Санкт-Петербургский Политехнический Университет имени Петра Великого
Институт Прикладной Математики и Механики
Кафедра "Прикладная Математика"

Отчет по лабораторной работе №3
по дисциплине
"Математическая Статистика"

Выполнил студент:
Тырыкин Я. А.
группа 3630102/80401
Проверил:
к.ф.-м.н., доцент
Баженов А. Н.

Санкт-Петербург
2021

Содержание

1 Постановка задачи **5**

2 Теория **5**

2.1 Распределения 5

2.2 Боксплот Тьюки 5

2.2.1 Определение 5

2.2.2 Описание 6

2.2.3 Построение 6

2.3 Теоретическая вероятность выбросов 6

3 Модульная структура программы **7**

4 Результаты **7**

4.1 Боксплот Тьюки 7

4.2 Доля выбросов 9

4.3 Теоретическая вероятность выбросов 10

5 Обсуждение **10**

5.1 Гистограммы и графики распределений 10

6 Ресурсы **11**

Список иллюстраций

1 Нормальное распределение (1) 7
2 Распределение Коши (2) 8
3 Распределение Лапласа (3) 8
4 Распределение Пуассона (4) 9
5 Равномерное распределение (5) 9

1 Постановка задачи

Для 5 распределений:

1. Нормальное распределение $N(x, 0, 1)$
2. Распределение Коши $C(x, 0, 1)$
3. Распределение Лапласа $L(x, 0, \frac{1}{\sqrt{2}})$
4. Распределение Пуассона $P(k, 10)$
5. Равномерное распределение $U(x, -\sqrt{3}, \sqrt{3})$

Необходимо сгенерировать выборки размером 10, 50 и 100 элементов. Построить на одном рисунке гистограмму и график плотности распределения.

2 Теория

2.1 Распределения

- Нормальное распределение:

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

- Распределение Коши:

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

- Распределение Лапласа:

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

- Распределение Пуассона:

$$P(k, 10) = \frac{10^k}{k!} e^{-10} \tag{4}$$

- Равномерное распределение:

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

2.2 Боксплот Тьюки

2.2.1 Определение

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

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.

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

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

где X_1 – нижняя граница уса, X_2 – верхняя граница 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 R в среде разработки RStudio можно вычислить теоретические первый и третий квартили распределений (Q_1^T и Q_3^T соответственно). По формуле (6) можно вычислить теоретические нижнюю и верхнюю границы уса (X_1^T и X_2^T соответственно). Выбросами считаются величины x такие, что:

$$\left[\begin{array}{lcl} x & < & X_1^T \\ x & > & X_2^T \end{array} \right. \tag{7}$$

Теоретическая вероятность выбросов для непрерывных распределений

$$P_B^T = P(x < X_1^T) + P(x > X_2^T) = F(X_1^T) + (1 - F(X_2^T)), \tag{8}$$

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

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

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

3 Модульная структура программы

Лабораторная работа выполнена с применением средств языка Python версии 3.7 в среде разработки PyCharm IDE (в частности, с применением встроенных методов библиотеки SciPy и MatPlotLib). Исходной код лабораторной работы находится по ссылке в приложении к отчёту.

4 Результаты

4.1 Боксплот Тьюки

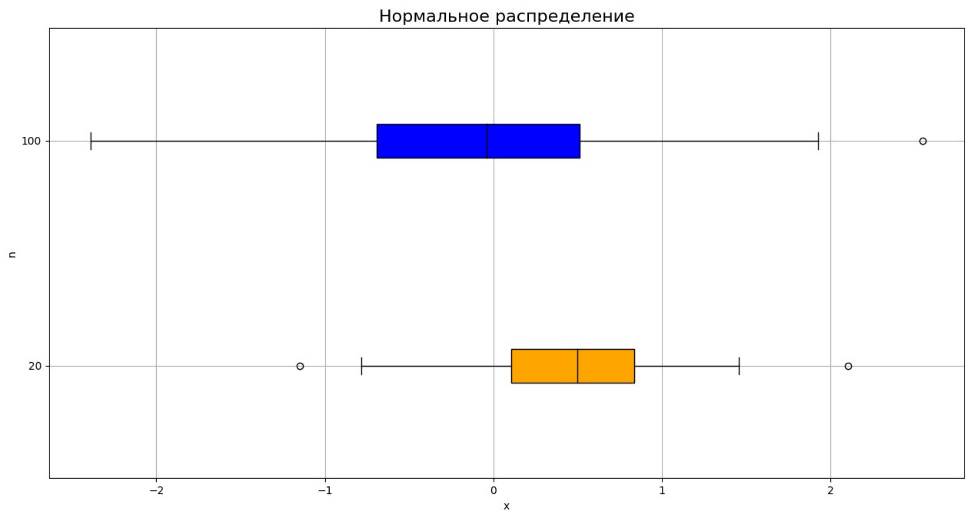

Рис. 1: Нормальное распределение (1)

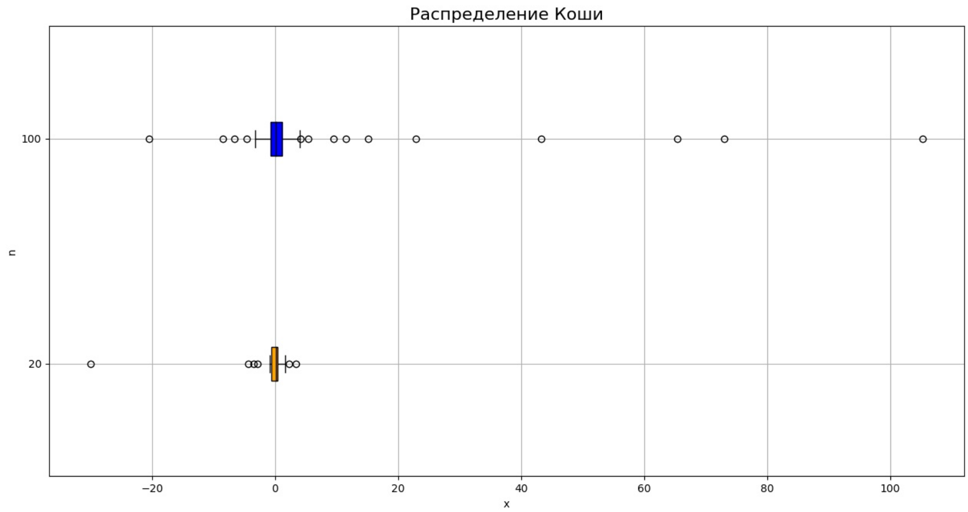

Рис. 2: Распределение Коши (2)

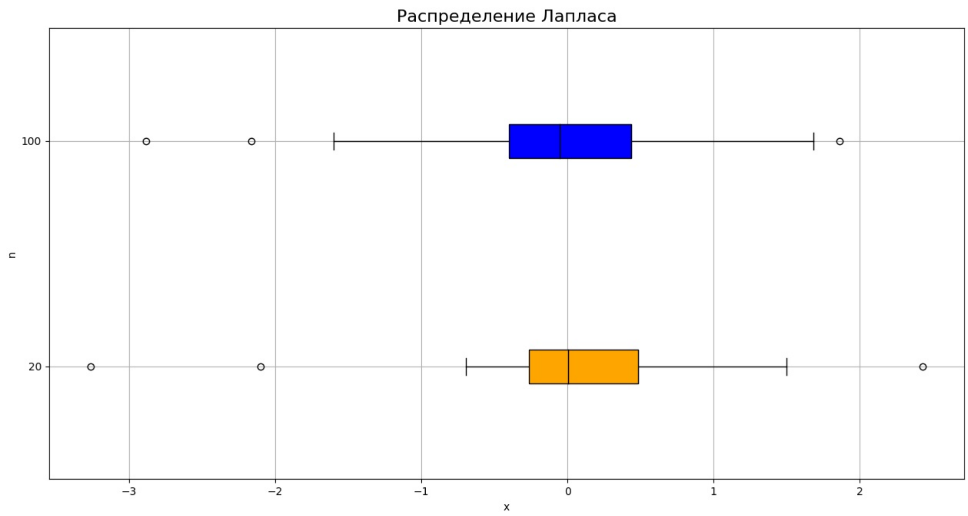

Рис. 3: Распределение Лапласа (3)

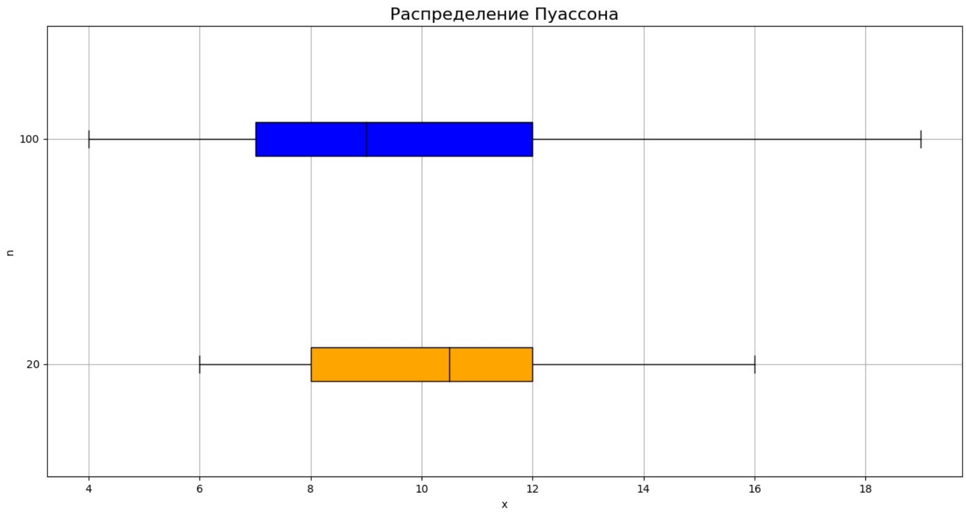

Рис. 4: Распределение Пуассона (4)

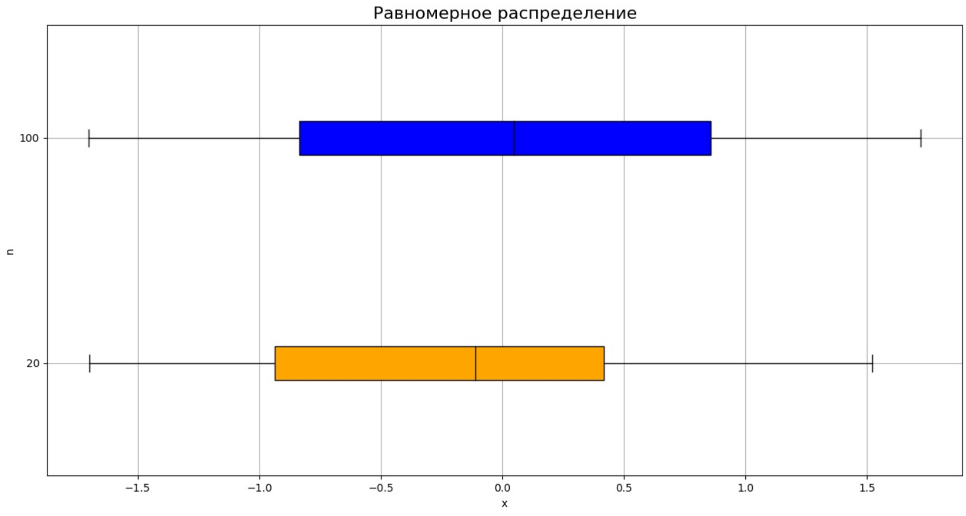

Рис. 5: Равномерное распределение (5)

4.2 Доля выбросов

Округление доли выбросов:

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

Доля $p_n = D_n/n = 1/\sqrt{n}$

Для $n = 20 : p_n = 1/\sqrt{20}-$ 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.

Выборвка	Доля выбросов
Normal n = 20	0.023
Normal n = 100	0.01
Cauchy n = 20	0.149
Cauchy n = 100	0.155
Laplace n = 20	0.075
Laplace n = 100	0.064
Poisson n = 20	0.024
Poisson n = 100	0.0099
Uniform n = 20	0.002
Uniform n = 100	0.0

Таблица 1: Доля выбросов

4.3 Теоретическая вероятность выбросов

Распределение	Q_1^T	Q_3^T	X_1^T (6)	X_2^T (6)	P_B^T (8), (9)
Нормальное распределение (1)	-0.674	0.674	-2.698	2.698	0.007
Распределение Коши (2)	-1	1	-4	4	0.156
Распределение Лапласа (3)	-0.490	0.490	-1.961	1.961	0.063
Распределение Пуассона (4)	8	12	2	18	0.008
Равномерное распределение (5)	-0.866	0.866	-3.464	3.464	0

Таблица 2: Теоретическая вероятность выбросов

5 Обсуждение

5.1 Гистограммы и графики распределений

По данным, приведенным в таблице, можно сказать, что чем больше выборка, тем ближе доля выбросов будет к теоретической оценке. Снова доля выбросов для распределения Коши значительно выше, чем для остальных распределений. Равномерное распределение же в точности повторяет теоретическую оценку - выбросов мы не получали.

Боксплоты Тьюки действительно позволяют более наглядно и с меньшими усилиями оценивать важные характеристики распределений. Так, исходя из полученных рисунков, наглядно видно то, что мы довольно трудоёмко анализировали в предыдущих частях.

6 Ресурсы

Код программы, реализующей отрисовку обозначенных распределений:

https://github.com/YaroslavAggressive/Mathematical-statistics-lab-works