

Санкт-Петербургский политехнический университет
Петра Великого

Институт прикладной математики и механики
Кафедра «Прикладная математика»

ОТЧЁТ ПО ЛАБОРАТОРНЫМ РАБОТАМ
ПО ДИСЦИПЛИНЕ
«МАТЕМАТИЧЕСКАЯ СТАТИСТИКА»

Выполнил студент
Войнова Алёна Игоревна
группы 3630102/80201

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

Санкт-Петербург
2021

Содержание

| | | |
|----------|---|----------|
| 1 | Постановка задачи | 2 |
| 1.1 | Задание 1 | 2 |
| 2 | Теория | 2 |
| 2.1 | Распределения | 2 |
| 2.2 | Гистограмма | 3 |
| 2.2.1 | Определение | 3 |
| 2.2.2 | Графическое описание | 3 |
| 2.2.3 | Использование | 3 |
| 3 | Реализация | 4 |
| 4 | Результаты | 4 |
| 4.1 | Гистограммы и графики плотности распределения | 4 |
| 5 | Обсуждение | 6 |
| 5.1 | Гистограмма и график плотности распределения | 6 |
| 6 | Приложения | 6 |
| | Литература | 7 |

Список иллюстраций

| | | |
|---|-------------------------------------|---|
| 1 | Нормальное распределение | 4 |
| 2 | Распределение Коши | 4 |
| 3 | Распределение Лапласа | 5 |
| 4 | Распределение Пуассона | 5 |
| 5 | Равномерное распределение | 6 |

Список таблиц

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

1.1 Задание 1

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

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

2.2 Гистограмма

2.2.1 Определение

Гистограмма в математической статистике — это функция, приближающая плотность вероятности некоторого распределения, построенная на основе выборки из него [1].

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 [1].

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 [1].

Они являются мощным инструментом для исследования неизвестных распределений. В частности, на этом построены различные способы обработки сигналов, изображений и других статистических объектов. Применение гистограмм к обработке экспериментальных данных позволяет устранять артефакты - шумы и выбросы, мешающие работе с данными и не являющиеся содержательными.

Шумы определяются как горизонтальные участки сигнала. Присутствуют они только до и после полезного сигнала, на нем они быть не могут. Определяются участки шума так: находятся границы 2-ух самых больших столбцов гистограммы полученного сигнала, затем путем прохода по сигналу в обе стороны окном определённого размера и сравнения процентного соотношения значений внутри границ большого столбца с выбранным порогом, принимается решение о пометке участка как шумового при превышении этого порога. [1].

3 Реализация

Лабораторная работа выполнена с помощью средств языка программирования **Python** в среде разработки **Jupyter**. Исходный код лабораторной работы приведён в приложении.

4 Результаты

4.1 Гистограммы и графики плотности распределения

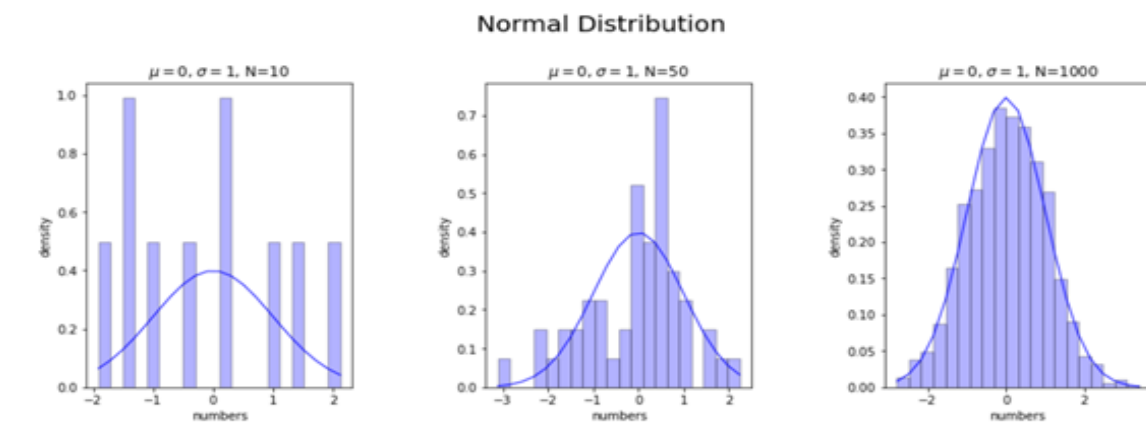


Рис. 1: Нормальное распределение

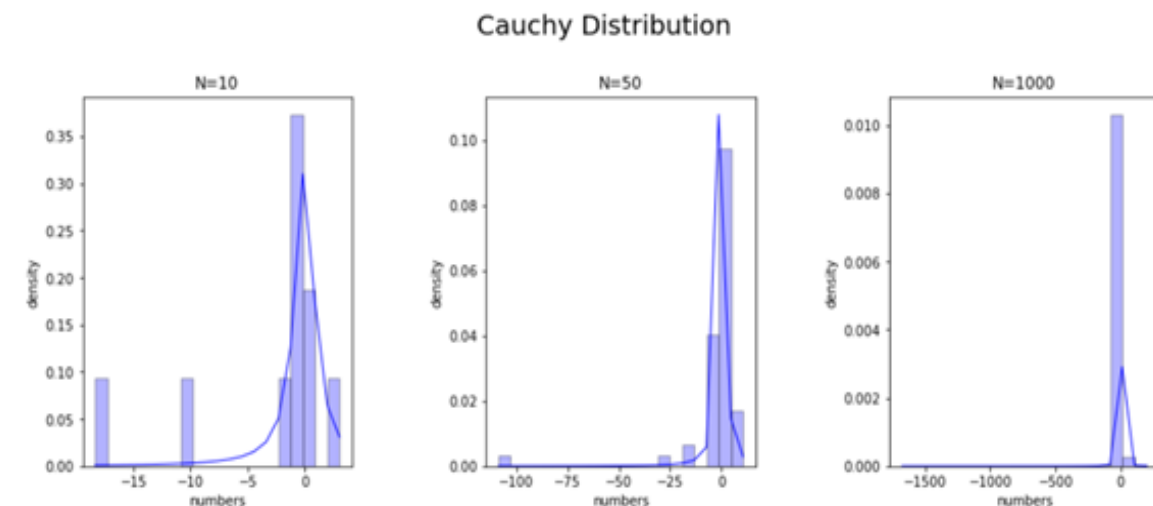


Рис. 2: Распределение Коши

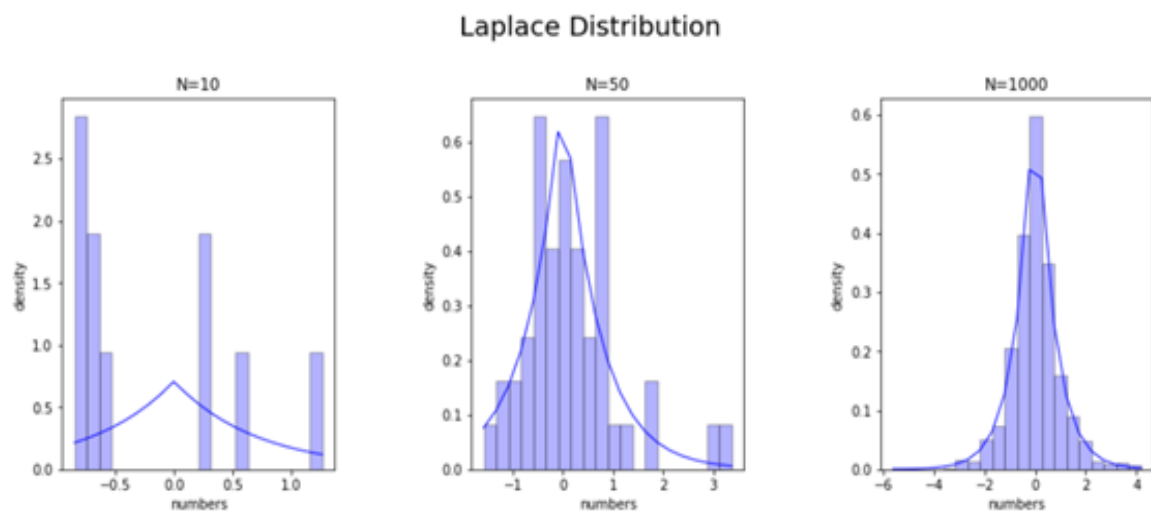


Рис. 3: Распределение Лапласа

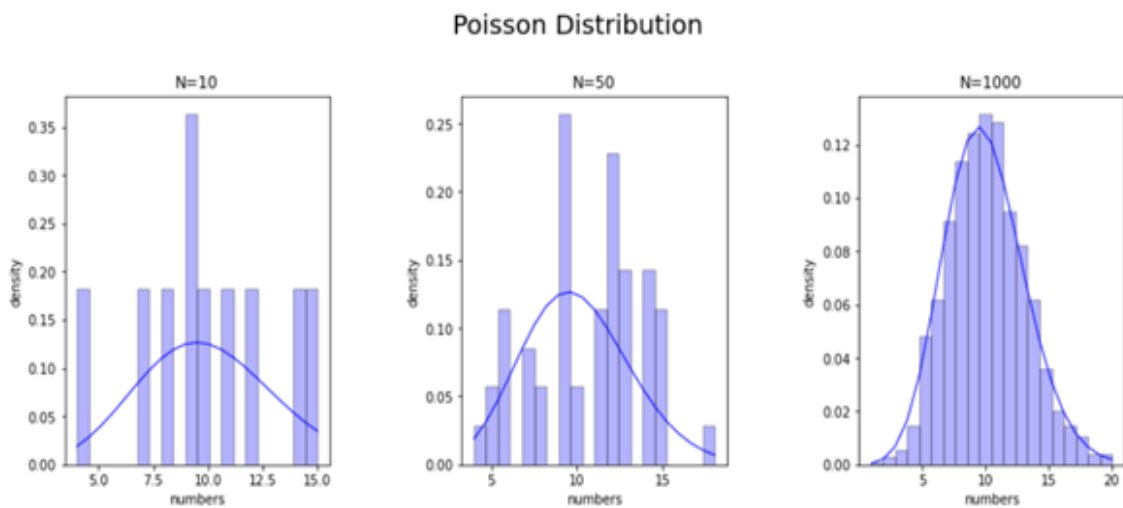


Рис. 4: Распределение Пуассона

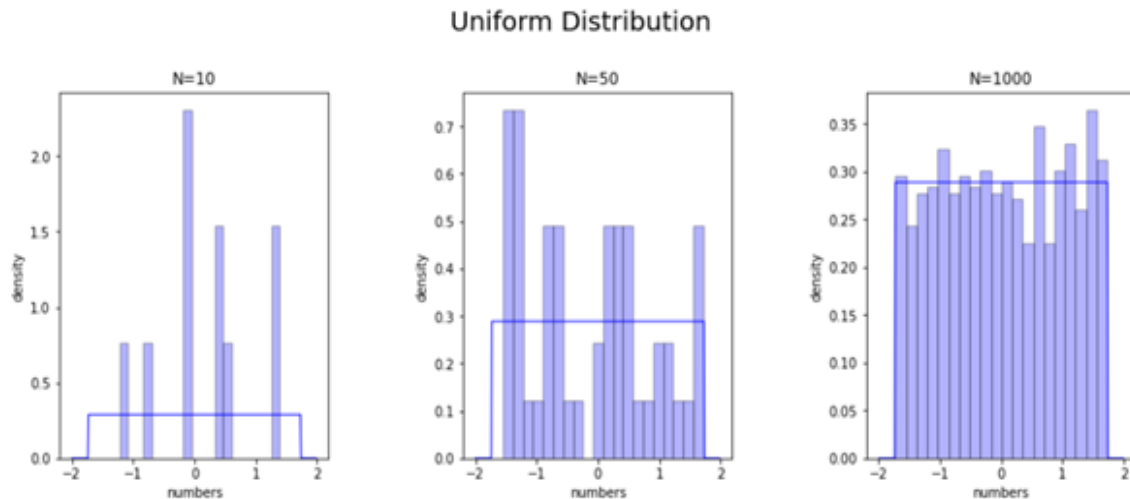


Рис. 5: Равномерное распределение

5 Обсуждение

5.1 Гистограмма и график плотности распределения

В ходе проведения лабораторной работы, были построены графики для наглядного выявления зависимостей внешнего вида гистограмм и функций плотности вероятности от количества данных в выборке.

Заметим, что чем больше данных, тем ближе гистограмма к графику.

Аналогично, чем больше данных, тем более заметен характер распределения.

При достаточно большом объеме данных имеет место быть визуальная оценка закона распределения переменной, что позволяет выдвинуть гипотезу о том, какой функции распределения подчиняется полученный результат. Чем меньше выборка, тем менее она показательна.

6 Приложения

URL: Выполненная лабораторная работа на GitHub

<https://github.com/pikabol88/Math-Statistics/blob/main/Lab1.ipyn>

Список литературы

[1] Histogram. URL: <https://en.wikipedia.org/wiki/Histogram>