

Санкт-Петербургский Политехнический университет
Петра Великого

Институт прикладной математики и механики
Высшая школа прикладной математики и вычислительной физики

Отчёт
по лабораторной работе 3
по дисциплине
"математическая статистика"

Выполнил студент:

Аникин Александр Алексеевич,
группа 3630102\80201

Проверил:

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

Санкт-Петербург
2021

Содержание

1	Постановка задачи	4
2	Теория	5
2.1	Рассматриваемые распределения	5
2.2	Боксплот Тьюки	5
2.2.1	Построение	5
2.3	Теоретическая доля выбросов	6
3	Реализация	7
4	Результаты	8
4.1	Боксплоты	8
4.2	Эмпирическая и теоретическая доли выбросов	13
	Литература	14

Список иллюстраций

1	Боксплоты для нормального распределения (1)	8
2	Боксплоты для распределения Коши (2)	9
3	Боксплоты для распределения Лапласа (3)	10
4	Боксплоты для распределения Пуассона (4)	11
5	Боксплоты для равномерного распределения (5)	12

Список таблиц

1	Теоретические доли выбросов	6
2	Эмпирические и теоретические доли выбросов	13

1 Постановка задачи

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

2 Теория

2.1 Рассматриваемые распределения

Плотности:

- 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:

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

2.2 Боксплот Тьюки

2.2.1 Построение

Границами ящика – первый и третий квартили, линия в середине ящика — медиана. Концы усов — края статистически значимой выборки (без выбросов). Длина «усов»:

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

где X_1 - нижняя граница, X_2 - верхняя граница, Q_1 - первый квартиль, Q_2 - нижний квартиль.

Данные, выходящие за границы усов (выбросы), отображаются на графике в виде маленьких кружков [1].

2.3 Теоретическая доля выбросов

Можно вычислить теоретические первый и третий квартили распределений - Q_1^T, Q_2^T . По формуле (6) можно вычислить теоретические нижнюю и верхнюю границы уса - X_1^T, X_2^T . Под выбросами понимаются такие величины x , что

$$\begin{cases} x < X_1^T \\ x > X_2^T \end{cases} \quad (7)$$

Теоретическая вероятность выбросов:

- для непрерывных распределений:

$$P_B^T = P(x < X_1^T) + P(x > X_2^T) = F(X_1^T) + (1 - F(X_2^T)) \quad (8)$$

- для дискретных распределений:

$$P_B^T = P(x < X_1^T) + P(x > X_2^T) = (F(X_1^T) - P(x = X_1^T)) + (1 - F(X_2^T)) \quad (9)$$

В формулах выше $F(X) = P(x \leq X)$ - функция распределения.

Теоретические доли выбросов для данных распределений - постоянные величины. Их значения приведены в следующей таблице.

Нормальное распределение	0.007
Распределение Коши	0.156
Распределение Лапласа	0.063
Распределение Пуассона	0.008
Равномерное распределение	0

Таблица 1: Теоретические доли выбросов

3 Реализация

Лабораторная работа выполнена на языке Python 3.8 с помощью загружаемых пакетов SciPy, Matplotlib, Pandas, Seaborn. Исходный код лабораторной работы находится на GitHub репозитории.

4 Результаты

4.1 Боксплоты

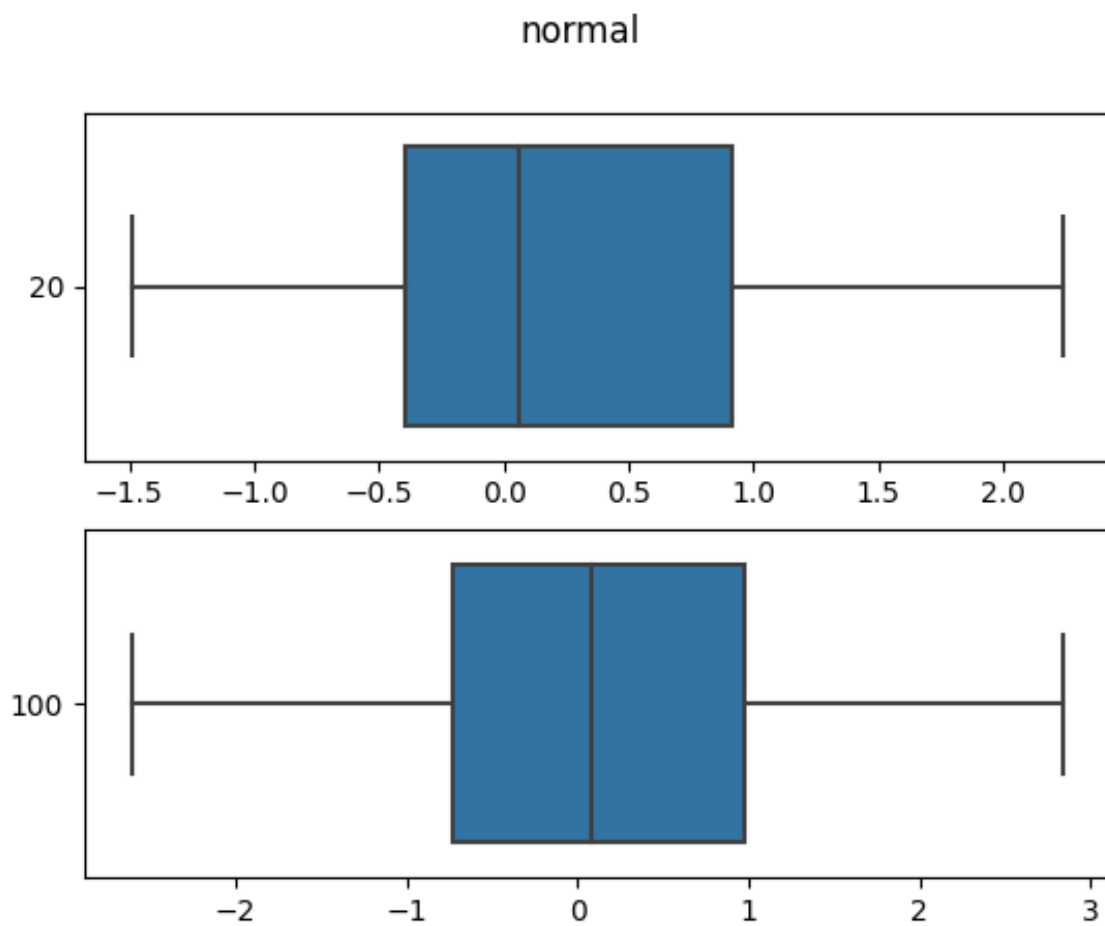


Рис. 1: Боксплоты для нормального распределения (1)

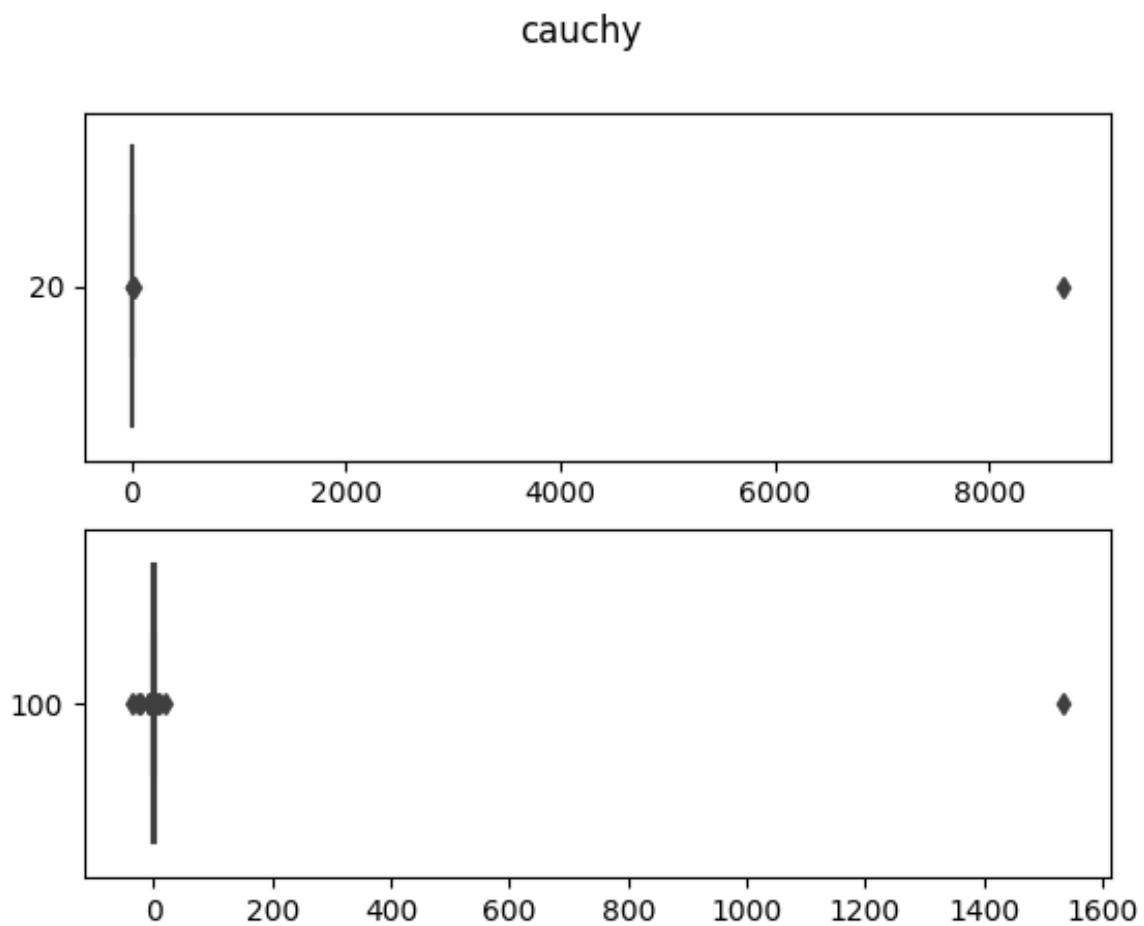


Рис. 2: Боксплоты для распределения Коши (2)

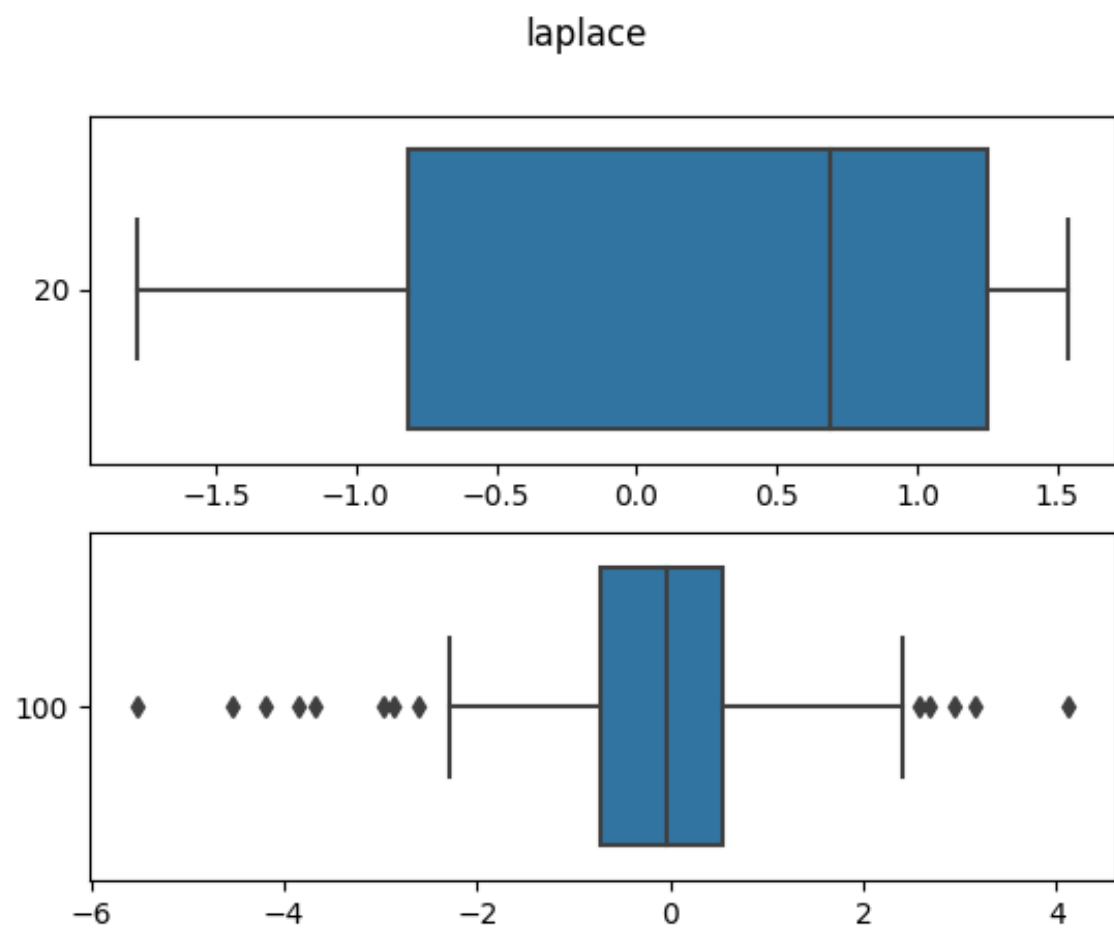


Рис. 3: Боксплоты для распределения Лапласа (3)

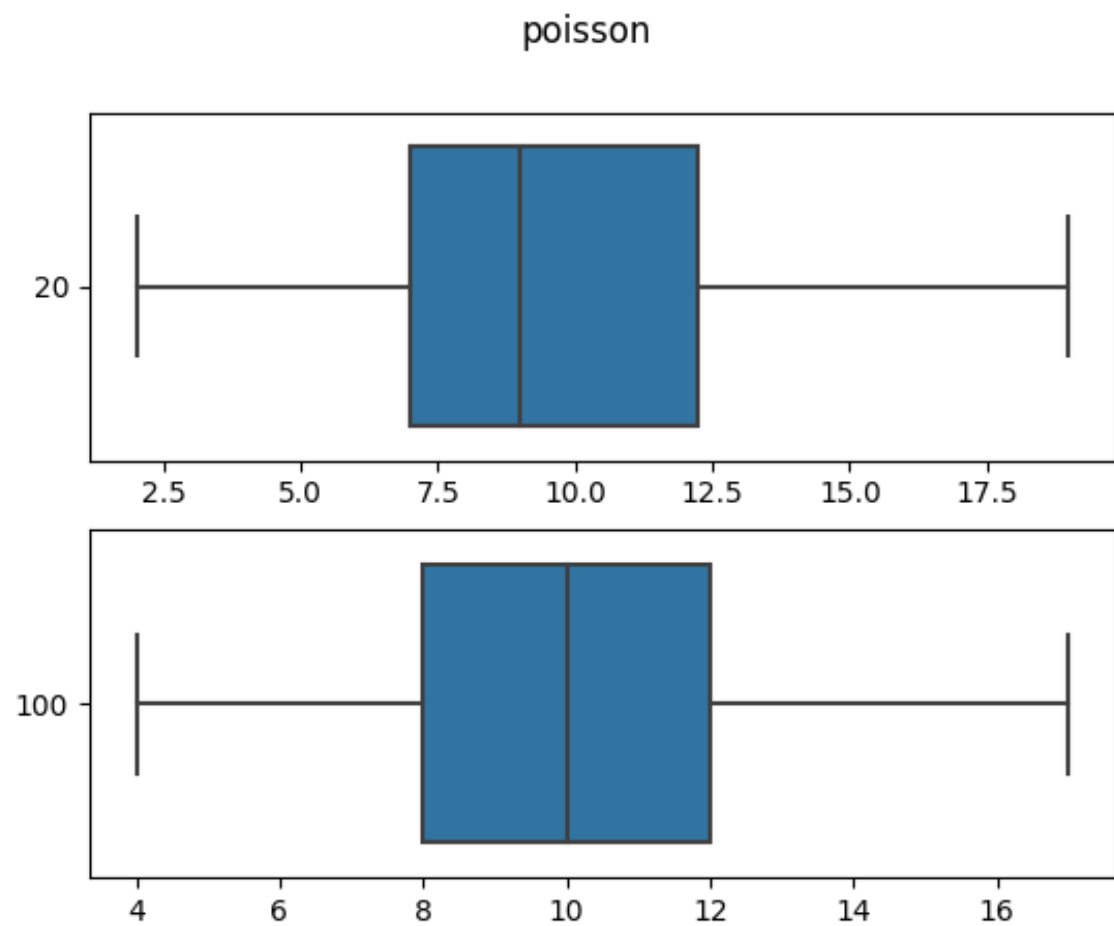


Рис. 4: Боксплоты для распределения Пуассона (4)

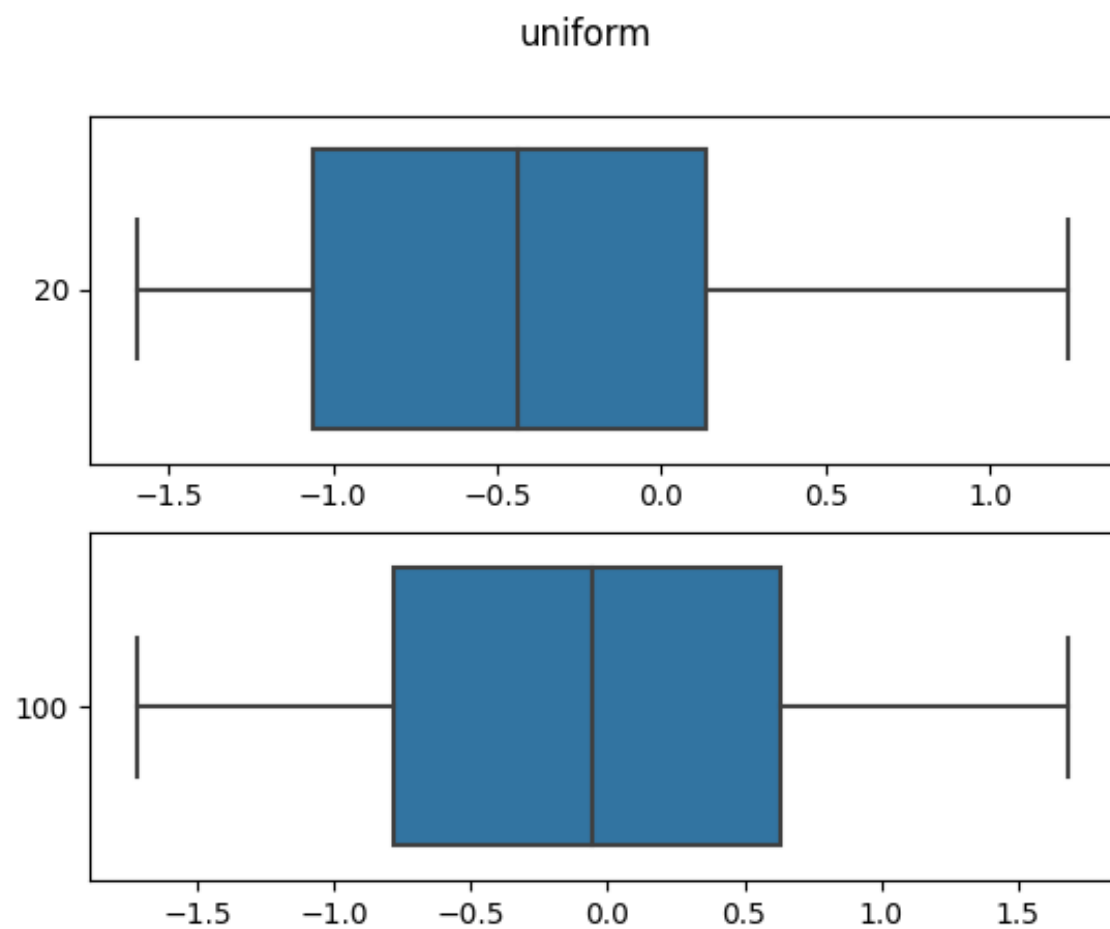


Рис. 5: Боксплоты для равномерного распределения (5)

4.2 Эмпирическая и теоретическая доли выбросов

Распределение	Эмп. ДВ, $n = 20$	Эмп. ДВ, $n = 100$	Теор. ДВ
Нормальное	0.0248	0.0101	0.007
Коши	0.1509	0.1564	0.156
Лапласа	0.0725	0.0646	0.063
Пуассона	0.0646	0.0110	0.008
Равномерное	0.0018	0.0	0

Таблица 2: Эмпирические и теоретические доли выбросов

Нетрудно заметить, что при увеличении размера выборки эмпирическая доля выбросов стремится к теоретической.

Список литературы

[1] Box plot. URL: https://en.wikipedia.org/wiki/Box_plot