Санкт-Петербургский политехнический университет Петра Великого

Институт прикладной математики и механики

Высшая школа прикладной математики

ОТЧЁТ ПО ЛАБОРАТОРНОЙ РАБОТЕ №3

по дисциплине «Математическая статистика»

Выполнила
студентка гр.3630102/80101

А.А. Тимофеева

Руководитель доцент, к.ф.-м.н.

А.Н.Баженов

Санкт-Петербург 2021

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 $L(x, 0, \frac{1}{\sqrt{2}})$
- 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 выборов) 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 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.

2.1.3 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

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

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.

2.2 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_1^T соответственно). Выбросами считаются величины x, такие что:

$$\left[\begin{array}{l} x < X_1^T \\ x > X_2^T \end{array}\right. \tag{2}$$

Теоретическая вероятность выбросов для непрерывных распределений

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

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

3 ПРОГРАММНАЯ РЕАЛИЗАЦИЯ

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

4 РЕЗУЛЬТАТЫ

4.1 Боксплот Тьюки

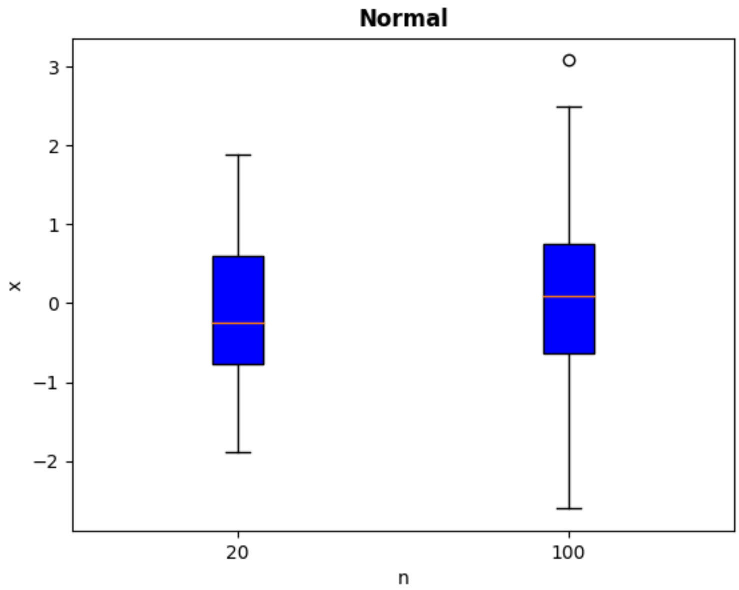

Рис.1 Нормальное распределение

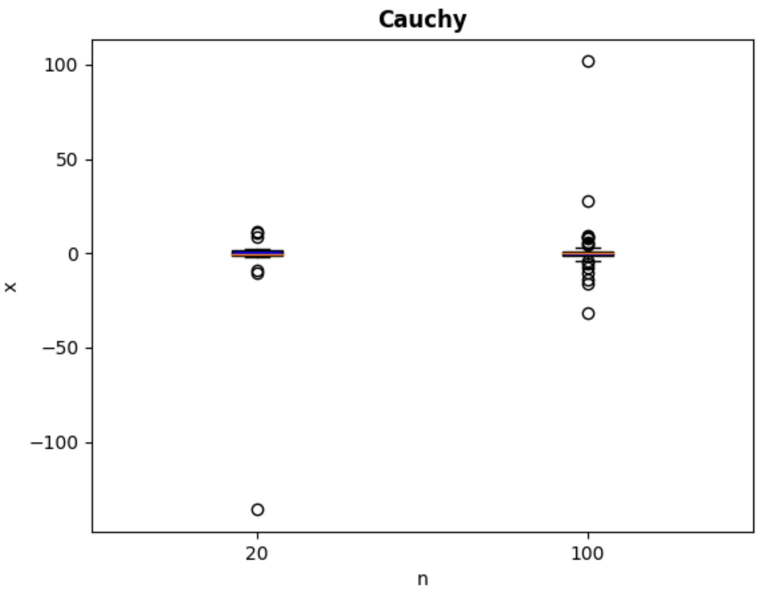

Рис.2 Распределение Коши

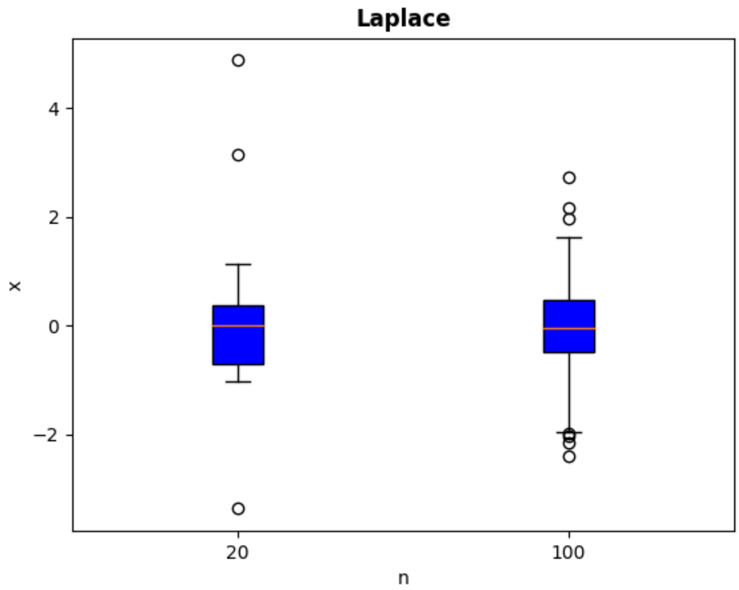

Рис.3 Распределение Лапласа

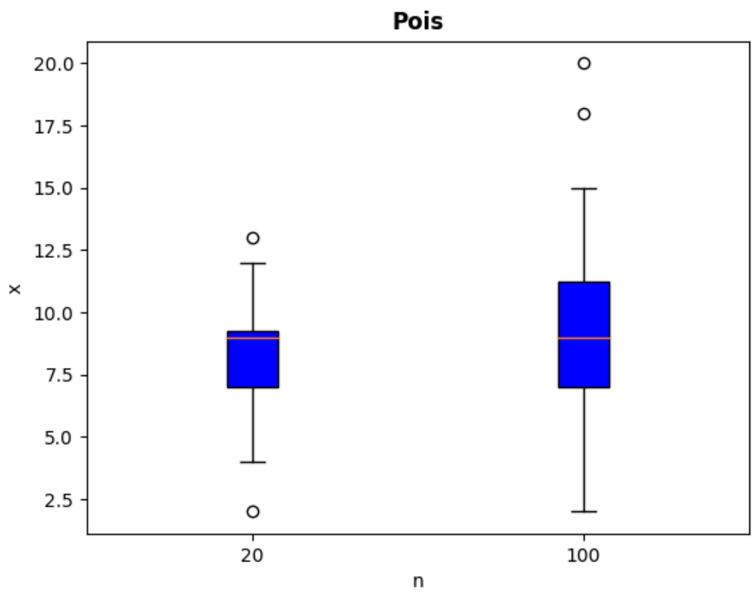

Рис.4 Распределение Пуассона

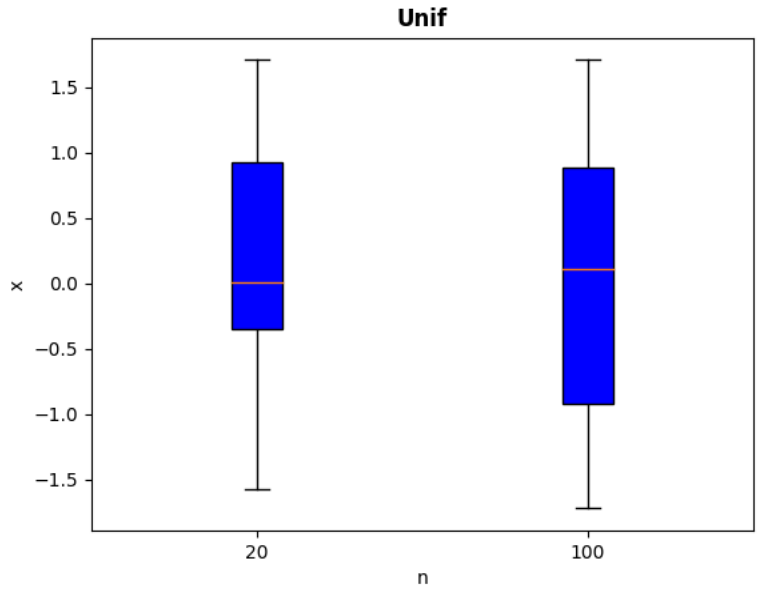

Рис.5 Равномерное распределение

4.2 Доля выбросов

Округление доли выбросов:

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

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
Normal, n = 20	0.02
Normal, n = 100	0.01
Cauchy, n = 20	0.15
Cauchy, n = 100	0.16
Laplace, n = 20	0.07
Laplace, n = 100	0.06
Pois, n = 20	0.02
Pois, n = 100	0.01
Unif, n = 20	0
Unif, n = 100	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
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

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 ПРИЛОЖЕНИЕ

Код программы URL:https://github.com/tmffv/MathStat/blob/master/lab3/lab3.py