Санкт-Петербургский политехнический университет
Петра Великого

Институт прикладной математики и механики

Кафедра «Прикладная математика»

Отчет по лабораторной работе №1

по дисциплине
"Математическая статистика"

Выполнил студент
Группы 3630102/80101 Шао Цзяци

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

Санкт-Петербург
2021 г.

Содержание

1. Постановка задачи **4**

2. Теория **4**

2.1. Рассматриваемые распределения 4

2.2. Гистограмма 5

3. Реализация **5**

4. Результаты **5**

5. Обсуждение **7**

Список иллюстраций

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

1. Постановка задачи

Даны 5 распределений:

- Нормальное распределение $N(0, 1)$
- Распределение Коши $C(0, 1)$
- Распределение Лапласа $L(0, \frac{1}{\sqrt{2}})$
- Распределение Пуассона $P(k, 10)$
- Равномерное распределение $U(-\sqrt{3}, \sqrt{3})$

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

2. Теория

2.1. Рассматриваемые распределения

Плотности распределений

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

2.2. Гистограмма

Гистограмма - это графическое представление распределения частот для количественного признака, образуемое соприкасающимися прямоугольниками, основаниями которых служат интервалы классов, а площади пропорциональны частотам этих классов.

Пусть $X_1, ..., X_n$ - выборка из некоторого распределения. Определим разбиение числовой прямой $-\infty < a_0 < a_1 < ... < a_m < \infty$.

Пусть $n_i = \sum\limits_{j=1}^{m} 1_{X_i \in (a_{i-1}, a_i]}$ - число элементов выборки, попавших в i-й интервал.

Тогда кусочно-постоянная функция $\overline{h} = \frac{n_i}{n \Delta a_i}, \Delta a_i = a_i - a_{i-1}, i = 1, ..., m.$

3. Реализация

Лабораторная работа выполнена с помощью встроенных средств языка программирования python в среде разработки Pycharm с дополнительными библиотеками.

- scipy
- numpy
- matplotlib
- math

Исходный код лабораторной работы размещен в Github-репозитории.
URL: https://github.com/Shaots/shaoMathStatistic/tree/master/Lab1

4. Результаты

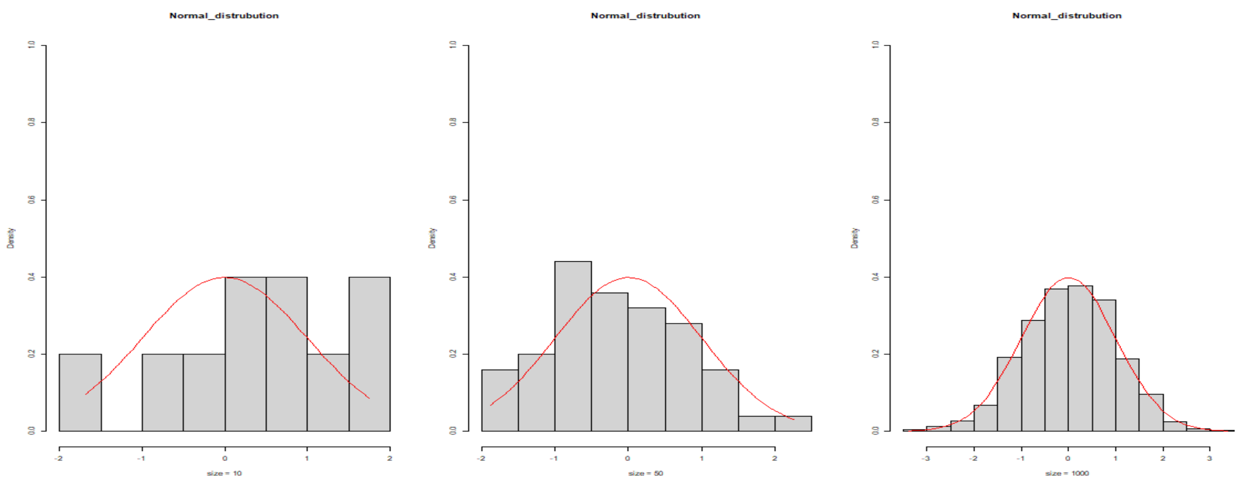

Рис. 1. Распределение 1: Нормальное распределение

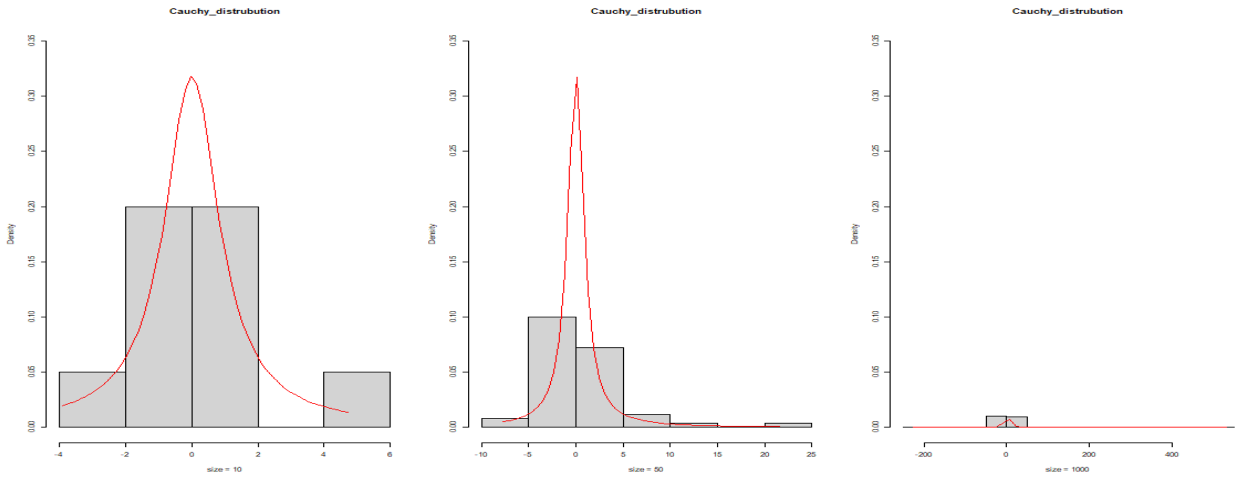

Рис. 2. Распределение 2: Распределение Коши

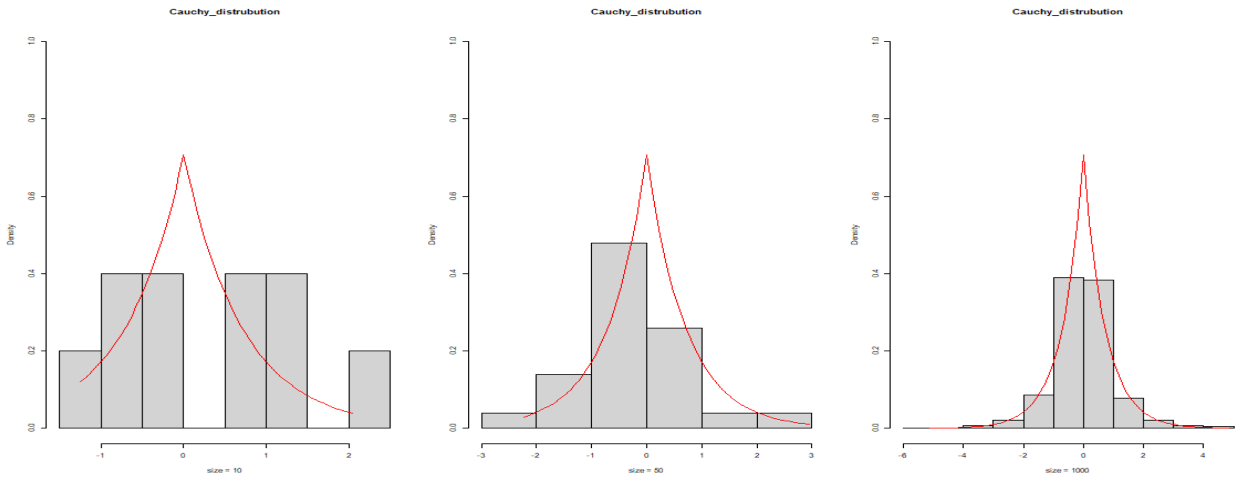

Рис. 3. Распределение 3: Распределение Лапласа

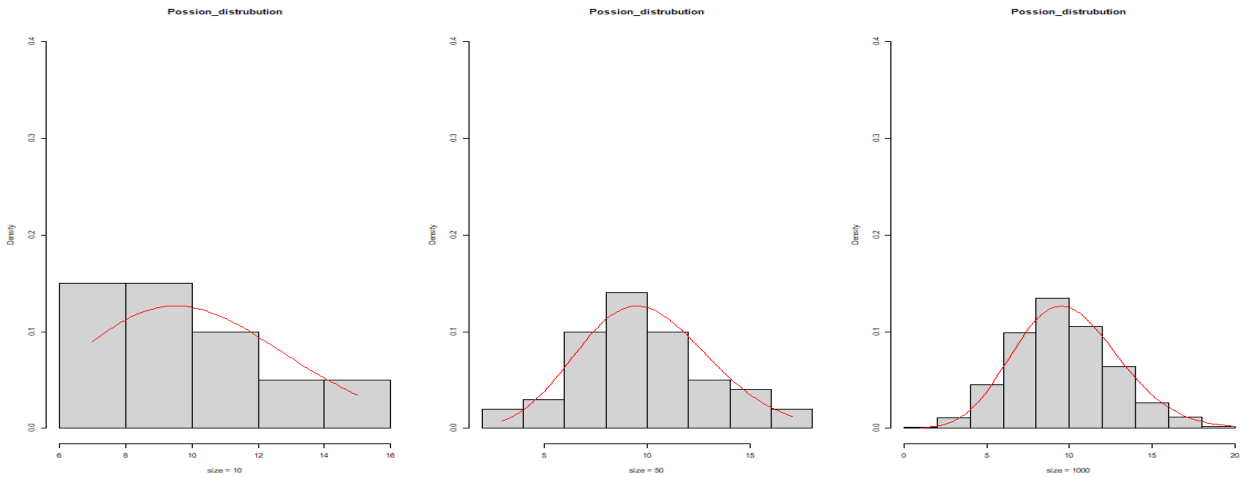

Рис. 4. Распределение 4: Распределение Пуассона

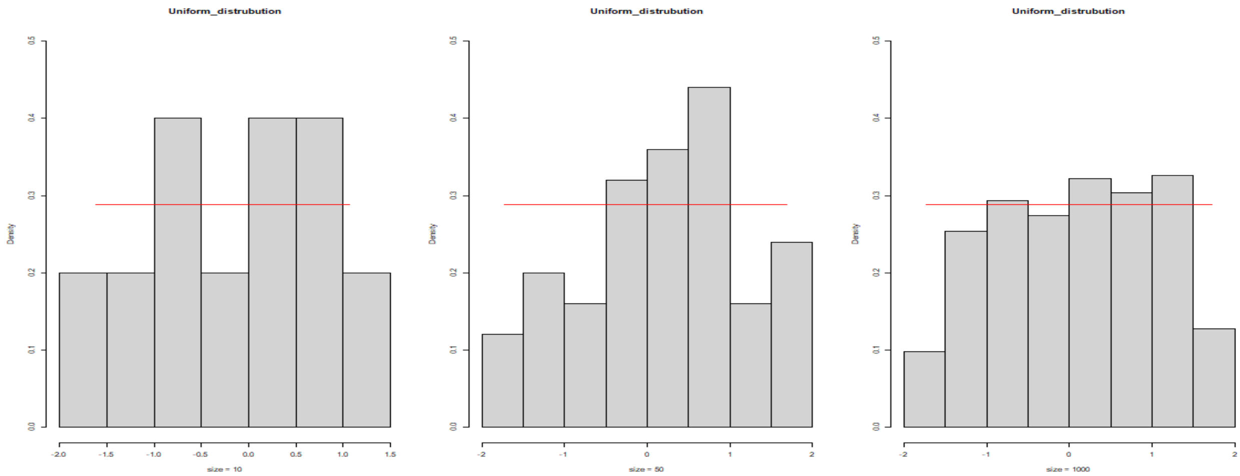

Рис. 5. Распределение 5: Равномерное распределение

5. Обсуждение

По полученным результатам видно, что увеличение выборки из распределения приближает гистограмму к графику плотности. Таким образом, рассмотрение выборки из большего числа элементов является необходимым.

Также формы гистограмм, особенно для маленьких выборок, очень похожи друг над друга. Это объясняется тем, что чем меньше выборка. тем меньше гистограмма похожа на график плотности, и имеет более общий вид.

Также большая часть распределений различается по высоте пиковой точки и по степени близости прохождения ветвей графиков плотностей распределения к асимптоте и оси симметрии, не считая равномерного распределения.

Стоит отметить, что распределение Пуассона является более широким, что можно выделить, как отличительную черту распределения. Аналогично для равномерного распределения характерно наличие столбцов более или менее однородной высоты. Касательно распределения Коши можно сказать, что высота пиковой точки гистограммы много меньше высоты пиковой точки графика распределения. На гистограммах наблюдаются всплески, явно превышающие исходное значение плотности распределения в заданной точке. Но они стихают по мере увеличения размеров выборки.