

Санкт-Петербургский политехнический университет  
Петра Великого

Институт прикладной математики и механики  
Кафедра «Прикладная математика»

## ОТЧЕТ ПО ЛАБОРАТОРНОЙ РАБОТЕ №3

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

<b>1. Постановка задачи</b>	<b>4</b>
<b>2. Теория</b>	<b>4</b>
2.1. Рассматриваемые распределения . . . . .	4
2.2. Боксплот Тьюки . . . . .	4
2.3. Теоретическая вероятность выбросов . . . . .	5
<b>3. Реализация</b>	<b>5</b>
<b>4. Результаты</b>	<b>6</b>
4.1. Боксплот Тьюки . . . . .	6
4.2. Доля выбросов . . . . .	10
<b>5. Обсуждение</b>	<b>11</b>

## Список таблиц

1	Практическая доля выбросов . . . . .	10
2	Теоретическая доля выбросов . . . . .	11

# 1. Постановка задачи

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 распределений

- 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

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

### 2.2. Боксплот Тьюки

Боксплот Тьюки или ящик с усами – график, использующийся в описательной статистике, компактно изображающий одномерное распределение вероятностей. Этот вид диаграммы в удобной форме показывает медиану (или, если нужно, среднее), нижний и верхний квартили, минимальное и максимальное значение выборки и выбросы.

Построение боксплота производится по следующим параметрам:

- границы боксплота -  $Q_1$ ,  $Q_3$  - первый и третий квартили соответственно
- линия середины боксплота - медиана
- концы "усов" края статистически значимой выборки (без выбросов). Их длина определяется по формуле:

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

, где  $X_1$  — нижняя граница уса,  $X_2$  — верхняя граница уса

- выбросы - данные, выходящие за границы усов, отображающиеся на графике в виде кружков

## 2.3. Теоретическая вероятность выбросов

После вычисления первого и третьего квартилей и нижней и верхней границы уса по формуле (6), можно определить выбросы  $x$  :

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

Пусть  $F(X) = P(x \leq X)$  - функция распределения. 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:

$$P_B^T = P(x < X_1^T) + P(x > x_2^T) = (F(X_1^T) - P(x = X_1^T)) + (1 - F(X_2^T)) \quad (9)$$

## 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/Lab3>

## 4. Результаты

### 4.1. Боксплот Тьюки

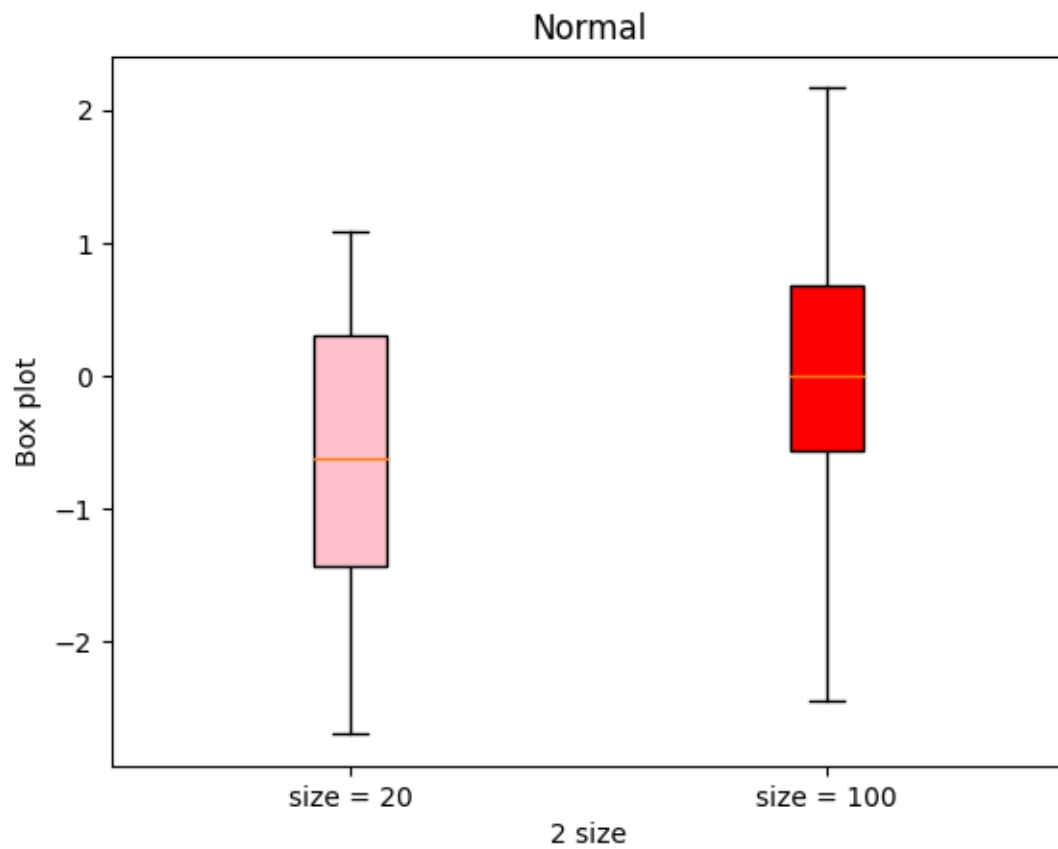


Рис. 1. Нормальное распределение(1)

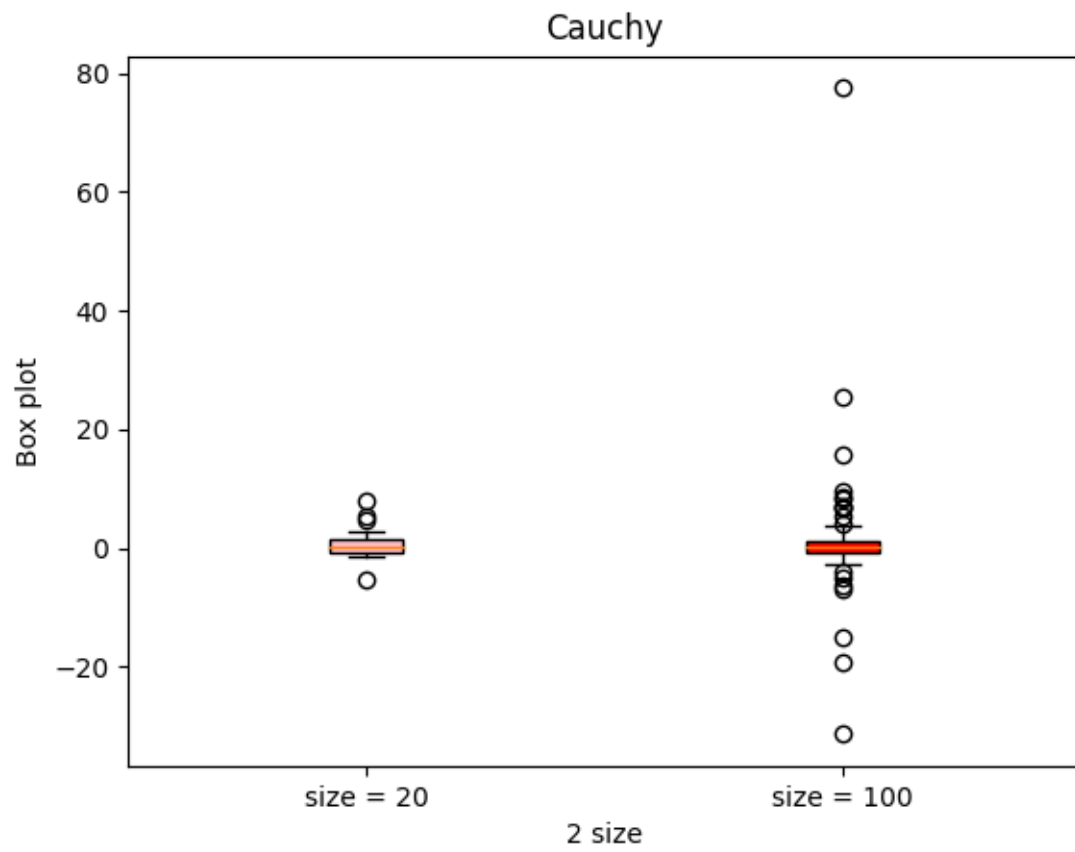


Рис. 2. Нормальное распределение(2)

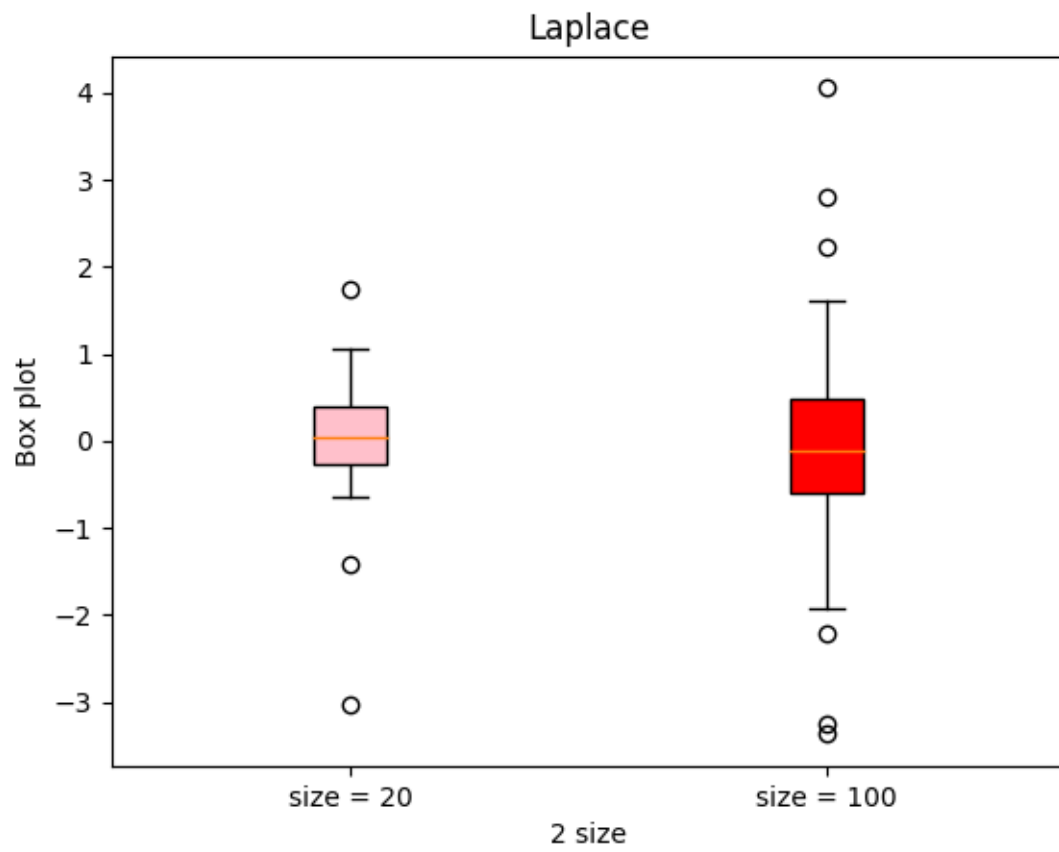


Рис. 3. Нормальное распределение(3)



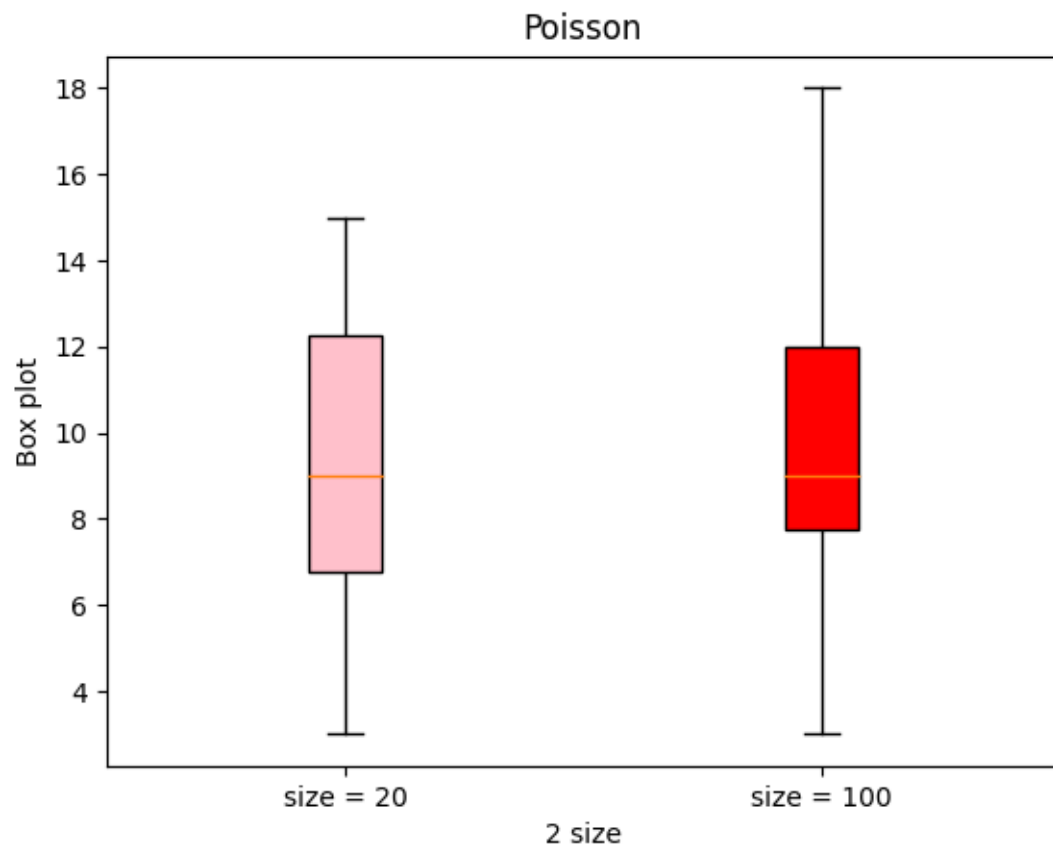


Рис. 4. Нормальное распределение(4)

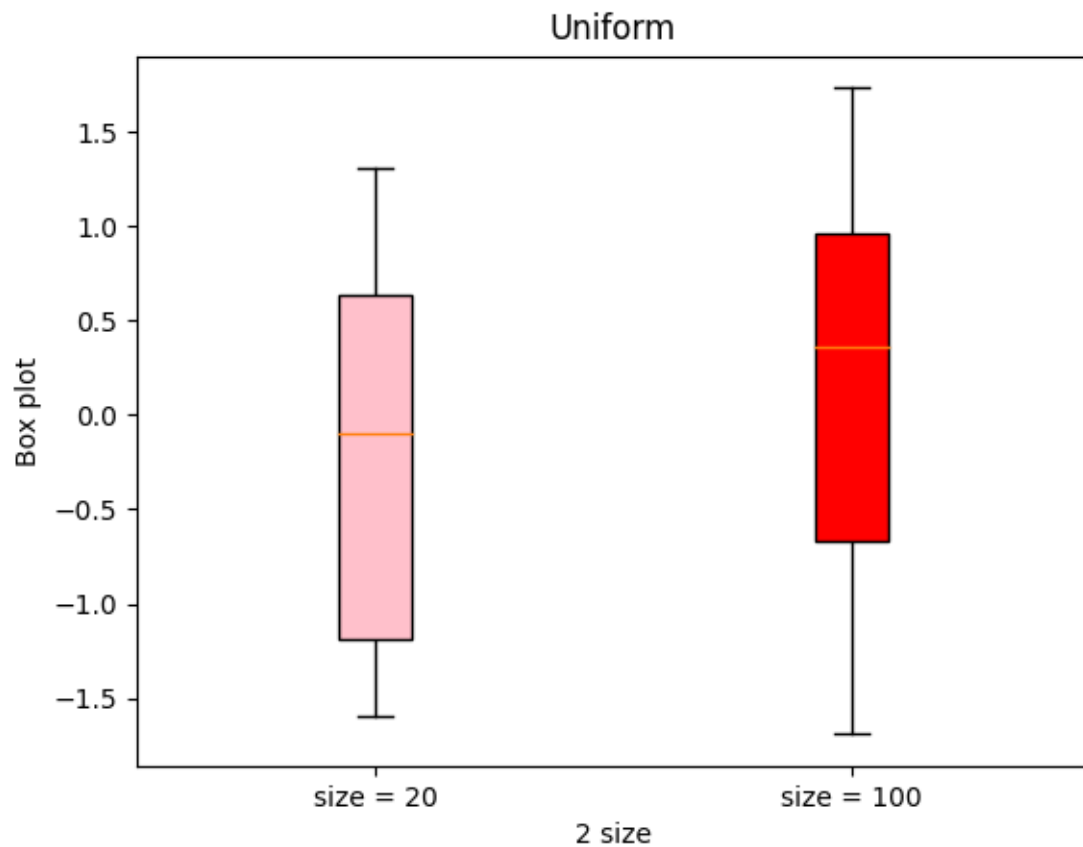


Рис. 5. Нормальное распределение(5)

## 4.2. Доля выбросов

Выборка	Размер	Доля выбросов
Normal	20	0.02
Normal	100	0.01
Cauchy	20	0.15
Cauchy	100	0.15
Laplace	20	0.08
Laplace	100	0.07
Poisson	20	0.02
Poisson	100	0.01
Uniform	20	0.00
Uniform	100	0.00

Таблица 1. Практическая доля выбросов

Распределение	$Q_1^T$	$Q_3^T$	$X_1^T(6)$	$X_2^T(6)$	$P_B^T(8)$
Normal	-0.674	0.674	-2.698	2.698	0.007
Cauchy	-1	1	-4	4	0.156
Laplace	-0.490	0.490	-1.960	1.960	0.063
Poisson	8	12	2	18	0.008
Uniform	-0.866	0.866	-34.64	34.64	0

Таблица 2. Теоретическая доля выбросов

## 5. Обсуждение

Боксплот Тьюки позволяет наглядно представить характеристики заданного распределения, такие как медиана, первый и третий квартили, наличие выбросов.

Самым большим преимуществом боксплота Тьюки является то, что они не подвержены влиянию выбросов и могут описывать дискретное распределение данных относительно стабильным образом.

Одновременно, недостатком боксплота Тьюки является то, что невозможно точно измерить степень асимметрии и хвостовой вес распределения данных; для больших пакетов данных отраженная информация более расплывчата, и существуют определенные ограничения в использовании медианы для представления общего уровня оценки.