

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:

Трощенко Константин
группа: 3630102/80301

Проверил:

Баженов Александр Николаевич

Санкт-Петербург
2021

Содержание

1	Постановка задачи	2
2	Теория	2
2.1	Боксплот Тьюки	2
2.1.1	Определение	2
2.1.2	Описание	2
2.1.3	Построение	2
2.2	Теоретическая вероятность выбросов	3
3	Реализация	3
4	Результаты	4
4.1	Боксплот Тьюки	4
4.2	Доля выбросов	6
4.3	Теоретическая вероятность выбросов	7
5	Обсуждение	7
6	Приложение	8

Список иллюстраций

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

Список таблиц

1	Доля выбросов	7
2	Теоретическая вероятность выбросов	7

1 Постановка задачи

Для 5 распределений:

- Нормальное распределение $N(x, 0, 1)$
- Распределение Коши $C(x, 0, 1)$
- Распределение Лапласа $L\left(x, 0, \frac{1}{\sqrt{2}}\right)$
- 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 (сгенерировать выборку, соответствующую распределению 1000 раз, и вычислив среднюю долю выбросов) и сравнить с результатами, полученными теоретически.

2 Теория

2.1 Боксплот Тьюки

2.1.1 Определение

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

2.1.2 Описание

Такой вид диаграммы в удобной форме показывает медиану, нижний и верхний квартили и выбросы. Несколько таких ящиков можно нарисовать бок о бок, чтобы визуальнo 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 данных и выявить выбросы.

2.1.3 Построение

Границами ящика служат первый и третий квартили, линия в середине ящика — медиана. Концы усов — края статистически значимой выборки (без выбросов). Длину «усов» определяют разность первого квартиля и полутора

межквартильных расстояний и сумма третьего квартиля и полутора межквартильных расстояний. Формула имеет вид

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

где X_1 – нижняя граница уса, X_2 – верхняя граница уса, Q_3 – третий квартиль. Данные, выходящие за границы усов (выбросы), отображаются на графике в виде маленьких кружков.

2.2 Теоретическая вероятность выбросов

Встроенными средствами языка программирования Python в среде разработки PyCharm можно вычислить теоретические первый и третий квартили распределений (Q_1^T и Q_3^T соответственно). По формуле (1) можно вычислить теоретические нижнюю и верхнюю границы уса (X_1^T и X_2^T соответственно). Выбросами считаются величины x , такие что:

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

Теоретическая вероятность выбросов для непрерывных распределений

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

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

3 Реализация

Лабораторная работа выполнена с помощью средств языка программирования Python 3 в среде разработки PyCharm. Исходный код лабораторной работы приведен в приложении.

4 Результаты

4.1 Боксплот Тьюки

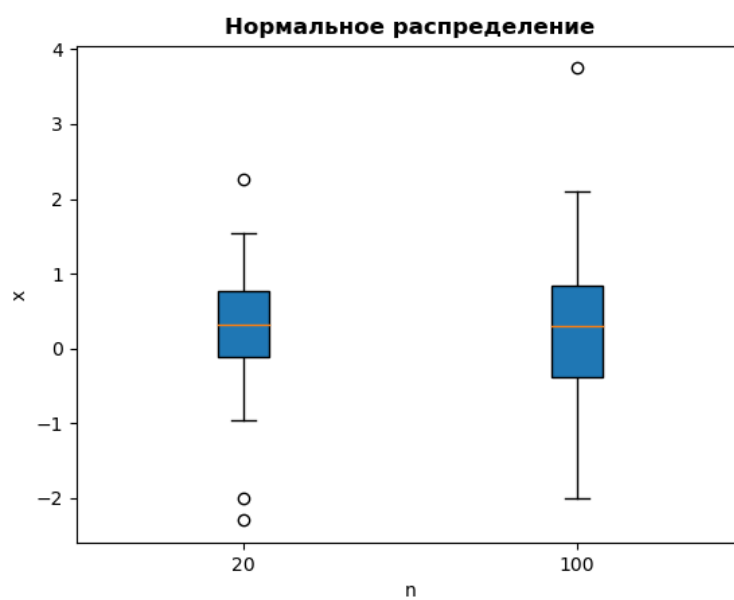


Рис. 1: Нормальное распределение

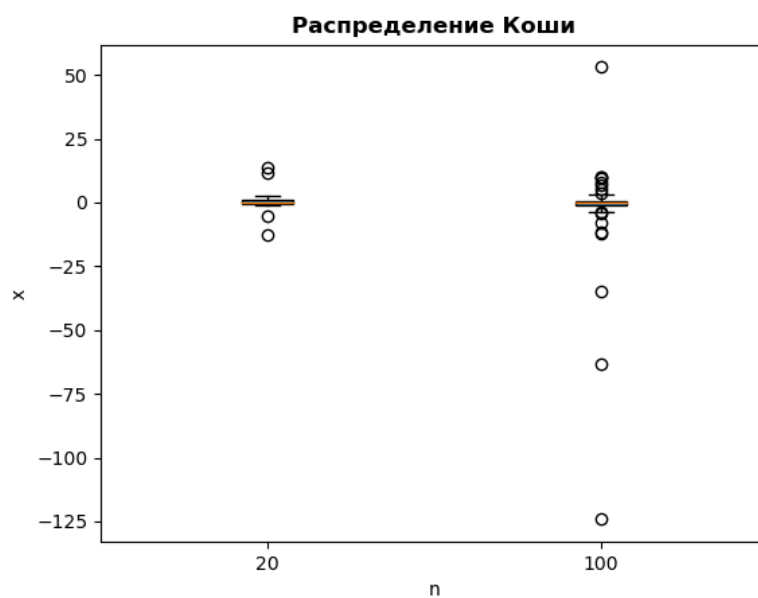


Рис. 2: Распределение Коши

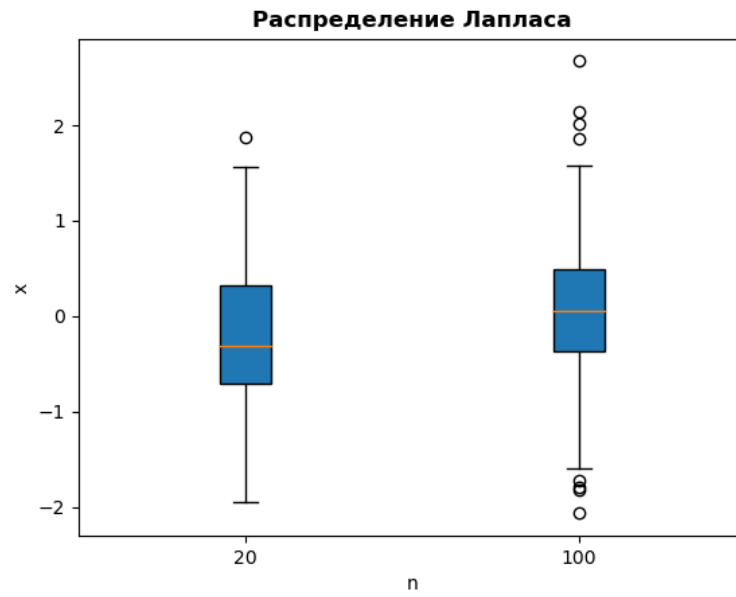


Рис. 3: Распределение Лапласа

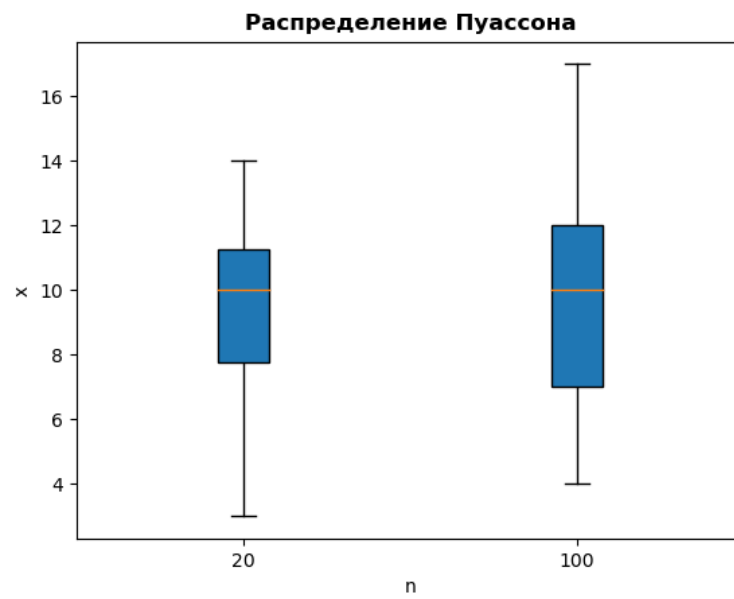


Рис. 4: Распределение Пуассона

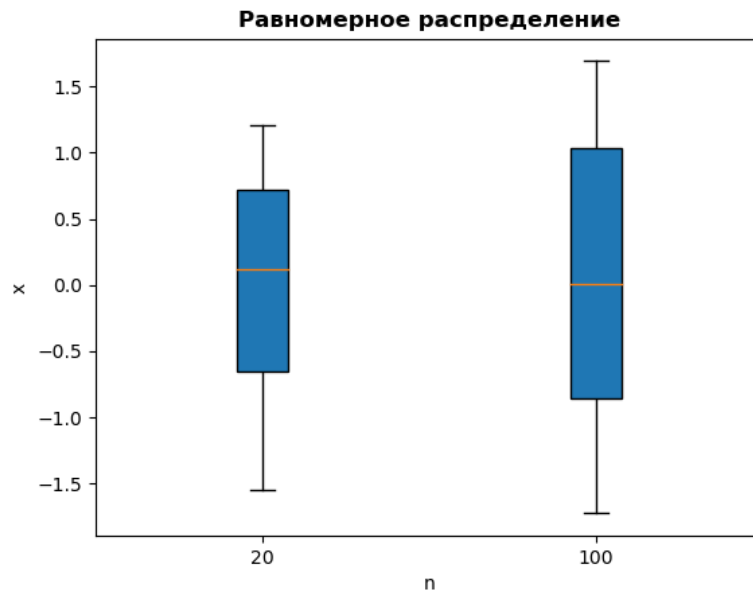


Рис. 5: Равномерное распределение

4.2 Доля выбросов

Округление доли выбросов:

Выборка случайна, поэтому в качестве оценки рассеяния можно взять дисперсию пуассоновского потока: $D_n \approx \sqrt{n}$

$$\text{Доля } p_n = \frac{D_n}{n} = \frac{1}{\sqrt{n}}$$

Для $n = 20$: $p_n = \frac{1}{\sqrt{n}}$ — примерно 0.2 или 20%

Для $n = 100$: $p_n = 0.1$ или 10% Исходя из этого можно решить, сколько знаков оставлять в доле выбросов.

Выборка	Доля выбросов
Нормальное распределение, $n = 20$	0.02
Нормальное распределение, $n = 100$	0.01
Распределение Коши, $n = 20$	0.15
Распределение Коши, $n = 100$	0.16
Распределение Лапласа, $n = 20$	0.08
Распределение Лапласа, $n = 100$	0.06
Распределение Пуассона, $n = 20$	0.02
Распределение Пуассона, $n = 100$	0.01
Равномерное распределение, $n = 20$	0.00
Равномерное распределение, $n = 100$	0.00

Таблица 1: Доля выбросов

4.3 Теоретическая вероятность выбросов

Распределение	Q_1^T	Q_1^T	Q_1^T	Q_1^T	Q_1^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
Нормальное Пуассона	8	12	2	18	0.008
Равномерное распределение	-0.866	0.866	-3.464	3.464	0

Таблица 2: Теоретическая вероятность выбросов

5 Обсуждение

По данным из таблиц можно сказать о том, что чем больше выборка, тем ближе экспериментальные данные доли выбросов к теоретической оценке. Доля выбросов для распределения Коши значительно выше, чем для других распределений. Для равномерного распределения, в точности как и при теоретической оценке, выбросов не наблюдается.

Боксплоты Тьюки действительно позволяют более наглядно и с меньшими усилиями оценивать важные характеристики распределений. Исходя из полученных рисунков мы смогли наглядно увидеть то, на анализ чего ушло довольно много сил в предыдущих работах.

6 Приложение

Код программы: <https://github.com/FaceHuntern/MatStat/tree/main/lab3>.