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
В. А. Рыженко

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

Санкт-Петербург, 2020 г.

Содержание

Список иллюстраций

Список таблиц

1. Постановка задачи

Для 5 распределений:

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

Сгенерировать выборки размером 20 и 100 элементов. Построить для них бокс-плот 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

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

- 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 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.

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

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:

$$\left[\begin{array}{l} x < X_1^T \\ x \leq X_2^T \end{array}\right. \tag{7}$$

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 $F(X) = P(x > X)$ - 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 $F(X) = P(x > X)$ - функция распределения.

3. Реализация

Лабораторная работа выполнена с помощью встроенных средств языка программирования Python в среде разработки Visual Code. Исходный код лабораторной работы приведён в приложении.

4. Результаты

4.1. Боксплот Тьюки

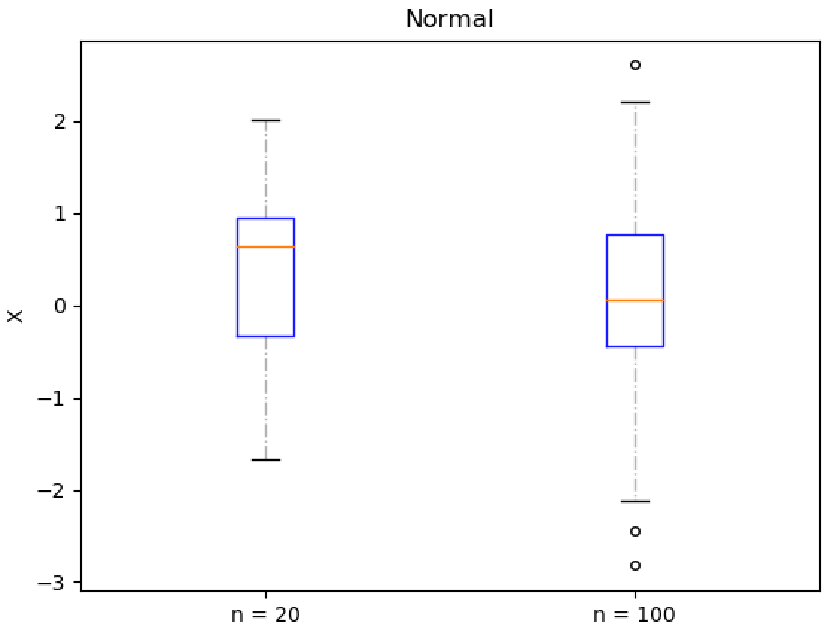

Рис. 1. Нормальное распределение (1)

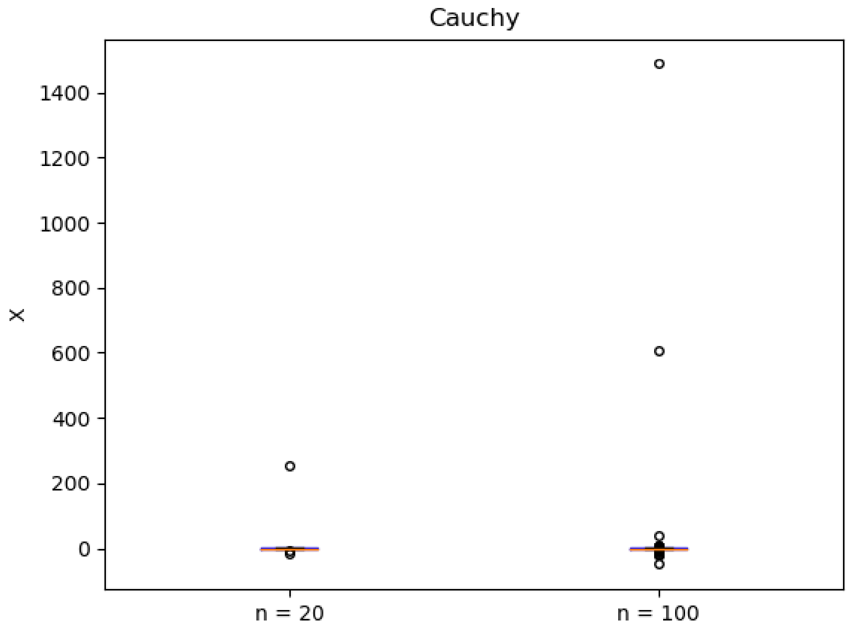

Рис. 2. Распределение Коши (2)

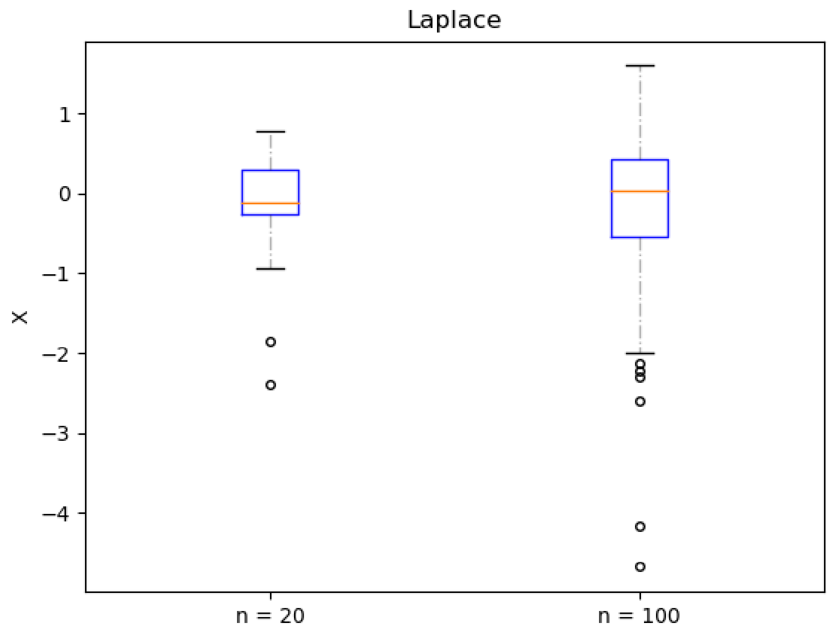

Рис. 3. Распределение Лапласа (3)

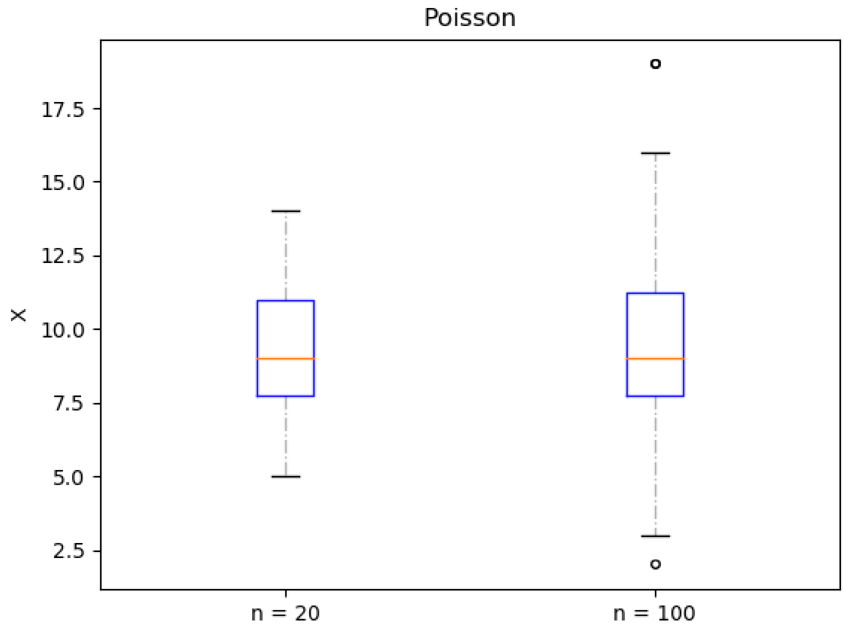

Рис. 4. Распределение Пуассона (4)

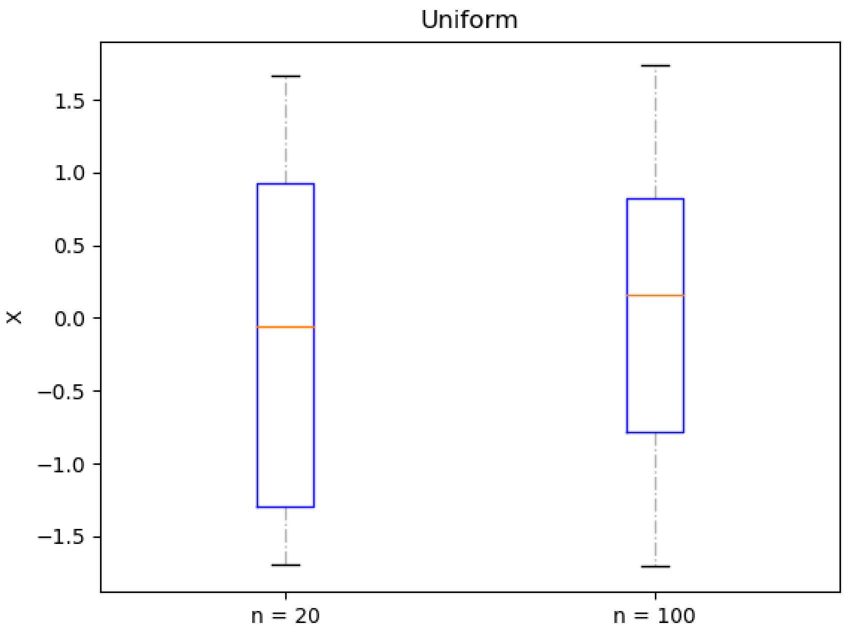

Рис. 5. Равномерное распределение (5)

4.2. Доля выбросов

Выборка	Среднее	Дисперсия
Normal, n = 20	0.017	0.001
Normal, n = 100	0.0095	0.0001
Cauchy, n = 20	0.139	0.005
Cauchy, n = 100	0.154	0.001
Laplace, n = 20	0.065	0.004
Laplace, n = 100	0.0642	0.0009
Poisson, n = 20	0.016	0.002
Poisson, n = 100	0.0101	0.0003
Uniform, n = 20	0	0
Uniform, n = 100	0	0

Таблица 1. Доля выбросов

4.3. Теоретическая вероятность выбросов

Распределение	Q_1^T	Q_3^T	X_1^T	X_2^T	P_B^T (8) (9)
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

Из полученных данных видно, что средняя доля выбросов (для 1000 эксперементов) стремится к теоретической оценке при увеличении размера выборки, для всех распределений, кроме распределения Пуассона (4). Дисперсия в свою очередь стремится к нулю уже для всех распределений. Также можно заметить, что для равномерного распределения отсутвуют выбросы, а вероятность их появления равна 0.

6. Приложения

Репозиторий на GitHub с релизацией: github.com.