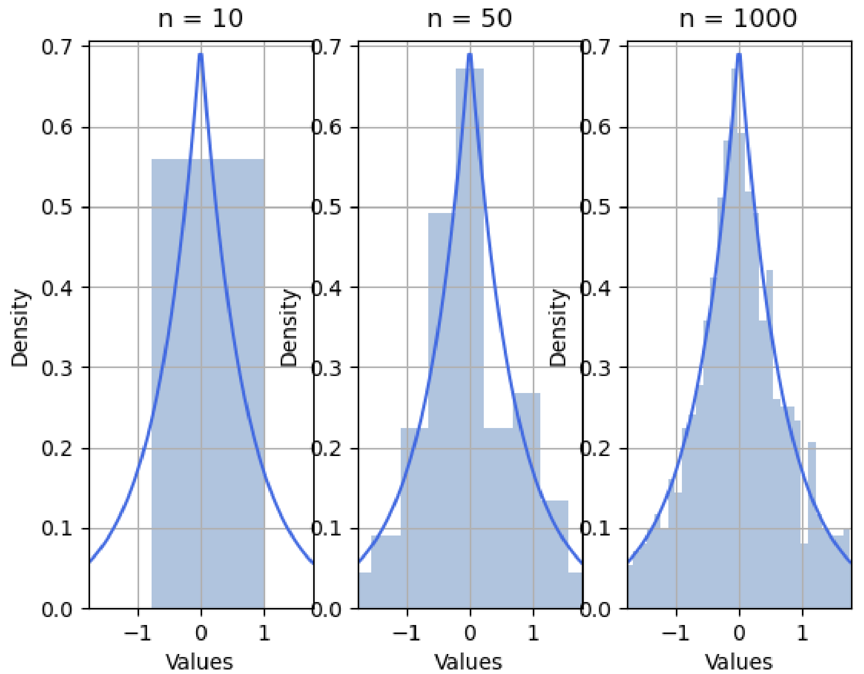

Рис. 3: Распределение Лапласа

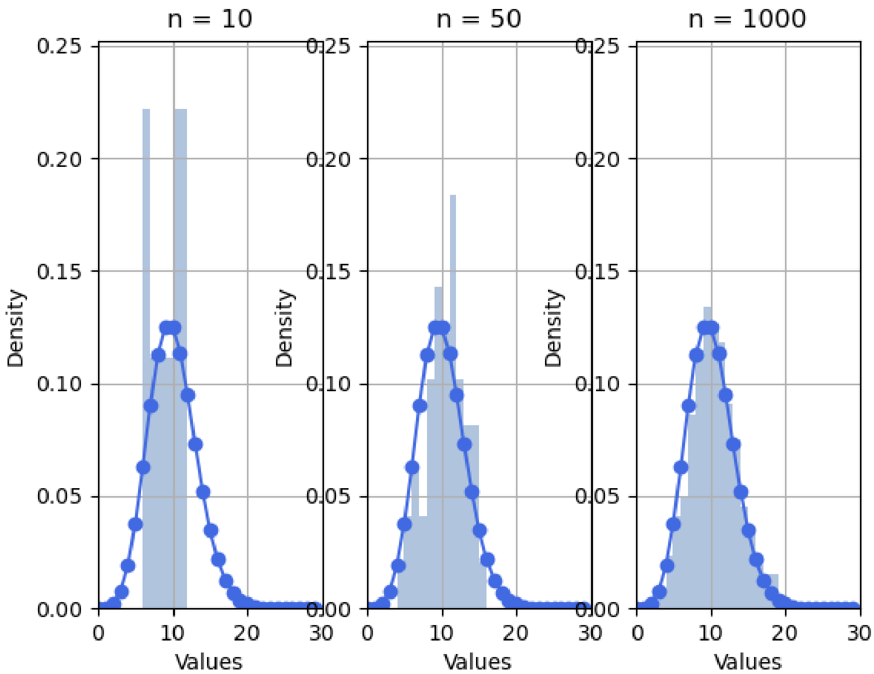

Рис. 4: Распределение Пуассона

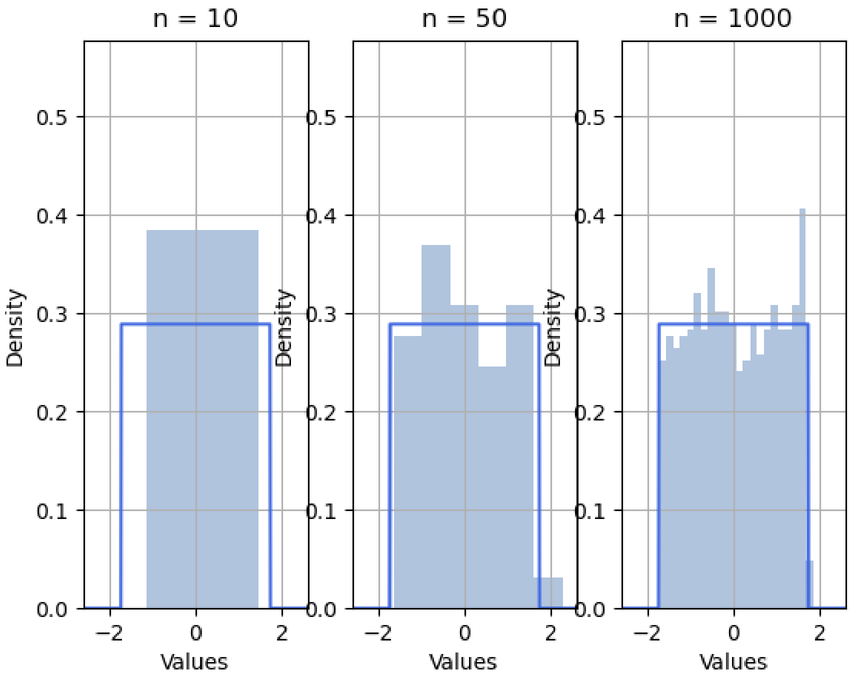

Рис. 5: Равномерное распределение

5 Обсуждение

Проведенный эксперимент подтверждает **утверждение**: *пусть плотность распределения по которому построена выборка является непрерывной функцией. Если число интервалов гистограммы* $k(n)$ *стремится к бесконечности таким образом что* $\lim\limits_{n\to\infty} \frac{k(n)}{n} = 0$, *то имеет место сходимость по вероятности гистограммы к плотности.* [2] Действительно мы взяли $k(n) = \lceil \sqrt{n} \rceil$ и очевидно условие утверждения в таком случае выполнено, при этом гистограмма при увеличении n заполняет площадь под графиком плотности (кусочно-линейной функции вероятности для распределения Пуассона), а это и означает сходимость по вероятности.

6 Приложения

1. Исходный код лабораторной `https://github.com/zhenyatos/statlabs/tree/master/Lab1`

Список литературы

[1] А. Н. Ширяев, *Вероятность-1*. Изд. МЦНМО, Москва, 2017. 551 стр.

[2] Н. И. Чернова, *Математическая статистика: Учеб. пособие*. Новосиб. гос. ун-т. Новосибирск, 2007. 148 стр.