Санкт-Петербургский политехнический университет Петра Великого
Физико-механический институт
Высшая школа прикладной математики и вычислительной физики

О т ч ё т

по лабораторной работе №1
по дисциплине
«Математическая статистика»

Выполнил
Студент гр.5030102/90201 Сачук А. С.

Руководитель
Доцент, к.ф.-м.н. Баженов А. Н.

Санкт-Петербург
2022

Оглавление

Список иллюстраций

1. Постановка задачи

Построить выборку размерами 10, 100 и 1000 элементов и построить на одном графике гистограмму и график плотности распределения для следующих 4 распределений:

- Нормальное распределение N(x, 0, 1);
- Распределение Коши C(x, 0, 1);
- Распределение Пуассона P(k, 10);
- Равномерное распределение U(x, $-\sqrt{3}$, $\sqrt{3}$).

2. Теория

2.1 Рассматриваемые распределения

Плотности распределений:

- Нормальное распределение:

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

- Распределение Коши: C(x, 0, 1):

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

- Распределение Пуассона: P(k, 10):

$$\mathrm{P}(k, 10) = \frac{10^k}{k!} e^{-10} \tag{3}$$

- Равномерное распределение: U(x, $-\sqrt{3}$, $\sqrt{3}$):

$$U\left(\mathrm{x}, -\sqrt{3}, \sqrt{3}\right) = \begin{cases} \frac{1}{2\sqrt{3}}, & \text{при } |x| \le \sqrt{3} \\ 0, & \text{при } |x| > \sqrt{3} \end{cases} \tag{4}$$

2.2 Гистограммы

2.2.1 Определение и описание

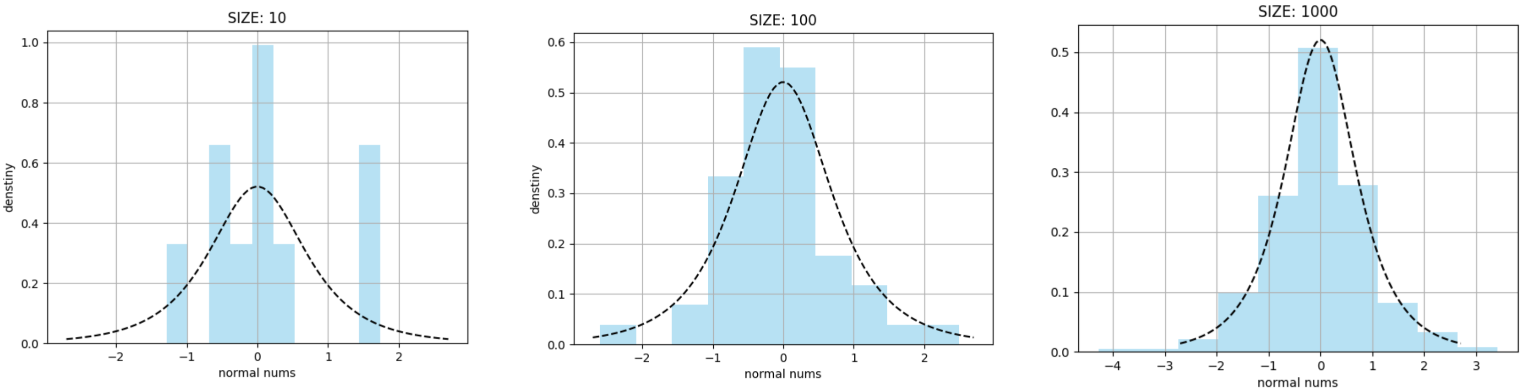
Гистограмма – функция, приближающая плотность вероятности некоторого распределения, построенная на основе выборки из него. Используются гистограммы для визуализации данных на начальном этапе статистической обработки. Построение гистограмм используется для получения эмпирической оценки плотности распределения случайной величины.

2.2.2 Построение

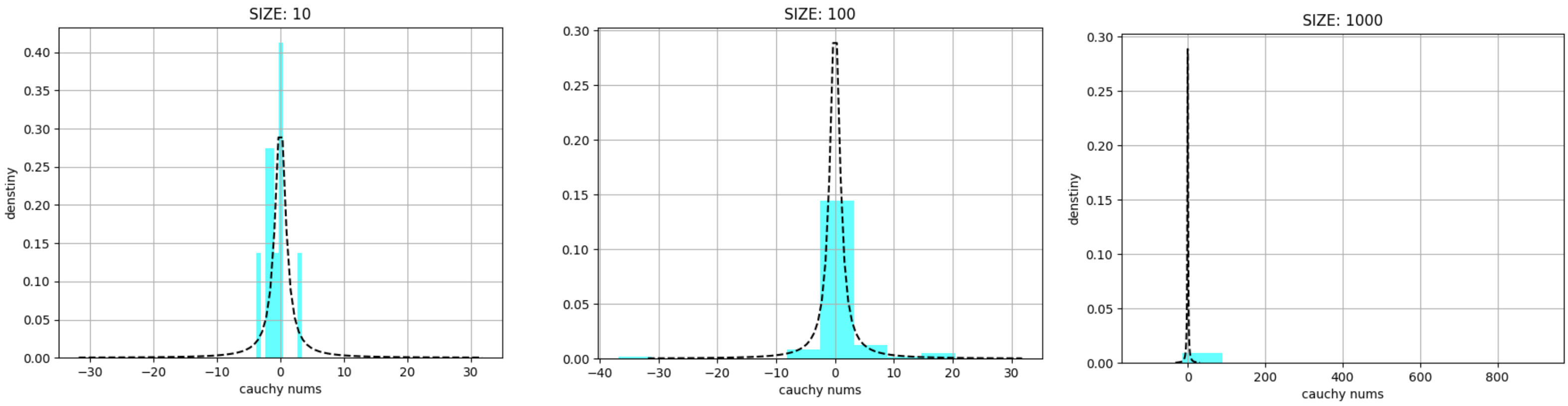
Гистограммы строятся следующим образом: все множество значений, которые могут принимать элементы выборки, разбивается на несколько интервалов. Чаще всего, эти интервалы делают одинакового размера, но это не обязательно (в данной лабораторной работе интервалы будут одинакового размера). Если интервалы одинакового размера, то высота каждого прямоугольника гистограммы будет прямо пропорционален числу элементов выборки, попавших в этот интервал. Если же интервалы разного размера, то высота прямоугольников выбирается так, чтоб их площадь была пропорциональна числу элементов выборки, попавших в этот интервал.

3. Результаты

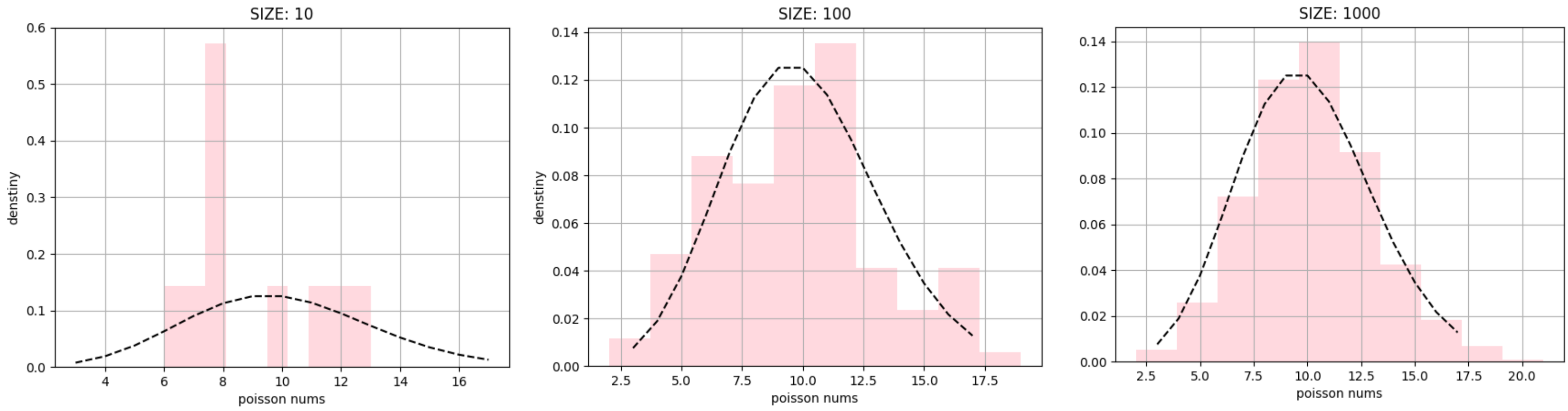
3.1 Нормальное распределение N(x, 0, 1)



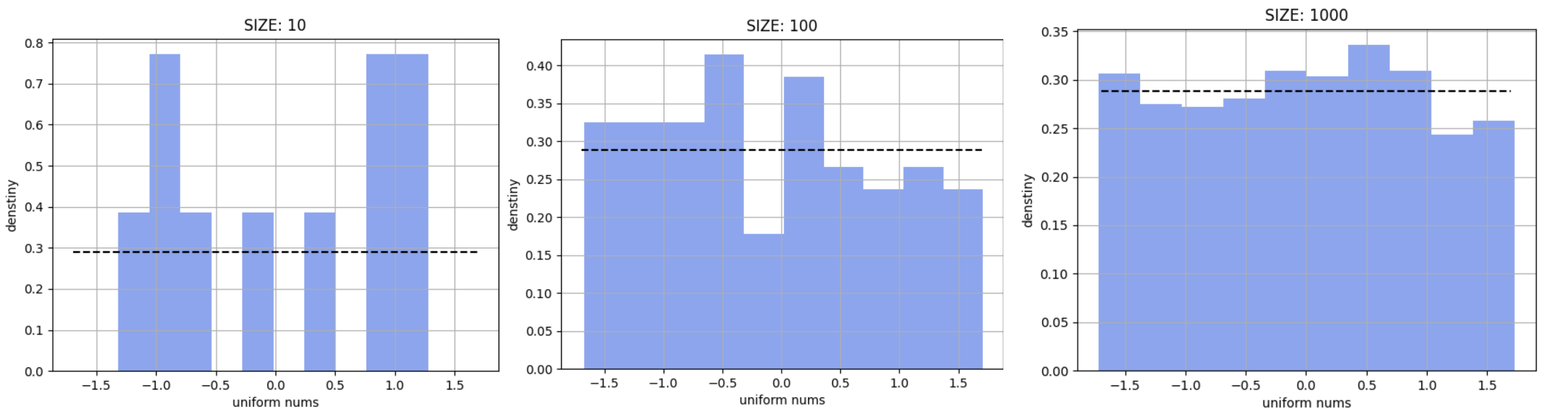
3.2 Распределение Коши C(x, 0, 1)



3.3 Распределение Пуассона P(k, 10)



3.4 Равномерное распределение



4. Реализация

Данная лабораторная работа была выполнена с использованием языка программирования Python 3.10 в среде разработки Visual Studio Code с использованием следующих библиотек:

- scipy версии 1.8.0
- numpy версии 1.22.0
- matplotlib версии 3.5.1

5. Обсуждение

Полученные результаты работы говорят о том, что при увеличении размеров выборок, гистограммы все ближе к графику плотности вероятности того закона, по которому были сгенерированы элементы выборок. Верно и обратно: чем меньше выборка, тем хуже по ней можно определить закон, по которой эта выборка генерировалась.

Также одним из ключевых выводов является тот факт, что по маленькому размеру выборки (n = 10) очень трудно отличить гистограммы, а, следовательно, и определить закон, по которой генерировалась выборка. Действительно, гистограмма выборки, построенной по распределению Пуассона при n = 10, могла бы с тем же успехом описывать график равномерного распределения (если не учитывать один единственный всплеск гистограммы, который вообще мог остаться незамеченным при более широких интервалах боксов гистограммы).

При выборках n = 1000 видно, что гистограммы уже достаточно неплохо приближаются к графикам плотностей соответствующих законов распределения: в равномерном распределении отклонения гистограммы от графика незначительны, а в нормальном распределении уже наблюдаются «хвосты», которые позволяют отличить треугольное распределение от нормального.

6. Ссылки на библиотеки

https://scipy.org/ - SciPy

https://numpy.org/ - NumPy

https://matplotlib.org/ - Matplotlib

7. Ссылка на GitHub

https://github.com/AS2/Mathematical-statistics - GitHub