

Санкт-Петербургский политехнический университет Петра Великого
Институт прикладной математики и механики
Высшая школа прикладной математики и вычислительной физики

ОТЧЁТ ПО ЛАБОРАТОРНОЙ РАБОТЕ №1

по дисциплине
«Математическая статистика»

Выполнил студент
группы 3630102/80401

Веденичев Дмитрий Александрович

Проверил
Доцент, к.ф.-м.н.

Баженов Александр Николаевич

Санкт-Петербург
2021

Содержание

Список иллюстраций	3
1 Постановка задачи	4
2 Теория	4
2.1 Распределения	4
2.2 Гистограмма	4
2.2.1 Определение	4
2.2.2 Графическое описание	5
2.2.3 Использование	5
3 Программная реализация	5
4 Результаты	6
4.1 Гистограммы и графики плотности распределения	6
5 Обсуждение	8
6 Приложение	8

Список иллюстраций

1	Нормальное распределение (1)	6
2	Распределение Коши (2)	6
3	Распределение Лапласа (3)	7
4	Распределение Пуассона (4)	7
5	Равномерное распределение (5)	7

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. $U(x, -\sqrt{3}, \sqrt{3})$ – равномерное 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

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

- Распределение Коши

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

- Распределение Лапласа

$$L(x, 0, \frac{1}{\sqrt{2}}) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}|x|} \quad (3)$$

- Распределение Пуассона

$$P(k, 10) = \frac{10^k}{k!} e^{-10} \quad (4)$$

- Равномерное распределение

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

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 версии 3.7 в среде разработки PyCharm. Использовались дополнительные библиотеки:

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

В приложении находится ссылка на GitHub репозиторий с исходным кодом.

4 Результаты

4.1 Гистограммы и графики плотности распределения

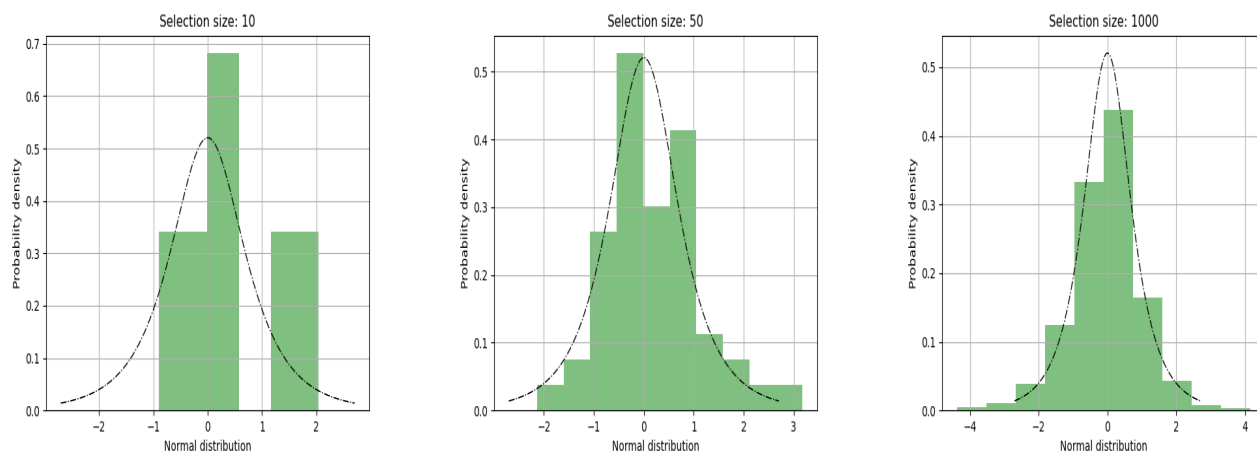


Рис. 1: Нормальное распределение (1)

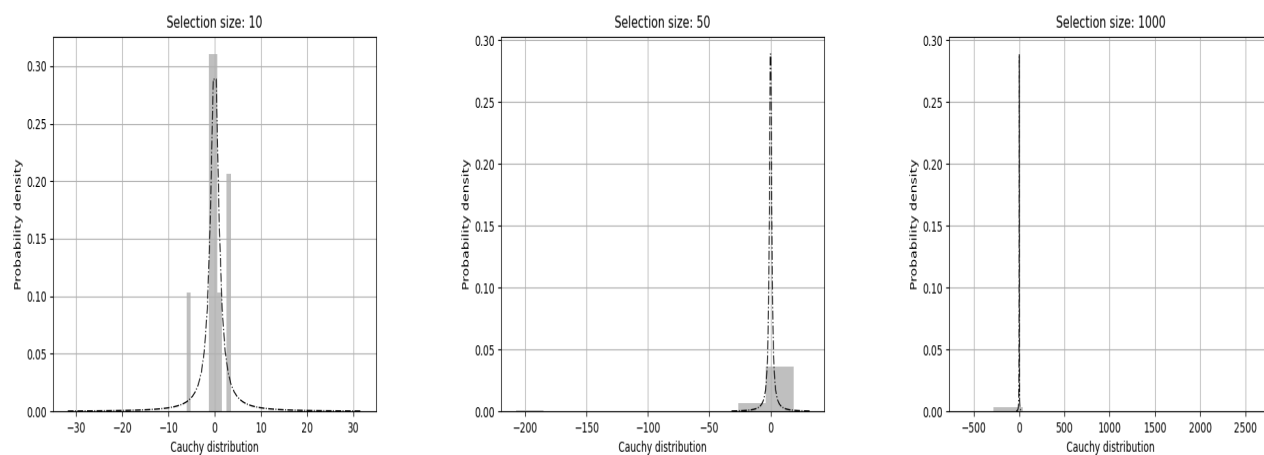


Рис. 2: Распределение Коши (2)

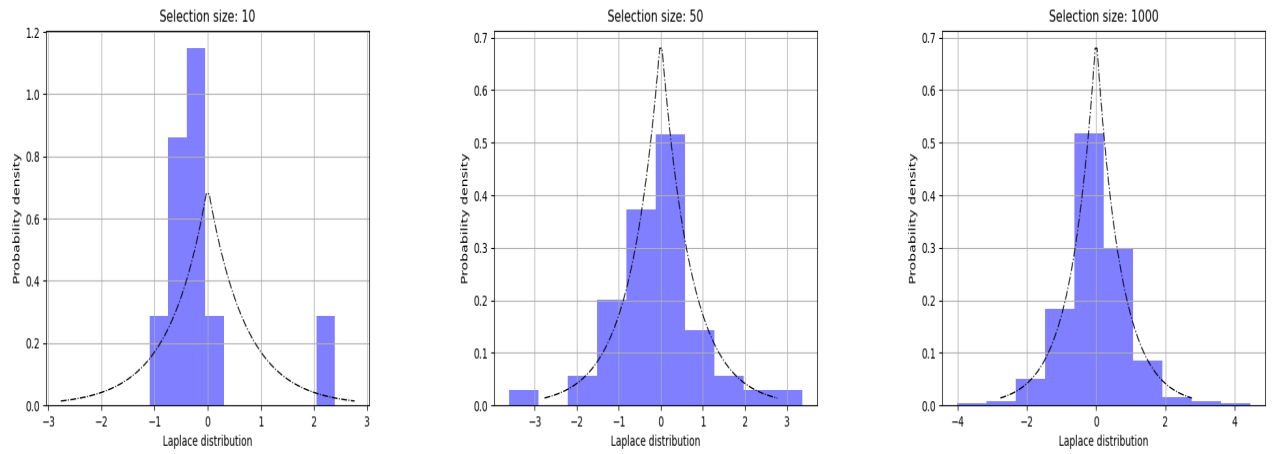


Рис. 3: Распределение Лапласа (3)

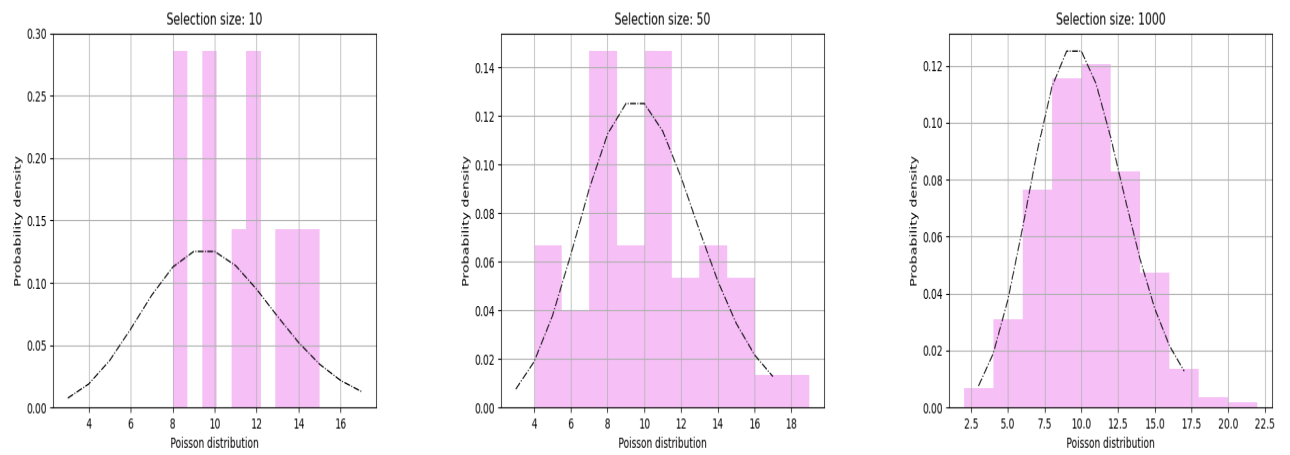


Рис. 4: Распределение Пуассона (4)

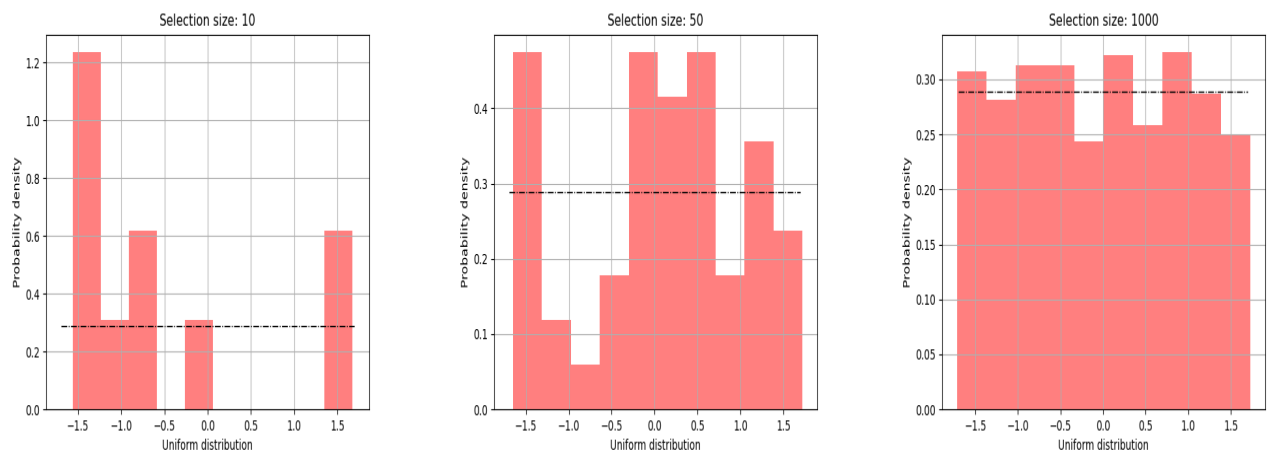


Рис. 5: Равномерное распределение (5)

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.

Гистограммы составленные по маленькой выборке из 10 элементов не информативны. По ним нельзя определить по какому распределению они построены. С увеличением выборки до 50 элементов некоторые гистограммы (Лаплас, Нормальное, Пуассон) начинают принимать форму близкую к описываемым ими распределениям. На выборке из 1000 элементов можно однозначно определить равномерное распределение, распределение Пуассона. Гистограммы нормального распределения и распределения Лапласа вышли зеркальными.

Также можно заметить, что максимумы гистограмм и плотностей распределения почти нигде не совпали. Из полученных гистограмм, ближе всего к совпадению находятся гистограммы Лапласа(1000) и Пуассона(1000). Также наблюдаются всплески гистограмм, что наиболее хорошо прослеживается на распределении Коши.

6 Приложение

Код программы GitHub URL:

https://github.com/PopeyeTheSailorsCat/math_stat_2021/tree/main/lab1/src