

Санкт-Петербургский политехнический университет
Петра Великого

Институт прикладной математики и механики
Кафедра «Прикладная математика»

ОТЧЕТ ПО ЛАБОРАТОРНОЙ РАБОТЕ №3

по дисциплине
"Математическая статистика"

Выполнила студентка
группы 3630102/80201

Деркаченко Анна Олеговна

Проверил
доцент, к.ф.-м.н.

Баженов Александр Николаевич

Санкт-Петербург
2021 г.

Содержание

1. Постановка задачи	4
2. Теория	4
2.1. Боксплот Тьюки	4
2.2. Теоретическая вероятность выбросов	5
3. Реализация	5
4. Результаты	6
4.1. Боксплот Тьюки	6
4.2. Доля выбросов	8
5. Обсуждение	9

Список иллюстраций

1.	Нормальное распределение	6
2.	Распределение Коши	6
3.	Распределение Лапласа	7
4.	Распределение Пуассона	7
5.	Равномерное распределение	8

1. Постановка задачи

Даны распределения:

- 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 $U(x, -\sqrt{3}, \sqrt{3})$

Необходимо:

- 1) Сгенерировать выборки размером 20 и 100 элементов
- 2) Построить для них боксплот Тьюки
- 3) 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. Боксплот Тьюки

Боксплот Тьюки - график, использующийся в описательной статистике, компактно изображающий одномерное распределение вероятностей. Данный вид диаграммы в удобной форме показывает множество характеристик положения случайной величины.

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:

$$X_1 = Q_1 - \frac{3}{2}(Q_3 - Q_1) \quad X_2 = Q_3 + \frac{3}{2}(Q_3 - Q_1) \quad (1)$$

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

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

2.2. Теоретическая вероятность выбросов

После вычисления первого и третьего квартилей и нижней и верхней границы уса по формуле (1), можно определить выбросы x :

$$\begin{cases} x < X_1^T \\ x > X_2^T \end{cases} \quad (2)$$

Пусть $F(X) = P(x \leq X)$ - функция распределения. Теоретическая вероятность выбросов

- для непрерывных распределений:

$$P_B^T = P(x < X_1^T) + P(x > X_2^T) = F(X_1^T) + (1 - F(X_2^T)) \quad (3)$$

- для дискретных распределений:

$$P_B^T = P(x < X_1^T) + P(x > x_2^T) = (F(X_1^T) - P(x = X_1^T)) + (1 - F(X_2^T)) \quad (4)$$

3. Реализация

Реализация лабораторной работы проводилась на языке Python в среде разработки PyCharm с использованием дополнительных библиотек:

- scipy
- numpy
- math
- matplotlib
- seaborn

Исходный код лабораторной работы размещен в GitHub-репозитории.

URL: <https://github.com/derkanw/Mathstat/tree/main/lab3>

4. Результаты

4.1. Боксплот Тьюки

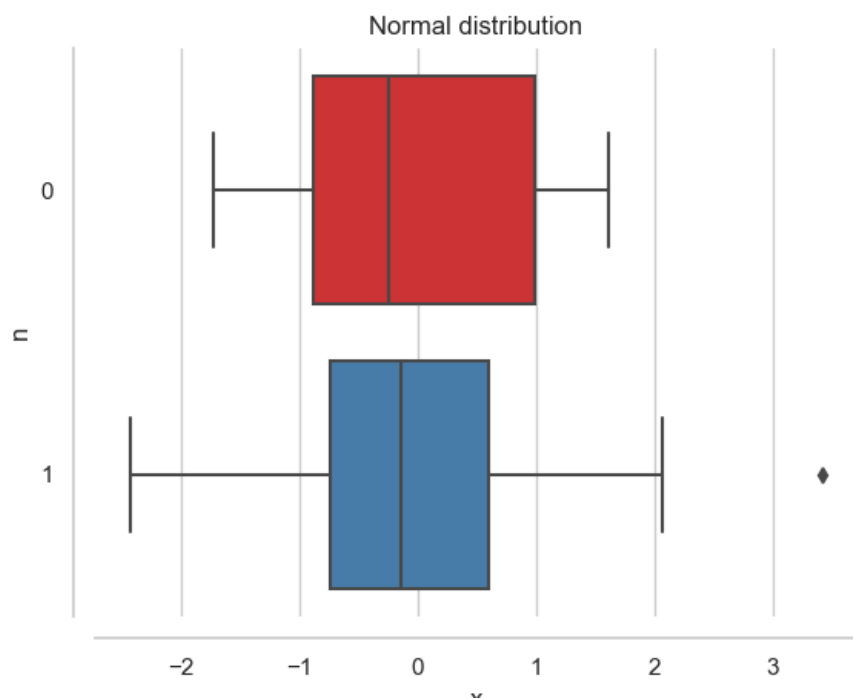


Рис. 1. Нормальное распределение

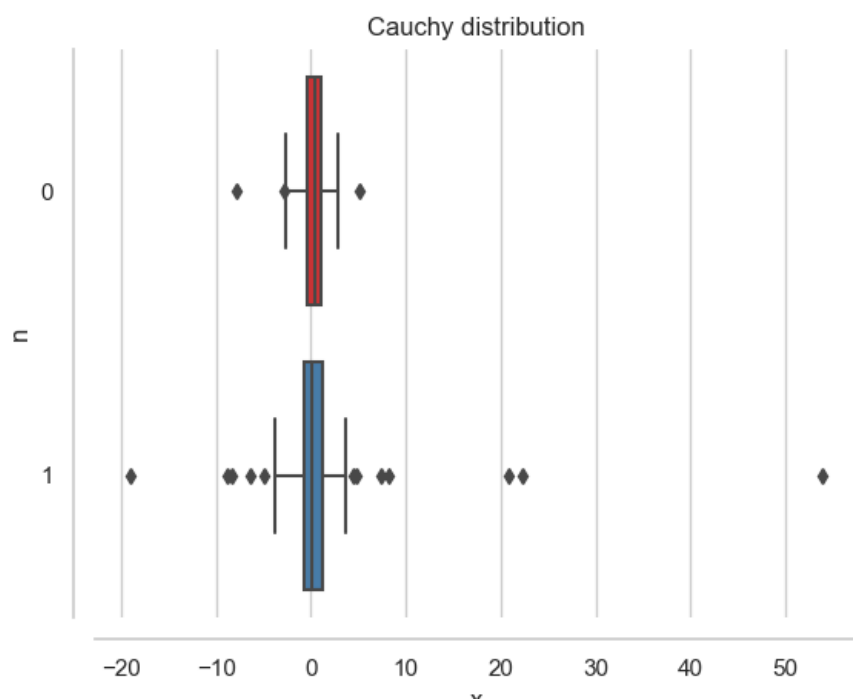


Рис. 2. Распределение Коши

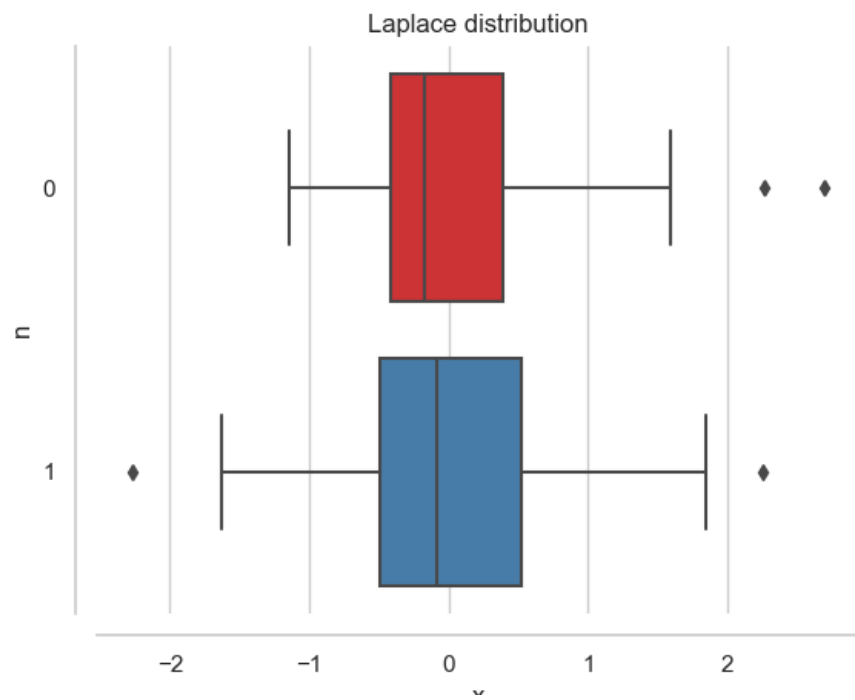


Рис. 3. Распределение Лапласа

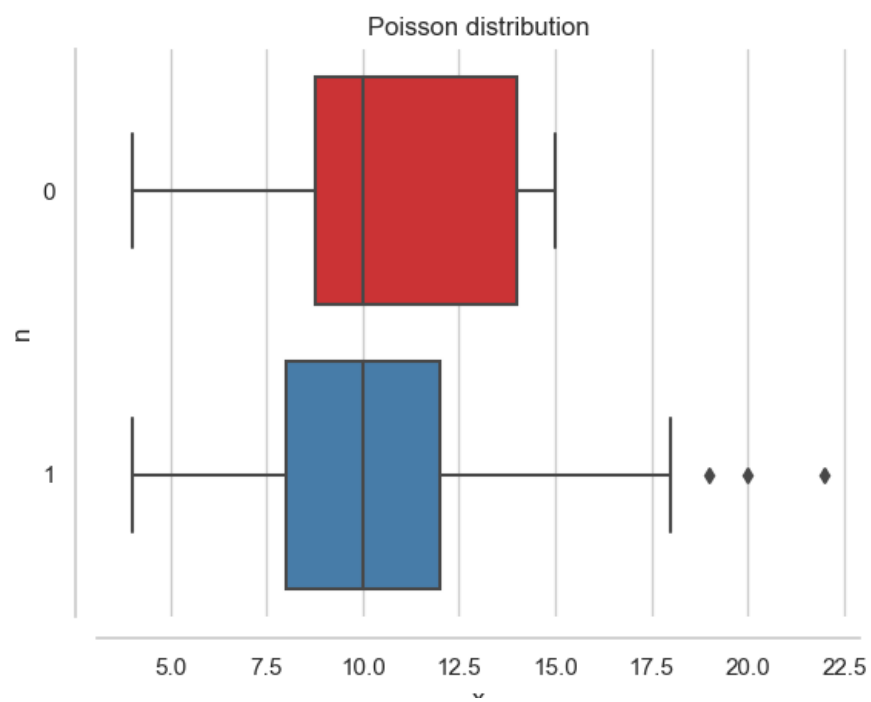


Рис. 4. Распределение Пуассона

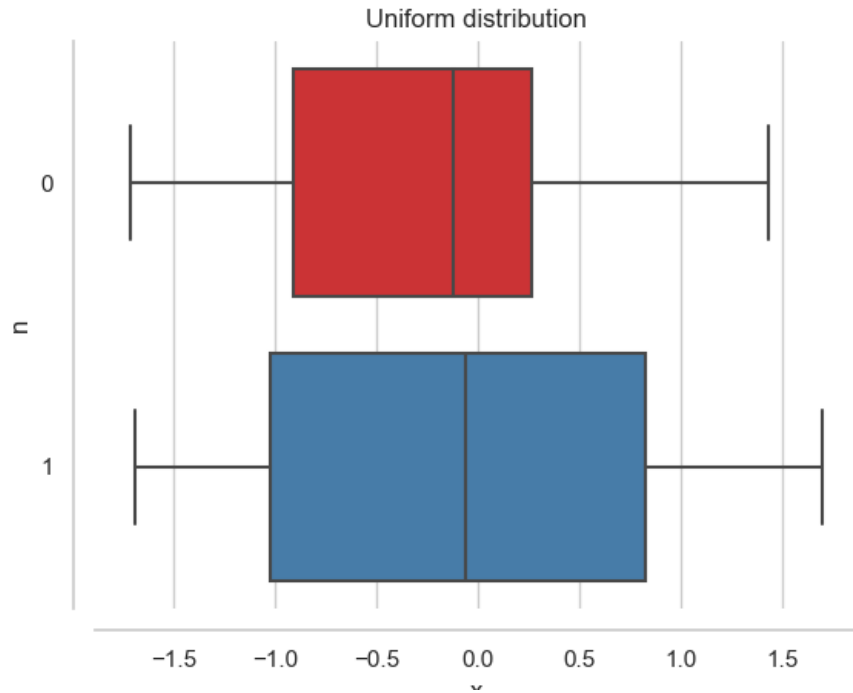


Рис. 5. Равномерное распределение

4.2. Доля выбросов

Выборка	Доля выбросов	P_B^T
Normal n = 20	0.02285	0.007
Normal n = 100	0.00959	0.007
Cauchy n = 20	0.1526	0.156
Cauchy n = 100	0.15549	0.156
Laplace n = 20	0.0744	0.063
Laplace n = 100	0.06626	0.063
Poisson n = 20	0.02375	0.008
Poisson n = 100	0.00956	0.008
Uniform n = 20	0.00265	0
Uniform n = 100	0	0

Таблица 1. Практическая доля выбросов

Распределение	Q_1^T	Q_3^T	X_1^T	X_2^T	P_B^T
Нормальное	-0.674	0.674	-2.698	2.698	0.007
Коши	-1	1	-4	4	0.156
Лапласа	-0.490	0.490	-1.961	1.961	0.063
Пуассона	8	12	2	18	0.008
Равномерное	-0.866	0.866	-3.464	3.464	0

Таблица 2. Теоретическая вероятность выбросов

5. Обсуждение

Боксплот Тьюки позволяет наглядно представить характеристики заданного распределения, такие как медиана, первый и третий квартили, наличие выбросов. К тому же процесс анализа данной диаграммы удобнее, нежели полный аналитический расчет.

Судя по данным таблиц для различных выборок, в большинстве случаев наблюдается закономерность: чем больше размерность выборки, тем ближе найденная доля выбросов к теоретической оценке. Практическая и теоретическая доля выбросов для распределения Коши имеют значения, значительно превышающие аналогичные для остальных распределений, что говорит о относительно частом возникновении выбросов при использовании данного распределения. Стоит сказать, что у равномерного распределения, наоборот, выбросы почти не наблюдаются.