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»

Выполнила студентка
группы 5030102/20202 Чинь Тхи Тху Хоай

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

Санкт-Петербург
2025

СОДЕРЖАНИЕ

1 Постановка задачи

Для 4 распределений:

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

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

2. Сгенерировать выборки размером 10, 100 и 1000 элементов.
   Для каждой выборки вычислить следующие статистические характеристики положения данных:
   $\overline{x}, medx, z_R, z_Q$. Повторить такие вычисления 1000 раз для каждой выборки и найти среднее характеристик положения и их квадратов:

   $$E(z) = \overline{z} \tag{1}$$

   Вычислить оценку дисперсий по формуле:

   $$D(z) = \overline{z^2} - \overline{z}^2 \tag{2}$$

   Представить полученные данные в виде таблиц

2 Цель работы

- Изучение и сравнение статистических характеристик различных распределений: Нормальное (Normal), Коши (Cauchy), Пуассона (Poisson), и Равномерное (Uniform).
- Исследование влияния размера выборки на статистические характеристики, такие как среднее значение, медиана и значение z_Q.
- Оценка дисперсии статистических характеристик.

3 Теория

3.1 Рассматриваемые распределения

- Нормальное распределение

$$N(x,0,1) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} \tag{3}$$

- Распределение Коши

$$C(x,0,1) = \frac{1}{\pi} \frac{1}{x^2+1} \tag{4}$$

- Распределение Пуассона

$$P(k,10) = \frac{10^k}{k!} e^{-10} \tag{5}$$

- Равномерное распределение

$$U(x,-\sqrt{3},\sqrt{3}) = \begin{cases} \frac{1}{2\sqrt{3}} & , \text{|x|} \leq \sqrt{3} \\ 0 & , \text{|x|} > \sqrt{3} \end{cases} \tag{6}$$

3.2 Гистограмма

3.2.1 Построение гистограммы

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.

3.3 Вариационный ряд

Вариационным ряд - последовательность элементов выборки, расположенных в неубывающем порядке. Одинаковые элементы повторяются.

3.4 Выборочные числовые характеристики

3.4.1 Характеристики положения

- Выборочное среднее

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{7}$$

- Выборочная медиана

$$medx = \begin{cases} x_{(l+1)} & , n = 2l+1 \\ \frac{x_{(l)}+x_{(l+1)}}{2} & , n = 2l \end{cases} \tag{8}$$

- Полусумма экстремальных выборочных элементов

$$z_R = \frac{x_{(1)} + x_{(n)}}{2} \tag{9}$$

- Полусумма квартилей

  Выборочная квартиль z_p порядка p определяется формулой:

$$z_p = \begin{cases} x_{(\lfloor np \rfloor+1)} & , np \text{ дробное} \\ x_{(np)} & , np \text{ целое} \end{cases} \tag{10}$$

  Полусумма квартилей

$$z_Q = \frac{z_{1/4} + z_{3/4}}{2} \tag{11}$$

$$z_{tr} = \frac{1}{n-2r} \sum_{i=r+1}^{n-r} x_{(i)}, r \approx \frac{n}{4} \tag{12}$$

3.4.2 Характеристики рассеяния

Выборочная дисперсия

$$D = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2 \tag{13}$$

4 Реализация

Лабораторная работа выполнена на языке Python версии 3.12.6 в среде разработки Visual Studio Code. Использовались дополнительные библиотеки:

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

В приложении находится ссылка на GitHub репозиторий с исходным кодом.

5 Результаты

5.1 Гистограммы и графики плотности распределения

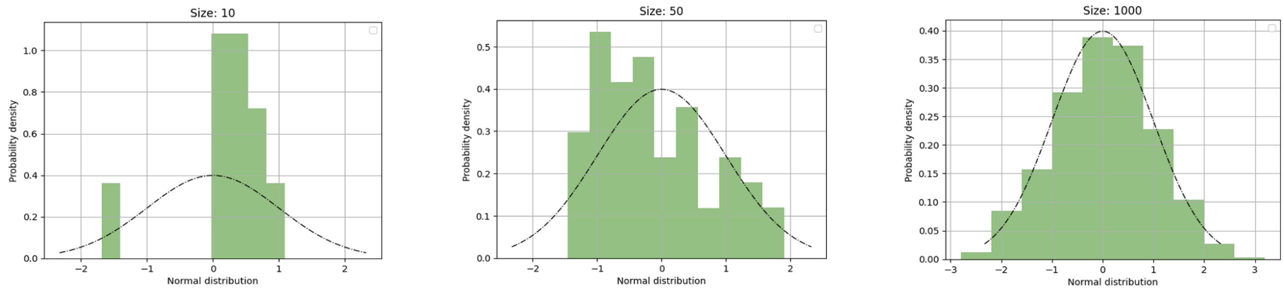

Рис. 1: Нормальное распределение

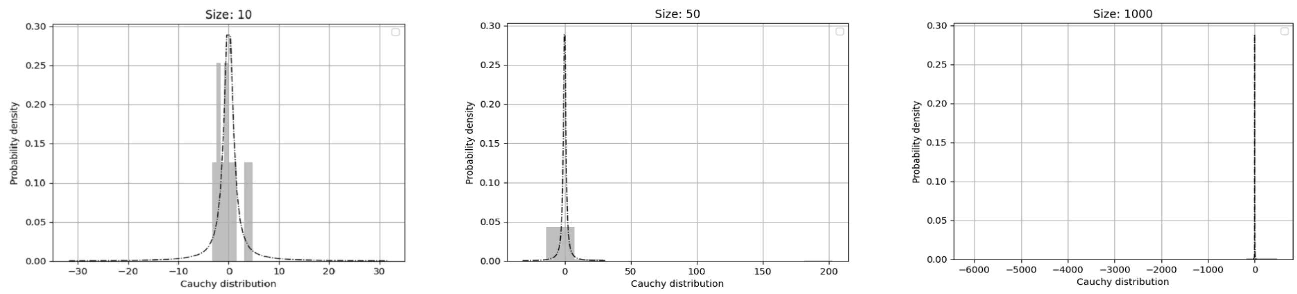

Рис. 2: Распределение Коши

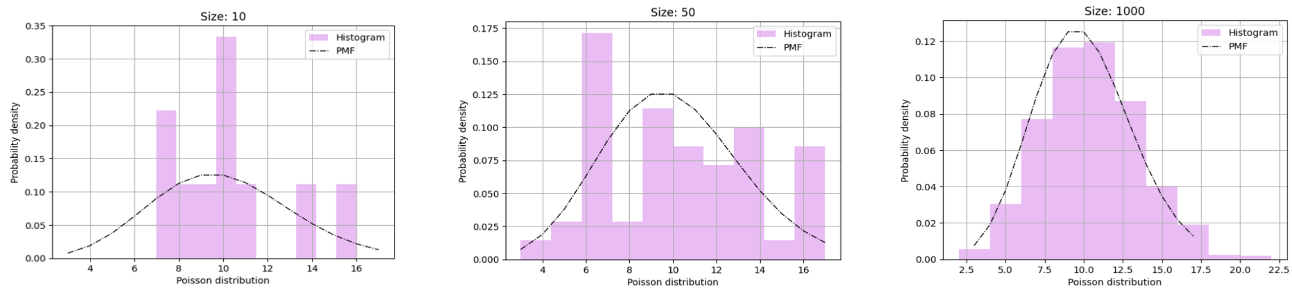

Рис. 3: Распределение Пуассона

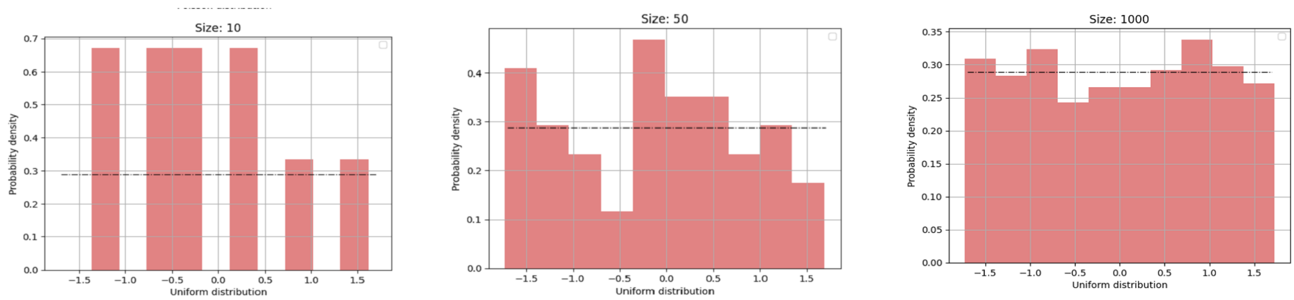

Рис. 4: Равномерное распределение

5.2 Характеристики положения и рассеяния

Results for Normal distribution:

Sample Size	E(mean)	E(median)	E(zQ)	D(mean)	D(median)	D(zQ)
10	-0	0	-0	0.1	0.1	0.1
100	0	0	0	0	0	0
1000	-0	-0	-0	0	0	0

Рис. 5: Нормальное распределение

Results for Cauchy distribution:

Sample Size	E(mean)	E(median)	E(zQ)	D(mean)	D(median)	D(zQ)
10	-3.8	-0	-0.1	24279.2	0.3	0.9
100	-0.2	0	0	122930	0	0.1
1000	0.7	-0	-0	900.1	0	0

Рис. 6: Распределение Коши

Results for Poisson distribution:

Sample Size	E(mean)	E(median)	E(zQ)	D(mean)	D(median)	D(zQ)
10	10	9.9	9.9	1	1.5	1.2
100	10	9.8	9.9	0.1	0.2	0.2
1000	10	10	10	0	0	0

Рис. 7: Распределение Пуассона

Results for Uniform distribution:

Sample Size	E(mean)	E(median)	E(zQ)	D(mean)	D(median)	D(zQ)
10	0	0	0	0.1	0.2	0.1
100	-0	-0	0	0	0	0
1000	0	0	0	0	0	0

Рис. 8: Равномерное распределение

6 Обсуждение

6.1 Гистограммы

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.
Также можно заметить, что максимумы гистограмм и плотностей распределения почти нигде не совпали. Также наблюдаются всплески гистограмм, что наиболее хорошо прослеживается на распределении Коши.

6.2 Характеристики положения и рассеяния

Для нормального и равномерного распределений все оценки (среднее, медиана, квартиль) сходятся к теоретическим значениям при увеличении выборки, а дисперсия уменьшается. В распределении Коши среднее нестабильно, тогда как медиана и квартильная оценка дают более надежные результаты. В распределении Пуассона среднее является наиболее точной характеристикой, но медиана и квартиль также показывают хорошую стабильность. В целом, среднее эффективно для обычных распределений, но для распределений с тяжелыми хвостами (Коши) предпочтительнее медиана.

7 Приложение

Код программы GitHub: `https://github.com/Akira1707/Math-Statistic/tree/main/Lab1`