

САНКТ-ПЕТЕРБУРГСКИЙ ПОЛИТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ
ИМ. ПЕТРА ВЕЛИКОГО

ИНСТИТУТ ПРИКЛАДНОЙ МАТЕМАТИКИ И МЕХАНИКИ
КАФЕДРА "ПРИКЛАДНАЯ МАТЕМАТИКА"

ОТЧЁТ
ЛАБОРАТОРНАЯ РАБОТА № 3
ПО ДИСЦИПЛИНЕ
"МАТЕМАТИЧЕСКАЯ СТАТИСТИКА"

ВЫПОЛНИЛ СТУДЕНТ:
ШАРАПОВ СЕРГЕЙ АНДРЕЕВИЧ
ГРУППА: 3630102/70401

ПРОВЕРИЛ:
К.Ф-М.Н., ДОЦЕНТ
БАЖЕНОВ АЛЕКСАНДР НИКОЛАЕВИЧ

САНКТ-ПЕТЕРБУРГ
2020 год

Содержание

Стр.

1. Постановка задачи	5
2. Теория	5
2.1. Распределения	5
2.2. Гистограмма	5
2.2.1 Определение	5
2.2.2 Описание	5
2.2.3 Использование	6
3. Реализация	6
4. Результаты	7
4.1. Боксплот Тьюки	7
5. Обсуждение	13
5.1. Анализ данных	13
5.2. Сравнение с теоретическими значениями	13
6. Литература	13
7. Приложения	13

Список иллюстраций

1	Нормальное распределение	7
2	Распределение Коши	8
3	Распределение Лапласа.....	9
4	Распределение Пуассона	10
5	Равномерное распределение.....	11

Список таблиц

1	Выбросы различных распределений в зависимости от выборки	12
---	--	----

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 сравнить с результатами полученными теоретически.

2 Теория

2.1 Распределения

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

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

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

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

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

2.2 Гистограмма

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 Использование

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:

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

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.

3 Реализация

Для генерации случайных чисел с различными распределениями был использован модуль *random* библиотеки *numpy* языка *Python 3.7*. Боксплот Тьюки был построен средствами *matplotlib*.

4 Результаты

4.1 Боксплот Тьюки

Введём на оси y следующие обозначения:
1 - соответствует выборке из 20-ти элементов
2 - соответствует выборке из 100 элементов

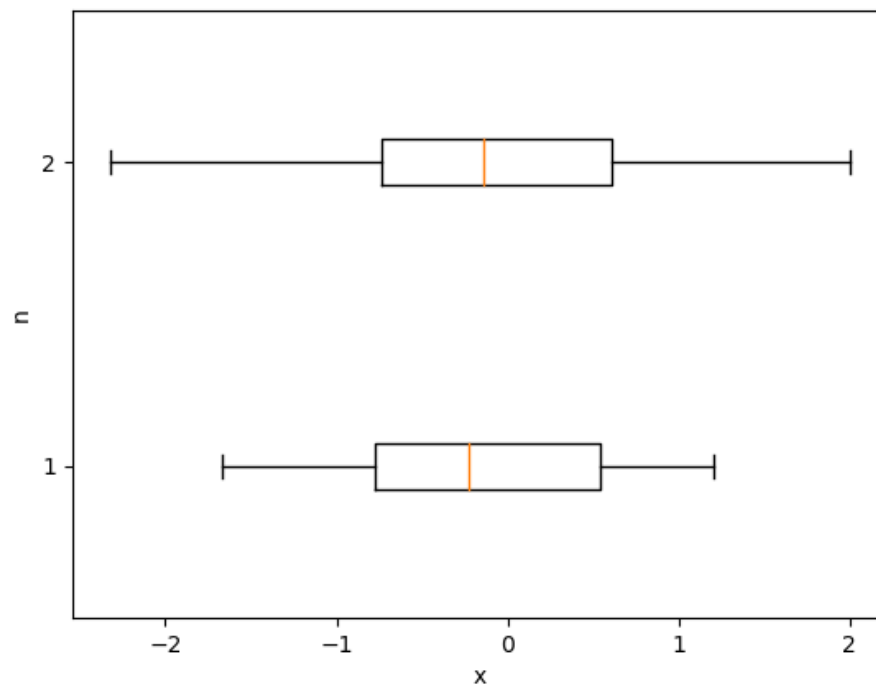


Рис. 1: Нормальное распределение

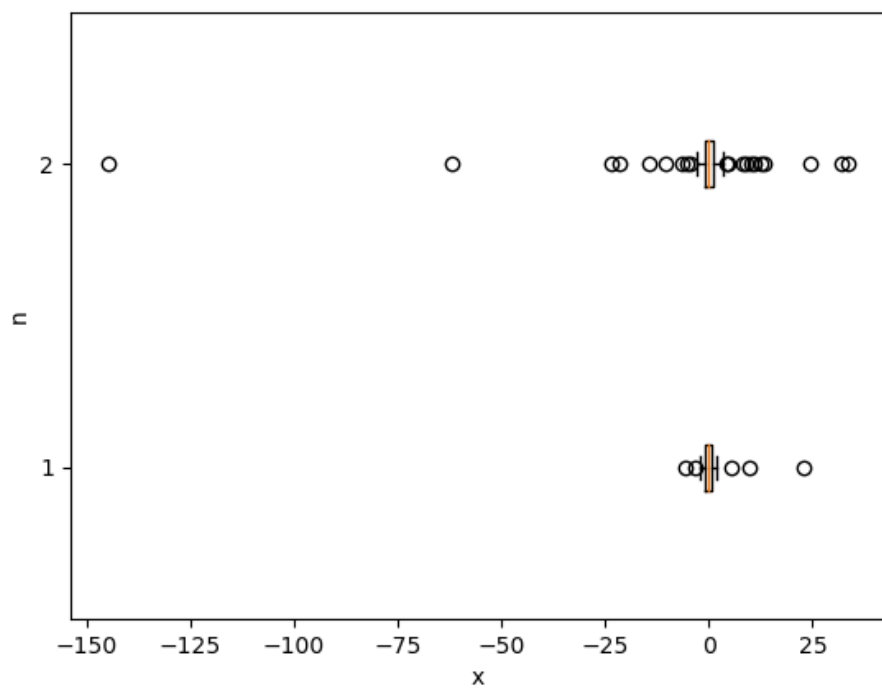


Рис. 2: Распределение Коши

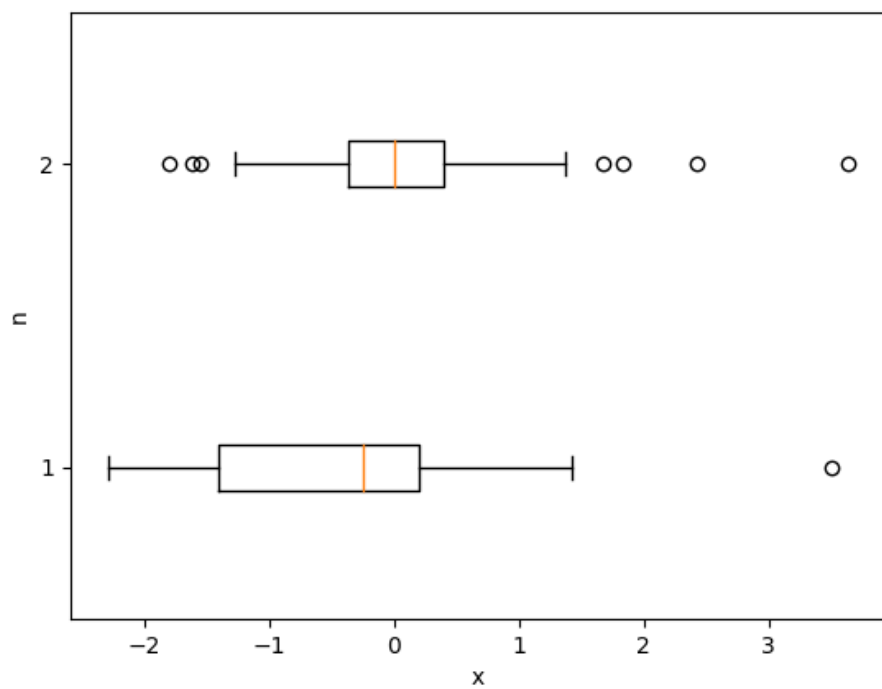


Рис. 3: Распределение Лапласа

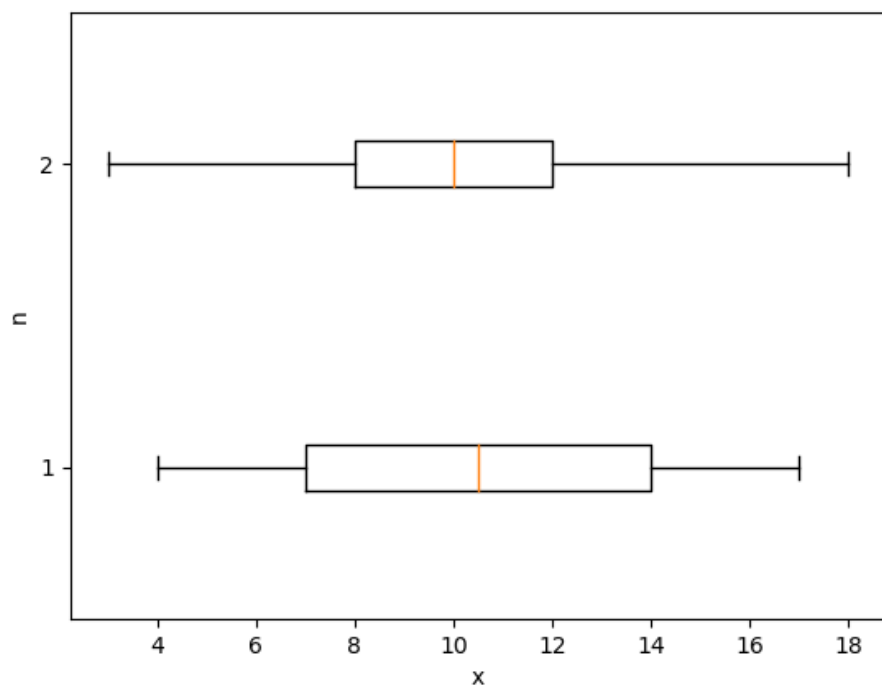


Рис. 4: Распределение Пуассона

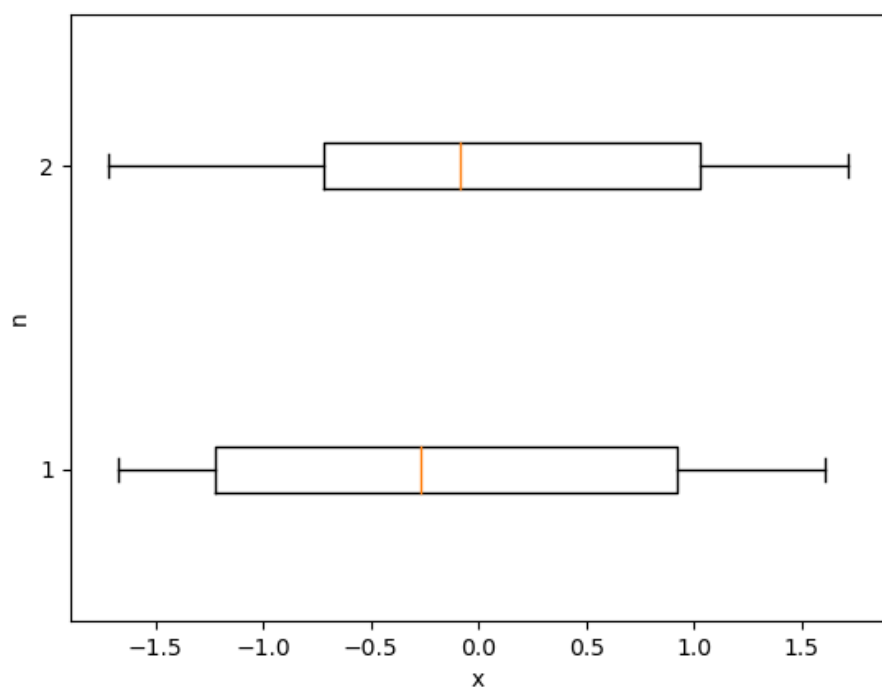


Рис. 5: Равномерное распределение

Таблица 1: Выбросы различных распределений в зависимости от выборки

Выборка	Процент выбросов	Дисперсии
Нормальное		
n = 20	2	0.0018
n = 100	1	0.0001
Коши		
n = 20	15	0.0051
n = 100	16	0.0012
Лапласа		
n = 20	7	0.0043
n = 100	6	0.0009
Пуассона		
n = 20	3	0.0023
n = 100	1	0.0002
Равномерное		
n = 20	0	0.0003
n = 100	0	0.0000

5 Обсуждение

5.1 Анализ данных

Из экспериментально полученных данных можно вывести соотношение между процентами выбросов: равномерное распределение < нормальное распределение < распределение Пуассона < распределение Лапласа < распределение Коши.

5.2 Сравнение с теоретическими значениями

Полученные экспериментально данные близки к теоретическим и видно, что наименьший процент выбросов у равномерного распределения ,а наибольший у распределения Коши.

6 Литература

Модуль `numpy`
`matplotlib` `boxplot`

7 Приложения

Код лабораторной