Санкт-Петербургский Политехнический
университет имени Петра Великого

Отчёт по лабораторной работе №2

Характеристики положения и рассеяния

Студент: Расторгуев Михаил Павлович
Группа: 5030102/30201

Санкт-Петербург
2026

1 Постановка задачи

В рамках данной лабораторной работы было необходимо:

- Исследовать сходимость выборочных характеристик к теоретическим при росте n.
- Исследовать оценки характеристик положения на устойчивость к выбросам.

2 Теоретическая часть

В данной работе требовалось изучить следующие распределения случайных величин:

- Нормальное распределение $N(x; 0, 1)$

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R}$$

- Распределение Коши $C(x; 0, 1)$

$$f(x) = \frac{1}{\pi(1 + x^2)}, \quad x \in \mathbb{R}$$

- Распределение Лапласа $L\left(x; 0, \frac{1}{\sqrt{2}}\right)$

$$f(x) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}|x|}, \quad x \in \mathbb{R}$$

- Распределение Пуассона $P(k; 5)$

$$P(X = k) = \frac{5^k e^{-5}}{k!}, \quad k = 0, 1, 2, \ldots$$

- Равномерное распределение $U(x; -\sqrt{3}, \sqrt{3})$

$$f(x) = \frac{1}{2\sqrt{3}}, \quad x \in [-\sqrt{3}, \sqrt{3}]$$

3 Реализация

Практическая часть работы была реализована на языке программирования `Python` с использованием библиотек: `scipy,seaborn,numpy,pandas`.
Код писался в интерактивном блокноте `Jupyter Notebook`.

Были реализованы следующие функции:

`get_five_values(n,dist,**param)` - функция, генерирующая набор из случайных данных и затем считает требуемые выборочные характеристик. На вход подаётся `n` - размер генерируемых данных; `dist` - распределение, по которому генерируются данные; `**param` - параметры распределения.

`show_esstimates(dist,real_values,**params)` - функция - обёртка над `get_five_values`. Дополнительный параметр `real_values` - кортеж теоретических значений для распределения `dist`.

4 Результаты

Ниже представлены результаты подсчёта характеристик, с указанием погрешностей и теоретических значений.

	n = 10	n = 100	n = 1000	Теоретическое значение
Выборочное среднее	-0.0 ± 0.3	-0.0 ± 0.1	0.0 ± 0.0	0
Медиана	-0.0 ± 0.4	0.0 ± 0.1	-0.0 ± 0.0	0
Полусумма экстремальных элементов	-0.0 ± 0.8	-0.0 ± 0.6	-0.0 ± 0.5	0
Полусумма квартилей	-0.0 ± 0.3	0.0 ± 0.1	0.0 ± 0.0	0.0
Усечённое среднее	-0.0 ± 0.3	0.0 ± 0.1	0.0 ± 0.0	0

Рис. 1: $N(0,1)$

	n = 10	n = 100	n = 1000	Теоретическое значение
Выборочное среднее	-0.0 ± 15.4	6.6 ± 225.8	37.2 ± 1157.0	Не существует
Медиана	-0.0 ± 0.6	0.0 ± 0.2	0.0 ± 0.0	0
Полусумма экстремальных элементов	0.1 ± 151.9	664.7 ± 22580.4	37210.2 ± 1156975.9	Не существует
Полусумма квартилей	-0.0 ± 1.0	0.0 ± 0.2	0.0 ± 0.1	0.0
Усечённое среднее	-0.1 ± 1.2	0.0 ± 0.2	0.0 ± 0.1	0

Рис. 2: $C(0,1)$

	n = 10	n = 100	n = 1000	Теоретическое значение
Выборочное среднее	0.0 ± 0.5	0.0 ± 0.1	0.0 ± 0.0	0
Медиана	-0.0 ± 0.4	0.0 ± 0.1	0.0 ± 0.0	0
Полусумма экстремальных элементов	0.0 ± 1.8	-0.0 ± 1.8	-0.0 ± 1.8	0
Полусумма квартилей	-0.0 ± 0.4	0.0 ± 0.1	0.0 ± 0.0	0.0
Усечённое среднее	-0.0 ± 0.4	0.0 ± 0.1	0.0 ± 0.0	0

Рис. 3: $L(0, \frac{1}{\sqrt{2}})$

	n = 10	n = 100	n = 1000	Теоретическое значение
Выборочное среднее	5.0 ± 0.7	5.0 ± 0.2	5.0 ± 0.1	5
Медиана	4.8 ± 0.8	4.9 ± 0.3	5.0 ± 0.0	5
Полусумма экстремальных элементов	10.5 ± 1.9	11.9 ± 1.4	13.6 ± 1.1	∞
Полусумма квартилей	4.9 ± 0.8	4.9 ± 0.4	4.7 ± 0.3	4.5
Усечённое среднее	4.9 ± 0.7	4.9 ± 0.2	4.9 ± 0.1	5

Рис. 4: $P(5)$

	n = 10	n = 100	n = 1000	Теоретическое значение
Выборочное среднее	0.0 ± 0.3	0.0 ± 0.1	-0.0 ± 0.0	0
Медиана	0.0 ± 0.5	0.0 ± 0.2	0.0 ± 0.1	0
Полусумма экстремальных элементов	-0.0 ± 0.4	0.0 ± 0.0	-0.0 ± 0.0	0
Полусумма квартилей	0.0 ± 0.4	0.0 ± 0.1	-0.0 ± 0.0	0.0
Усечённое среднее	0.0 ± 0.4	0.0 ± 0.1	-0.0 ± 0.0	0

Рис. 5: $U(-\sqrt{3}, \sqrt{3})$

5 Обсуждение

Из представленных выше таблиц видно, что если характеристика определена для распределения, то с увеличением размера выборки - её среднее стремится к теоретическому значению, а дисперсия уменьшается.

Для распределения Коши наблюдается следующее - выборочное среднее для сгенерированных данных по нему данных отказывается сходиться при увеличении n. При этом медиана лишена этого недостатка. Это связано с тем, что последняя характеристика менее чувствительна к выбросам, чем среднее арифметическое.

6 Выводы

1. С увеличением размера выборки, характеристики сходятся к теоретическим (если таковы определены).

2. Среднее арифметическое не всегда хорошо описывает статистическую зависимость, так как оно подвержено выбросам в данных.

7 Список литературы

1. Документация `scipy` - https://docs.scipy.org/doc/scipy/

2. Документация `seaborn` - https://seaborn.pydata.org/

8 Приложение

https://github.com/haskell-md2/MatStat/blob/main/lab1-4/lab1-4.ipynb - Параграф "Лабораторная 2"