Санкт-Петербургский политехнический университет
Петра Великого

Институт прикладной математики и механики
Кафедра «Прикладная математика»

ОТЧЁТ ПО ЛАБОРАТОРНЫМ РАБОТАМ ПО ДИСЦИПЛИНЕ **«МАТЕМАТИЧЕСКАЯ СТАТИСТИКА»**

Выполнил студент
Войнова Алёна Игоревна
группы 3630102/80201

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

Санкт-Петербург
2021

Содержание

Список иллюстраций

Список таблиц

1 Постановка задачи

Для 5 распределений:

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

1.1 Задание 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 сравнить с результатами, полученными теоретически.

2 Теория

2.1 Распределения

- Нормальное распределение

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

- Распределение Коши

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

- Распределение Лапласа

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

- Распределение Пуассона

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

- Равномерное распределение

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

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

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

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 [1].

2.3 Теоретическая вероятность выбросов

По формуле (15) можно вычислить теоретические нижнюю и верхнюю границы уса (X_1^T, 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

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

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

3 Реализация

Лабораторная работа выполнена с помощью средств языка программирования **Python** в среде разработки **Jupyter**. Исходный код лабораторной работы приведён в приложении.

4 Результаты

4.1 Боксплот Тьюки

Для каждого распределения представлен боксплот Тьюки для выборок размером 20 и 100 элементов.

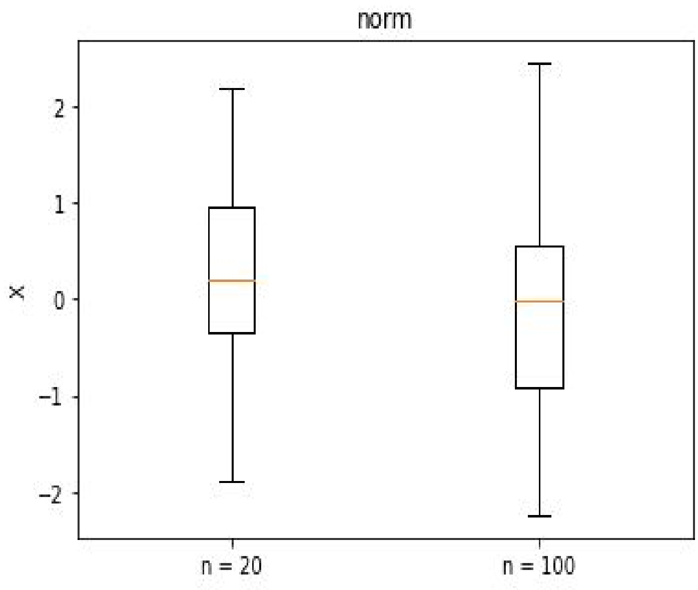

Рис. 1: Нормальное распределение

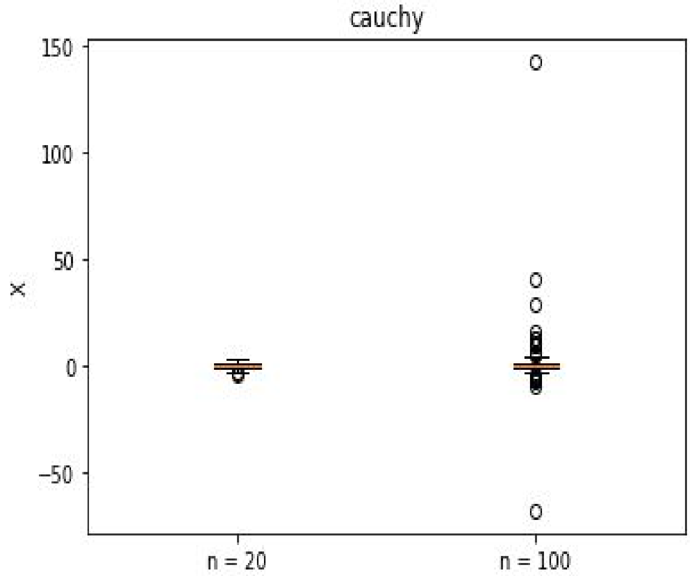

Рис. 2: Распределение Коши

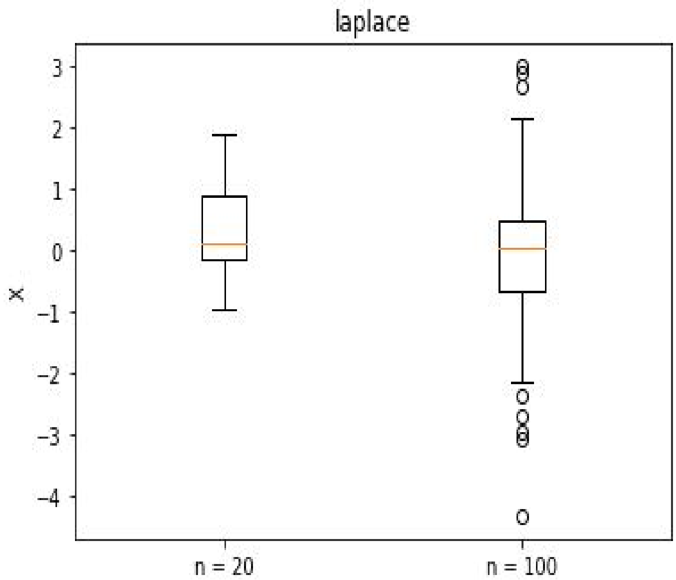

Рис. 3: Распределение Лапласа

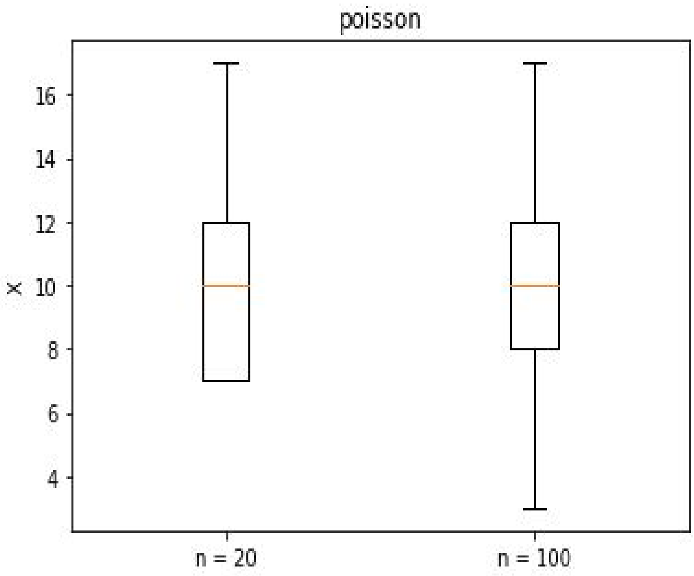

Рис. 4: Распределение Пуассона

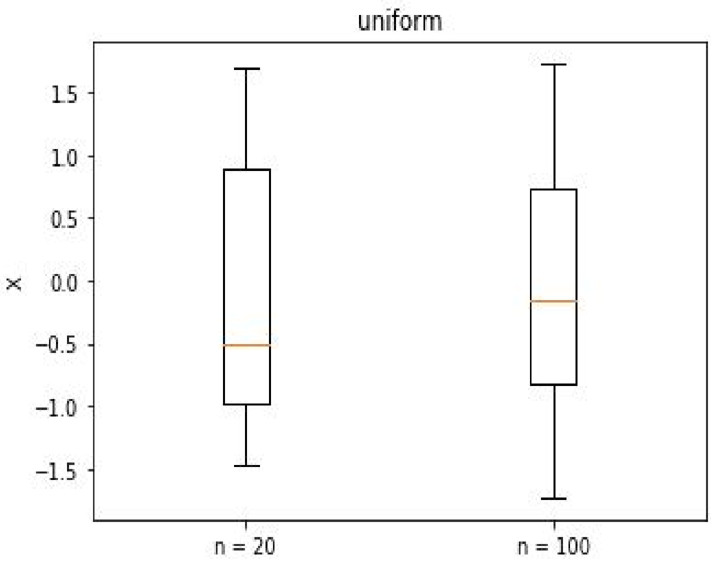

Рис. 5: Равномерное распределение

4.2 Доля выбросов

Выборка $n = 20$	Доля выбросов
norm	0.1
cauch	0.28
laplace	0.15
poisson	0.08
uniform	0.15

Выборка $n = 100$	Доля выбросов
norm0	0.12
cauch	0.27
laplac	0.15
poisso	0.05
uniform	0.04

Таблица 1: Доля выбросов

4.3 Теоретическая вероятность выбросов

Распределение	Q_1^T	Q_2^T	X_1^T	X_2^T	P_B^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
Распределение 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

Сравним долю выбросов определенную экспериментально с результатами, полученными теоретически. Видим точное соответствие с теорией для равномерного распределения - вероятность нулевая и выбросов мы не получили.
Результаты для выборок, сгенерированных в соответствии с законами распределения Лапласа и Коши, оказались близкими к теории, а доля выбросов для распределений Пуассона и Нормального ниже соответствующих теоретических оценок.
Заметим, что все распределения дают для большей выборки (100 элементов) результат ближе к теории, чем для меньшей выборки (20 элементов). Следовательно чем больше выборка, тем результыат ближе к теоретическим.

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

6 Приложения

URL: Выполненная лабораторная работа на GitHub
https://github.com/pikabol88/Math-Statistics/blob/main/labs/Lab3.ipynb

Список литературы

[1] Box plot. URL: https://en.wikipedia.org/wiki/Box_plot