Санкт-Петербургский Политехнический университет

Петра Великого

Институт Прикладной Математики и Механики

Кафедра «Прикладная Математика и Информатика»

Отчет

По лабораторной работе № 3

По Дисциплине «Математическая статистика»

Выполнил:

Студент Селянкин Федор

Группа 3630102/70301

Проверил:

к.ф. – м.н., доцент

Баженов Александр Николаевич

Санкт-Петербург
2020 г.

Содержание

Список иллюстраций:

Список таблиц:

Постановка задачи

Для 5 распределений:

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

Сгенерировать выборки размером 20, 100 элементов.
Построить для них боксплот Тьюки.
Для каждого распределения определить долю выбросов экспериментально (сгенерировав выборку, соответствующую распределению 1000 раз, и вычислил среднюю долю выбросов) и сравнить с результатами, полученными теоретически.

Теория

Распределения

- Нормальное распределение

1

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

- Распределение Коши

2

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

- Распределение Лапласа

3

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

- Распределение Пуассона

4

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

- Равномерное распределение

5

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

Боксплот Тьюки

Определение

Боксплот (англ. box plot) – график, использующийся в описательной статистике, компактно изображающий одномерное распределение вероятностей.

Описание

Такой вид диаграммы в удобной форме показывает медиану, нижний и верхний квартили и выбросы. Несколько таких ящиков можно нарисовать бок о бок, чтобы визуально сравнивать одно распределение с другим; их можно располагать как горизонтально, так и вертикально. Расстояния между различными частями ящика позволяют определить степень разброса (дисперсии) и асимметрии данный и выявить выбросы.

Построение

Границами ящика служат первый и третий квартили, линия в середине ящика – медиана. Концы усов – крася статистически значимой выборки (без выбросов). Длину «усов» определяют разность первого квартиля и полутора межквартильных расстояний и сумма третьего и полутора межквартильных расстояний. Формула имеет вид

6

$$\boldsymbol{X_1 = Q_1 - \frac{3}{2}(Q_3 - Q_1),\ X_2 = Q_3 + \frac{3}{2}(Q_3 - Q_1),} \tag{6}$$

Где X_1 – нижняя граница уса, X_2 – верхняя граница уса, Q_1 – первый квартиль, Q_3 – третий квартиль.

Данные, выходящие за границы усов (выбросы), отображаются на графике в виде маленьких кружков.

Теоретическая вероятность выбросов

Зная, первый и третий квартили по формуле (6) можно вычислить теоретически нижнюю и верхнюю границы уса (X_1^T и X_2^T соответственно). Выбросами считаются такие величины x, что

7

$$\left[\begin{array}{l} \boldsymbol{x < X_1^T} \\ \boldsymbol{x > X_2^T} \end{array}\right. \tag{7}$$

Теоретическая вероятность выбросов для непрерывных распределений

8

$$\boldsymbol{P_{\text{в}}^T = P(x < X_1^T) + P(x > X_2^T) = F(X_1^T) + (1 - F(X_2^T)),} \tag{8}$$

где $F(X) = P\ (x \le X)$ – функция распределения

Теоретическая вероятность выбросов для дискретных распределений

9

$$\boldsymbol{P_{\text{в}}^T = P(x < X_1^T) + P(x > X_2^T) = (F(X_1^T) - P(x = X_1^T) + (1 - F(X_2^T)),} \tag{9}$$

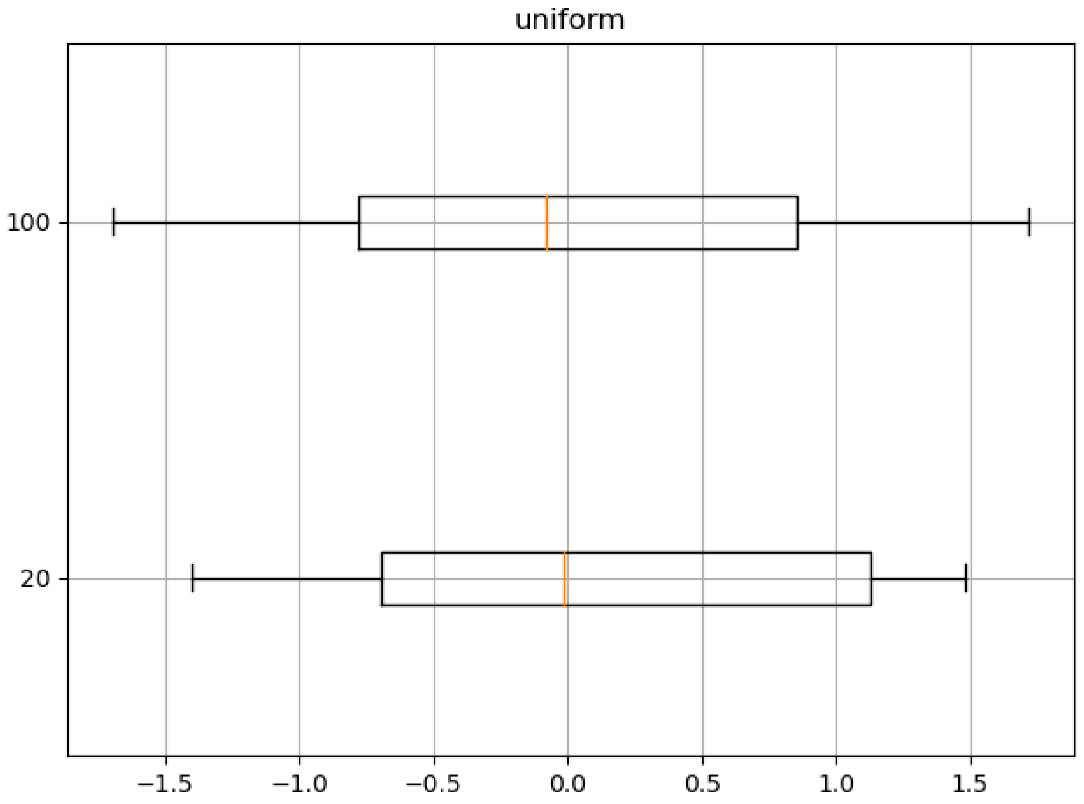
где $F(X) = P\ (x \le X)$ – функция распределения

Реализация

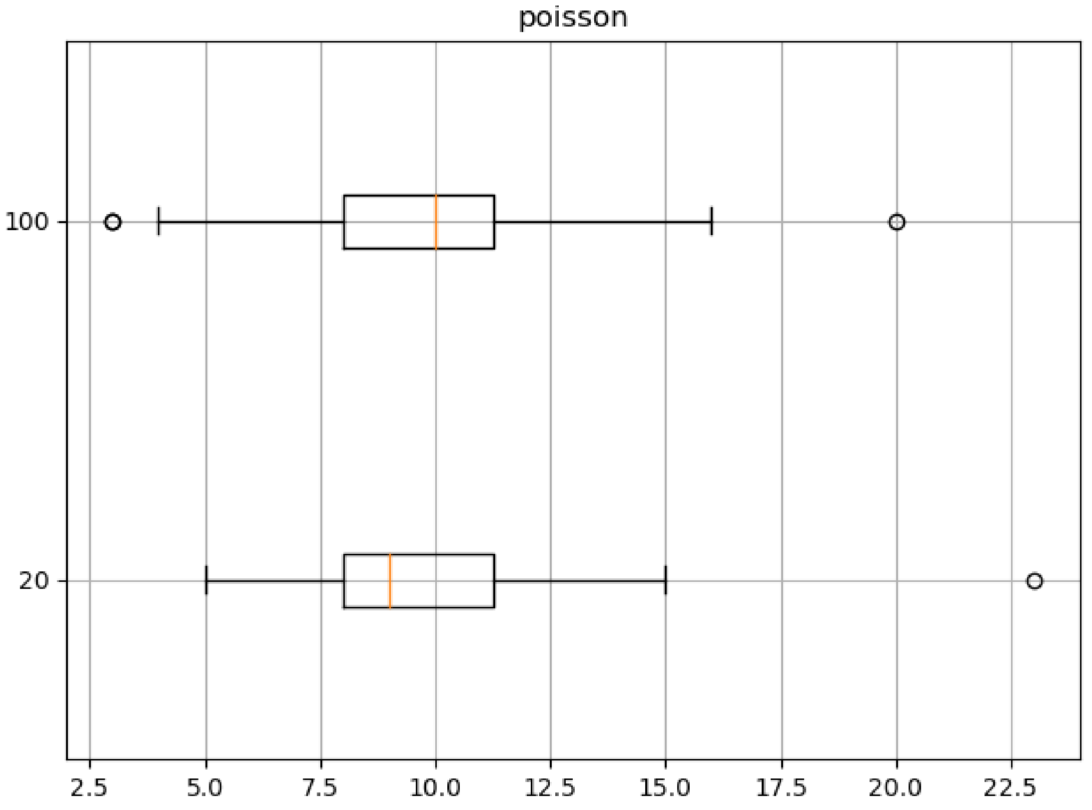
Лабораторная работа выполнена с помощью встроенных средств языка программирования Python в среде разработки PyCharm, с использованием дополнительных библиотек для отображения и расчетов. Исходный код лабораторной выложен на веб-сервисе GitHub [2].

Результаты

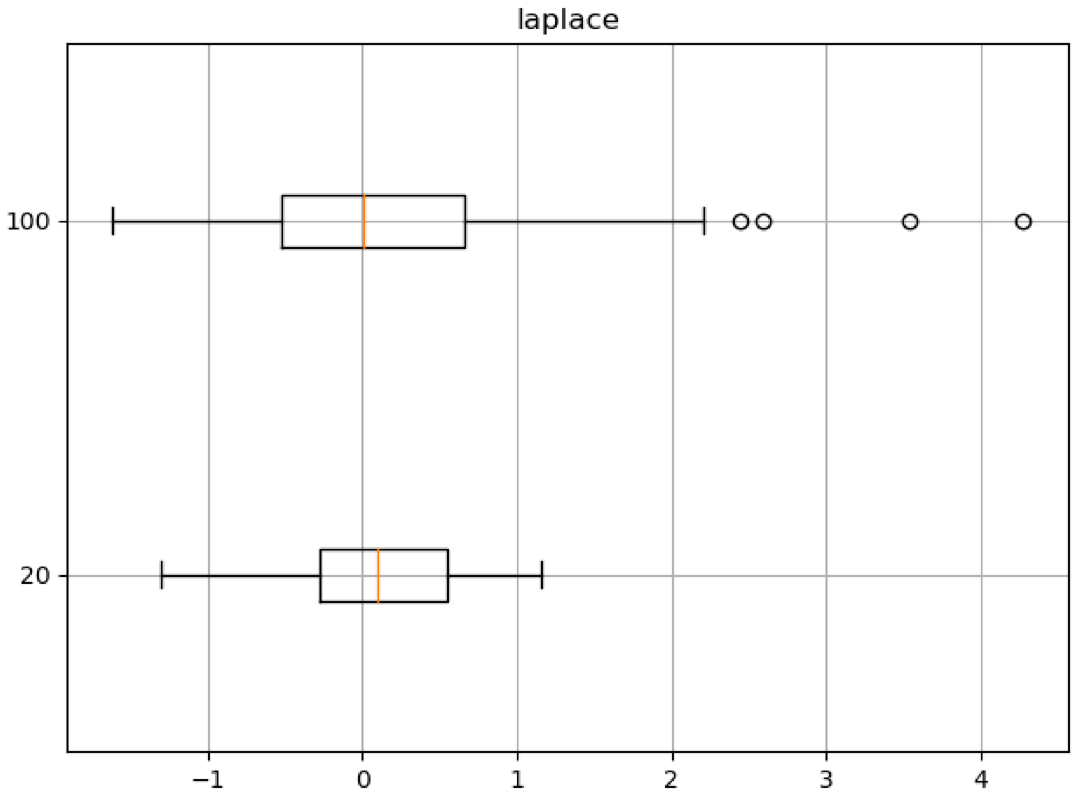
Боксплот Тьюки



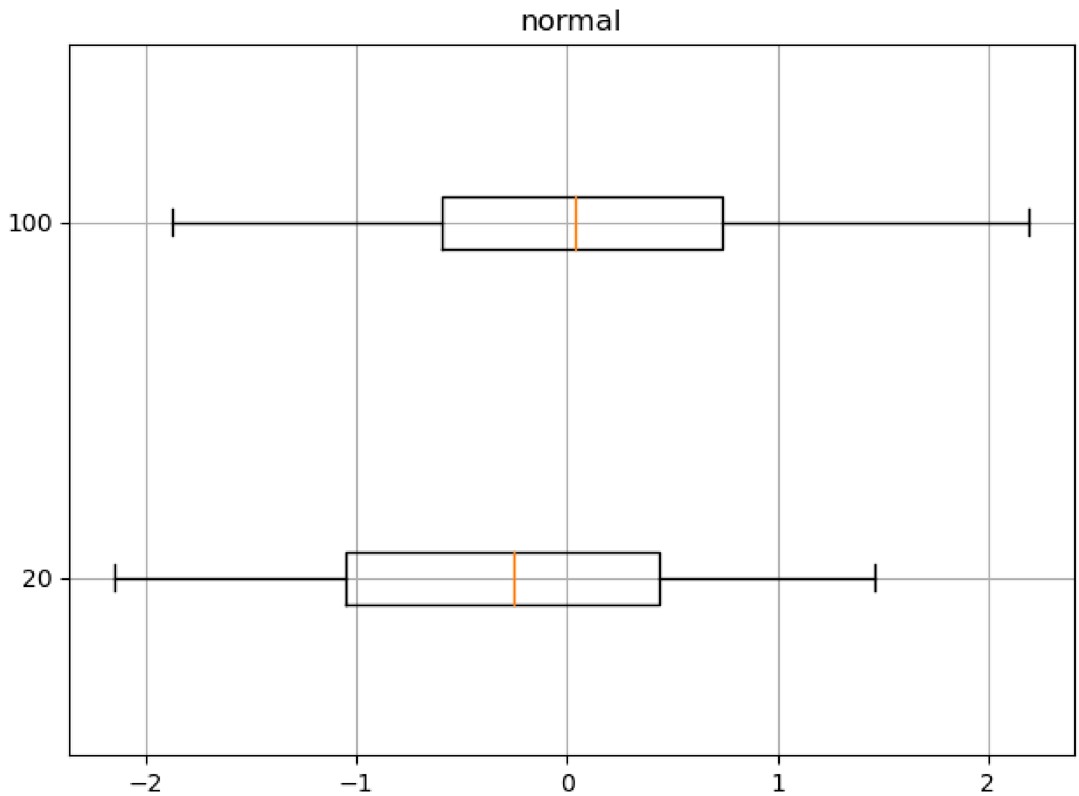
1Равномерное распределение



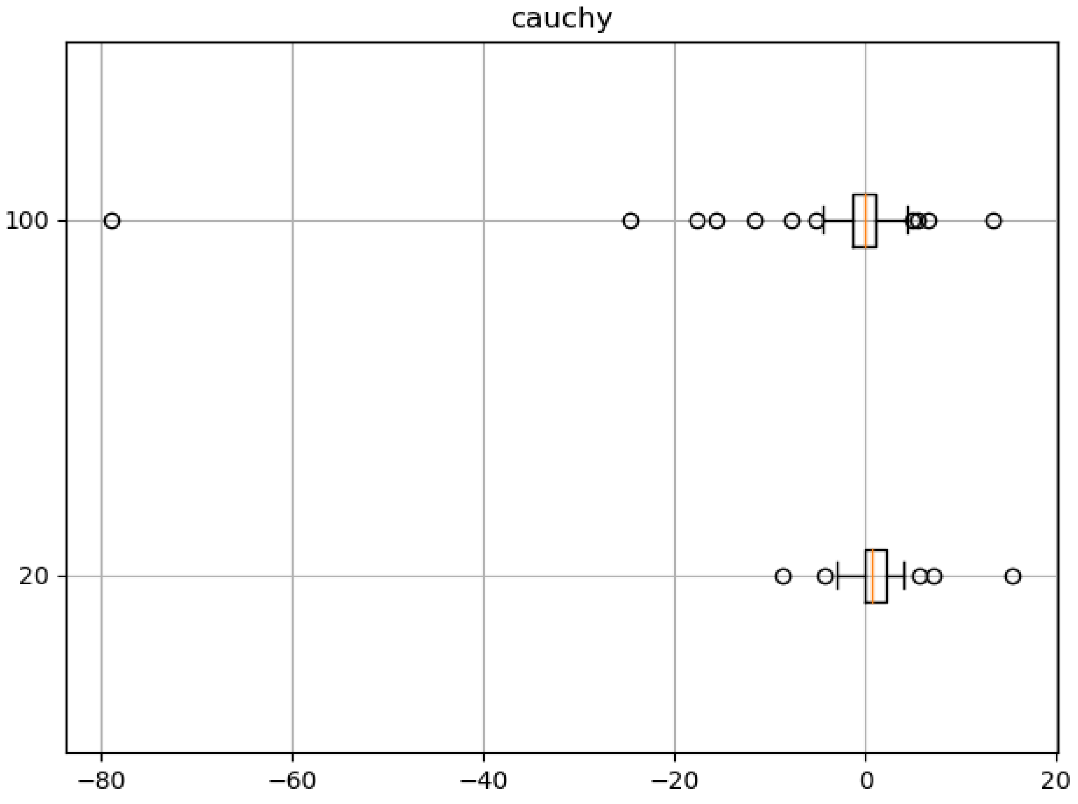
2Распределение Пуассона



3Распределение Лапласа



4Нормальное распределение



5Распределение Коши

Доля Выбросов

Выборка	Доля выбросов
Нормальное распределение n = 20	0.02
Нормальное распределение n = 100	0.01
Распределение Коши n = 20	0.15
Распределение Коши n = 100	0.15
Распределение Лапласа n = 20	0.08
Распределение Лапласа n = 100	0.07
Распределение Пуассона n = 20	0.02
Распределение Пуассона n = 100	0.01
Равномерное распределение n = 20	0.00
Равномерное распределение n = 100	0.00

1Доля Выбросов

Теоретическая вероятность выбросов

Распределение	Q_1^T	Q_3^T	X_1^T	x_2^T	$P_в^T$
Нормальное распределение	-0.674	0.674	-2.698	2.698	0.007
Распределение Коши	-1	1	-4	4	0.156
Распределение Лапласа	-0.490	0.490	1.961	1.961	0.063
Распределение Пуассона	8	12	2	18	0.008
Равномерное распределение	-0.866	0. 866	3.464	3.464	0

2Теоретическая вероятность выбросов

Литература

1. Боксплот
   https://en.wikipedia.org/wiki/Box_plot
2. Ссылка на репозиторий GitHub
   https://github.com/SelyankinFyodor/math-statistics/tree/master/Lab3

Обсуждения

В распределении Коши доля выбросов меньше, чем в распределении Лапласа или нормальном распределении, но из боксплота видно, что разброс значений намного больше.