Санкт-Петербургский политехнический университет
Петра Великого

Институт прикладной математики и механики
Кафедра «Прикладная математика»

Отчёт
по лабораторной работе №3
по дисциплине
«Математическая статистика»

Выполнил студент:

Кондратьев Д. А.
группа: 3630102/70301

Проверил:

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

Санкт-Петербург
2020 г.

Содержание

Список таблиц

Список иллюстраций

1. Постановка задачи

Для 5-ти рапределений:

- 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 их дисперсии) и сравнить с результатами, полученными теоретически.
Средняя доля выбросов:

$$E(z) = \overline{z} \tag{1}$$

Дисперсия:

$$D(z) = \overline{z^2} - \overline{z}^2 \tag{2}$$

2. Теория

2.1. Распределения

- Нормальное распределение

$$N(x, 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \tag{3}$$

- Распределение Коши

$$C(x, 0, 1) = \frac{1}{\pi(1 + x^2)} \tag{4}$$

- Распределение Лапласа

$$L\left(x, 0, \frac{1}{\sqrt{2}}\right) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}|x|} \tag{5}$$

- Распределение Пуассона

$$P(k, 10) = \frac{10^k}{k!} e^{-10} \tag{6}$$

- Равномерное Распределение

$$U(x, -\sqrt{3}, \sqrt{3}) = \begin{cases} \frac{1}{2\sqrt{3}} & |x| \leqslant \sqrt{3} \\ 0 & |x| > \sqrt{3} \end{cases} \tag{7}$$

2.2. Боксплот Тьюки

2.2.1. Определение

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

2.2.2. Описание

Такой вид диаграммы в удобной форме показывает медиану, нижний и верхний квартили и выбросы. Несколько таких ящиков можно нарисовать бок о бок, чтобы 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 данных и выявить выбросы [1].

2.2.3. 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:

$$X_1 = Q_1 - \frac{3}{2}(Q_3 - Q_1), \; X_2 = Q_3 + \frac{3}{2}(Q_3 - Q_1) \tag{8}$$

где X_1 — нижняя граница уса, X_2 — верхняя граица уса, Q_1 — первый квартиль, Q_3 — третий квартиль.

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

2.3. Теоретическая вероятность выбросов

Встроенными средствами языка программирования *Python* можно вычислить теоретические первый и третий квартили распределений (Q_1^T и Q_3^T соответственно). По формуле (8) можно вычислить теоретические нижнюю и верхнюю границы уса (X_1^T и X_3^T соответственно). Выбросами считаются величины x, такие что:

$$\left[\begin{array}{l} x < X_1^T \\ x > X_3^T \end{array}\right. \tag{9}$$

Теоретическая вероятность выбросов для непрерывных распределений:

$$P_{out}^T = P(x < X_1^T) + P(x > X_2^T) = F(X_1^T) + (1 - F(X_2^T)), \tag{10}$$

где $F(X) = P(x \leq X)$ — функция распределения. Теоретическая вероятность выбросов для дискретных распределений:

$$P_{out}^T = P(x < X_1^T) + P(x > X_2^T) = (F(X_1^T) - P(x = X_1^T)) + (1 - F(X_2^T)), \tag{11}$$

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

3. Реализация

Лабораторная работа выполнена на программном языке *Python 3.8* в среде разработки *Jupyter Notebook 6.0.3*. В работе использовались следующие пакеты языка *Python*:

- *numpy* — для генерации выборки и работы с массивами;
- *matplotlib.pyplot* — для построения боксплотов Тьюки;
- *scipy.stats* — содержит все необходимые распределения, а также именно с помощью него можно получить теоретические оценки.

Ссылка на исходный код лабораторной работы приведена в приложении.

4. Результаты

4.1. Боксплот Тьюки

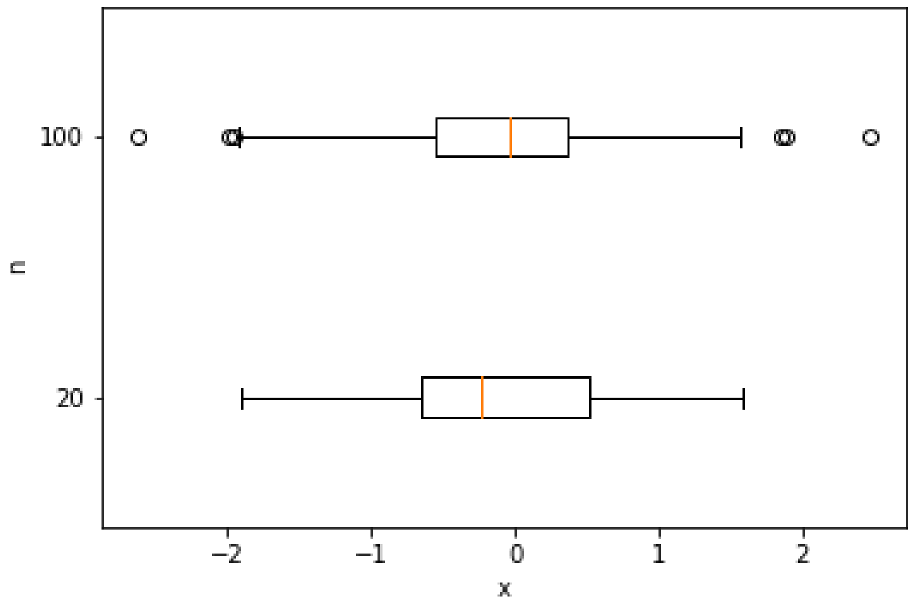

Рис. 1. Нормальное распределение

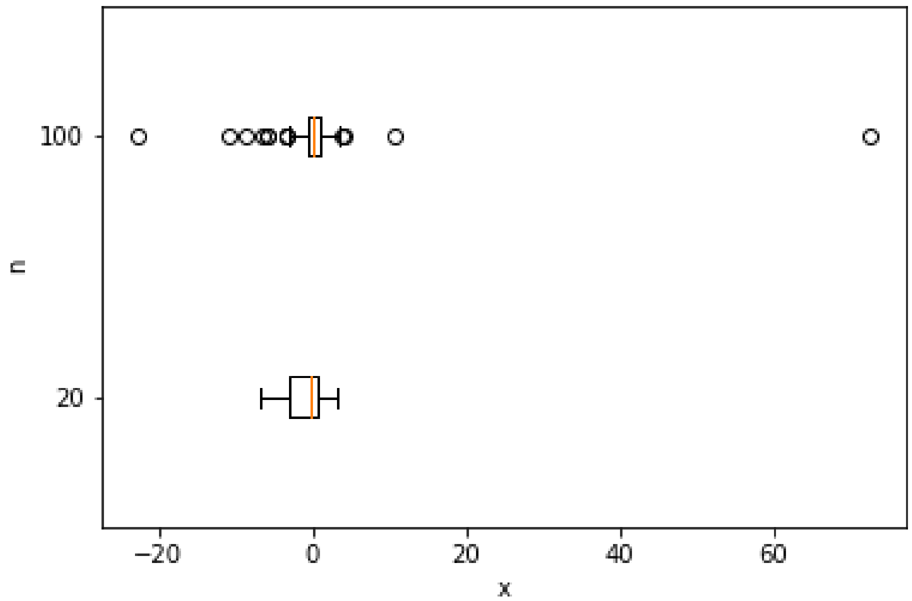

Рис. 2. Распределение Коши

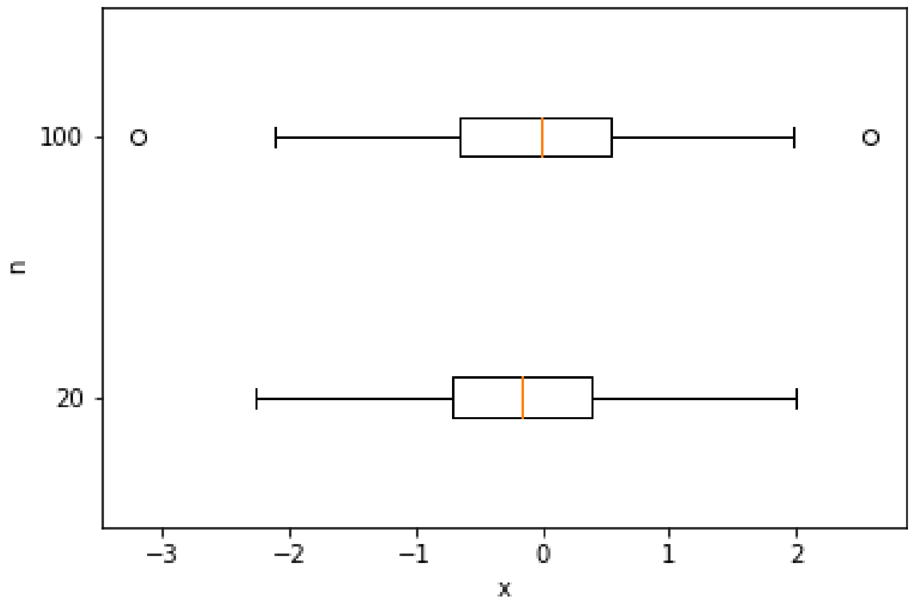

Рис. 3. Распределение Лапласа

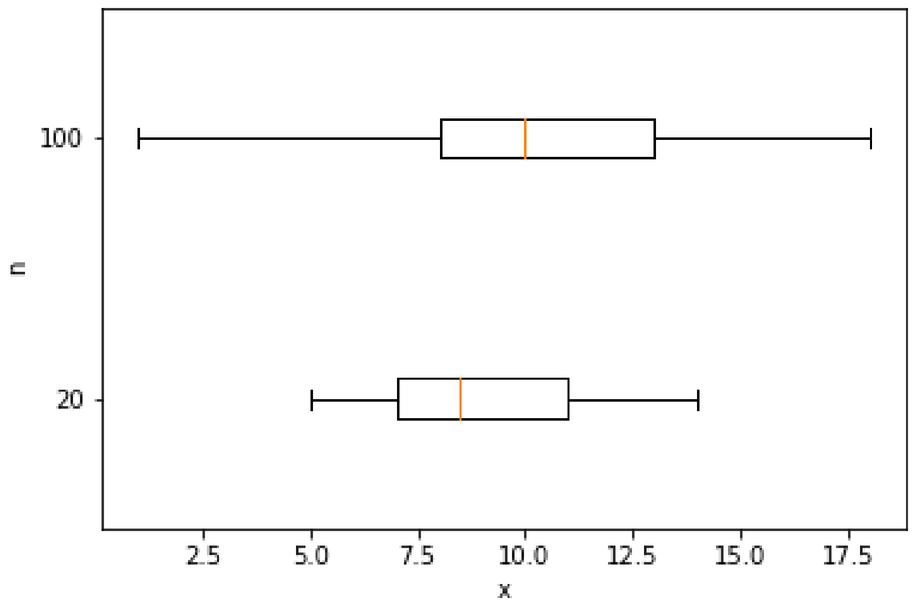

Рис. 4. Распределение Пуассона

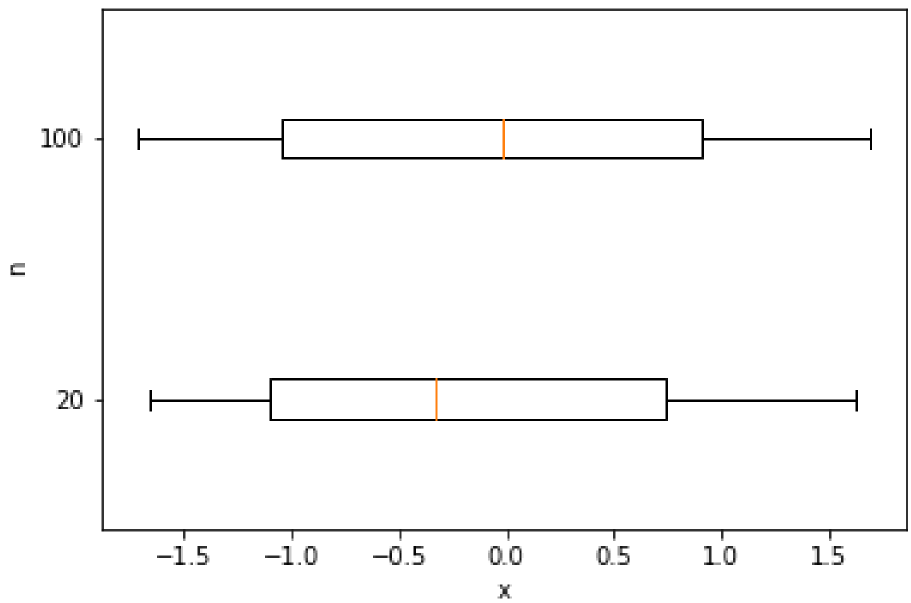

Рис. 5. Равномерное распределение

4.2. Доля выбросов

Таблица 1. Доля выбросов

Распределение	$n = 20$		$n = 100$	
	E (1)	D (2)	E	D
Нормальное распределение	0.024	0.0019	0.0101	0.0002
Распределение Коши	0.152	0.0049	0.154	0.0011
Распределение Лапласа	0.074	0.0044	0.0632	0.0009
Распределение Пуассона	0.024	0.0022	0.0105	0.0002
Равномерное распределение	0.0020	0.0002	0.0	0.0

4.3. Теоретическая вероятность выбросов

Таблица 2. Теоретическая вероятность выбросов

Распределение	Q_1^T	Q_3^T	X_1^T (8)	X_2^T (8)	P_{out}^T (10), (11)
Нормальное распределение	-0.6745	0.6745	-2.698	2.698	0.007
Распределение Коши	-1	1	-4	4	0.156
Распределение Лапласа	-0.4901	0.4901	-1.9605	1.9605	0.0625
Распределение Пуассона	8	12	2	18	0.0077
Равномерное распределение	-0.866	0.866	-3.4641	3.4641	0

5. Обсуждение

Исходя из полученных результатов можно сделать следующие выводы:

- Справедливо следующее соотношение для долей выбросов: равномерное распределение $<$ нормальное распределение $<$ распределение Пуассона $<$ распределение Лапласа $<$ распределение Коши.

- Доля выбросов, полученная экспериментально, близка с результатами, полученными теоретически.

6. Литература

1) Box plot. URL: `https://en.wikipedia.org/wiki/Box_plot`

7. Приложение

1) Код лабораторной. URL: `https://github.com/DmitriiKondratev/MatStat/blob/master/Lab_3/Lab_3.ipynb`

2) Код отчёта. URL: `https://github.com/DmitriiKondratev/MatStat/blob/master/Lab_3/Lab_report_3.tex`