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

Отчёт по лабораторной работе №1

по дисциплине
«Математическая статистика»

Выполнил студент
группа 5030102/90101 Лаэтин Андрей Алексеевич

Проверил
Доцент, к.ф.-м.н. Баженов Александр Николаевич

Санкт-Петербург
2022

СОДЕРЖАНИЕ

СПИСОК ИЛЛЮСТРАЦИЙ

1 Постановка задачи

Для 5 распределений:

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

Сгенерировать выборки размером 10, 50 и 1000 элементов.
Построить на одном рисунке гистограмму и график плотности распределения.

2 Теория

2.1 Распределения

- Нормальное распределение
$$N(x, 0, 1) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} \tag{1}$$
- Распределение Коши
$$C(x, 0, 1) = \frac{1}{\pi} \frac{1}{x^2 + 1} \tag{2}$$
- Распределение Лапласа
$$L(x, 0, \frac{1}{\sqrt{2}}) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}|x|} \tag{3}$$
- Распределение Пуассона
$$P(k, 10) = \frac{10^k}{k!} e^{-10} \tag{4}$$
- Равномерное распределение

$$U(x, -\sqrt{3}, \sqrt{3}) = \begin{cases} \frac{1}{2\sqrt{3}} & \text{при} |x| \leq \sqrt{3} \\ 0 & \text{при} |x| > \sqrt{3} \end{cases} \tag{5}$$

2.2 Гистограмма

2.2.1 Определение

Гистограмма в математической статистике — это функция, приближающая плотность вероятности некоторого распределения, построенная на основе выборки из него.

2.2.2 Графическое описание

Графически гистограмма строится следующим образом. Сначала множество значений, которое может принимать элемент выборки, разбивается на несколько интервалов. Чаще всего эти интервалы берут одинаковыми, но это не является строгим требованием. Эти интервалы откладываются на горизонтальной оси, затем над каждым рисуется прямоугольник. Если все интервалы были одинаковыми, то высота каждого прямоугольника пропорциональна числу элементов выборки, попадающих в соответствующий интервал. Если интервалы разные, то высота прямоугольника выбирается таким образом, чтобы его площадь была пропорциональна числу элементов выборки, которые попали в этот интервал.

2.2.3 Использование

Гистограммы применяются в основном для визуализации данных на начальном этапе статистической обработки.
Построение гистограмм используется для получения эмпирической оценки плотности распределения случайной величины. Для построения гистограммы наблюдаемый диапазон изменения случайной величины разбивается на несколько интервалов и подсчитывается доля от всех измерений, попавшая в каждый из интервалов. Величина каждой доли, отнесенная к величине интервала, принимается в качестве оценки значения плотности распределения на соответствующем интервале.

3 Программная реализация

Лабораторная работа выполнена на языке Python вресии 3.9 в среде разработки PyCharm. Использовались дополнительные библиотеки:

1. scipy
2. numpy
3. matplotlib
4. math

В приложении находится ссылка на GitHub репозиторий с исходныи кодом.

4 Результаты

4.1 Гистограммы и графики плотности распределения

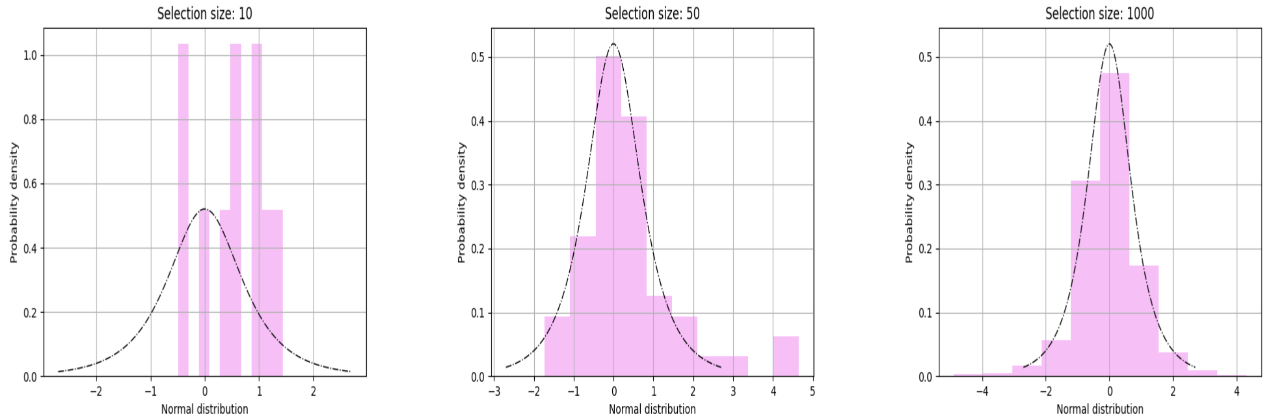

Рис. 1: Нормальное распределение (1)

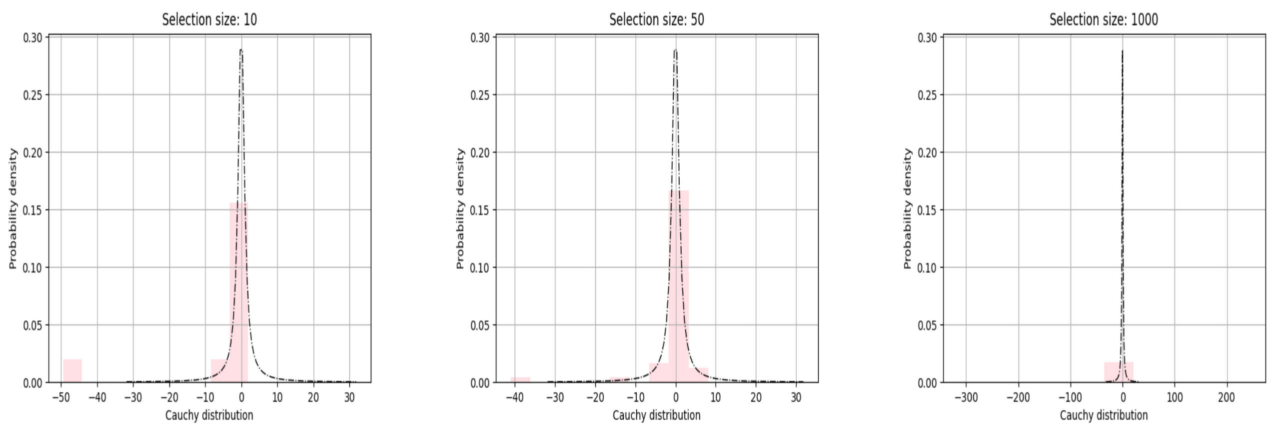

Рис. 2: Распределение Коши (2)

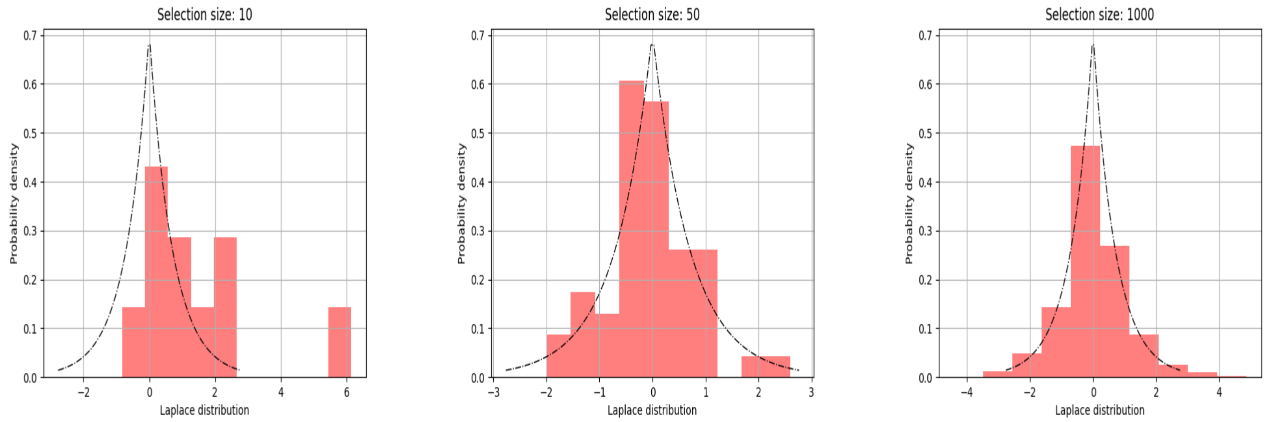

Рис. 3: Распределение Лапласа (3)

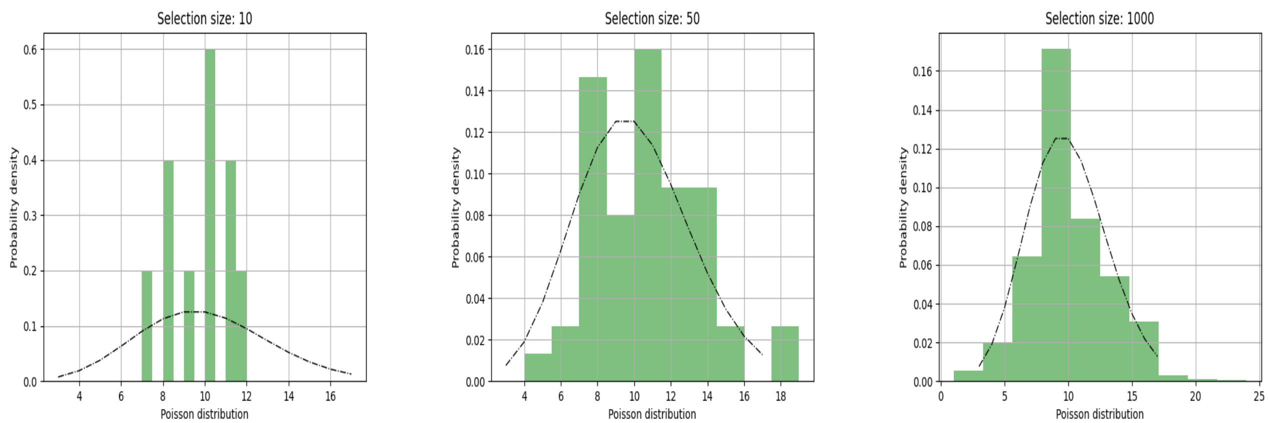

Рис. 4: Распределение Пуассона (4)

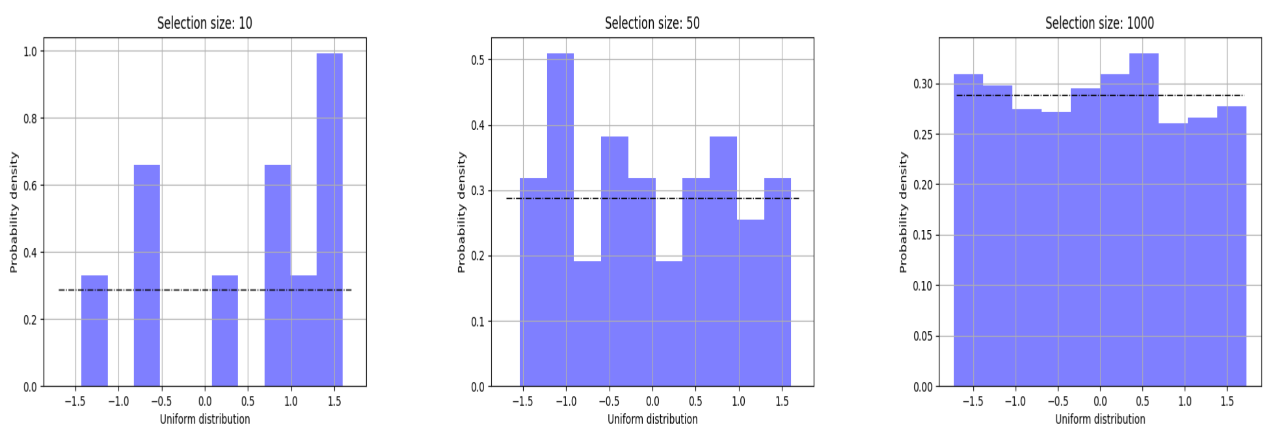

Рис. 5: Равномерное распределение (5)

5 Обсуждение

По результатам проделанной работы можем сделать вывод о том, что чем больше выборка для каждого из распределений, тем ближе ее гистограмма к графику плотности вероятности того закона, по которому распределены величины сгенерированной выборки. Чем меньше выборка, тем менее она показательна - тем хуже по ней определяется характер распределения величины.

Визуально очень трудно отличить гистограммы друг от друга, тем более при маленьких выборках. При выборке из 10 элементов вид гистрограммы много сильно отличается от плотности распределения. Чем больше выборка, тем точнее становится гистограмма. На выборке из 1000 элементов можем отличить и распознать с большей вероятностью равномерное распределение (все прямоугольники примерно на одном уровне), а также распределение Пуассона (оно визуально шире чем распределение Лапласа и нормальное). Однако отличить между собой распределение Лапласа и нормальное тяжело. Так как визульно гистограммы получились похожи друг на друга и без подписей отличить их почти не представляется возможным.

Также можно заметить, что максимумы гистограмм и плотностей распределения почти нигде не совпали. Из полученных графиков можно увидеть, что только при распределении Пуассона на выборке из 1000 элементов, максимум графика плотности вероятности совпал с максимумом гистограммы. Также наблюдаются всплески гистограмм, что наиболее хорошо прослеживается на распределении Коши.

6 Приложение

Код программы GitHub URL:

https://github.com/A21l63/math-prob-stat